

**Psychologisches Institut der Universität Tübingen**

**Übung: Statistik II, Sommersemester 2003**

**Skriptum zur Arbeit mit dem Computer**

**Prof. Dr. Wilhelm R. Glaser**

Dieses Skriptum wurde ursprünglich für die Studentenversion von SPSS 9 englisch verfaßt. Die späteren Vollversionen, englisch oder deutsch, weichen gelegentlich davon ab. Die wichtigsten Änderungen der Vollversion 11.5 englisch wurden eingearbeitet. Die deutschen Termini des Glossars entsprechen SPSS 10 deutsch.

## **Vorbemerkung**

Dieses Skriptum kann und soll kein Einführungsbuch in SPSS und andere Statistikprogramme ersetzen. Es zeigt die für die Statistik I und die Statistik II der Psychologen grundlegenden Auswertungen an Beispielen aus der Vorlesung und der Übung, die Schritt für Schritt nachvollzogen werden können. Die Übertragung auf andere, gleichartige Probleme dürfte nach dem gründlichen Durcharbeiten dieser Aufgaben am Rechner kein Problem mehr sein. Dabei konnte Vollständigkeit natürlich weder im Stoff der Vorlesung, noch in den Funktionen von SPSS erreicht werden. Die Lektüre dieses Skriptums sollte durch ausgedehntes, selbständiges Explorieren des SPSS-Programmes und ein Bücherstudium ergänzt werden. Nutzen Sie die Selbsterklärungsfähigkeit und die Hilfefunktionen von SPSS für ein aktives Selbststudium aus!

## **Literatur**

- Bühl, A. & Zöfel, P. (2000). *SPSS Version 9. Einführung in die moderne Datenanalyse unter Windows*. 6. überarb. u. erw. Aufl., mit CD-ROM. München: Addison-Wesley. (Für die neueren Versionen von SPSS erscheinen regelmäßig Neuauflagen.)
- Diehl, J. M. & Staufenbiel, T. (2001). *Statistik mit SPSS Version 10.0*. Eschborn: Verlag Dietmar Klotz.
- Kinnear, P. R. & Gray, C. D. (1999). *SPSS for Windows made simple*. 3<sup>rd</sup> Edition. Hove, East Sussex, UK: Psychology Press.

## Aufgabe 01

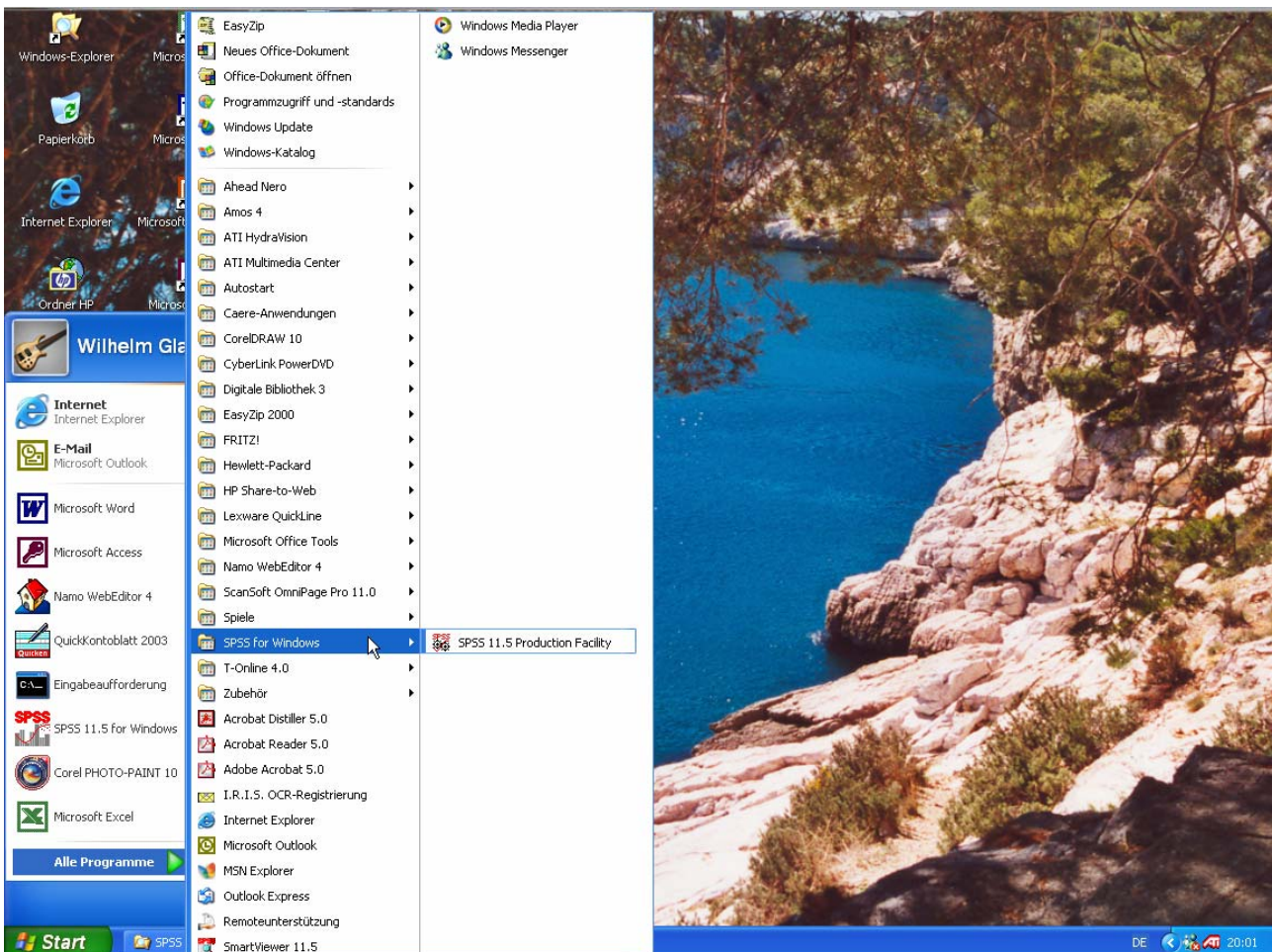
Diese Aufgabe basiert auf der univariaten Urliste, die den Erklärungen in der Vorlesung „Statistik I“ zugrundelag. Sie lautet (*stat130.doc*):

24 17 29 21 19 22 23 19 26 21 17 19 18 20 25 17 18  
 20 21 24 21 25 23 17 20 19 20 15 18 17 14 19 19 15  
 12 18 22 16 16 17 21 14 20 22 22 16 16 22 20 21 18  
 20 16 19 23 20 23 22 25 23 20 22 20 13 22 19 19 21  
 14

- Geben Sie diese Urliste in SPSS ein. Nennen Sie die abhängige Variable  $x$  und geben Sie ihr den Bezeichner *Abhängige Variable*. Speichern Sie die Datenmatrix in *Aufgabe01.sav*.
- Erzeugen Sie die Häufigkeitsverteilung, die prozentuale Häufigkeitsverteilung und die kumulierte prozentuale Häufigkeitsverteilung als Tabelle.
- Berechnen Sie Mittelwert, Median, Modus, Standardabweichung, Varianz, Schiefe, Exzeß, absoluten Streubereich, erstes und drittes Quartil.
- Stellen Sie die Häufigkeitsverteilung und die kumulierte Häufigkeitsverteilung als Histogramm graphisch dar.

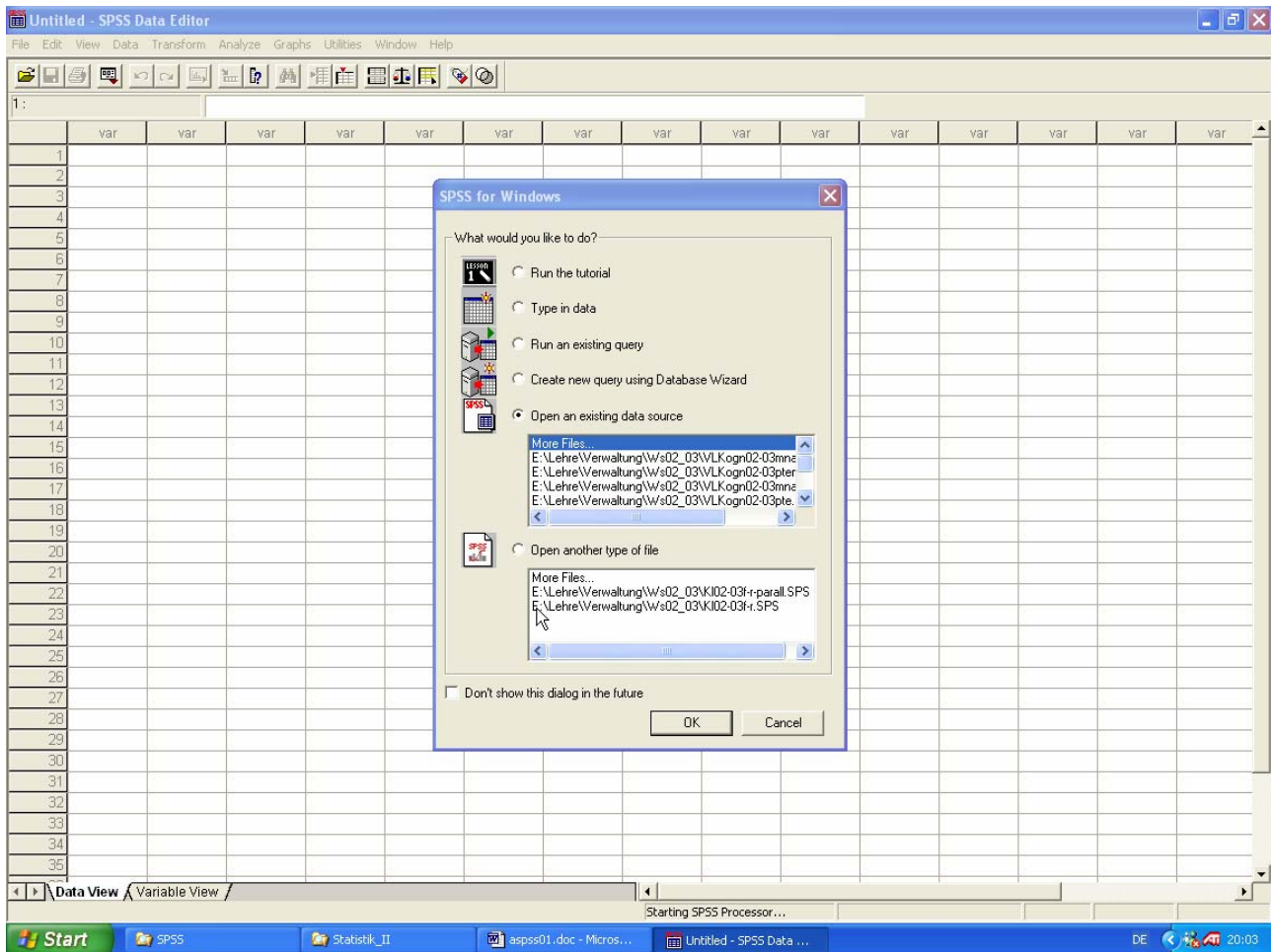
## Lösung

- Starten Sie SPSS 11.5 for WINDOWS



mit *Start* → *Programme* → *SPSS 11.5 for Windows*.

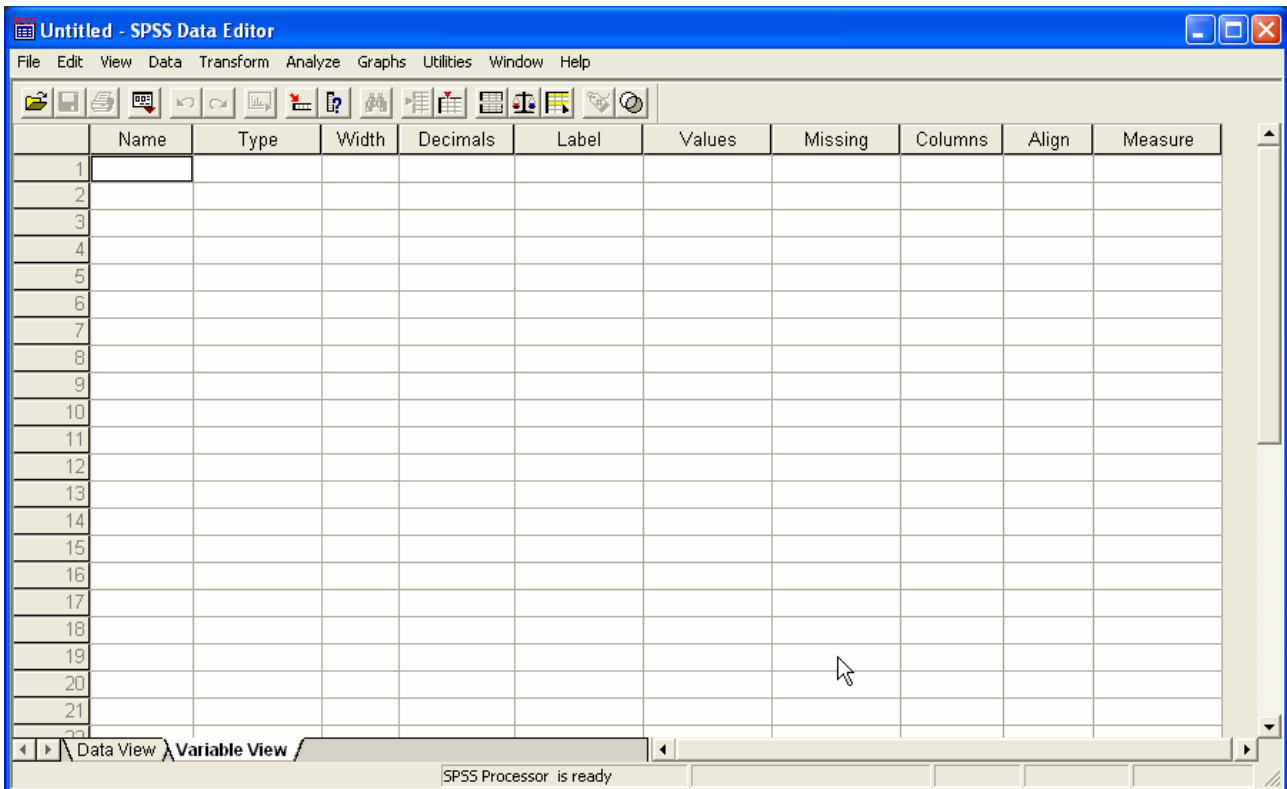
Sie erhalten den folgenden Bildschirm:



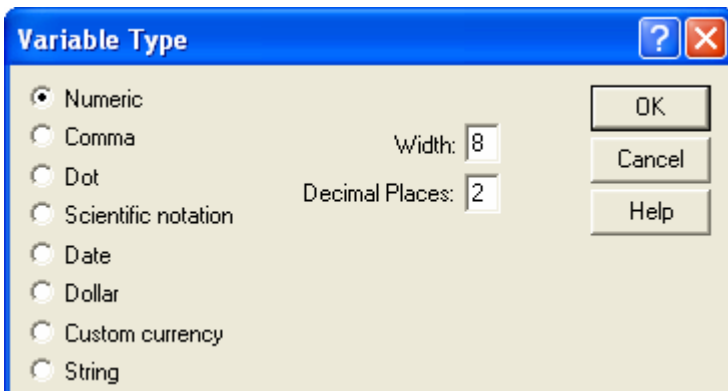
Löschen Sie das innere Fenster durch *Cancel*. (Wenn Sie später mit SPSS geläufig arbeiten, werden Sie es nicht mehr schließen, sondern seine Optionen nutzen.)

Das verbleibende äußere Fenster nimmt die Rohdatei auf. Die Zeilen sind grundsätzlich die Fälle, meistens entspricht eine Zeile einer Person. Die Spalten entsprechen den Variablen, eine Spalte ist immer eine Variable. Unser Beispiel enthält nur eine Variable. Die Dateneingabe beginnt mit der Variablendefinition.

Öffnen Sie durch Anklicken von *Variable View* das Variablendefinitionsfenster. Sie erhalten den folgenden Bildschirm:

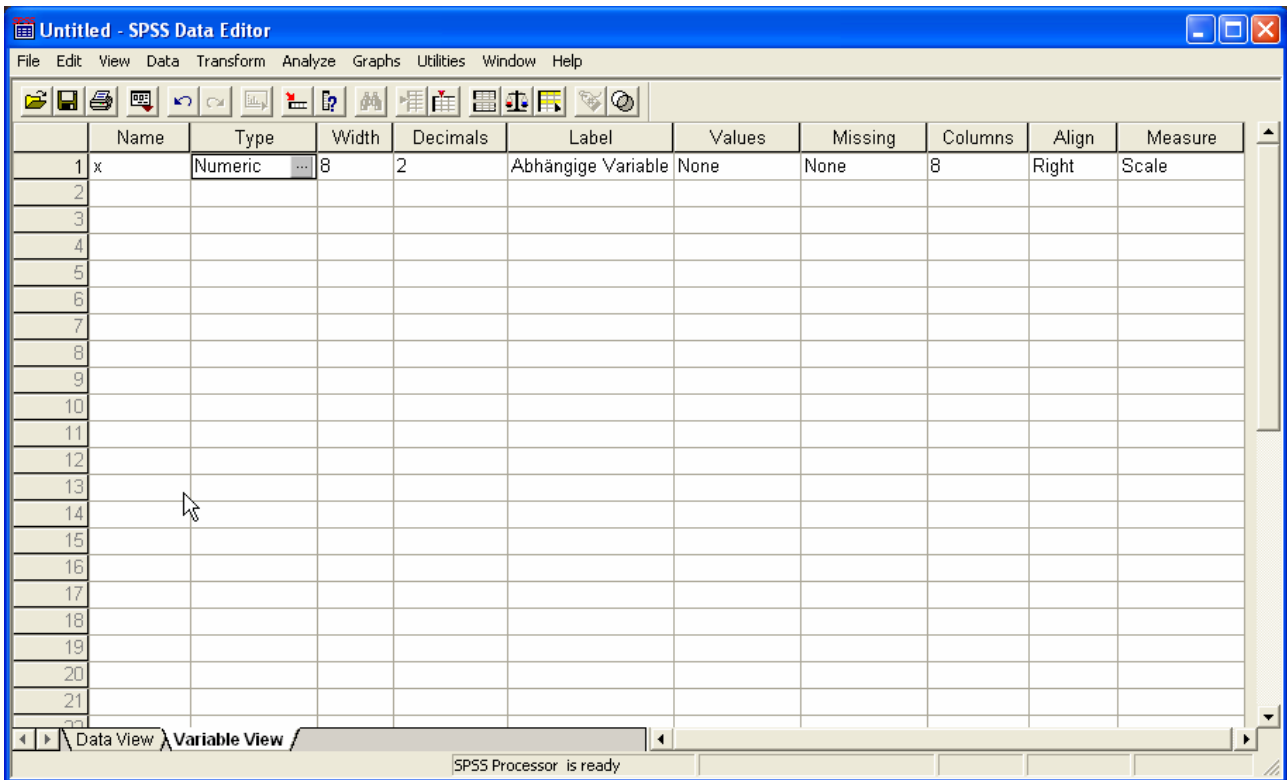


Tragen Sie unter *Name* den Variablennamen *x* ein und klicken Sie dann in der gleichen Zeile in das Feld für *Type*. Sie erhalten das folgende Typendefinitionsfenster:



Sie können hier verschiedene Variablentypen wählen. Die meisten Variablen, so auch unser *x*, sind vom Typ *Numeric*. In diesem Fall brauchen Sie nichts zu ändern. Studieren Sie einmal die Alternativen, die Sie hier haben. Besonders interessant ist auch der Typ *Date*, der für Kalenderdaten und Uhrzeiten vorgesehen ist. Die damit verbundenen Zeit- und Datumsfunktionen des SPSS sind für die Praxis sehr nützlich. Wir werden sie später im einzelnen kennenlernen. Schließen Sie dieses Fenster mit *OK*.

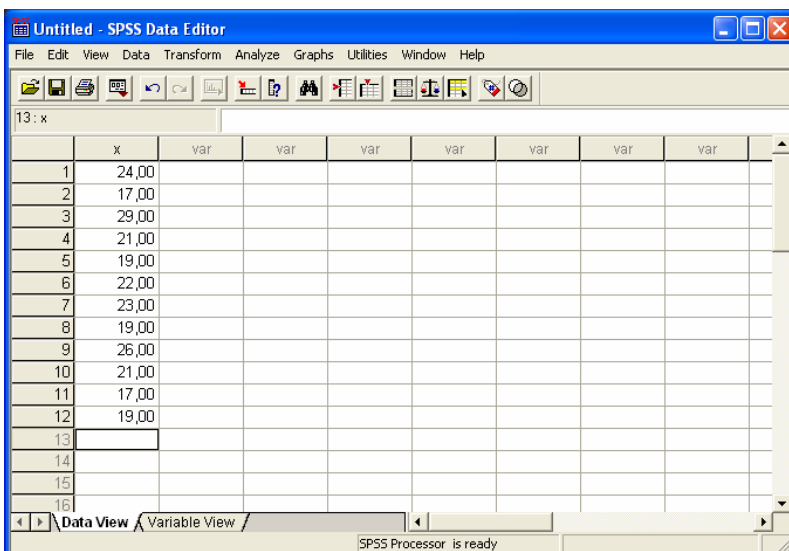
Damit kehren Sie zum Variablendefinitionsfenster zurück:



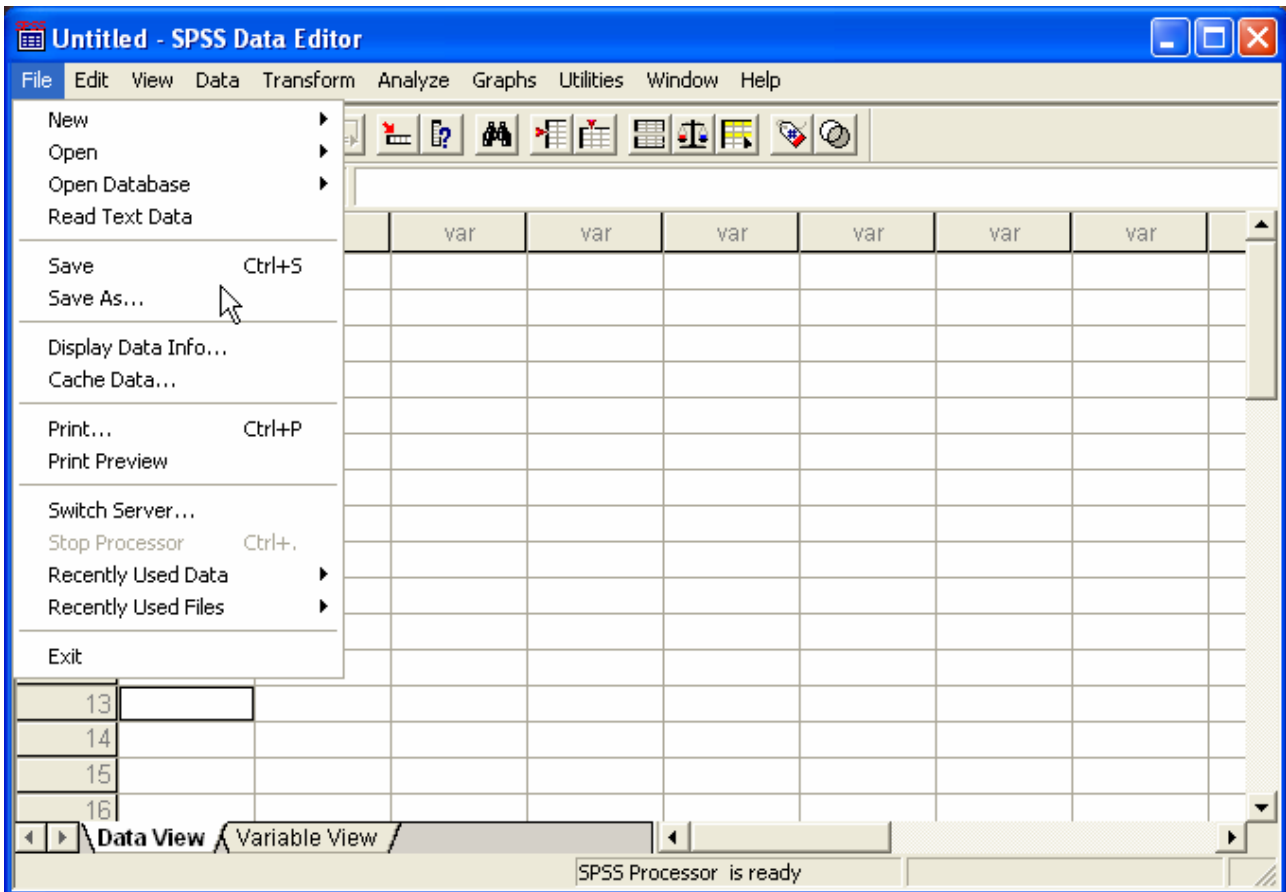
Sie stellen fest, daß die erste Zeile von SPSS mit Voreinstellungen gefüllt wurde. Nur das Feld *Label* ist noch frei. Tragen Sie hier *Abhängige Variable* ein, da wir unsere Variable *x* mit diesem Bezeichner versehen wollen. Die Labels unter *Value* benutzen wir in diesem Beispiel nicht. Im Feld *Missing* können wir Zahlen definieren, die fehlende Werte kennzeichnen, und in *Columns* können wir die Zahl der Spalten festlegen, mit denen Zahlenwerte angezeigt werden. Wir lassen es bei den Voreinstellungen. Schließlich können Sie unter *Measure* das Skalenniveau eingeben. Wir bleiben bei der Voreinstellung *Scale*, die der Intervall- und Verhältnisskala entspricht.

Schließen Sie das Variablendefinitionsfenster mit *DataView*. Sie kehren damit zum Dateneingabefenster zurück. Die Kopfzeile trägt jetzt für die erste Variable den Namen *x*.

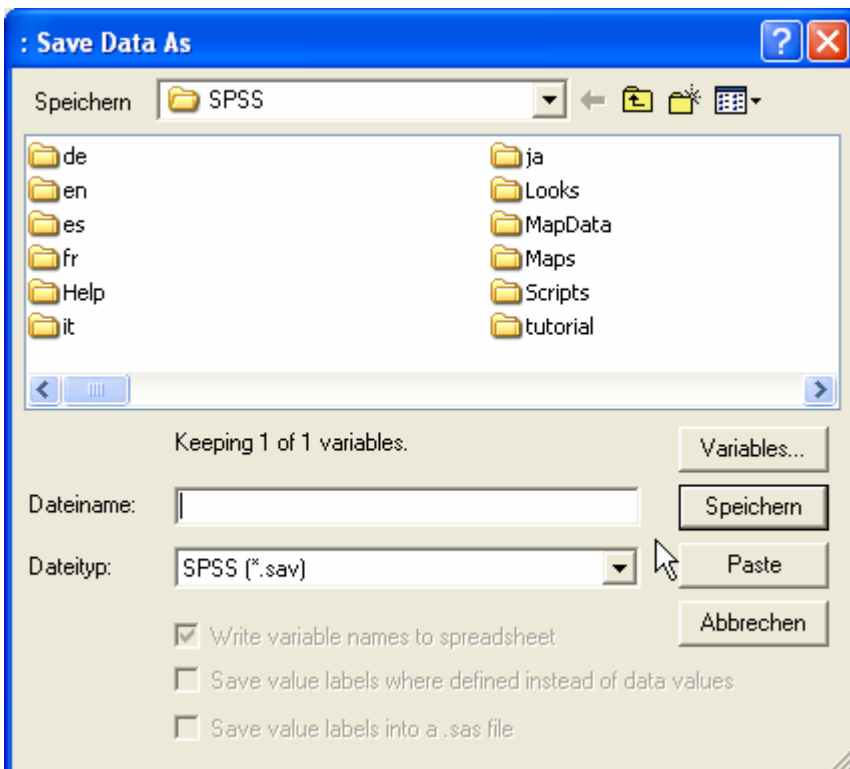
Nunmehr können Sie die Urliste eingeben. Aktivieren Sie durch Anklicken die Zelle in der ersten Zeile zu *x* und tippen Sie die erste Zahl, 24, ein. Schließen Sie die Eingabe mit der Eingabetaste ab. Die Zahl 24,00 erscheint in der ersten Zelle der ersten Spalte und die zweite Zelle wird aktiv. Geben Sie die zweite Zahl, 17, ein und wiederholen Sie die Eingaben, bis zur 69. Zahl, der 14. Nach der Eingabe der 12. Zahl erhalten Sie den folgenden Bildschirm:



Nach der letzten Eingabe müssen Sie Ihre Rohdatei speichern. Klicken Sie dazu zunächst *File* an:



Nach dem Anklicken von *Save As ...* erhalten Sie einen Bildschirm ähnlich wie diesen:



Im Auswahlfenster hinter *Speichern*, wo im Beispiel *SPSS* steht, müssen Sie jetzt Ihr Arbeitsverzeichnis wählen. Hinter *Dateiname* müssen Sie danach den Namen Ihrer Datei, im

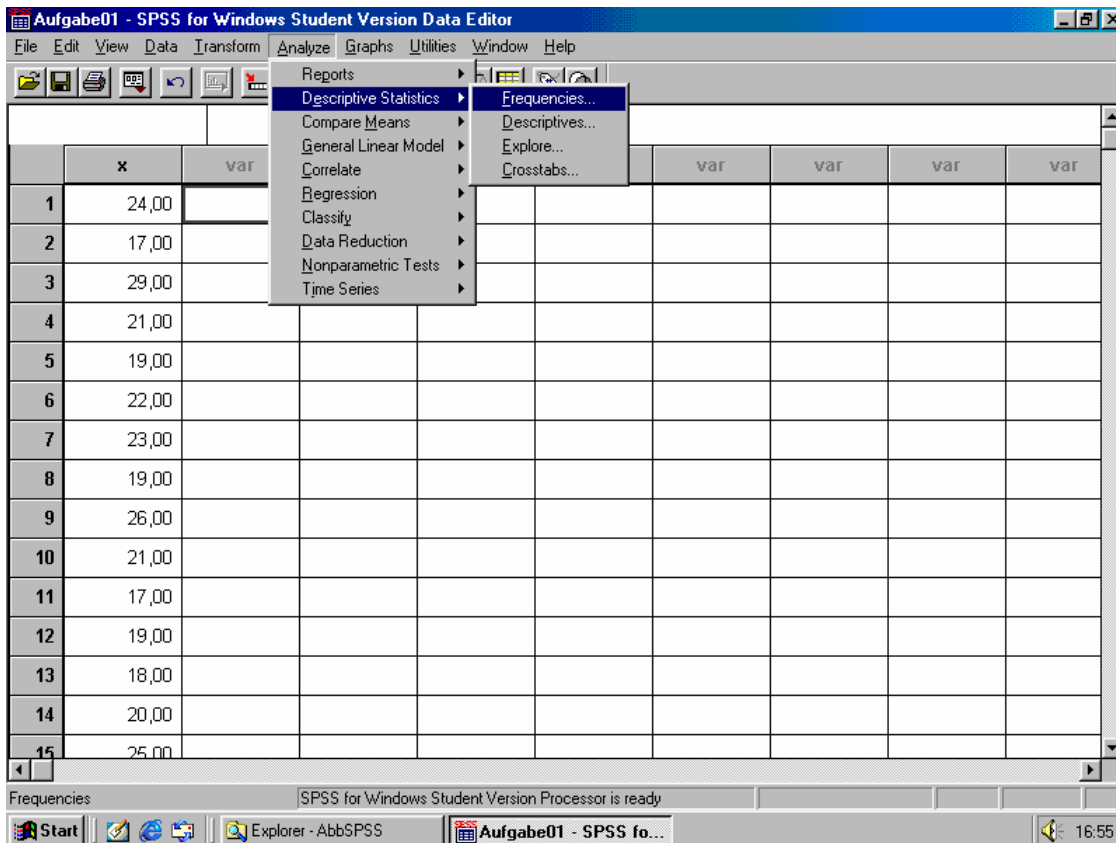
jetzigen Beispiel *Aufgabe01*, eintragen (ohne das Komma). SPSS ergänzt den Dateinamen zu *Aufgabe01.sav*. Verlassen Sie dieses Fenster mit *Speichern*.

Damit kehren Sie wieder zum Dateneingabefenster zurück und werden feststellen, daß ganz am oberen Rand Ihres Fensters statt *Untitled* der Name der Datei, *Aufgabe01*, steht.

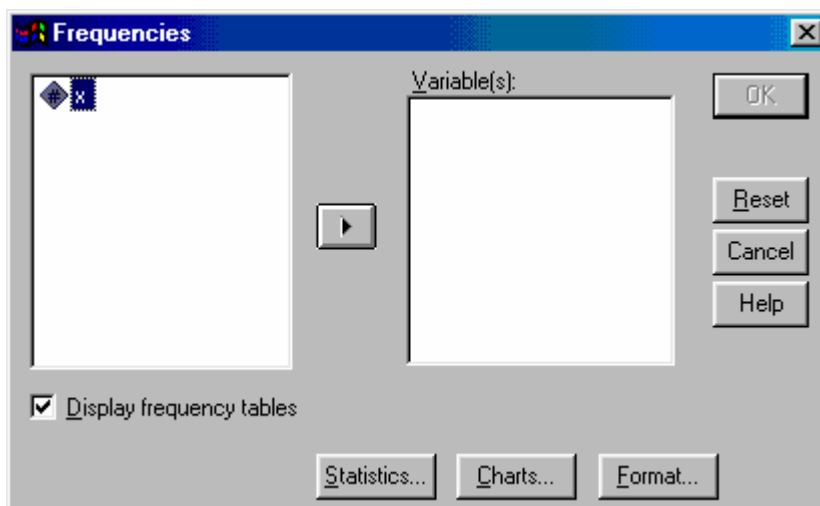
Sie können nun noch, was dieses Beispiel aber nicht verlangt, weitere Variablen und Daten in der gleichen Weise, wie gerade gezeigt, eingeben. Am Ende dieser Eingaben müssen Sie *File* → *Save* anklicken, um den aktuellen Stand Ihrer Eingaben in Ihrer Datei abzuspeichern.

Sie haben jetzt alle Vorarbeiten geleistet, um die geforderten statistischen Aufgaben mit SPSS lösen zu können.

b, c) Diese beiden Teilaufgaben lösen Sie in einem Arbeitsgang, indem Sie im Dateneingabefenster



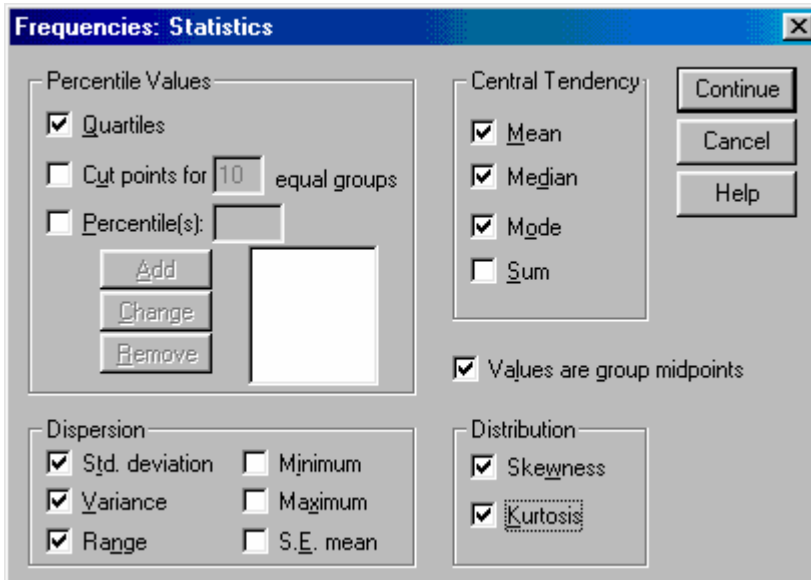
*Analyze* → *Descriptive Statistics* → *Frequencies* anklicken. Sie erhalten den Bildschirm:





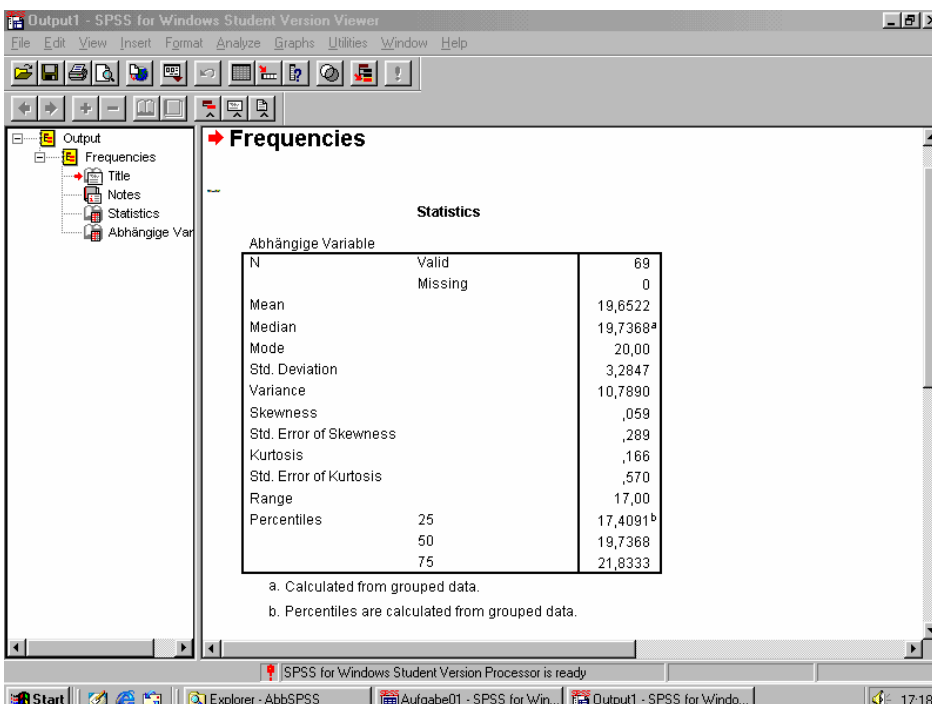
Zunächst müssen Sie im linken Variablenfeld  $x$ , die Variable, die Sie auswerten wollen, durch Anklicken aktivieren, was hier schon geschehen ist. Danach müssen Sie sie durch Anklicken des Pfeils in der Fenstermitte in das rechte Variablenfeld bringen. Das Schaltfeld *Display frequency tables* müssen Sie aktivieren, falls es nicht schon aktiv ist.

Danach klicken Sie das Feld *Statistics* an. Sie erhalten das Auswahlfenster für die Statistiken, die Sie berechnen wollen:



Klicken Sie der Aufgabe gemäß Mittelwert, Median, Modus, Standardabweichung, Varianz, absoluten Streubereich, Schiefe und Exzeß, schließlich noch Quartile, an. Die englischen Bezeichnungen dürften hier kein Problem darstellen. Sehr wichtig ist, daß Sie das Schaltfeld *Values are group midpoints* aktivieren, da nur so SPSS den Median und die Quartile nach den Gleichungen Glaser (2), (3) und (4) rechnet. Verlassen Sie das Fenster mit *Continue*. Damit kehren Sie zum Fenster *Frequencies* zurück. Hier sehen Sie unten noch die Schaltknöpfe *Charts* und *Format*. Diese lassen Sie jetzt aus und starten die Berechnungen mit *OK*.

SPSS meldet sich jetzt mit einem neuen Bildschirm, der Ausgabedatei.



Wie Sie sehen, sind in der Tabelle *Statistics* alle gewünschten Berechnungen enthalten. Am oberen linken Rand erscheint der Bezeichner unserer Variablen, *Abhängige Variable*. Bei der Standardabweichung und der Varianz rechnet SPSS mit den Gleichungen (3.2) und (3.3), also mit  $n - 1$  im Nenner. Schiefe und Exzeß werden mit den Gleichungen (1.30) und (1.31) berechnet. Die Dispersionsmaße *Interquartilbereich* und *mittlerer Quartilabstand* liefert SPSS nicht, auf der Basis der Quartile lassen sie sich jedoch sehr leicht „von Hand“ berechnen.

Weiter unten werden schließlich in einer zweiten Tabelle die Häufigkeitsverteilung, die prozentuale Häufigkeitsverteilung und die kumulierte prozentuale Häufigkeitsverteilung dargestellt.

Die prozentualen relativen Häufigkeiten sind in zwei Spalten wiedergegeben, *Percent* und *Valid Percent*, die sich in unserem Beispiel nicht unterscheiden. Das liegt daran, daß hier, anders als meistens sonst im Leben, keine Daten fehlen (vgl. oben *missing data* bei der Variablendefinition). Fälle ohne Datum werden in der Spalte *Percent* mitgerechnet, bleiben aber in der Spalte *Valid Percent* unberücksichtigt.

The screenshot shows the SPSS Output Viewer window. The main area displays a table titled "Abhängige Variable" with the following data:

|       | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----------|---------|---------------|--------------------|
| Valid | 12,00     | 1,4     | 1,4           | 1,4                |
|       | 13,00     | 1,4     | 1,4           | 2,9                |
|       | 14,00     | 3       | 4,3           | 7,2                |
|       | 15,00     | 2       | 2,9           | 10,1               |
|       | 16,00     | 5       | 7,2           | 17,4               |
|       | 17,00     | 6       | 8,7           | 26,1               |
|       | 18,00     | 5       | 7,2           | 33,3               |
|       | 19,00     | 9       | 13,0          | 46,4               |
|       | 20,00     | 10      | 14,5          | 60,9               |
|       | 21,00     | 7       | 10,1          | 71,0               |
|       | 22,00     | 8       | 11,6          | 82,6               |
|       | 23,00     | 5       | 7,2           | 89,9               |
|       | 24,00     | 2       | 2,9           | 92,8               |
|       | 25,00     | 3       | 4,3           | 97,1               |
|       | 26,00     | 1       | 1,4           | 98,6               |
|       | 29,00     | 1       | 1,4           | 100,0              |
| Total | 69        | 100,0   | 100,0         |                    |

The left sidebar shows the output structure: Output > Frequencies > Title, Notes, Statistics, and Abhängige Var. A red arrow points to the 'Abhängige Var' folder. The taskbar at the bottom shows the Start button and several open applications: Explorer - AbbSPSS, Aufgabe01 - SPSS for Win..., and Output1 - SPSS for ...

Damit Sie auch bei längeren Auswertungen die Übersicht behalten, enthält die linke Seite des Bildschirms die Gliederung der Ausgabedatei. Diese Datei läßt sich ebenfalls, wie die Urdatei, abspeichern. Wir wählen wieder den Namen *Aufgabe01*, SPSS ergänzt jetzt das Suffix *.spo*, so daß die Datei *Aufgabe01.spo* entsteht. Klicken Sie *File* → *Save as ...* an und geben Sie in das erscheinende Fenster Ihr Arbeitsverzeichnis und den Dateinamen *Aufgabe01* ein. Der Vorteil dieser Speicherung ist es, daß Sie bei einem neuen Start von SPSS die Ergebnisdatei wieder laden und weiter verwenden können, ohne die Berechnungen erneut laufen lassen zu müssen.

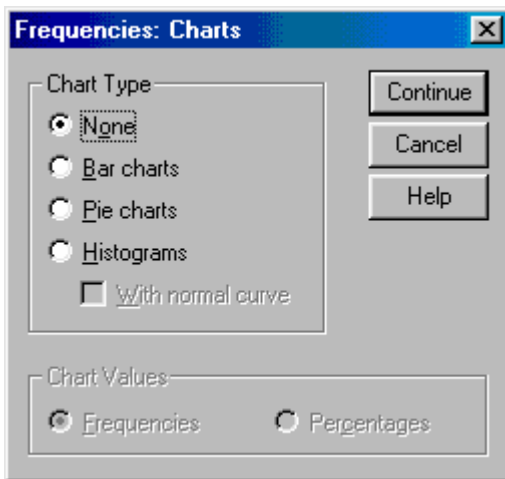
Bei näherem Hinsehen werden Sie einen „Schönheitsfehler“ der von SPSS erzeugten Häufigkeitsverteilungen finden. In unserem Beispiel haben die Maßzahlklassen  $x = 27$  und  $x = 28$  die Häufigkeit 0. SPSS läßt solche Klassen ganz aus, so daß die entsprechenden Lücken in der Tabelle nicht vorkommen.

Die Tabellen in der Ausgabedatei von SPSS lassen sich durch Anklicken aktivieren. Sie zeigen dann einen zweiten Rahmen und auf der linken Seite einen roten Pfeil. Sie lassen sich jetzt durch *Edit* → *Copy* in die Zwischenablage kopieren, von wo sie in andere Programme durch *Edit* → *Paste*

bzw. *Bearbeiten* → *Einfügen* übernommen werden können. Das ist wichtig, wenn man die Ergebnisse einer SPSS-Auswertung in ein Textdokument aufnehmen oder mit einer Tabellenkalkulation wie EXCEL weiterbearbeiten möchte.

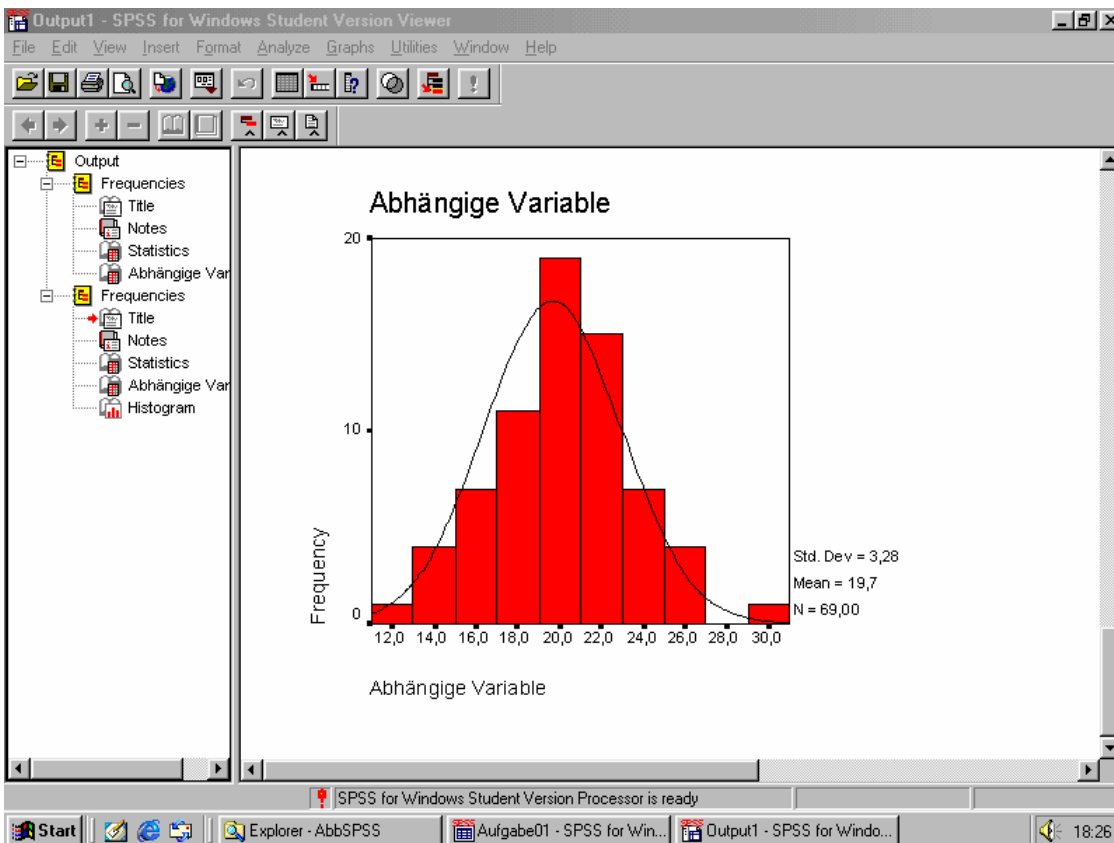
Mit dem Anklicken eines der beiden Symbole für die Rohdatei (schwarze Pfeile in der letzten Abbildung) kehren wir zur Rohdatei zurück.

- d) Mit *Analyze* → *Descriptive Statistics* → *Frequencies* holen wir uns erneut den Auswertungsbildschirm. Die gewünschte Graphik wählen wir mit *Charts* aus. Wir erhalten den Bildschirm:



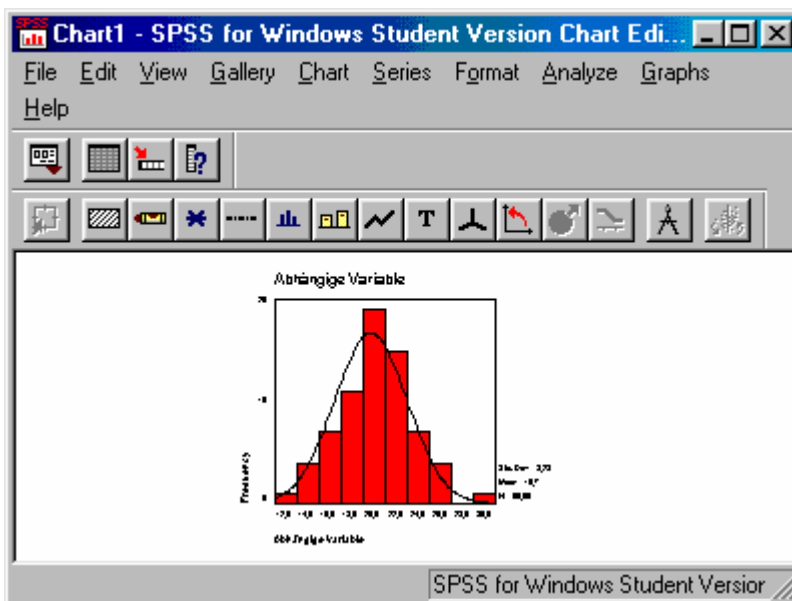
*Pie charts* bedeutet dabei *Kreisdiagramm*, was hier nicht in Frage kommt. SPSS unterscheidet dann noch zwischen *Bar charts* und *Histograms*, zu übersetzen mit *Balkendiagramm* und *Histogramm*. Nach dem Sprachgebrauch in der Vorlesung handelt es sich in beiden Fällen um Histogramme; der Unterschied liegt darin, daß bei Nominal- und Ordinalskalenniveau von  $x$  das *Balkendiagramm*, bei Intervall- und Verhältnisskalenniveau das *Histogramm* zu wählen ist. Probieren Sie beides aus und machen Sie sich den Unterschied dessen, was SPSS zeichnet, klar! Wir wählen *Histograms*. Dabei erhalten wir noch die Wahl, ob eine Normalverteilung mit Mittelwert und Standardabweichung der empirischen Verteilung in die Zeichnung aufgenommen werden soll. Wir wählen diese Option und verlassen den Bildschirm mit *Continue*. Die Auswertung starten wir nun mit *OK*. Dabei werden unsere unter *Statistics* früher gewählten Berechnungen erneut ausgeführt. Wenn uns das stört, können wir das nach einem erneuten Anklicken von *Statistics* vor dem *OK* noch ändern.

SPSS meldet sich jetzt wieder mit dem Ausgabebildschirm. Für jede neu gestartete Auswertung wird das Ergebnis an die vorgehenden Ergebnisse angehängt, so daß die Auswertungsdatei unter Umständen recht lang werden kann. Sie läßt sich verkürzen, indem man nicht mehr benötigte Teile durch Anklicken aktiviert und mit *Edit* → *Delete* löscht. Das Ende unseres Ausgabebildschirms sieht jetzt so aus:

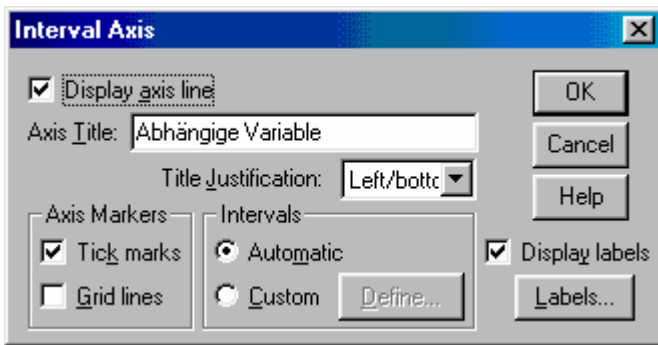


Wir haben ein Histogramm mit eingezeichneter Normalverteilung bekommen. Die Ordinate enthält die absoluten, nicht die prozentualen Häufigkeiten. SPSS hat eine Maßzahlklassenzusammenfassung gewählt, die unseren Vorstellungen eher nicht entspricht: Immer zwei alte Klassen wurden zu einer neuen zusammengefaßt, so daß sich jeweils eine gerade Zahl als Klassenmitte ergab. Wir können das abändern.

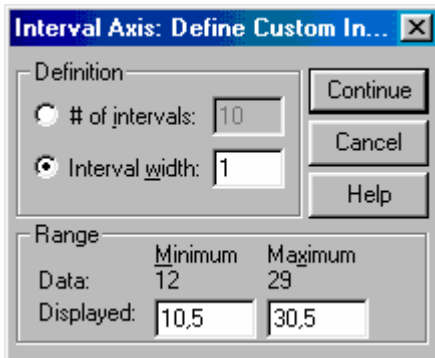
Durch Doppelklicken auf die Abbildung öffnet sich der Chart Editor:



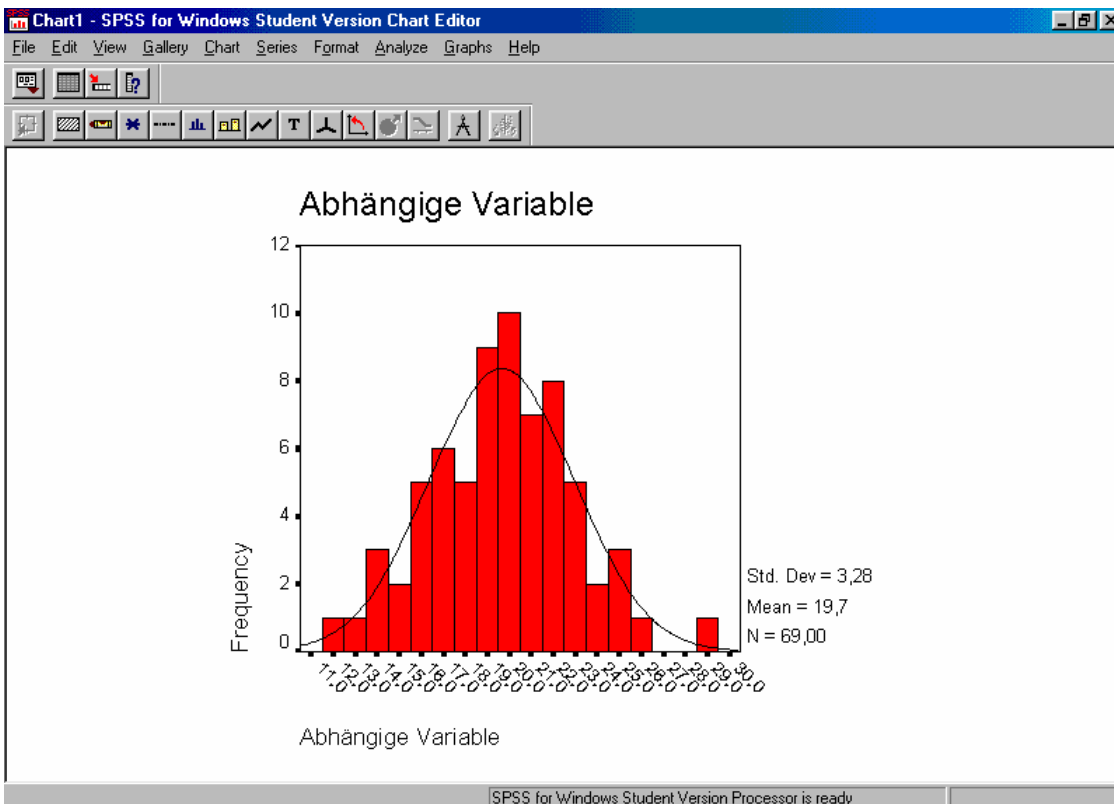
Sie haben nun eine Fülle von Möglichkeiten, die Graphik nach Ihren Wünschen zu ändern. Probieren Sie sie aus, spielen Sie mit der Abbildung! Nur die wichtigsten Optionen können in dieses Skriptum aufgenommen werden. Mit *Chart* → *Axis* → *Interval* → *OK* erhalten Sie den folgenden Bildschirm:



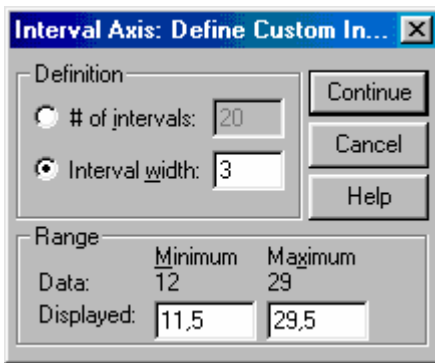
Mit *Custom* → *Define* öffnen Sie das Fenster:



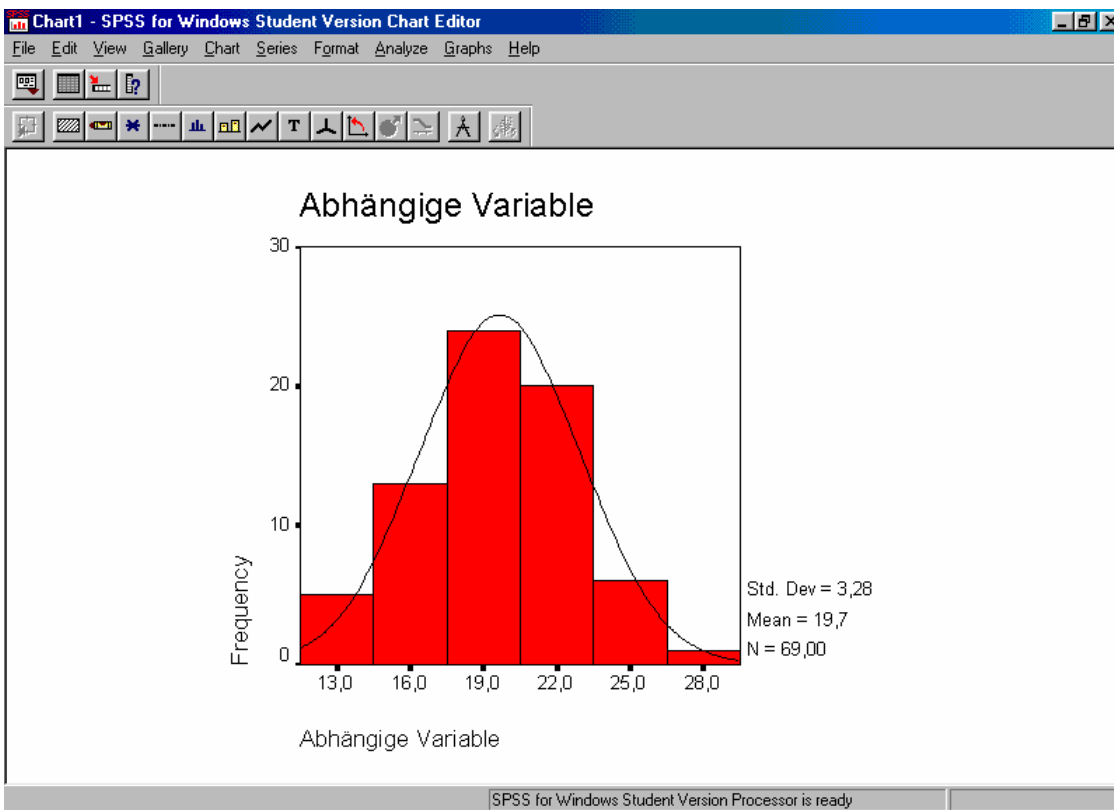
Mit der Wahl der Maßzahlklassenbreite (*Interval width*) 1 und der unteren und oberen Klassengrenze der Verteilung von 10,5 und 30,5 wählen Sie das gewünschte Histogramm, das Sie nach *Continue* → *OK* auch ansehen können:



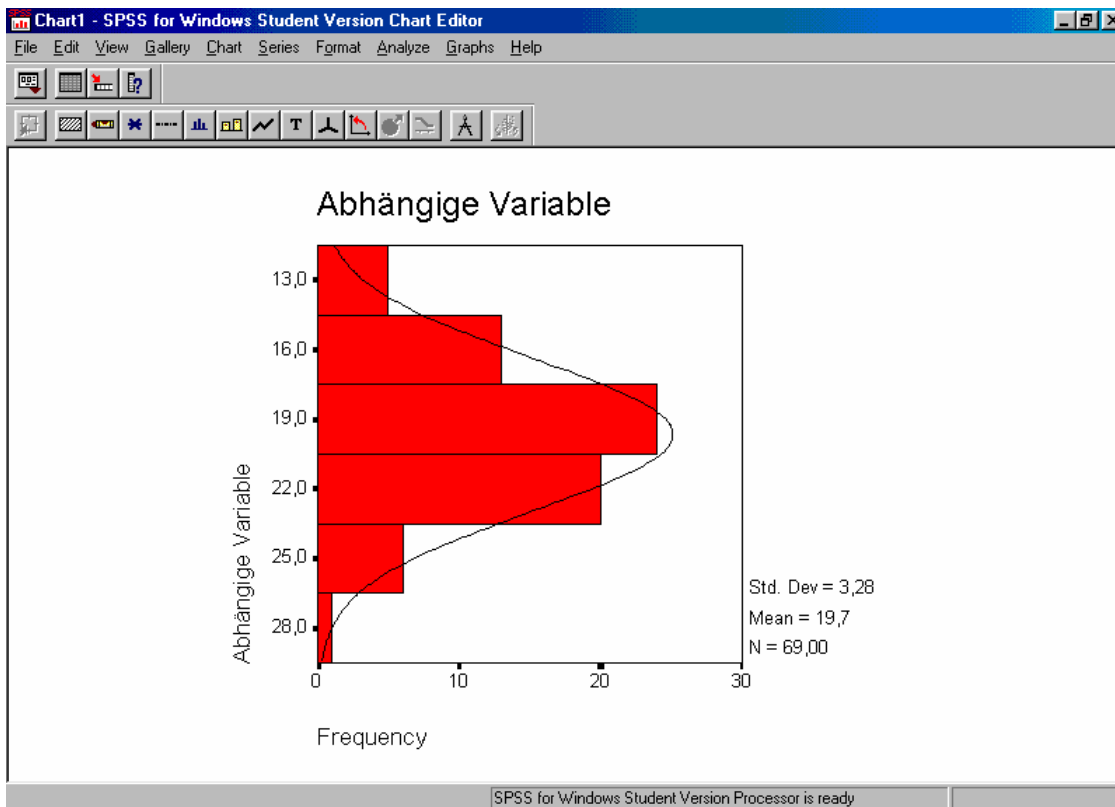
Die in der Vorlesung ebenfalls gezeigte Maßzahlklassenbreite 3 erhalten Sie auf genau die gleiche Weise: *Chart* → *Axis* → *Interval* → *OK* → *Custom* → *Define* ergibt den Bildschirm



mit dessen Einstellungen für *Interval width*, *Minimum* und *Maximum* nach Bild Sie die gewünschte Verteilung mit den Klassenmitten 13, 16, ... erhalten, wie sie das nächste Bild zeigt.

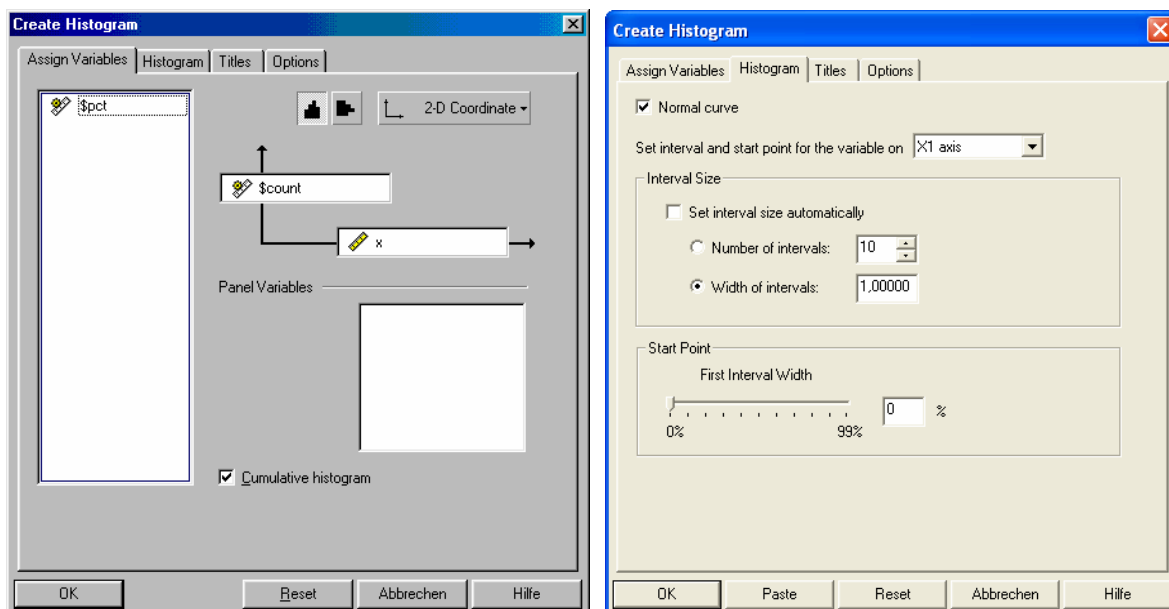


Nochmals: Probieren Sie die Fülle der Möglichkeiten, die Sie im Chart Editor haben, selbst aus. Denken Sie auch an Farbe und Füllmuster der Säulen sowie die vielen Möglichkeiten der Beschriftung. Hier soll nur noch eine Option gezeigt werden: *Format* → *Swap Axes*. Das Ergebnis ist eine Vertauschung der beiden Achsen, die in der Psychologie, vor allem bei nominalen Variablen mit langen Bezeichnungen, oft höchst wünschenswert ist:

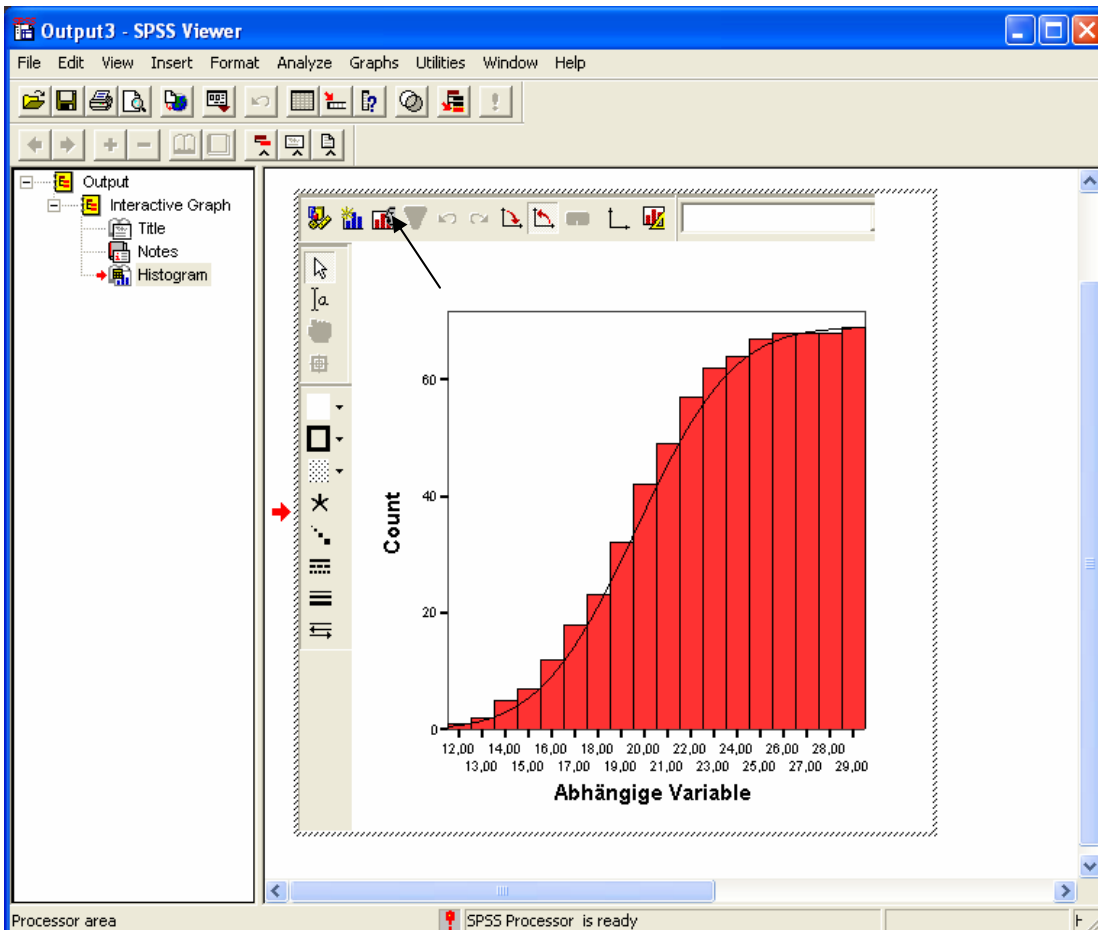


Wenn Ihre Abbildung Ihren Wünschen entspricht, schließen Sie den Chart Editor mit *File* → *Close* und kehren damit in die Ausgabedatei zurück.

Leider erlaubt der Chart Editor nicht, auch die kumulierte Häufigkeitsverteilung zu erzeugen. Dafür ist wie folgt vorzugehen. Im Datenbildschirm klicken Sie *Graphs* → *Interactive* → *Histogram* an. In dem jetzt erscheinenden Bildschirm



ziehen Sie *\$count* in die Ordinate, die Variable, im Beispiel *x*, in die Abszisse. Die kumulierte Verteilung fordern Sie durch das Schaltfeld *Cumulative histogram* an. Nach dem Anklicken von *Histogram* kann im rechts gezeigten Fenster die Klassenbreite oder die Anzahl der Klassen gewählt werden. Durch Probieren ist die Einstellung zu suchen, die zu einem brauchbaren Ergebnis führt. Doppelklicken auf die Abbildung in der Ausgabe führt dann zu folgendem Bildschirm:



Anklicken der mit dem schwarzen Pfeil markierten Stelle öffnet eine andere Version des Chart Editors, in dem sich mit *Scale Axis* → *Edit* der folgende Bildschirm holen läßt:

The screenshot shows the 'Scale Axis - Abhängige Variable' dialog box. The 'Scale' tab is active. The 'Scale' section includes 'Data Min/Max: 12,00 / 29,00', 'Minimum: 11,50', 'Maximum: 29,50', 'Tick Interval: 1,00', and 'Number of Ticks: 18'. There are also checkboxes for 'Auto', 'Ticks originate from zero', and 'Reverse scale'. The 'Tick Format' section has 'Floating' selected. The 'Display Secondary Axis' checkbox is unchecked. The 'OK' button is highlighted.

Hier müssen nun Einstellungen gesucht werden, die ein einwandfreies Histogramm der kumulierten Häufigkeitsverteilung liefern, nachdem man diesen Bildschirm mit *OK* wieder verlassen und den Chart Editor mit *X* geschlossen hat. Die hier eingetragenen Zahlen für *Minimum* und *Maximum* liefern das gewünschte Resultat.



## Aufgabe 02

Diese Aufgabe basiert auf der bivariaten Urliste, die den Erklärungen in der Vorlesung „Statistik I“ zugrundelag. Sie lautet (*stat130.doc*):

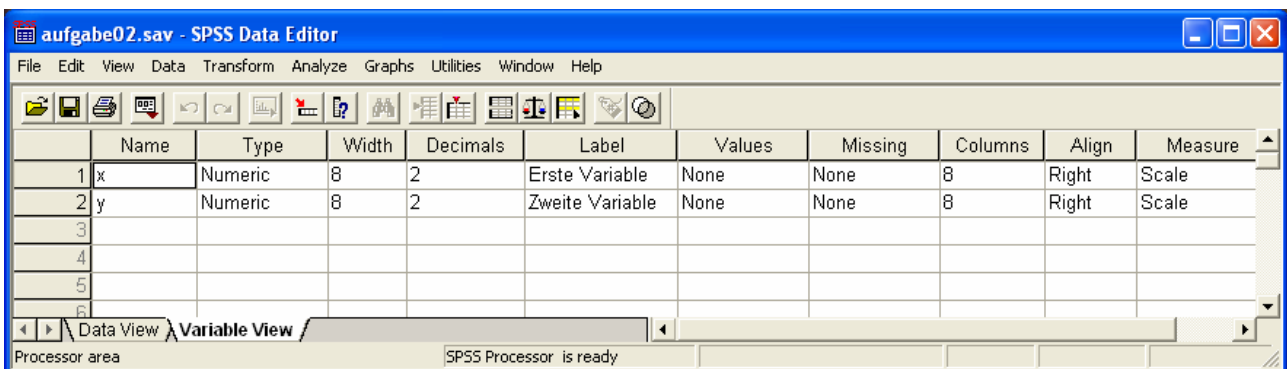
```
X 31 35 28 30 28 26 33 31 30 29 29 33 29 28 27 30 31 32 34 31
Y 18 25 16 17 17 14 23 21 18 17 19 21 18 16 15 19 19 20 24 19

X 27 30 30 31 32 31 29 30 30 33 32 30 29 30 30 33 31 33 29 29
Y 15 20 17 20 21 20 18 19 18 22 22 19 17 17 19 22 20 21 19 18
```

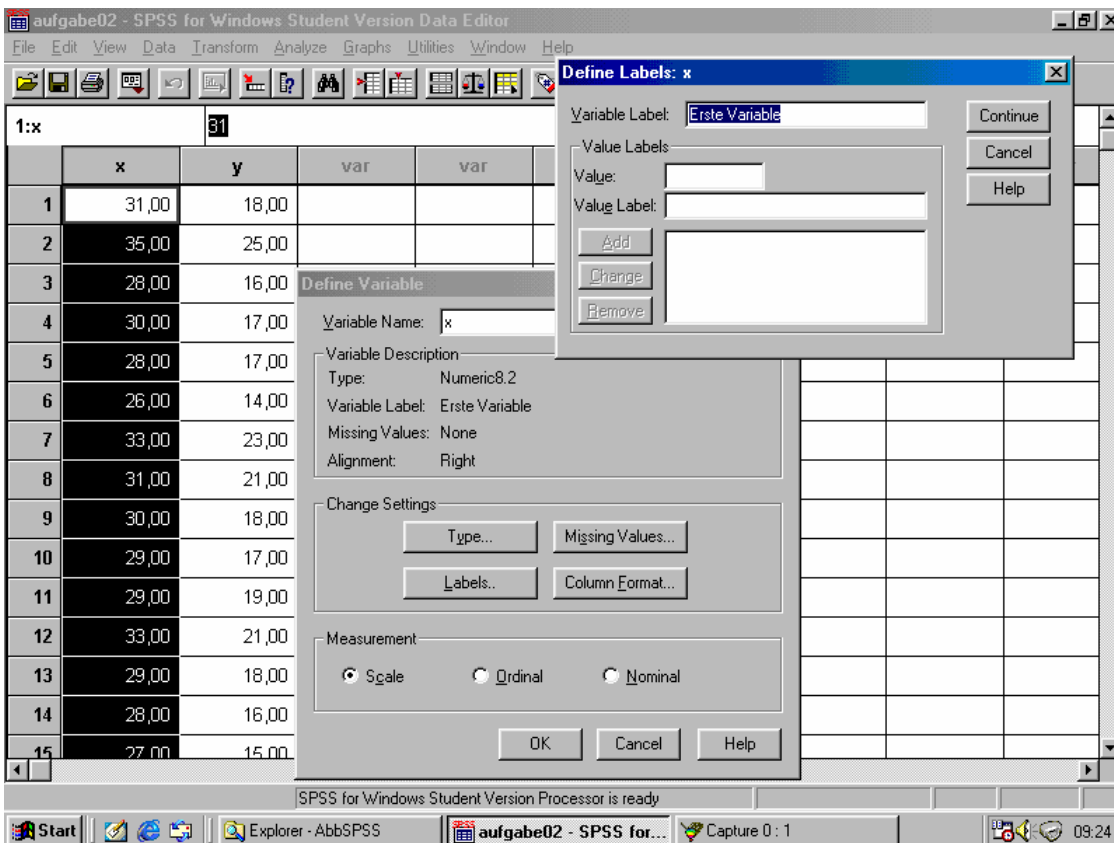
- Geben Sie diese Urliste in SPSS ein. Nennen Sie die abhängigen Variablen  $x$  und  $y$  und geben Sie ihnen die Bezeichnungen *Erste Variable* und *Zweite Variable*. Speichern Sie die Datenmatrix in *Aufgabe02.sav* ab.
- Erzeugen Sie für die univariaten Randverteilungen Häufigkeitsverteilung, prozentuale Häufigkeitsverteilung und kumulierte prozentuale Häufigkeitsverteilung als Tabelle.
- Berechnen Sie für die univariaten Randverteilungen Mittelwert, Median, Modus, Standardabweichung, Varianz, Schiefe, Exzeß, absoluten Streubereich, erstes und drittes Quartil.
- Stellen Sie die beiden univariaten Randverteilungen als Histogramme graphisch dar.
- Berechnen Sie die Produktmomentkorrelation und die Spearmansche Rangkorrelation.
- Berechnen Sie die Gleichungen für die Regression von  $y$  auf  $x$  und von  $x$  auf  $y$ .
- Stellen Sie die bivariate Häufigkeitsverteilung mit den Regressionsgeraden graphisch dar.

## Lösung

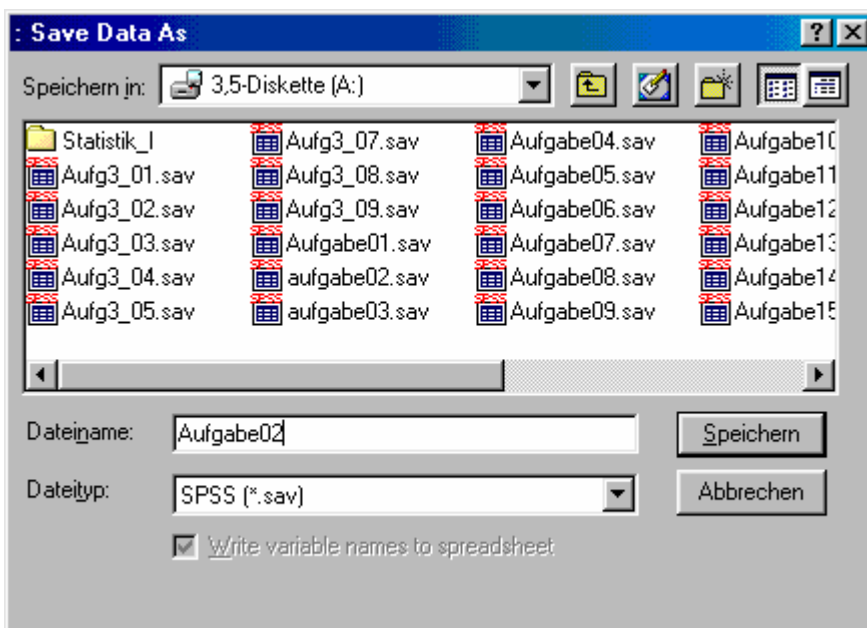
- Hier verfahren Sie, wie bei Aufgabe 01 genauer beschrieben. Die Fenster für die Definition von Variablennamen und Variablenbezeichnungen sehen für die Variable  $x$  und  $y$  wie folgt aus:



Die nächste Seite zeigt noch die Bildschirme zur Variablendefinition aus Version 9 von SPSS.

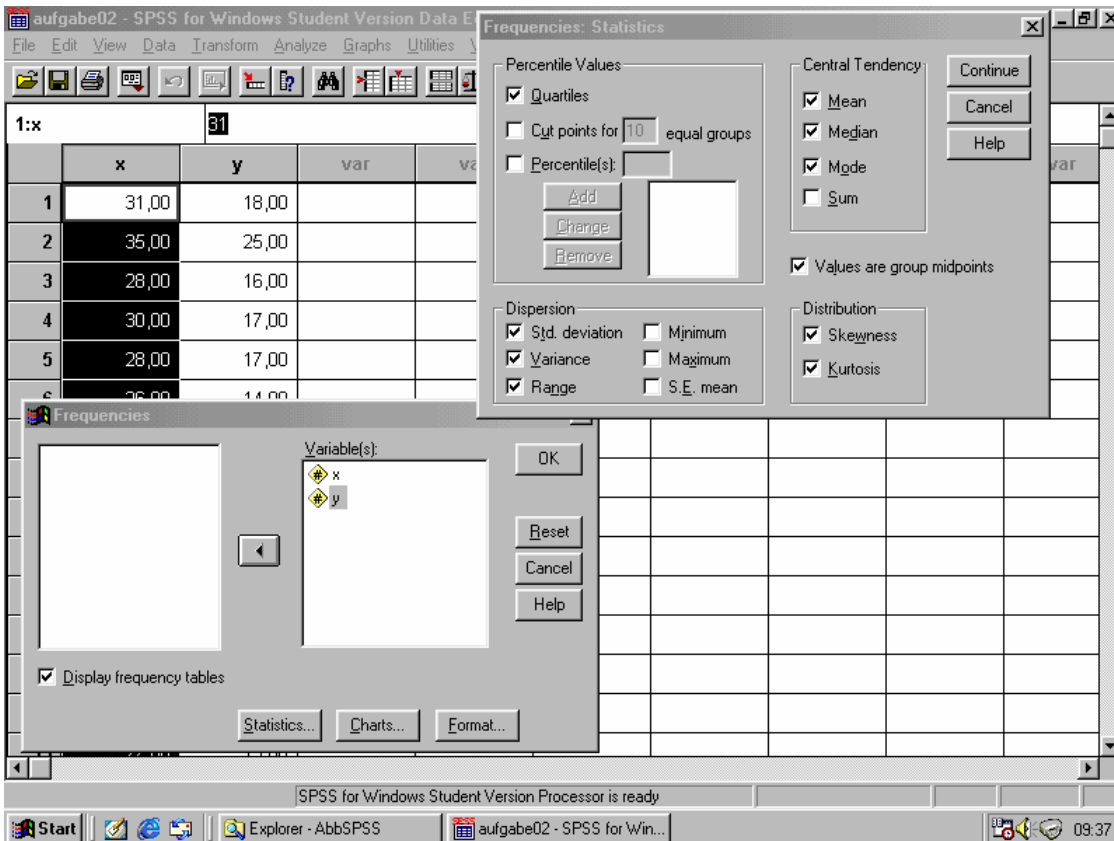


Bei der zweiten Variablen muß es dann entsprechend  $y$  und *Zweite Variable* heißen. Mit *File* → *Save As* erhalten Sie das Fenster,

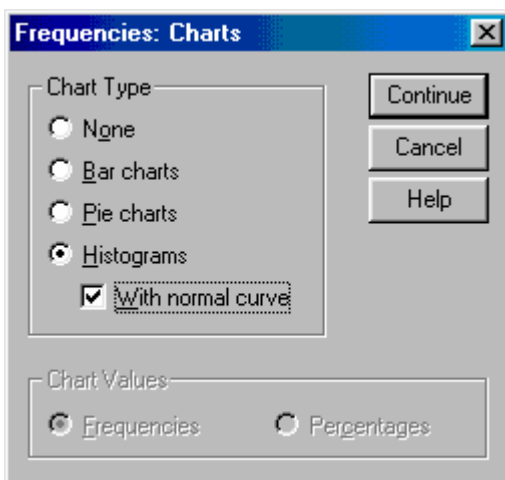


bei dem Sie unter *Speichern in* Ihr Arbeitsverzeichnis und unter *Dateiname* *Aufgabe02* eintragen. Sie verlassen es mit *Speichern*.

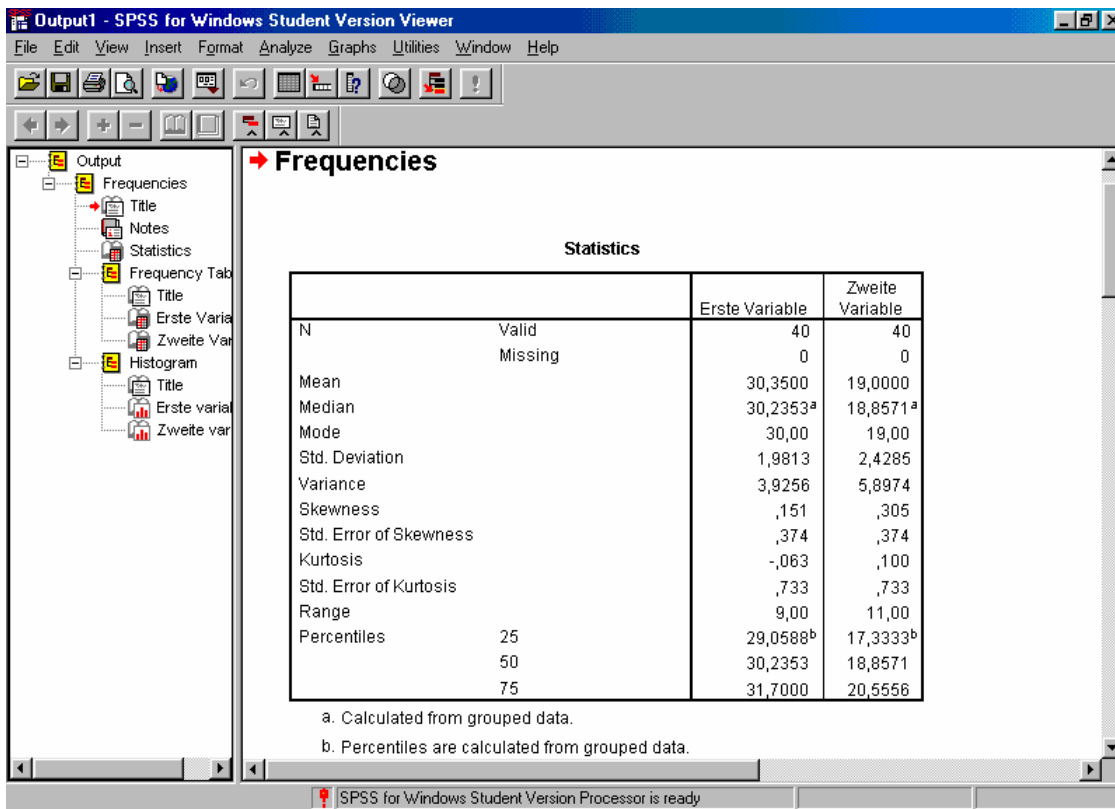
b, c, d) Diese drei Teilaufgaben erledigen wir in einem Gang. Mit *Analyse* → *Descriptive Statistics* → *Frequencies* holen wir uns das Auswertefenster, in dem wir jetzt beide Variablen,  $x$  und  $y$ , nach rechts unter *Variable(s)* bringen. Mit *Statistics* holen wir das Auswahlfenster für die nötigen Berechnungen, in dem wir die geforderten Maße anklicken. Wir dürfen nicht vergessen, *Values are group midpoints* zu aktivieren.



Mit *Continue* schließen wir das Auswahlfenster für die Berechnungen, mit *Charts* öffnen wir dann das Auswahlfenster für die Graphiken:



Hier wählen wir *Histograms* sowie *With normal curve* aus und verlassen das Fenster mit *Continue*. Die Auswertung starten wir mit *OK*. SPSS meldet sich mit dem Ausgabebildschirm. Die statistischen Berechnungen finden sich jetzt für beide Variablen in einer gemeinsamen Tabelle:



Die beiden univariaten Häufigkeitsverteilungen benötigen auch zwei einzelne Tabellen. Wir zeigen jetzt nicht das Schirmbild, sondern aktivieren die Tabellen in SPSS einzeln mit einem Mausklick und kopieren sie mit *Edit* → *Copy* in die Zwischenablage. In das vorliegende Word-Dokument holen wir sie mit *Bearbeiten* → *Einfügen*. Danach sahen sie in den früheren Versionen von SPSS und WORD meistens erst einmal unbrauchbar aus, was wir aber mit den Funktionen unter *Tabelle* in Ordnung bringen konnten. Bei den aktuellen Versionen – SPSS 11.5 und OFFICE XP – scheinen die früheren Probleme beim Kopieren von Tabellen über die Zwischenablage von SPSS in WORD behoben zu sein.

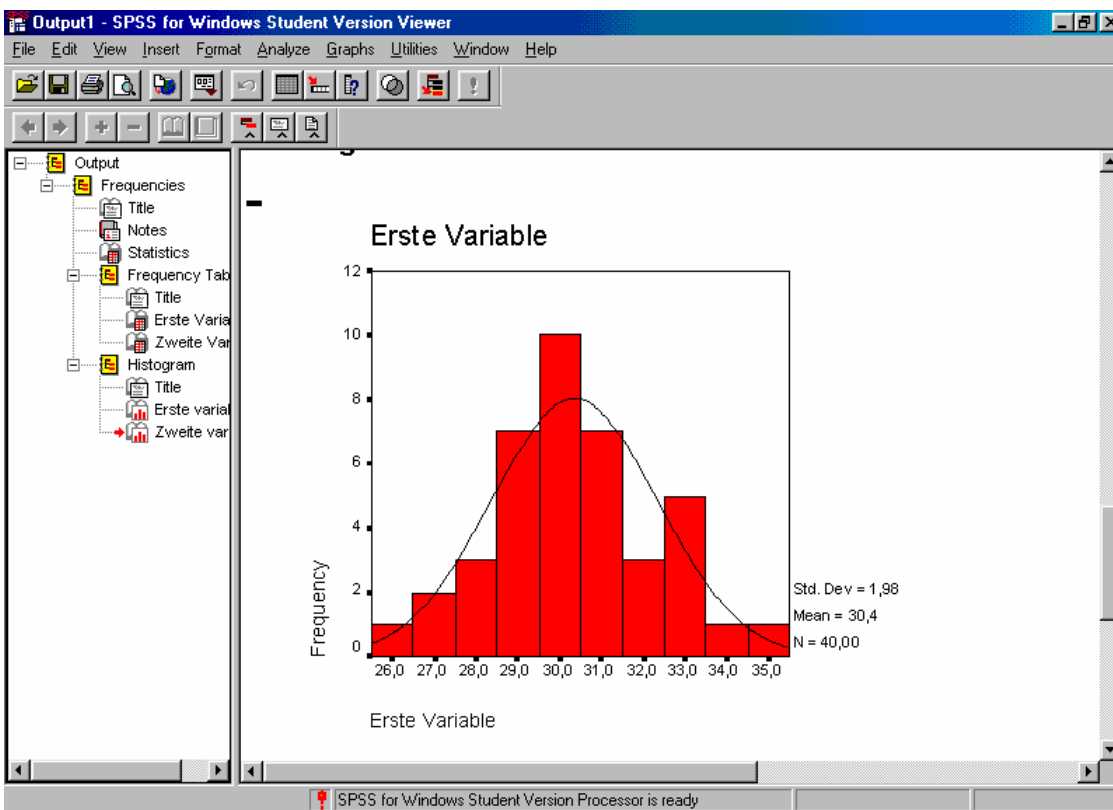
#### Erste Variable

|       |       | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | 26,00 | 1         | 2,5     | 2,5           | 2,5                |
|       | 27,00 | 2         | 5,0     | 5,0           | 7,5                |
|       | 28,00 | 3         | 7,5     | 7,5           | 15,0               |
|       | 29,00 | 7         | 17,5    | 17,5          | 32,5               |
|       | 30,00 | 10        | 25,0    | 25,0          | 57,5               |
|       | 31,00 | 7         | 17,5    | 17,5          | 75,0               |
|       | 32,00 | 3         | 7,5     | 7,5           | 82,5               |
|       | 33,00 | 5         | 12,5    | 12,5          | 95,0               |
|       | 34,00 | 1         | 2,5     | 2,5           | 97,5               |
|       | 35,00 | 1         | 2,5     | 2,5           | 100,0              |
| Total | 40    | 100,0     | 100,0   |               |                    |

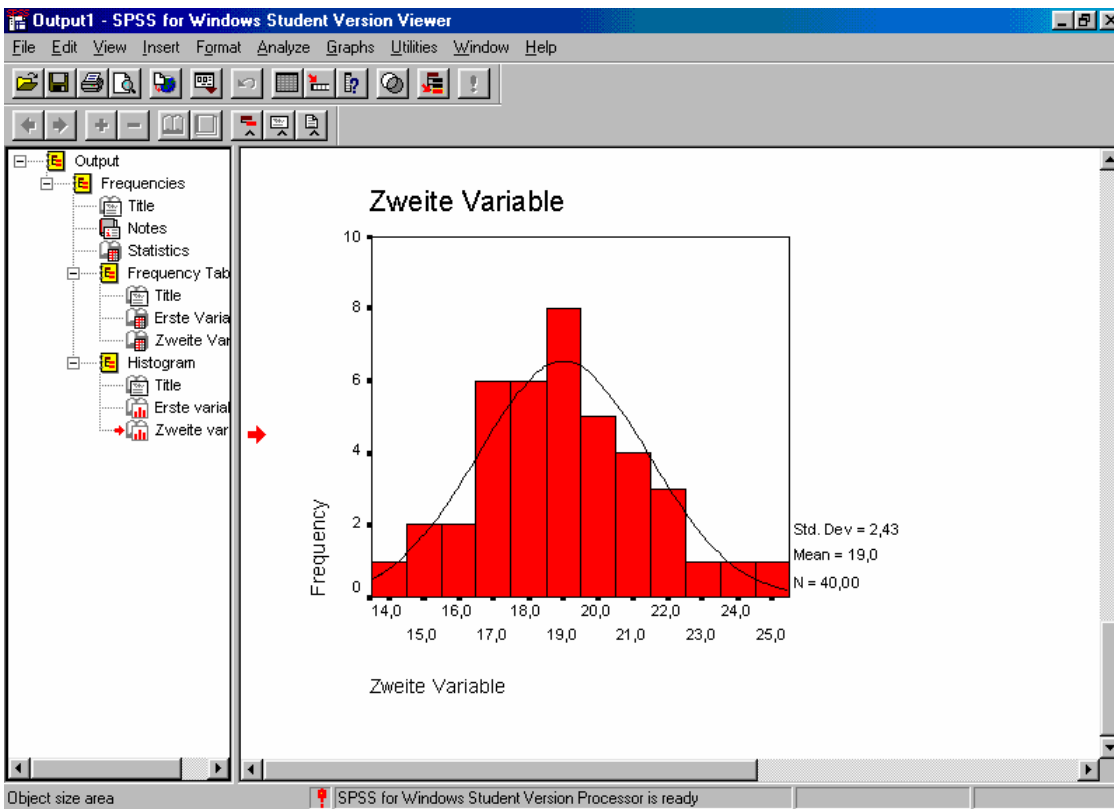
### Zweite Variable

|       |       | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | 14,00 | 1         | 2,5     | 2,5           | 2,5                |
|       | 15,00 | 2         | 5,0     | 5,0           | 7,5                |
|       | 16,00 | 2         | 5,0     | 5,0           | 12,5               |
|       | 17,00 | 6         | 15,0    | 15,0          | 27,5               |
|       | 18,00 | 6         | 15,0    | 15,0          | 42,5               |
|       | 19,00 | 8         | 20,0    | 20,0          | 62,5               |
|       | 20,00 | 5         | 12,5    | 12,5          | 75,0               |
|       | 21,00 | 4         | 10,0    | 10,0          | 85,0               |
|       | 22,00 | 3         | 7,5     | 7,5           | 92,5               |
|       | 23,00 | 1         | 2,5     | 2,5           | 95,0               |
|       | 24,00 | 1         | 2,5     | 2,5           | 97,5               |
|       | 25,00 | 1         | 2,5     | 2,5           | 100,0              |
| Total |       | 40        | 100,0   | 100,0         |                    |

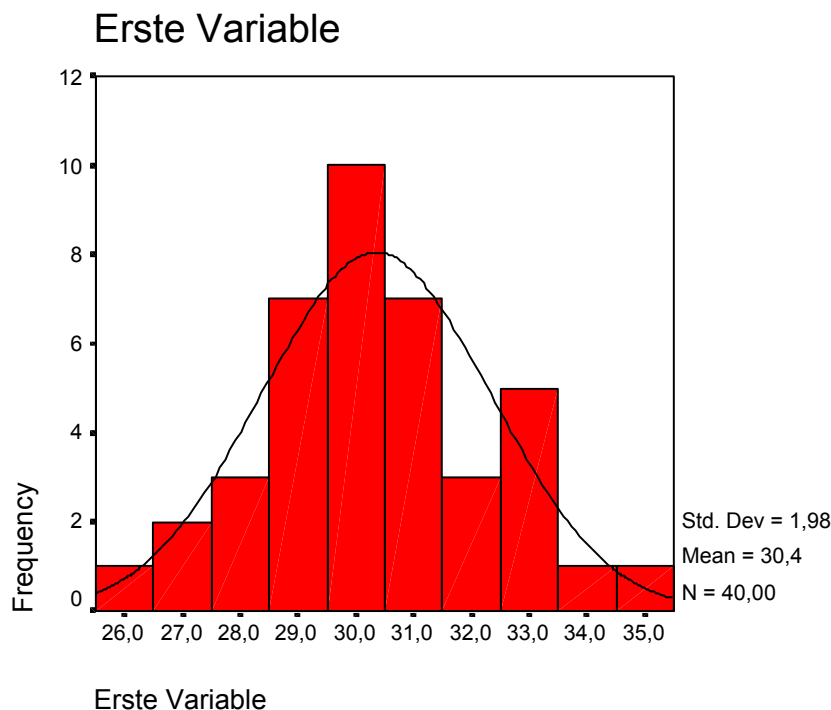
Unsere Histogramme sind in der Ausgabedatei von SPSS auch noch nicht brauchbar, wir müssen erst noch mit dem *Chart Editor*, wie in Aufgabe 01 gezeigt, die Maßzahlklassenbreite 1 herstellen. Nachdem das geschehen ist, erhalten wir:



und



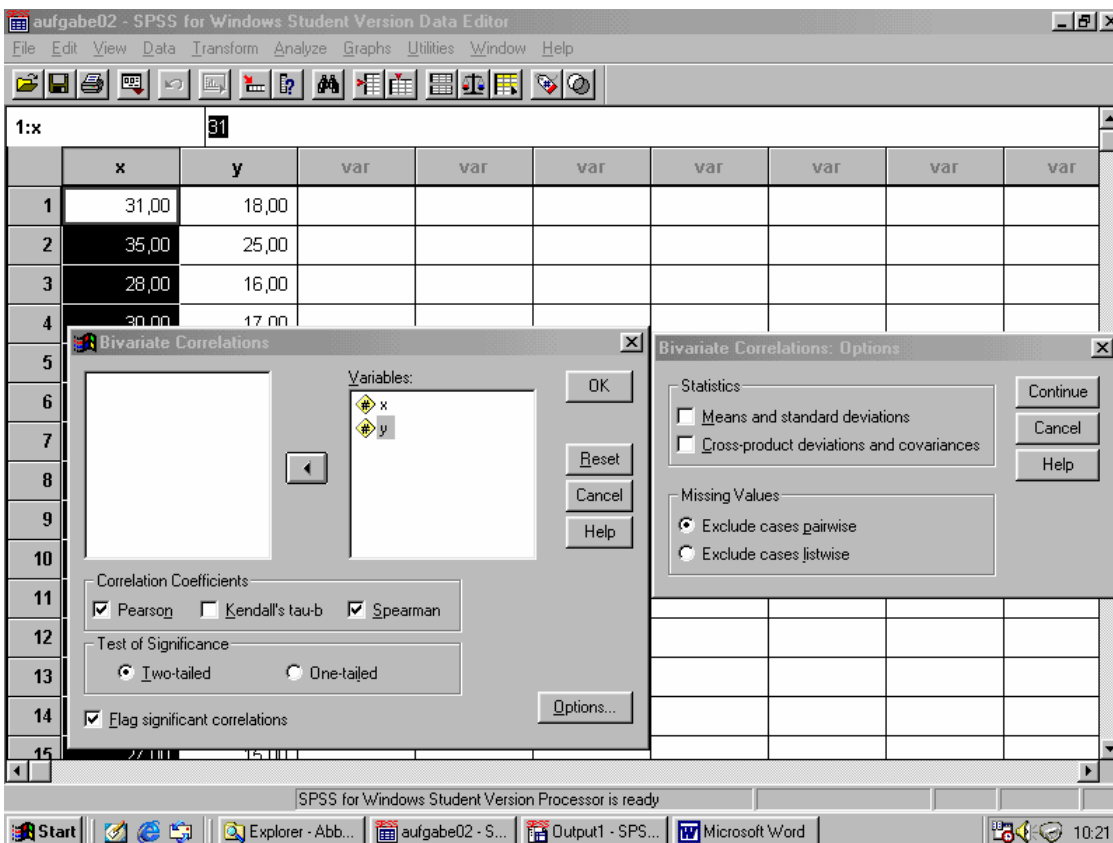
Wir können auch die Abbildung, nachdem sie auskorrigiert ist, mit *Edit* → *Copy* in die Zwischenablage bringen und sie mit *Bearbeiten* → *Einfügen* in das vorliegende Word-Dokument übertragen. Das Ergebnis sieht dann so aus:



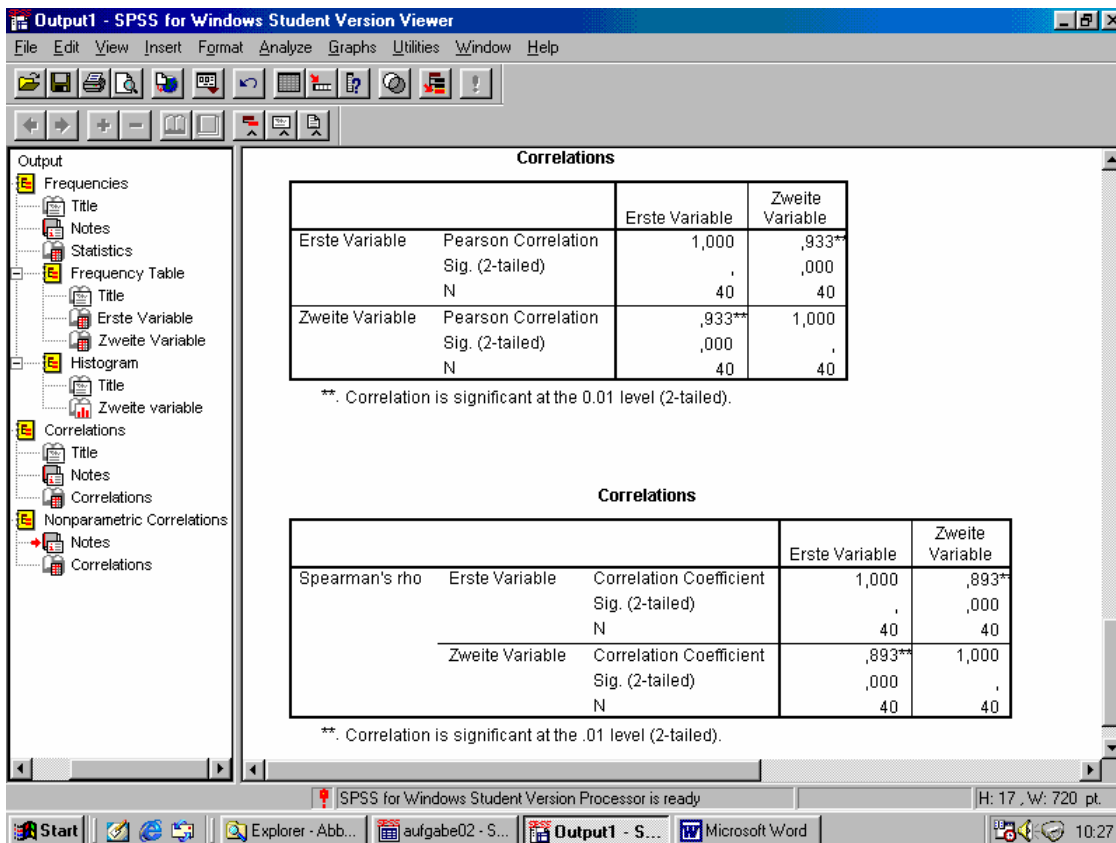
In älteren Versionen von SPSS (bis 8.0.0) und von WORD (vor WORD 97) haben diese Übertragungen aus dem Ausgabebildschirm von SPSS über die Zwischenablage in ein Word-Dokument oft nicht funktioniert und unvorhergesehene, unbrauchbare Resultate hervorgebracht. Das scheint sich wesentlich gebessert zu haben, aber stellen Sie sich zur Sicherheit dabei immer noch auf Überraschungen ein. Bei der Arbeit an diesem Skriptum habe ich gelegentlich noch den

folgenden Fehler festgestellt: Aus dem Ausgabebildschirm von SPSS in dieses Word-Dokument übernommene Tabellen erscheinen auf dem Bildschirm einwandfrei, zeigen im Druck aber die Verdoppelung einzelner Zeilen, allerdings ohne Verfälschung oder Veränderung der darin enthaltenen Daten. Dagegen half ein Trick, den zu kennen sich auch heute noch lohnt: Im Menü der Tabellenfunktionen wählt man *Tabelle umwandeln in Text* mit der Option *Trennzeichen Semikolon*. Den so erhaltenen Text ändert man nicht, markiert ihn oder läßt ihn markiert und wandelt ihn mit *Text umwandeln in Tabelle* in eine Tabelle zurück. Die Fehler sind beseitigt und man kann die Tabelle nach seinen Wünschen gestalten. Diese Fehler scheinen aber in den aktuellen Versionen von SPSS und WORD behoben worden zu sein.

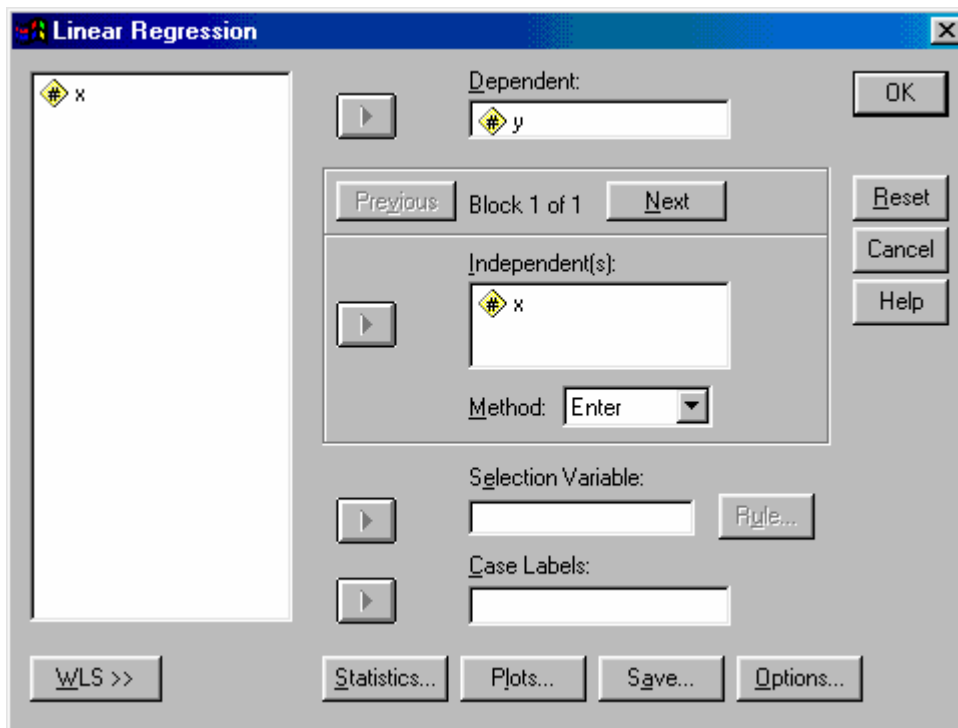
e) Für die Korrelationen holen wir uns aus dem Dateneingabefenster über *Analyze* → *Correlate* → *Bivariate* das entsprechende Auswahlfenster:



Die beiden Variablen  $x$  und  $y$  bringen wir mit dem Pfeil in der Mitte nach rechts, wie es im Bild schon geschehen ist, und darunter wählen wir die gewünschten Korrelationen aus. Mit *Optionen* können wir ein Optionenfenster öffnen, in dem wir weitere uni- und bivariate Statistiken anfordern und die Behandlung fehlender Daten steuern können. Wir verlassen das Optionenfenster mit *Continue* und das Korrelationsfenster mit *OK*. SPSS zeigt uns jetzt im Ausgabebildschirm die gewünschten Korrelationen zusammen mit den Ergebnissen der Signifikanztests, die wir erst im Laufe der Statistik II kennenlernen.

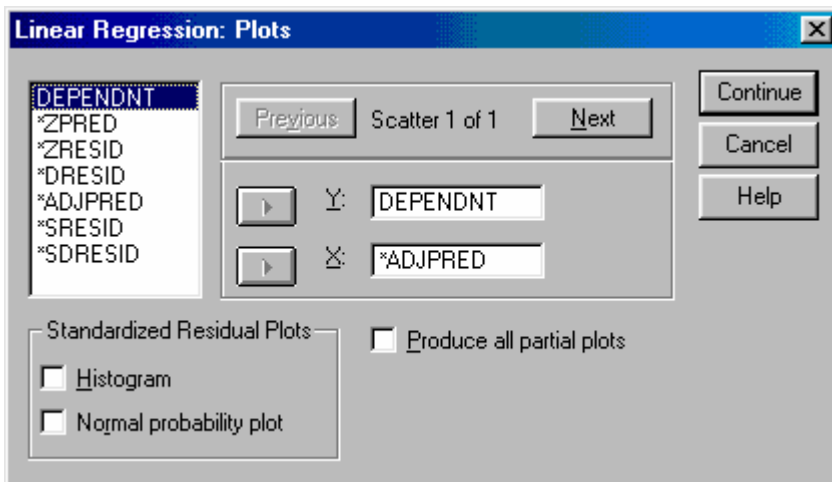


f) Für die Regressionsrechnung kehren wir zunächst in den Dateneingabebildschirm zurück. Mit *Analyse* → *Regression* → *Linear* öffnen wir das Regressionsfenster. Das Kriterium heißt hier *Dependent*, der Prädiktor *Independent(s)*. Der Plural besagt, daß hier auch eine Regression mit mehr als einer Prädiktorvariablen berechnet werden kann, was in Statistik I nicht behandelt wurde. Für die Regression von  $y$  auf  $x$  wird das Regressionsfenster wie folgt ausgefüllt:





Der Schalter *Statistics* öffnet ein Fenster, mit dem wir eine Reihe von Berechnungen auswählen können. Wir belassen es hier bei den Voreinstellungen und aktivieren ihn nicht. Mit dem Schalter *Plots* öffnen wir ein Auswahlfenster, das wir wie folgt ausfüllen.



Nach *Continue* → *OK* erhalten wir einige Tabellen im Ausgabebildschirm, von denen uns vor allem die der Regressionskoeffizienten interessiert:

**Coefficients(a)**

| Model |                | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|----------------|-----------------------------|------------|---------------------------|--------|------|
|       |                | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant)     | -15,691                     | 2,183      |                           | -7,189 | ,000 |
|       | Erste Variable | 1,143                       | ,072       | ,933                      | 15,926 | ,000 |

a Dependent Variable: Zweite Variable

Unter *(Constant)* enthält sie den Ordinatenabschnitt der Regressionsgeraden  $a_{yx}$ , unter *Erste Variable* das Steigungsmaß  $b_{yx}$ . Die Zahlenwerte stimmen mit den in der Vorlesung „von Hand“ gerechneten Werten überein.

Die Vertauschung von  $x$  und  $y$ , also die Regression von  $x$  auf  $y$ , liefert:

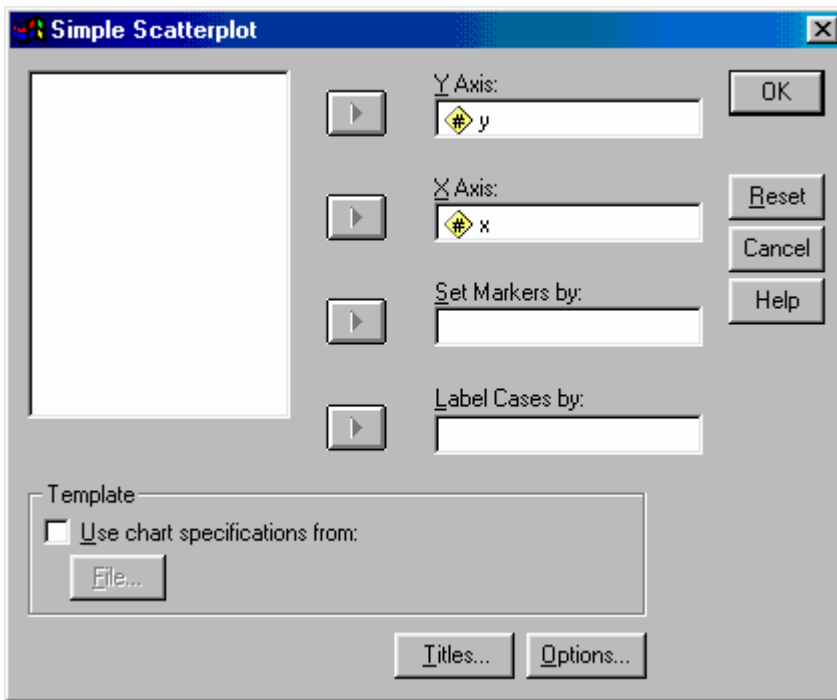
**Coefficients(a)**

| Model |                 | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|-----------------|-----------------------------|------------|---------------------------|--------|------|
|       |                 | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant)      | 15,893                      | ,915       |                           | 17,372 | ,000 |
|       | Zweite Variable | ,761                        | ,048       | ,933                      | 15,926 | ,000 |

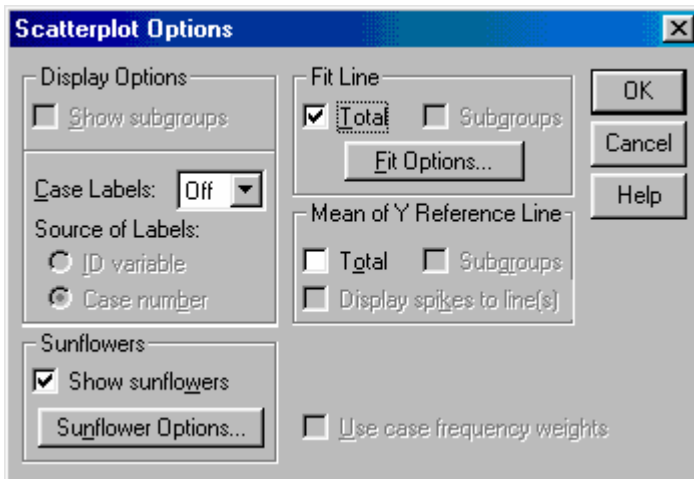
a Dependent Variable: Erste Variable

Auch diese Zahlenwerte stimmen mit den früher berechneten (bis auf Rundungsabweichungen) überein.

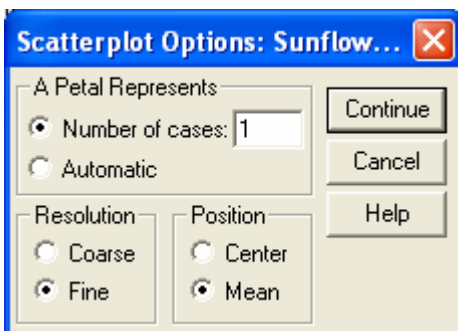
g) Mit den graphischen Möglichkeiten verwöhnt uns SPSS bei den bivariaten Verteilungen nicht besonders. Wir gehen wieder vom Dateneingabebildschirm aus. Mit *Graphs* → *Scatter* → *Simple* → *Define* bekommen wir das folgende Fenster:



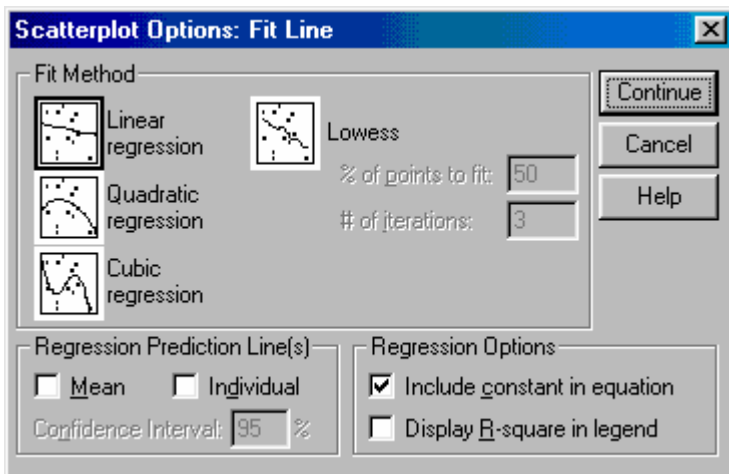
Nachdem wir die Variablen der Abszisse und der Ordinate zugeordnet haben, können wir unter *Titles* noch Titel eintragen, unter *Options* die Behandlung fehlender Werte steuern. Wir lassen beides aus und schließen das Fenster mit *OK*. Das Resultat enthält zunächst noch keine Information über die Häufigkeit der einzelnen Maßzahlpaare und keine Regressionsgerade. (Schauen Sie es sich daraufhin an, ich nehme es nicht in das Skriptum auf.) Das läßt sich jedoch im Chart Editor ändern. Rufen Sie ihn im Ausgabebildschirm durch Doppelklicken auf die Abbildung auf und wählen Sie dann *Chart* → *Options*. Sie erhalten ein Optionenfenster für die bivariate Häufigkeitsverteilung:



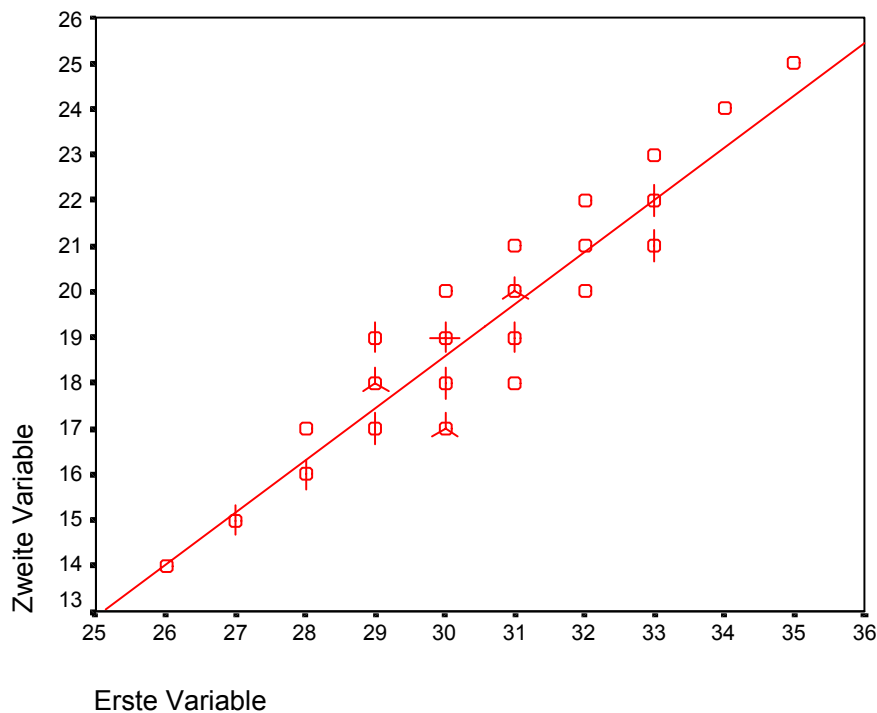
Sie müssen jetzt *Show sunflowers* aktivieren. *Sunflower options* öffnet das folgende Fenster,



das Ihnen mit den gezeigten Einstellungen ein brauchbares Resultat liefert, mit dem Sie aber auch etwas experimentieren sollten. Schließen Sie es mit *Continue*, aktivieren Sie unter *Fit Line* den Schalter *Total* und öffnen Sie das *Fit Options*-Fenster:



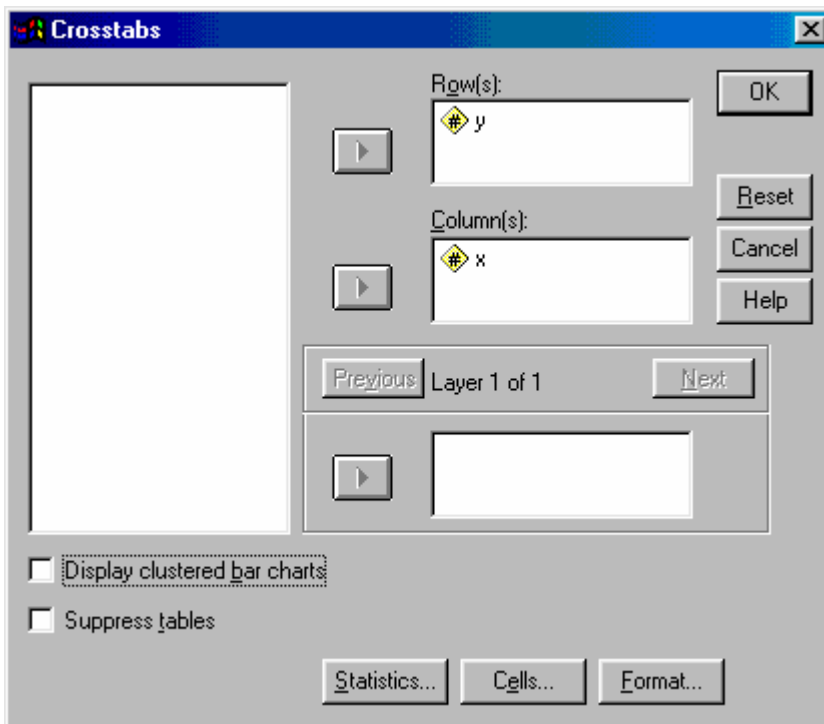
Hier kommen Sie ebenfalls mit den aktiven Alternativen gut zurecht. Schließen Sie das Fenster mit *Continue* und ändern Sie die Graphik mit *OK*. Das Ergebnis sieht folgendermaßen aus:



Sie sehen jetzt, was SPSS unter „Sonnenblumen“ versteht: Die Häufigkeit jedes Maßzahlpaares wird durch die Anzahl um das Symbol herum eingezeichneter Radien wiedergegeben. Beim Vergleich der Abbildung mit Ihrer Zeichnung zur Vorlesung werden Sie feststellen, daß die Abbildung korrekt ist. Das Paar (30, 19) hat 4 Radien, viele andere Paare, z. B. (27, 15), haben deren 2. Unschön ist, daß SPSS für Paare mit der Häufigkeit 1 nur das Symbol und keinen Radius zeichnet. Ein weiterer Nachteil dieser Darstellung liegt darin, daß die Radien sich mit den Gitterlinien decken, falls Sie solche einzeichnen, und damit in schwarz-weißen Darstellungen teilweise unsichtbar werden. Eine Zeichnung mit beiden Regressionsgeraden liefert SPSS nicht; um die Regressionsgerade von  $x$  auf  $y$  zu zeigen, ist eine neue Darstellung nötig, bei der  $x$  und  $y$  die Rolle tauschen.

Wahrscheinlich vermissen Sie jetzt noch die tabellarische Darstellung der bivariaten Häufigkeitsverteilung. Diese bekommen Sie mit einer sehr wichtigen Funktion von SPSS:

*Crosstabs*. Gehen Sie in den Dateieingabebildschirm zurück und wählen Sie *Analyze* → *Descriptive Statistics* → *Crosstabs*. Mit dem folgenden Fenster



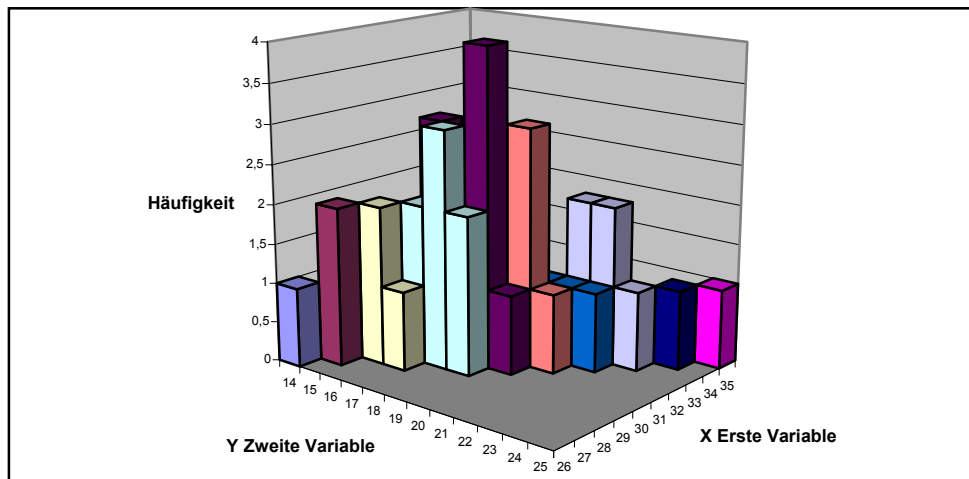
können Sie eine bivariate Tabelle aufbauen. Ordnen Sie *x* den Spalten, *y* den Zeilen zu, wie es die Abbildung zeigt. Mit *Statistics*, *Cells* und *Format* können Sie eine Reihe wichtiger Optionen eingeben, die wir jetzt noch nicht benötigen. Sie können aber schon damit experimentieren und sich anschauen, was die einzelnen Optionen bewirken. (Die Graphiken unter *Display clustered bar charts* sind von geringem Interesse.) Beachten Sie, daß auch die in der Statistik I eingehend besprochene Vierfeldertafel in SPSS unter *Crosstabs* fällt. Mit *OK* erhalten Sie Ihr Ergebnis:

**Zweite Variable \* Erste Variable Crosstabulation**

| Count           |       | Erste Variable |       |       |       |       |       |       |       |       |       | Total |
|-----------------|-------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                 |       | 26,00          | 27,00 | 28,00 | 29,00 | 30,00 | 31,00 | 32,00 | 33,00 | 34,00 | 35,00 |       |
| Zweite Variable | 14,00 | 1              | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 1     |
|                 | 15,00 | 0              | 2     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2     |
|                 | 16,00 | 0              | 0     | 2     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 2     |
|                 | 17,00 | 0              | 0     | 1     | 2     | 3     | 0     | 0     | 0     | 0     | 0     | 6     |
|                 | 18,00 | 0              | 0     | 0     | 3     | 2     | 1     | 0     | 0     | 0     | 0     | 6     |
|                 | 19,00 | 0              | 0     | 0     | 2     | 4     | 2     | 0     | 0     | 0     | 0     | 8     |
|                 | 20,00 | 0              | 0     | 0     | 0     | 1     | 3     | 1     | 0     | 0     | 0     | 5     |
|                 | 21,00 | 0              | 0     | 0     | 0     | 0     | 1     | 1     | 2     | 0     | 0     | 4     |
|                 | 22,00 | 0              | 0     | 0     | 0     | 0     | 0     | 1     | 2     | 0     | 0     | 3     |
|                 | 23,00 | 0              | 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 1     |
|                 | 24,00 | 0              | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 1     |
| 25,00           | 0     | 0              | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 1     | 1     |       |
| Total           |       | 1              | 2     | 3     | 7     | 10    | 7     | 3     | 5     | 1     | 1     | 40    |

Die Kreuztabelle läßt sich auch über die Zwischenablage von SPSS in ein EXCEL-Datenblatt übertragen, wo bessere Möglichkeiten zur graphischen Darstellung bestehen. Ein auf diese Weise erzielbares Ergebnis zeigt dieses Bild:

Bivariate Häufigkeitsverteilung



Mit den beiden Aufgaben, 01 und 02, haben Sie den größten Teil des Stoffes aus Statistik I mit SPSS bearbeitet. Einige Einzelheiten werden später noch nachzutragen oder zu ergänzen sein. Nehmen Sie sich jetzt weitere Aufgabenbeispiele vor, die Sie in Statistik I mit Papier und Bleistift bearbeitet haben, und üben Sie das über SPSS Gelernte gründlich ein. Mit der nächsten Aufgabe begeben wir uns sofort mitten in den Stoff von Statistik II.

### Aufgabe 03

Ein psychologischer Test auf Neurotizismus soll daraufhin überprüft werden, ob sich die Maßzahlen von Neurotikern von denen von Psychotikern stärker als nach Zufall zu erwarten voneinander unterscheiden. Dazu werden 16 Neurotiker und 14 Psychotiker (Diagnose aufgrund eines psychiatrischen Urteils, das als gültig vorausgesetzt wird) dem Test unterzogen. Man erhält die Maßzahlen:

Neurotiker: 7 9 13 10 8 11 13 9 12 11 8 12 10 14 9 15

Psychotiker: 10 16 12 11 14 9 10 13 11 17 15 10 12 9.

Kann man sagen, daß sich Neurotiker und Psychotiker im Mittelwert ihrer in diesem Test erhaltenen Maßzahlen unterscheiden?

(Quelle: Bartel, Glaser & Metzger, 1976, *Statistik II*, S. 142, Aufg. 7, Text geändert)

### Lösung

Öffnen Sie den Dateneingabebildschirm von SPSS. Geben Sie der ersten Variablen den Namen *x* und den Bezeichner *Maßzahlen*. Tasten Sie jetzt die Maßzahlen der Aufgabe ein, und zwar zuerst für die Neurotiker, dann, ohne Unterbrechung, für die Psychotiker. Bei den Psychotikern lautet die 8. Maßzahl 13. Nachdem Sie diese eingegeben haben, sieht Ihr Eingabebildschirm so aus:

|    | x     | var | var | var | var | var | var | var | var |
|----|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| 12 | 12,00 |     |     |     |     |     |     |     |     |
| 13 | 10,00 |     |     |     |     |     |     |     |     |
| 14 | 14,00 |     |     |     |     |     |     |     |     |
| 15 | 9,00  |     |     |     |     |     |     |     |     |
| 16 | 15,00 |     |     |     |     |     |     |     |     |
| 17 | 10,00 |     |     |     |     |     |     |     |     |
| 18 | 16,00 |     |     |     |     |     |     |     |     |
| 19 | 12,00 |     |     |     |     |     |     |     |     |
| 20 | 11,00 |     |     |     |     |     |     |     |     |
| 21 | 14,00 |     |     |     |     |     |     |     |     |
| 22 | 9,00  |     |     |     |     |     |     |     |     |
| 23 | 10,00 |     |     |     |     |     |     |     |     |
| 24 | 13,00 |     |     |     |     |     |     |     |     |
| 25 |       |     |     |     |     |     |     |     |     |
| 26 |       |     |     |     |     |     |     |     |     |

Beachten Sie, daß die 16. Maßzahl 15 lautet, also die letzte Maßzahl aus der Gruppe der Neurotiker darstellt. Die 17. Maßzahl lautet 10, ist also die erste Maßzahl der Gruppe der Psychotiker. Die letzte Maßzahl der Gruppe der Psychotiker, 9, wird schließlich die 30. Eingabe. Dieses Schema gilt grundsätzlich für die Eingabe von Daten, die zu mehreren Personengruppe gehören: Jede Person ist ein Fall, also eine Zeile im Eingabebildschirm von SPSS. Personen aus verschiedenen Gruppen werden fortlaufend untereinander eingetragen.

Natürlich „weiß“ SPSS bis jetzt noch nicht, welche Person zur Gruppe der Neurotiker, welche zur Gruppe der Psychotiker gehört. Zu diesem Zweck führen wir eine Indikatorvariable ein. Öffnen Sie *Variable View* und tragen Sie als Variablennamen *ind* ein:

|   | Name | Type    | Width | Decimals | Label     | Values                   | Missing | Columns | Align | Measur... |
|---|------|---------|-------|----------|-----------|--------------------------|---------|---------|-------|-----------|
| 1 | x    | Numeric | 8     | 2        | Testwert  | None                     | None    | 8       | Right | Scale     |
| 2 | ind  | Numeric | 8     | 2        | Indikator | {1,00, Neurotiker}... .. | None    | 8       | Right | Scale     |
| 3 |      |         |       |          |           |                          |         |         |       |           |
| 4 |      |         |       |          |           |                          |         |         |       |           |
| 5 |      |         |       |          |           |                          |         |         |       |           |
| 6 |      |         |       |          |           |                          |         |         |       |           |

Öffnen Sie dann unter *Values* in der Zeile für *ind* das Bezeichnerfenster:

**Value Labels**

Value Labels

Value:

Value Label:

Add 1,00 = "Neurotiker"

Change 2,00 = "Psychotiker"

Remove

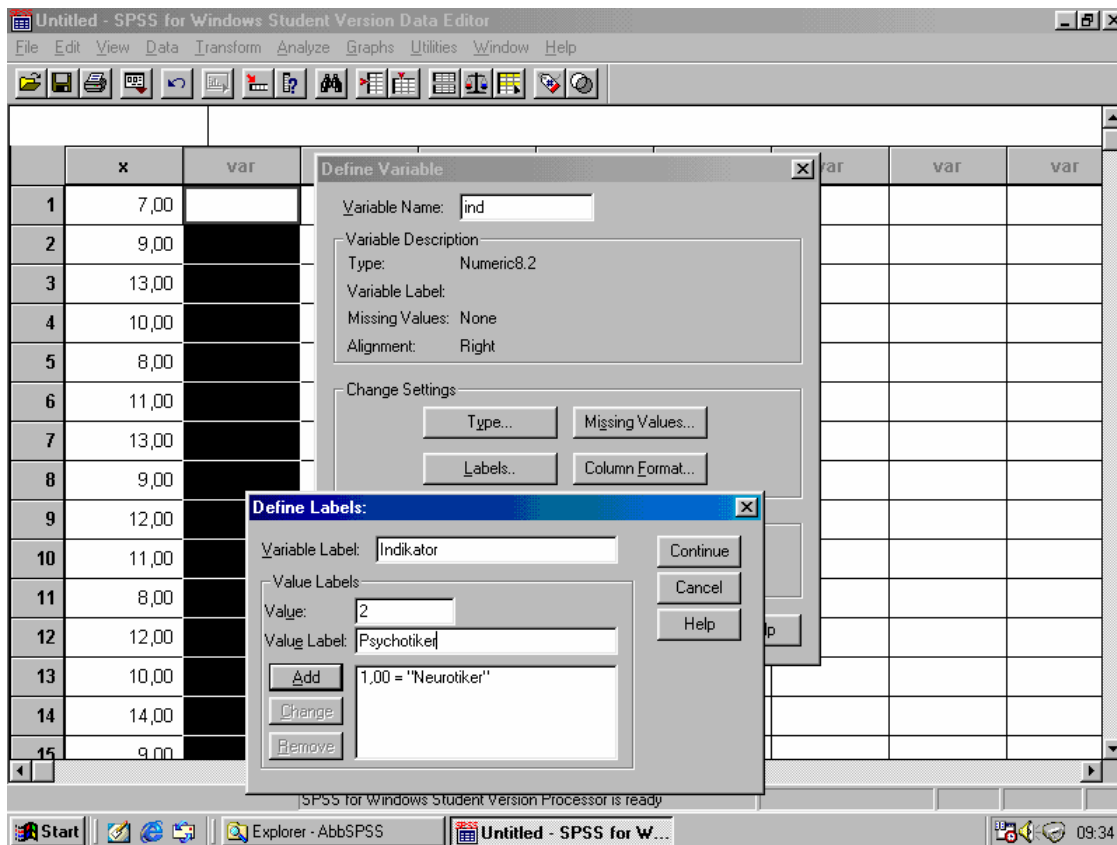
OK

Cancel

Help

Setzen Sie die Wertebezeichner *Neurotiker* und *Psychotiker* und schließen Sie das Fenster mit OK.

In den älteren Versionen von SPSS sieht das so aus:



Geben Sie der Variablen *x* den Bezeichner *Indikator*, ihrem Wert *1* den Wertebezeichner *Neurotiker* und ihrem Wert *2* den Wertebezeichner *Psychotiker*. Die Einzelheiten sehen Sie im Bild. Wenn Ihr Bildschirm exakt dem abgebildeten Fenster *Define Labels* entspricht, müssen Sie den Wertebezeichner *Psychotiker* noch mit *Add* in das untere Feld bringen und das Fenster mit *Continue* schließen. Mit *OK* schließen Sie das Variablendefinitionsfenster.

Nun müssen Sie die Werte der Indikatorvariablen eingeben. Personen 1 bis 16 gehören zur Gruppe der Neurotiker, erhalten also die Eingabe  $ind = 1$ . Entsprechend erhalten die Personen 17 bis 30 die Eingabe  $ind = 2$ . Nach dem Ende der Eingabe speichern Sie Ihre Datei unter *Aufgabe03.sav* ab.

Die Mitte des Dateneingabebildschirm sehen Sie im folgenden Bild:

1:x

|    | x     | ind  | var | var | var | var | var | var | var |
|----|-------|------|-----|-----|-----|-----|-----|-----|-----|
| 11 | 8,00  | 1,00 |     |     |     |     |     |     |     |
| 12 | 12,00 | 1,00 |     |     |     |     |     |     |     |
| 13 | 10,00 | 1,00 |     |     |     |     |     |     |     |
| 14 | 14,00 | 1,00 |     |     |     |     |     |     |     |
| 15 | 9,00  | 1,00 |     |     |     |     |     |     |     |
| 16 | 15,00 | 1,00 |     |     |     |     |     |     |     |
| 17 | 10,00 | 2,00 |     |     |     |     |     |     |     |
| 18 | 16,00 | 2,00 |     |     |     |     |     |     |     |
| 19 | 12,00 | 2,00 |     |     |     |     |     |     |     |
| 20 | 11,00 | 2,00 |     |     |     |     |     |     |     |
| 21 | 14,00 | 2,00 |     |     |     |     |     |     |     |
| 22 | 9,00  | 2,00 |     |     |     |     |     |     |     |
| 23 | 10,00 | 2,00 |     |     |     |     |     |     |     |
| 24 | 13,00 | 2,00 |     |     |     |     |     |     |     |
| 25 | 11,00 | 2,00 |     |     |     |     |     |     |     |

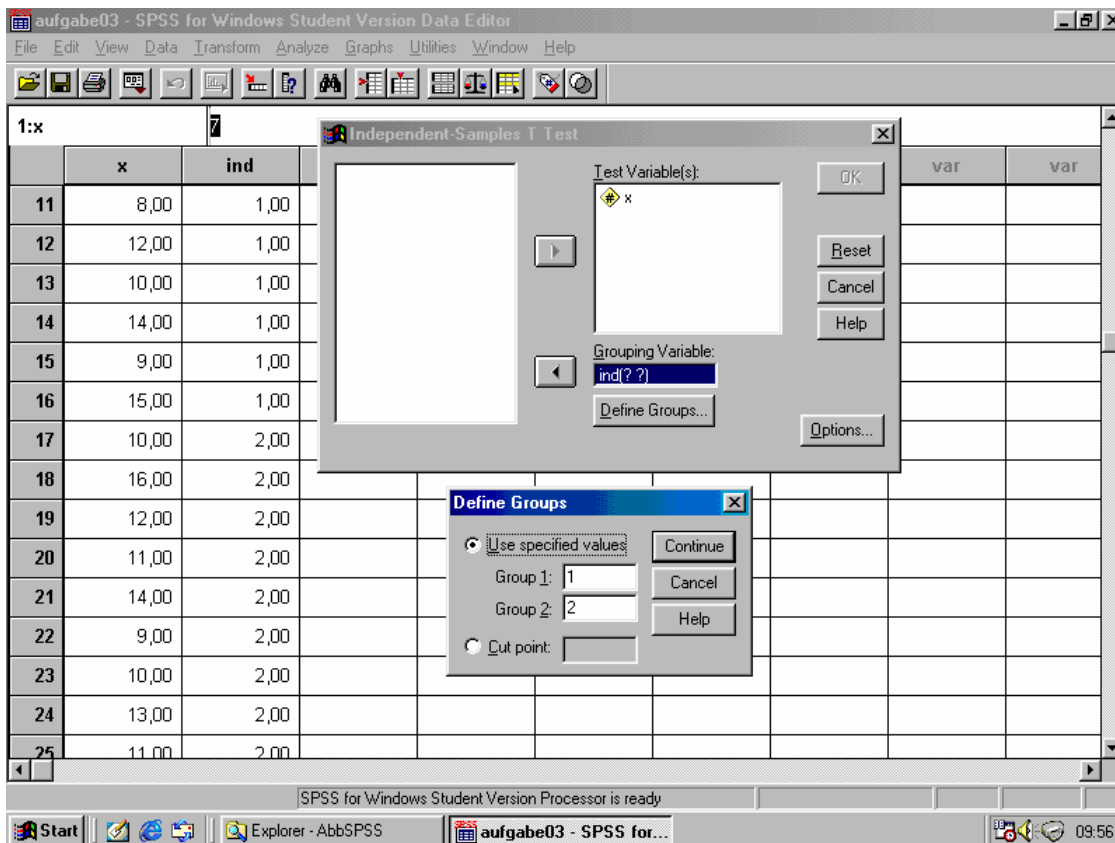
SPSS for Windows Student Version Processor is ready

Start | Explorer - AbbSPSS | aufgabe03 - SPSS for... | 09:46

Daß Sie erfolgreich abgespeichert haben, sehen Sie in der linken oberen Ecke am Dateinamen *Aufgabe03*. Die Variable *ind* enthält bis zur 16. Person durchgängig den Wert 1, ab der 17. Person den Wert 2.

Die Aufgabe verlangt den t-Test für unabhängige Stichproben. Mit *Analyze* → *Compare Means* → *Independent-Samples T Test* rufen Sie das Einstellungsfenster für diesen Test auf. Danach müssen Sie *x* in das Feld *Test Variable(s)* und *ind* in das Feld *Grouping Variable* bringen, wie es im folgenden Bild schon geschehen ist:





Als nächstes rufen Sie *Define Groups* auf. SPSS bildet die beiden Gruppen, deren Mittelwerte verglichen werden sollen, anhand der Werte der Indikatorvariablen *ind*. Da wir den Neurotikern *ind* = 1 und den Psychotikern *ind* = 2 zugewiesen haben, veranlassen wir SPSS mit der Eingabe der beiden Werte unter *Group 1* und *Group 2*, für diese beiden Personengruppen die Mittelwerte zu berechnen und mittels t-Test auf Signifikanz des Unterschiedes zu prüfen.

Mit *Continue* schließen Sie das Fenster für die Gruppendefinitionen. Unter *Options* können Sie nun noch das Signifikanzniveau und die Behandlung fehlender Werte wählen. Wir lassen es bei der Voreinstellung:  $\alpha = 5\%$  zweiseitig und lokaler Ausschluß von Fällen mit fehlenden Wert für *x*. Mit *OK* starten wir die Auswertung.

Der Ausgabebildschirm zeigt uns die beiden folgenden Tabellen:

#### Group Statistics

|          | Indikator   | N  | Mean    | Std. Deviation | Std. Error Mean |
|----------|-------------|----|---------|----------------|-----------------|
| Testwert | Neurotiker  | 16 | 10,6875 | 2,33006        | ,58251          |
|          | Psychotiker | 14 | 12,0714 | 2,58589        | ,69111          |

### Independent Samples Test

|          |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                       |   |        |
|----------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|--------|
|          |                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |        |
|          |                             |   |      |                              |        |                 |                 |                       | Lower                                     | Upper  |
| Testwert | Equal variances assumed     | ,106                                    | ,747 | -1,542                       | 28     | ,134            | -1,3839         | ,89740                | -3,22216                                  | ,45431 |
|          | Equal variances not assumed |   |      | -1,531                       | 26,459 | ,138            | -1,3839         | ,90385                | -3,24026                                  | ,47240 |

Die erste Tabelle enthält die wichtigsten deskriptiven Statistiken, Gruppengröße, Mittelwert, Standardabweichung (mit  $n - 1$  im Nenner) und Standardfehler des Mittelwertes. Die zweite Tabelle liefert das Ergebnis eines Präliminartests auf Varianzhomogenität, des Levene-Tests. Der t-Test wird für beide möglichen Ergebnisse, homogene und heterogene Varianzen, gerechnet. In beiden Fällen beträgt die zweiseitige Überschreitungswahrscheinlichkeit 13,4 % bzw. 13,8 %. Zwischen beiden Gruppen besteht also kein interpretierbarer Unterschied. Der gegebene Test zur Diagnose von Neurotizismus ist also nicht geeignet, zwischen Neurotikern und Psychotikern zu trennen.

#### Aufgabe 04

Unter dem Namen *Stroop-Phänomen* ist in der Psychologie seit rund 60 Jahren bekannt, daß die Zeit, die man benötigt, um eine Farbe zu benennen, wesentlich länger ist, wenn die zu benennende Farbe ein Farbwort ausschreibt, mit dessen Bedeutung sie nicht übereinstimmt.

Seit einigen Jahren weiß man, daß dies auch für Süchtige gilt, wenn man die Farben zum Schreiben von Wörtern verwendet, die in einem semantischen Zusammenhang zur Sucht stehen.

Ein Psychologe möchte diese Zusammenhänge mit Alkoholikern überprüfen und entwirft dazu das folgende Experiment. Auf den Kontrolltafeln werden 100 Wörter, angeordnet in 10 Zeilen und 10 Spalten, in den wechselnden Farben *rot*, *gelb*, *grün* und *blau* gedruckt. Die Wörter sind semantisch neutral, z. B. *Teller*, *Stuhl*, *Fenster*. Die Versuchstafeln werden genauso aufgebaut, nur daß die Wortbedeutungen jetzt einen semantischen Zusammenhang zum Alkoholkonsum haben, z. B. *Flasche*, *Bier*, *Trunkenheit* usw.

Als Versuchsgruppe wird eine Stichprobe von  $n_1 = 12$  in stationärer Behandlung befindlichen Alkoholikern untersucht. Sie haben die Aufgabe, so schnell wie möglich und unter möglicher Vermeidung von Fehlern die 100 Farben der Kontroll- und der Versuchstafel zu benennen. Die Zeit dafür wird mit einer Stoppuhr gemessen. Man erhält die folgenden Resultate, wobei untereinanderstehende Zahlen zur gleichen Person gehören:

|               |    |    |    |    |    |    |    |    |    |    |    |     |
|---------------|----|----|----|----|----|----|----|----|----|----|----|-----|
| Kontrolltafel | 55 | 70 | 42 | 61 | 49 | 46 | 62 | 65 | 44 | 57 | 52 | 60  |
| Versuchstafel | 70 | 52 | 89 | 65 | 76 | 79 | 61 | 83 | 58 | 73 | 66 | 64. |

Als Kontrollgruppe wird eine Stichprobe von  $n_2 = 12$  nicht süchtigen Personen gebildet. Man sorgt dafür, daß sich die Merkmale Geschlecht, Alter, Ausbildung und sozioökonomischer Status in beiden Stichproben etwa gleich verteilen. Bei der Kontrollgruppe werden die folgenden Zeiten gemessen:

|               |    |    |    |    |    |    |    |    |    |    |    |     |
|---------------|----|----|----|----|----|----|----|----|----|----|----|-----|
| Kontrolltafel | 50 | 52 | 47 | 69 | 39 | 61 | 42 | 63 | 38 | 55 | 45 | 56  |
| Versuchstafel | 36 | 63 | 50 | 52 | 46 | 58 | 44 | 65 | 37 | 62 | 41 | 57. |

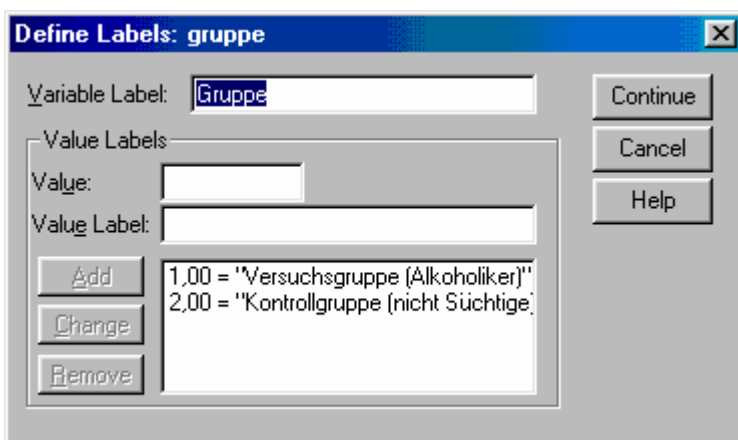
Bitte verdeutlichen Sie die Signifikanztests der folgenden Aufgaben a) und b) mit je einer Skizze. Testen Sie mit  $\alpha \leq 5\%$  zweiseitig.

- Zeigt die Versuchsgruppe wirklich eine interpretierbare Verlängerung der Zeit für das Farbenbenennen in der Versuchsbedingung mit den alkoholbezogenen Wörtern?
- Bitte prüfen Sie die Nullhypothese für den Unterschied zwischen Kontrolltafel und Versuchstafel bei der Kontrollgruppe. Was ist aus dem Resultat des Signifikanztests zu folgern?
- Der Psychologe hat die weitere Hypothese, daß der Alkoholismus zu einer allgemeinen Verminderung kognitiver Fähigkeiten führt, die auch an der verlangsamten Benennung der Farben in der Kontrollbedingung bei Alkoholikern zu erkennen ist. Bitte prüfen Sie die entsprechende Nullhypothese statistisch. Wie ist das Resultat dieses Tests inhaltlich zu interpretieren? (Hier ist **keine** Skizze verlangt;  $\alpha \leq 5\%$ , zweiseitig.)

(Quelle: Nachholklausur Statistik II Sommersemester 1998 v. 14.12.1998, Aufg. 5; Übungsblatt Statistik II Sommersemester 1999 *stat211.doc*)

## Lösung

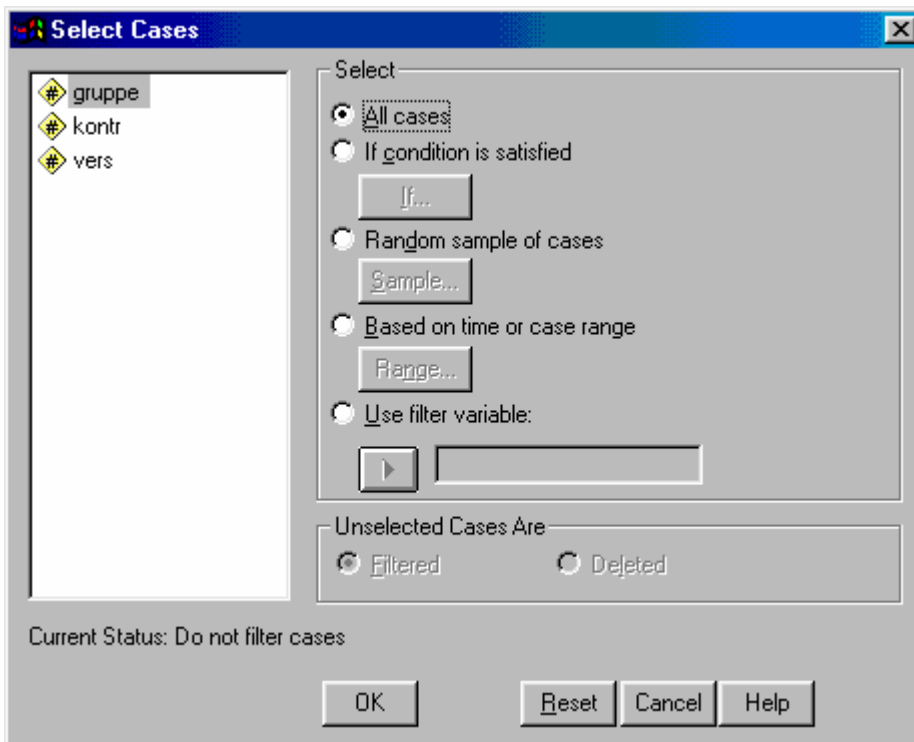
Öffnen Sie den Datenbildschirm von SPSS. Bei der Dateneingabe müssen Sie jetzt beachten, daß von jeder Person zwei Messungen vorliegen, eine für die Kontrolltafel und eine für die Versuchstafel. Richten Sie dafür die beiden Variablen *kontr* und *vers* mit den Variablenbezeichnern *Kontrolltafel* und *Versuchstafel* ein. Außerdem haben wir zwei unabhängige Gruppen von Personen, Versuchsgruppe und Kontrollgruppe. Die Indikatorvariable bekommt jetzt den Namen *gruppe* und den Variablenbezeichner *Gruppe*. Die folgende Abbildung zeigt das Fenster für die Bezeichner:



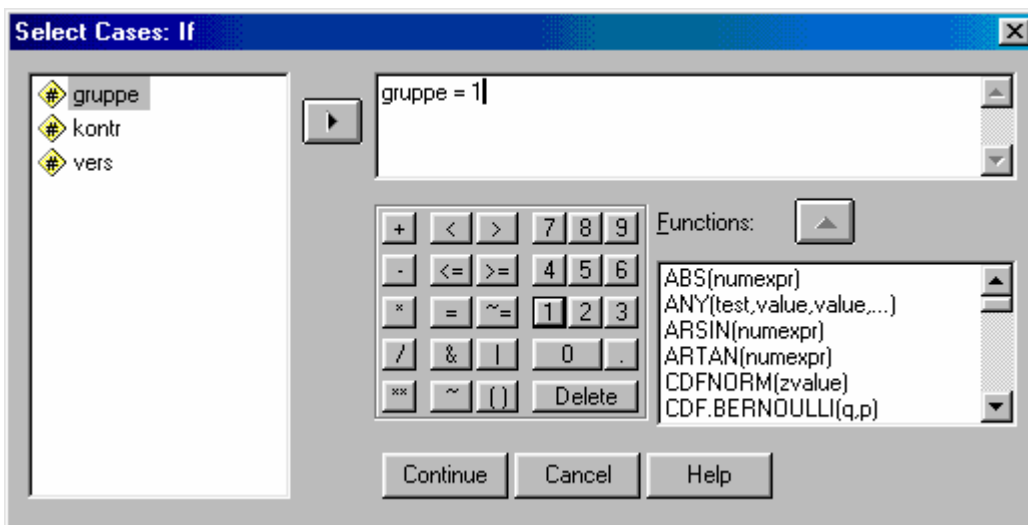
Bei der Indikatorvariablen *gruppe* vergeben wir die beiden Wertebezeichner *Versuchsgruppe (Alkoholiker)* und *Kontrollgruppe (nicht Süchtige)*.

Speichern Sie Ihre Datei nach Eingabe aller Daten unter *Aufgabe04.sav*.

- Diese Frage zielt auf den Unterschied zwischen den Mittelwerten für die Kontrolltafel und die Stroop-Tafel bei den gleichen Personen, und zwar denjenigen der Versuchsgruppe. Es wird also ein t-Test für abhängige Stichproben verlangt. Die Aufgabe bezieht sich aber nur auf eine Teilmenge der Daten, die Versuchsgruppe. Von der Kontrollgruppe ist hier überhaupt nicht die Rede. Wir müssen also zunächst SPSS dazu veranlassen, den Signifikanztest nur für die Versuchsgruppe auszuführen, die an dem Wert *gruppe = 1* zu erkennen ist. Dafür bietet uns SPSS eine Reihe von Filterfunktionen an. Öffnen Sie im Datenbildschirm mit *Data* → *Select Cases* den Filterbildschirm:



Sie sehen links eine alphabetische Liste Ihrer Variablen und rechts unter *Select* verschiedene Filteroptionen. Interessant ist beispielsweise, daß Sie hier eine Zufallsstichprobe aus Ihren Daten erzeugen können, indem Sie *Random sample of cases* anklicken. Wir wählen *If condition is satisfied* und erhalten durch Anklicken des darunter jetzt aktiv werdenden *If* das Bedingungs Fenster für die Filterung:

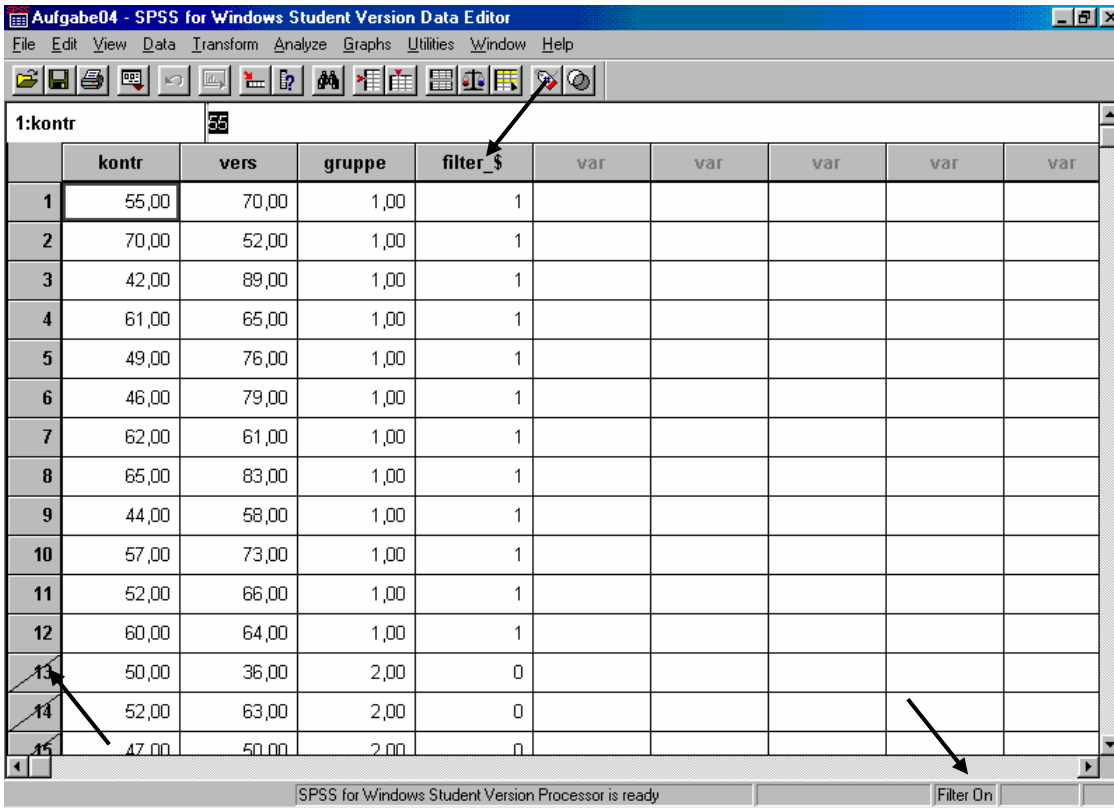


Mit den Schaltfeldern dieses Fensters können wir im oberen rechten Bedingungsfield eine Bedingung formulieren. Nur die Fälle, die diese Bedingung erfüllen, werden in die spätere Auswertung einbezogen. Wir haben *gruppe = 1* eingegeben und erreichen damit, daß sich die spätere Auswertung nur auf die Personen erstreckt, für die *gruppe = 1* ist, also die Versuchsgruppe der Alkoholiker. Mit *Continue* kehren wir zum Filterfenster zurück.

Beachten Sie, daß darin jetzt die Schalter in *Unselected Cases are* aktiv sind. Wir lassen es bei der Voreinstellung *Filtered*. Damit bleibt unsere Datenmatrix vollständig, auch wenn die Fälle, die die Filterbedingung nicht erfüllen, aus den folgenden Auswertungen ausgeschlossen bleiben. Durch eine andere Filterbedingung können wir später auch andere Teilmengen der Daten für die Auswertung auswählen. Gelegentlich kommt es vor, daß man die einmal weggefilterten Daten nicht mehr braucht. In diesem Fall kann man hier *Deleted* aktivieren, um eine kürzere und

übersichtlichere Datendatei zu bekommen. Das ist jedoch gefährlich, denn wenn Sie danach die aktuelle Datei mit *File* → *Save* speichern, ist der weggefilterte Teil Ihrer Daten unwiderruflich verloren, es sei denn, Sie haben noch eine Sicherheitskopie unter einem anderen Namen.

Mit *OK* schließen wir das Filterfenster und lassen die Filterung wirksam werden. Daß eine Filterung eingeschaltet ist, erkennen wir im Datenbildschirm an drei Stellen (schwarze Pfeile in der folgenden Abbildung):



Eine Variable *filter\_\$* wurde erzeugt, die bei den ausgewählten Fällen den Wert 1, bei den weggefilterten Fällen den Wert 0 hat, die Nummern der weggefilterten Fälle erscheint durchgestrichen und die Fußzeile des Bildschirm zeigt *Filter On*.

Für den t-Test mit Meßwiederholungen öffnen wir das Einstellungsfenster mit *Analyze* → *Compare Means* → *Paired-Samples T Test* und bringen das Variablenpaar *kontr* und *vers* in das Bearbeitungsfeld *Paired Variables*. Bei *Options* akzeptieren wir die Voreinstellungen, die die gleiche Bedeutung haben wie beim t-Test für unabhängige Gruppen (Aufgabe 03). Mit *OK* starten wir die Auswertung. Der Ausgabebildschirm zeigt die drei folgenden Tabellen:

**Paired Samples Statistics**

|        |               | Mean    | N  | Std. Deviation | Std. Error Mean |
|--------|---------------|---------|----|----------------|-----------------|
| Pair 1 | Kontrolltafel | 55,2500 | 12 | 8,81244        | 2,54393         |
|        | Versuchstafel | 69,6667 | 12 | 10,79001       | 3,11481         |

**Paired Samples Correlations**

|        |                               | N  | Correlation | Sig. |
|--------|-------------------------------|----|-------------|------|
| Pair 1 | Kontrolltafel & Versuchstafel | 12 | -,455       | ,137 |

**Paired Samples Test**

|        |                               | Paired Differences |                |                 |   |         | t     | df | Sig. (2-tailed) |
|--------|-------------------------------|--------------------|----------------|-----------------|---|---------|-------|----|-----------------|
|        |                               | Mean               | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference |         |       |    |                 |
|        |                               |                    |                |                 | Lower                                     | Upper   |       |    |                 |
| Pair 1 | Kontrolltafel - Versuchstafel | 14,4167            | 16,75198       | 4,83588         | -25,0604                                  | -3,7730 | 2,981 | 11 | ,012            |

Die Tafeln zeigen wieder die wichtigsten deskriptiven Statistiken einschließlich der Produktmomentkorrelation zwischen beiden Variablen. Die zweiseitige Überschreitungswahrscheinlichkeit beträgt im t-Test 1,2 %, auf dem vorgegebenen Signifikanzniveau  $\alpha = 5\%$  zweiseitig ist der Mittelwertsunterschied also signifikant. Die Stroop-Bedingung mit den alkoholbezogenen Distraktorwörtern führt also zu einer Verlängerung der Benennzeit für die Farben. Beachten Sie, an welchen Stellen der Tabellen die Variablen- und Wertebezeichner wieder auftauchen und welche Rolle sie bei Auswertungen mit SPSS demnach spielen.

b) Hier wird die Frage aus a) auf die Gruppe der nicht süchtigen Kontrollpersonen angewandt. Wir müssen also die Filterbedingung so ändern, daß nur die Kontrollgruppe ausgewertet wird, also auf *gruppe = 2*, und dann den t-Test für abhängige Stichproben in der gleichen Weise wie unter a) rechnen. Dabei stellen Sie fest, daß das Einstellungsfenster für den t-Test sogar noch die Eingaben aus der Aufgabe a) enthält. Das Ergebnis lautet:

**Paired Samples Statistics**

|        |               | Mean    | N  | Std. Deviation | Std. Error Mean |
|--------|---------------|---------|----|----------------|-----------------|
| Pair 1 | Kontrolltafel | 51,4167 | 12 | 9,79293        | 2,82698         |
|        | Versuchstafel | 50,9167 | 12 | 10,19321       | 2,94253         |

**Paired Samples Correlations**

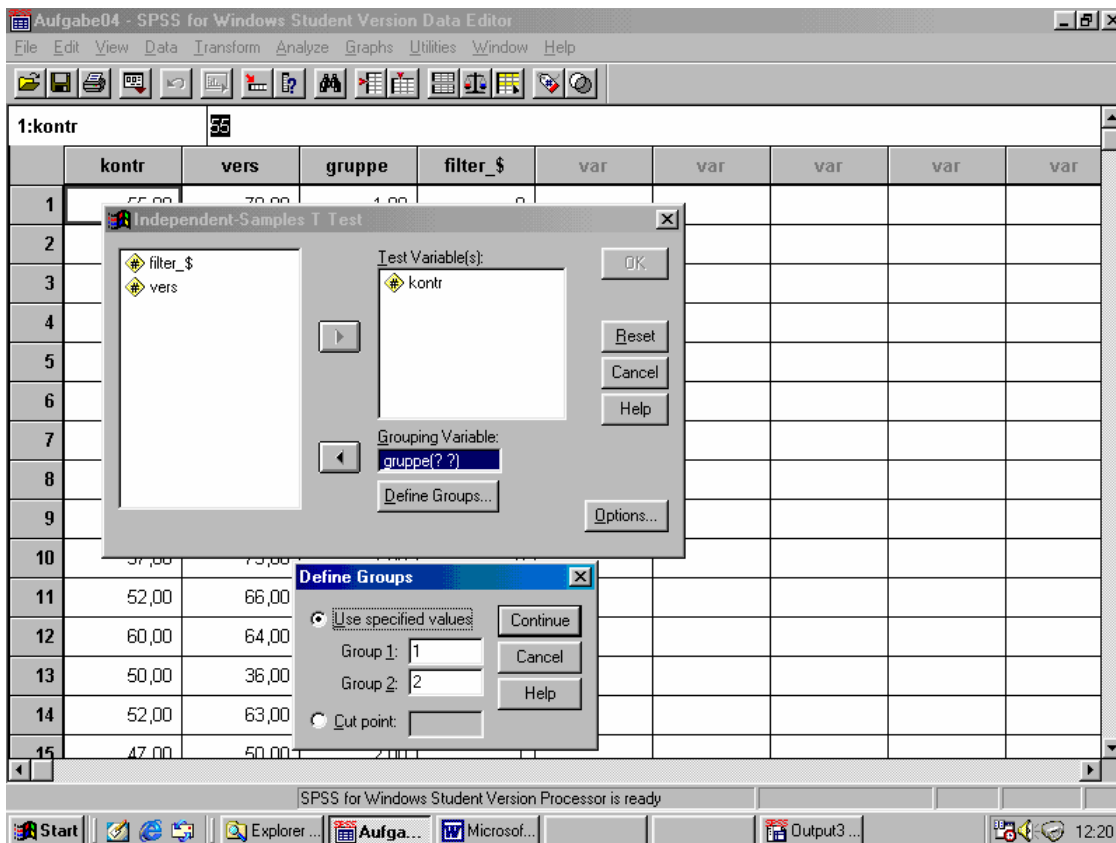
|        |                               | N  | Correlation | Sig. |
|--------|-------------------------------|----|-------------|------|
| Pair 1 | Kontrolltafel & Versuchstafel | 12 | ,662        | ,019 |

**Paired Samples Test**

|        |                               | Paired Differences |                |                 |   |        | t    | df | Sig. (2-tailed) |
|--------|-------------------------------|--------------------|----------------|-----------------|---|--------|------|----|-----------------|
|        |                               | Mean               | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference |        |      |    |                 |
|        |                               |                    |                |                 | Lower                                     | Upper  |      |    |                 |
| Pair 1 | Kontrolltafel - Versuchstafel | ,5000              | 8,22966        | 2,37570         | -4,7289                                   | 5,7289 | ,210 | 11 | ,837            |

Bei dieser Teilaufgabe zeigt sich kein signifikanter Unterschied zwischen beiden Mittelwerten, bei den nicht süchtigen Personen haben also die alkoholbezogenen Distraktorwörter keine Auswirkung auf die Geschwindigkeit beim Benennen der Farben.

c) Diese Teilaufgabe verlangt einen t-Test für unabhängige Gruppen, also auf der Basis aller Personen, nur für die Variable *kontr.* Zunächst müssen wir die Filterung wieder ausschalten. Das ist mit der Eingabe von *Data* → *Select Cases* → *All Cases* → *OK* erledigt. Beachten Sie, daß die Anzeichen für die Filterung im Datenbildschirm wieder verschwunden sind. Mit *Analyze* → *Compare Means* → *Independent-Samples T Test* rufen wir das nötige Einstellfenster auf:



Wir machen *kontr* zur Testvariablen und *gruppe* mit den Werten 1 und 2 zur Gruppierungsvariablen. Über Continue und OK erhalten wir das Ergebnis:

#### Group Statistics

|               | Gruppe                          | N  | Mean    | Std. Deviation | Std. Error Mean |
|---------------|---------------------------------|----|---------|----------------|-----------------|
| Kontrolltafel | Versuchsgruppe (Alkoholiker)    | 12 | 55,2500 | 8,81244        | 2,54393         |
|               | Kontrollgruppe (nicht Süchtige) | 12 | 51,4167 | 9,79293        | 2,82698         |

#### Independent Samples Test

|               |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                       |   |          |
|---------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|----------|
|               |                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |          |
|               |                             |   |      |                              |        |                 |                 |                       | Lower                                     | Upper    |
| Kontrolltafel | Equal variances assumed     | ,111                                    | ,742 | 1,008                        | 22     | ,324            | 3,8333          | 3,80308               | -4,05376                                  | 11,72043 |
|               | Equal variances not assumed |   |      | 1,008                        | 21,760 | ,325            | 3,8333          | 3,80308               | -4,05882                                  | 11,72548 |

Die zweiseitige Überschreitungswahrscheinlichkeit beträgt im t-Test für homogene Varianzen 32,4 %, für heterogene Varianzen 32,5 %. In beiden Fällen das gesetzte Signifikanzniveau verfehlt, es besteht also kein interpretierbarer Unterschied.

## Aufgabe 05

Eine Psychologin interessiert sich für die Frage, inwieweit die Fähigkeit von Versuchspersonen, die „klassischen“ Aufgaben der experimentellen Problemlöseforschung zu bewältigen, mit deren Fähigkeit, „komplexe“ Probleme in Wirtschaft oder Politik erfolgreich zu bearbeiten, zusammenhängt. Zu diesem Zweck nimmt eine Gruppe von  $n = 75$  Versuchspersonen jeweils an einem „klassischen“ Problemlöseversuch und an einer „komplexen“ Problemlöseaufgabe in Form eines Simulationsspieles auf dem Computer teil. Das „klassische“ Problem lösen 49 Personen, das „komplexe“ 41. Bei beiden Problemen haben 30 Personen Erfolg.

- Bitte stellen Sie die Vierfeldertafel für diese Daten auf und tragen Sie die im Text fehlenden Häufigkeiten nach.
- Berechnen Sie die Vierfelderkorrelation und prüfen Sie sie auf Signifikanz.
- Sind die beiden Variablen, Lösen der „klassischen“ und Lösen der „komplexen“ Aufgabe, stochastisch voneinander unabhängig?

(Quelle: Abschlußklausur Statistik I Wintersemester 1998/99 v. 09.02.1999, Aufg. 3 abgewandelt, Aufgabenblatt *stat123.doc*)

## Lösung

a) Vierfeldertafel:

|                            |      | Komplexes Problem gelöst |      |           |
|----------------------------|------|--------------------------|------|-----------|
|                            |      | ja                       | nein |           |
| Klassisches Problem gelöst | ja   | <b>30</b>                | 19   | <b>49</b> |
|                            | nein | 11                       | 15   | 26        |
|                            |      | <b>41</b>                | 34   | <b>75</b> |

Die im Text enthaltenen Häufigkeiten sind in der Vierfeldertafel fett gedruckt. Die normal gedruckten Zahlen müssen berechnet werden.

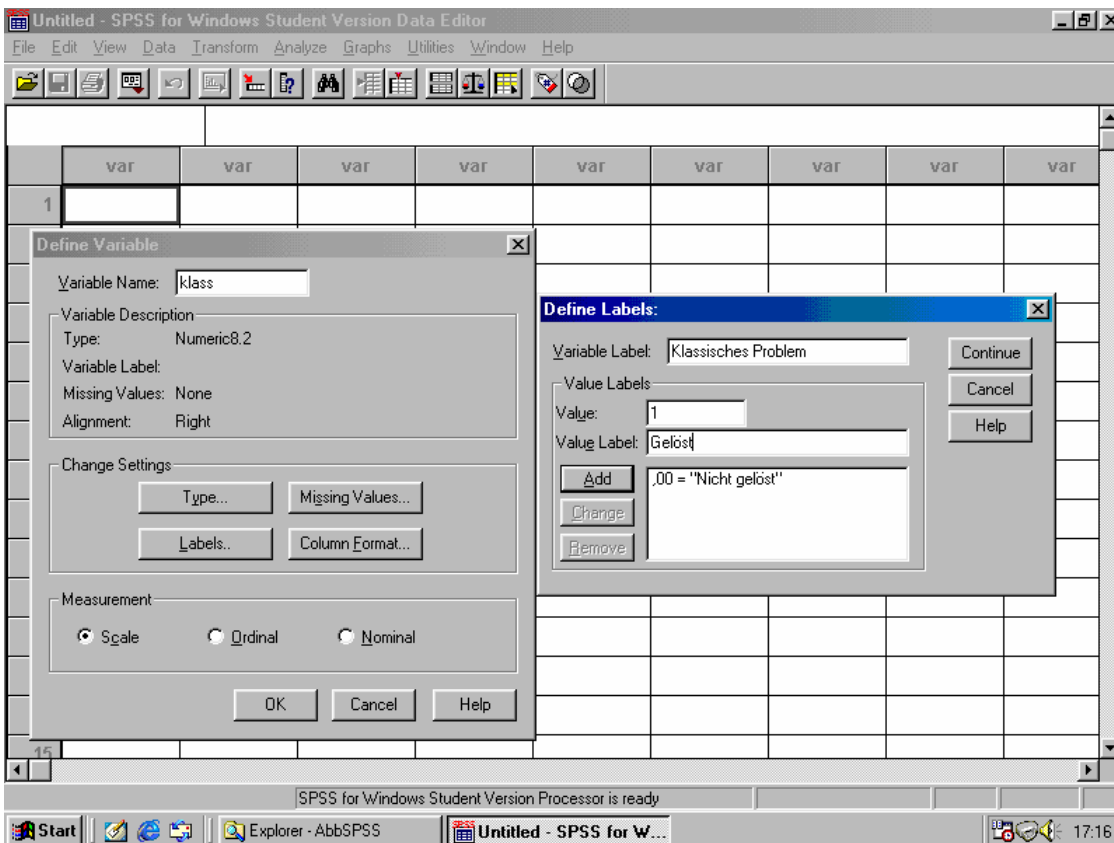
Vor der Eingabe in SPSS müssen Sie sich klarmachen, daß die Vierfeldertafel bereits eine bivariate Häufigkeitsverteilung für zwei dichotome Variablen darstellt, während SPSS normalerweise eine Urliste verlangt. Wir wollen der Variablen *Komplexes Problem gelöst* den Namen *kompl* geben, der Variablen *Klassisches Problem gelöst* den Namen *klass*. *Ja* codieren wir bei beiden Variablen mit 1, *nein* mit 0. Die Urliste liegt uns nicht vor. Wir können sie mit den gegebenen Informationen auch nicht rekonstruieren, uns aber ihre grundsätzliche Struktur klarmachen:

| Variable ↓                  | Wertepaar |        |        |        |
|-----------------------------|-----------|--------|--------|--------|
| <i>klass</i>                | 1         | 1      | 0      | 0      |
| <i>kompl</i>                | 1         | 0      | 1      | 0      |
| Häufigkeit des Wertepaars → | 30 mal    | 19 mal | 11 mal | 15 mal |

(Vergewissern Sie sich, daß Sie verstanden haben, daß es sich hierbei um das Schema einer Urliste für dieses Beispiel handelt. Machen Sie sich den Zusammenhang mit der Vierfeldertafel klar. Stellen Sie mit Papier und Bleistift eine Urliste auf, die zur Vierfeldertafel gehören könnte.)

b) Beginnen Sie die Eingabe in SPSS mit den Variablen- und Bezeichnerdefinitionen aus dem Datenbildschirm heraus. Für die erste Variable sieht das wie folgt aus:





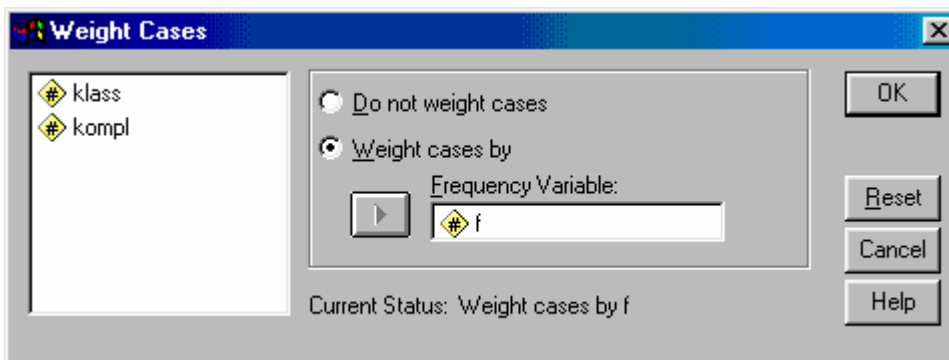
Mit *Add* → *Continue* → *OK* kehren Sie zum Dateneingabebildschirm zurück. Obwohl wir jetzt mit *0* und *1* codierte dichotome Variablen haben, lassen wir *Scale* der Voreinstellung entsprechend aktiv. In der gleichen Weise geben wir Variablennamen, Variablenbezeichner und Wertebezeichner für *kompl* ein. Danach sind die Vorarbeiten für die Eingabe der Urliste beendet. Da wir keine Urliste haben, geben wir nun als Daten jedes mögliche Wertepaar genau einmal ein. Das ist natürlich noch keine Urliste, und mit diesen Daten können wir auch nichts Sinnvolles rechnen.

Es gibt jedoch eine Möglichkeit, SPSS unsere Vierfeldertafel anders als auf dem Umweg über die Urliste einzugeben. Dazu richten wir eine weitere Variable *f* mit dem Variablenbezeichner *Häufigkeit* ein. Hier verzichten wir auf Wertebezeichner. Die Häufigkeiten der Vierfeldertafel geben wir schließlich in der Form ein, die das folgende Bild zeigt, und nach der letzten Eingabe speichern wir unter *Aufgabe05.sav*:

|    | klass | kompl | f     | var | var | var | var | var | var |
|----|-------|-------|-------|-----|-----|-----|-----|-----|-----|
| 1  | 1,00  | 1,00  | 30,00 |     |     |     |     |     |     |
| 2  | 1,00  | ,00   | 19,00 |     |     |     |     |     |     |
| 3  | ,00   | 1,00  | 11,00 |     |     |     |     |     |     |
| 4  | ,00   | ,00   | 15,00 |     |     |     |     |     |     |
| 5  |       |       |       |     |     |     |     |     |     |
| 6  |       |       |       |     |     |     |     |     |     |
| 7  |       |       |       |     |     |     |     |     |     |
| 8  |       |       |       |     |     |     |     |     |     |
| 9  |       |       |       |     |     |     |     |     |     |
| 10 |       |       |       |     |     |     |     |     |     |
| 11 |       |       |       |     |     |     |     |     |     |
| 12 |       |       |       |     |     |     |     |     |     |
| 13 |       |       |       |     |     |     |     |     |     |
| 14 |       |       |       |     |     |     |     |     |     |
| 15 |       |       |       |     |     |     |     |     |     |

Unsere SPSS-Datei enthält jetzt die Vierfeldertafel in einem charakteristischen Code: Die Häufigkeiten der einzelnen Zellen sind als Werte der Variablen *f* eingetragen.

Bevor wir unsere Vierfeldertafel auswerten können, müssen wir SPSS die besondere Bedeutung der Variablen *f* für diesen Zweck mitteilen. Dazu rufen wir aus dem Datenbildschirm *Data* → *Weight Cases* auf. Hier klicken wir *Weight cases by* an und bringen *f* in das Feld *Frequency Variable*, wie es das folgende Bild zeigt.



Mit *OK* kehren wir in den Datenbildschirm zurück, der uns jetzt mit *Weight On* in der Fußzeile (schwarzer Pfeil) anzeigt, daß eine Gewichtungsfunktion aktiv ist, die SPSS bei den folgenden Auswertungen als absolute oder relative Häufigkeiten interpretiert.

Aufgabe05 - SPSS for Windows Student Version Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

|    | klass | kompl | f     | var | var | var | var | var | var |
|----|-------|-------|-------|-----|-----|-----|-----|-----|-----|
| 1  | 1,00  | 1,00  | 30,00 |     |     |     |     |     |     |
| 2  | 1,00  | ,00   | 19,00 |     |     |     |     |     |     |
| 3  | ,00   | 1,00  | 11,00 |     |     |     |     |     |     |
| 4  | ,00   | ,00   | 15,00 |     |     |     |     |     |     |
| 5  |       |       |       |     |     |     |     |     |     |
| 6  |       |       |       |     |     |     |     |     |     |
| 7  |       |       |       |     |     |     |     |     |     |
| 8  |       |       |       |     |     |     |     |     |     |
| 9  |       |       |       |     |     |     |     |     |     |
| 10 |       |       |       |     |     |     |     |     |     |
| 11 |       |       |       |     |     |     |     |     |     |
| 12 |       |       |       |     |     |     |     |     |     |
| 13 |       |       |       |     |     |     |     |     |     |
| 14 |       |       |       |     |     |     |     |     |     |
| 15 |       |       |       |     |     |     |     |     |     |

SPSS for Windows Student Version Processor is ready

Weight On

Alle statistischen Rechnungen erledigen wir mit der schon aus Aufgabe 02 bekannten Eingabe *Analyze* → *Descriptive Statistics* → *Crosstabs*. Im jetzt offenen Kreuztabellefenster bringen wir *klass* in das Feld für die Zeilen, *kompl* in das Feld für die Spalten. Aus der großen Auswahl unter *Statistics* fordern wir *Chi-square*, *Correlations* und *Contingency coefficient* an, wie das nächste Bild zeigt:

Crosstabs: Statistics

Chi-square

Correlations

Nominal

Contingency coefficient

Phi and Cramér's V

Lambda

Uncertainty coefficient

Ordinal

Gamma

Somers' d

Kendall's tau-b

Kendall's tau-c

Nominal by Interval

Eta

Kappa

Risk

McNemar

Cochran's and Mantel-Haenszel statistics

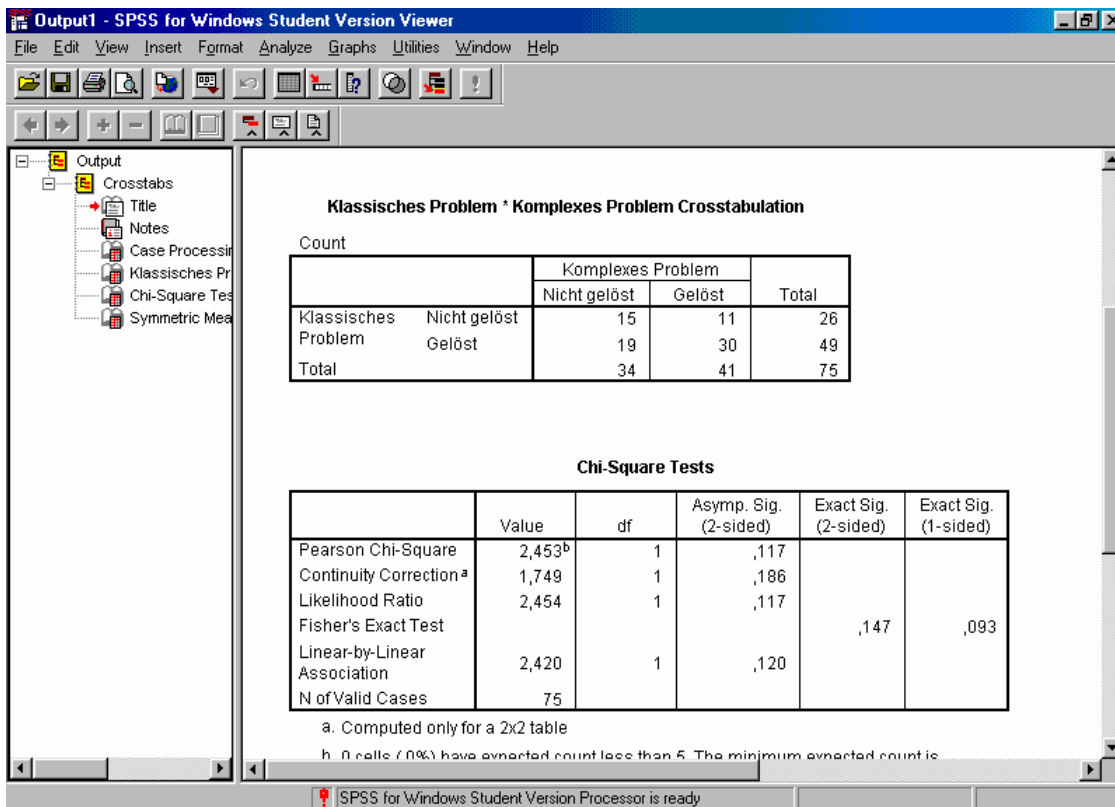
Test common odds ratio equals: 1

Continue

Cancel

Help

Nach *Continue* starten wir die Berechnung mit *OK*. Das Ergebnis zeigt der Ausgabebildschirm:



Unter den Tabellen befinden sich unsere Vierfeldertafel in der gewohnten Form und die Ergebnisse verschiedener Signifikanztests, darunter der Vierfelder-Chi-quadrat-Test (Pearson) mit und ohne Kontinuitätskorrektur. Dieser Test ist nicht signifikant, die Nullhypothese einer stochastischen Unabhängigkeit der beiden Aufgaben wird also beibehalten (Teilaufgabe c).

Eine weitere Tabelle enthält Maße für die Korrelation: den Kontingenzkoeffizienten, die Produktmomentkorrelation (Pearson's R), die hier, wie in der Vorlesung gezeigt, der Punkt-Vierfelderkorrelation entspricht, und die Rangkorrelation nach Spearman. (Warum muß diese hier den gleichen Zahlenwert liefern?)

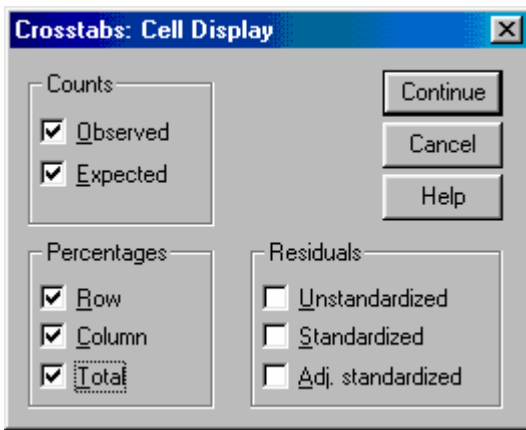
**Symmetric Measures**

|                      |                         | Value | Asymp. Std. Error | Approx. T | Approx. Sig. |
|----------------------|-------------------------|-------|-------------------|-----------|--------------|
| Nominal by Nominal   | Contingency Coefficient | ,178  |                   |           | ,117         |
| Interval by Interval | Pearson's R             | ,181  | ,114              | 1,571     | ,121         |
| Ordinal by Ordinal   | Spearman Correlation    | ,181  | ,114              | 1,571     | ,121         |
| N of Valid Cases     |                         | 75    |                   |           |              |

- a Not assuming the null hypothesis.
- b Using the asymptotic standard error assuming the null hypothesis.
- c Based on normal approximation.

Die Korrelationen sind niedrig und, ebenso wie Chi-quadrat, nicht signifikant. Es besteht also kein statistischer Zusammenhang zwischen dem Lösen „klassischer“ und „komplexer“ Probleme. Das ist in der Tat das Resultat, das Dietrich Dörner in vielen Untersuchungen zum komplexen Problemlösen gefunden hat.

Im Kreuztabellenfenster, das sich nach *Analyze* → *Descriptive Statistics* → *Crosstabs* öffnet, können wir *Cells* anklicken, was wir bisher bei dieser Aufgabe unterlassen haben. Damit erhalten wir das Fenster:



Wenn wir es ausfüllen, wie es hier abgebildet ist, erhalten wir nach *Continue* und *OK* eine sehr informative Vierfeldertafel:

Klassisches Problem \* Komplexes Problem Crosstabulation

|                     |              | Komplexes Problem            |        | Total  |        |
|---------------------|--------------|------------------------------|--------|--------|--------|
|                     |              | Nicht gelöst                 | Gelöst |        |        |
| Klassisches Problem | Nicht gelöst | Count                        | 15     | 11     | 26     |
|                     |              | Expected Count               | 11,8   | 14,2   | 26,0   |
|                     |              | % within Klassisches Problem | 57,7%  | 42,3%  | 100,0% |
|                     |              | % within Komplexes Problem   | 44,1%  | 26,8%  | 34,7%  |
|                     |              | % of Total                   | 20,0%  | 14,7%  | 34,7%  |
| Gelöst              |              | Count                        | 19     | 30     | 49     |
|                     |              | Expected Count               | 22,2   | 26,8   | 49,0   |
|                     |              | % within Klassisches Problem | 38,8%  | 61,2%  | 100,0% |
|                     |              | % within Komplexes Problem   | 55,9%  | 73,2%  | 65,3%  |
|                     |              | % of Total                   | 25,3%  | 40,0%  | 65,3%  |
| Total               |              | Count                        | 34     | 41     | 75     |
|                     |              | Expected Count               | 34,0   | 41,0   | 75,0   |
|                     |              | % within Klassisches Problem | 45,3%  | 54,7%  | 100,0% |
|                     |              | % within Komplexes Problem   | 100,0% | 100,0% | 100,0% |
|                     |              | % of Total                   | 45,3%  | 54,7%  | 100,0% |

Die einzelnen Zellen der Vierfeldertafel (grau hinterlegt), ebenso wie die Zellen der Randspalte und der Fußzeile, enthalten jetzt neben den beobachteten absoluten Häufigkeiten auch die erwarteten absoluten Häufigkeiten, die dem Chi-quadrat-Test zugrundeliegen, und verschiedene prozentuale relative Häufigkeiten, die natürlich auch Schätzer der zugrundeliegenden Wahrscheinlichkeiten sind. Vergleichen Sie diese Zahlen mit den verschiedenen Wahrscheinlichkeiten, die in der Vorlesung Statistik I an der Vierfeldertafel berechnet wurden! Sie werden feststellen, daß *% of Total* den Schnittmengenwahrscheinlichkeiten, *% within Klassisches Problemlösen* den zeilenweisen bedingten Wahrscheinlichkeiten und *% within Komplexes Problemlösen* den spaltenweisen bedingten Wahrscheinlichkeiten entsprechen.

## Aufgabe 06

Gegeben ist die Urliste

x 13 12 25 11 24 18 15 11 26 21 17 10 14 16 23 14  
10 27.

- Bitte stellen Sie die Häufigkeitsverteilung als Tabelle auf. Fassen Sie dabei immer drei der gegebenen Maßzahlklassen zusammen und transformieren Sie die Maßzahlen so, daß die neuen Maßzahlen in Einerschritten variieren. Legen Sie Ihre Maßzahlklassengrenzen so, daß aus der alten Maßzahl  $x = 17$  die neue Maßzahl  $x_t = 0$  wird. **Bearbeiten Sie alle folgenden Teilaufgaben ausschließlich mit den transformierten Maßzahlen  $X_t$ .**
- Zeichnen Sie die Häufigkeitsverteilung als Histogramm.
- Zeichnen Sie in einer weiteren Darstellung die kumulierte Häufigkeitsverteilung.
- Berechnen Sie die Quartile  $Q_1$ ,  $Q_2 = Md$  und  $Q_3$ . Berechnen Sie aufgrund der Werte, die SPSS liefert, den Interquartilabstand und den mittleren Quartilabstand.
- Berechnen Sie den arithmetischen Mittelwert und die Standardabweichung.
- Berechnen Sie ein Maß für die Schiefe der Verteilung, das auf der Methode der Momente beruht. Interpretieren Sie es kurz.
- Berechnen Sie ein Maß für den Exzeß (die Breitgipfligkeit) der Verteilung, das auf der Methode der Momente beruht. Interpretieren Sie es kurz.

(Quelle: Abschlußklausur Statistik I Wintersemester 1998/99 v. 09.02.1999, Aufg. 1 abgewandelt, *stat123.doc*)

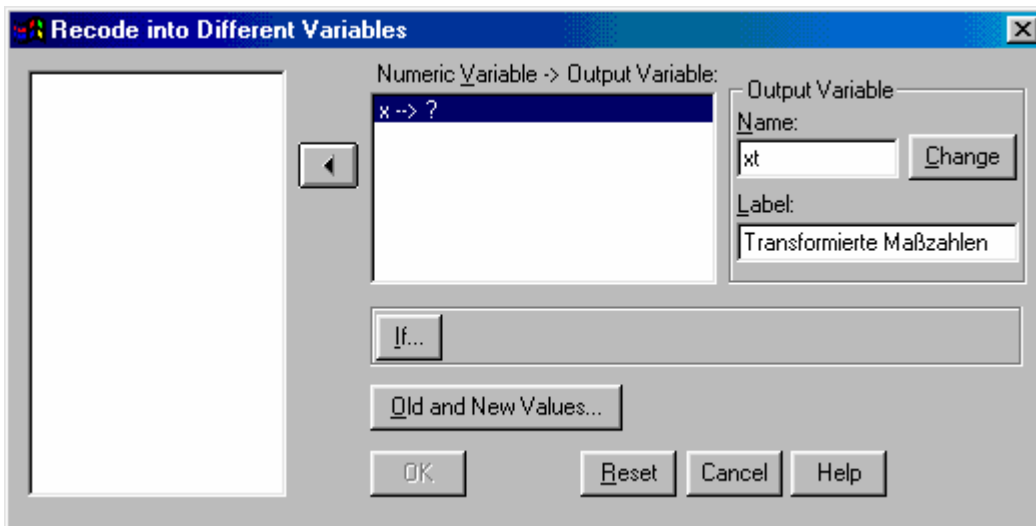
## Lösung

Die Aufgabe entspricht im wesentlichen der Aufgabe 01. Die dort gezeigten Lösungsschritte sollen hier nicht wiederholt werden. Neu ist aber die Maßzahlklassenzusammenfassung und die Maßzahlentransformation. Dafür gibt es in SPSS verschiedene Möglichkeiten. Wir benutzen eine davon.

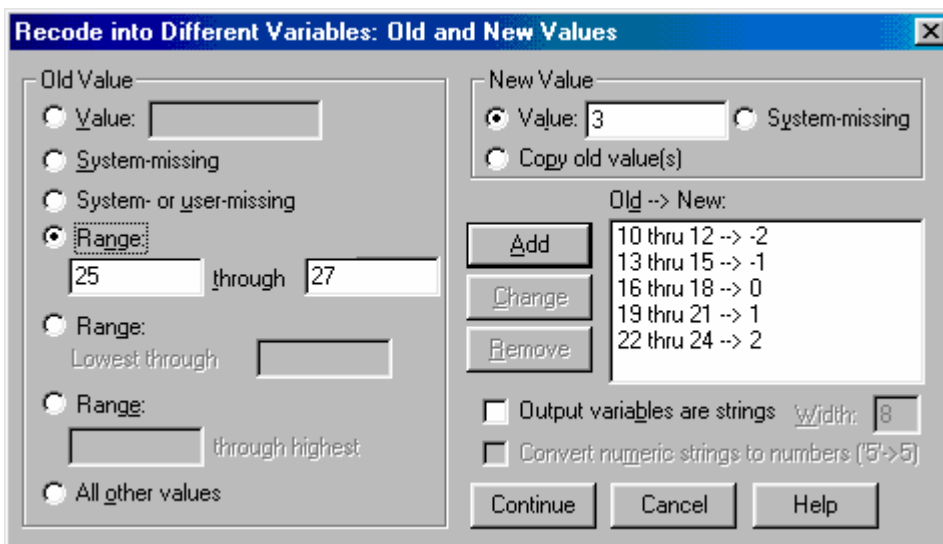
a) Wir beginnen mit der gewohnten Dateneingabe. Die Variable nennen wir  $x$  mit dem Variablenbezeichner *Maßzahl*. Die Datei speichern wir unter *Aufgabe06.sav* ab. Zur Vorbereitung der Maßzahlklassenzusammenfassung und Maßzahlentransformation stellen wir von Hand die folgende Tabelle für die Zuordnung der ursprünglichen zu den transformierten Maßzahlen auf:

| x     | 10-12 | 13-15 | 16-18 | 19-21 | 22-24 | 25-27 |
|-------|-------|-------|-------|-------|-------|-------|
| $x_t$ | -2    | -1    | 0     | 1     | 2     | 3     |

Im Datenbildschirm von SPSS klicken wir *Transform* → *Recode* → *Into Different Variables* an. Das sich öffnende Fenster füllen wir nach der Abbildung aus:



Wir bringen  $x$  in das Feld *Numeric Variable*  $\rightarrow$  *Output Variable*, geben für *Output Variable*  $xt$  und für *Label* *Transformierte Maßzahlen* ein. Diese Eingaben beenden wir mit *Change*. Mit *Old and New Values* öffnen wir das Recodierfenster:



Die Werte von  $x$ , denen ein Wert von  $xt$  zugeordnet werden soll, werden links unter *Old Value - Range* eingetragen, der Wert von  $xt$  rechts unter *New Value - Value*. Mit *Add* wird jeweils eine Zuordnung in das Feld *Old*  $\rightarrow$  *New* übertragen. Das Bild zeigt den Zustand des Fensters vor dem letzten *Add*. Nach dessen Eingabe kehren wir mit *Continue* in das vorangehende Fenster zurück, das wir mit *OK* verlassen. Der Datenbildschirm zeigt jetzt unter der Variablen  $xt$  die transformierten Werte. Wir speichern die Datei unter *Aufgabe06.sav*:

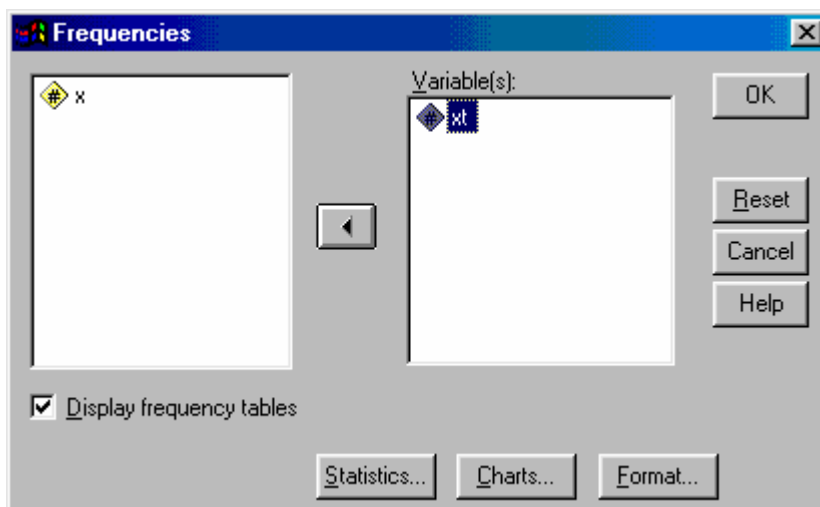
Aufgabe06 - SPSS for Windows Student Version Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

|    | x     | xt    | var | var | var | var | var | var | var |
|----|-------|-------|-----|-----|-----|-----|-----|-----|-----|
| 1  | 13,00 | -1,00 |     |     |     |     |     |     |     |
| 2  | 12,00 | -2,00 |     |     |     |     |     |     |     |
| 3  | 25,00 | 3,00  |     |     |     |     |     |     |     |
| 4  | 11,00 | -2,00 |     |     |     |     |     |     |     |
| 5  | 24,00 | 2,00  |     |     |     |     |     |     |     |
| 6  | 18,00 | ,00   |     |     |     |     |     |     |     |
| 7  | 15,00 | -1,00 |     |     |     |     |     |     |     |
| 8  | 11,00 | -2,00 |     |     |     |     |     |     |     |
| 9  | 26,00 | 3,00  |     |     |     |     |     |     |     |
| 10 | 21,00 | 1,00  |     |     |     |     |     |     |     |
| 11 | 17,00 | ,00   |     |     |     |     |     |     |     |
| 12 | 10,00 | -2,00 |     |     |     |     |     |     |     |
| 13 | 14,00 | -1,00 |     |     |     |     |     |     |     |
| 14 | 16,00 | ,00   |     |     |     |     |     |     |     |
| 15 | 23,00 | 2,00  |     |     |     |     |     |     |     |

SPSS for Windows Student Version Processor is ready

Die geforderten Berechnungen starten wir in einem Arbeitsgang über *Analyze* → *Descriptive Statistics* → *Frequencies*. Im sich öffnenden Fenster bringen wir *xt* unter *Variables*:



Wir rufen *Statistics* auf und fordern die gewünschten Statistiken an, wie in Aufgabe 01. Mit *Charts* können wir Graphiken erzeugen, was jetzt übergangen werden soll. Mit *Continue* und *OK* starten wir schließlich die Auswertung. Der Ausgabebildschirm zeigt in den beiden folgenden Tabellen die Statistiken und die Häufigkeitsverteilung, beides wie gewünscht für die transformierten Maßzahlen.

Statistics  
Transformierte Maßzahlen

|                        |         |        |
|------------------------|---------|--------|
| N                      | Valid   | 18     |
|                        | Missing | 0      |
| Mean                   |         | ,0000  |
| Median                 |         | -,4286 |
| Mode                   |         | -2,00  |
| Std. Deviation         |         | 1,8787 |
| Variance               |         | 3,5294 |
| Skewness               |         | ,539   |
| Std. Error of Skewness |         | ,536   |



|                        |    |         |
|------------------------|----|---------|
| Kurtosis               |    | -1,190  |
| Std. Error of Kurtosis |    | 1,038   |
| Range                  |    | 5,00    |
| Percentiles            | 25 | -1,5556 |
|                        | 50 | -,4286  |
|                        | 75 | 1,6667  |

- a Calculated from grouped data.  
b Percentiles are calculated from grouped data.

#### Transformierte Maßzahlen

|       |       | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | -2,00 | 5         | 27,8    | 27,8          | 27,8               |
|       | -1,00 | 4         | 22,2    | 22,2          | 50,0               |
|       | ,00   | 3         | 16,7    | 16,7          | 66,7               |
|       | 1,00  | 1         | 5,6     | 5,6           | 72,2               |
|       | 2,00  | 2         | 11,1    | 11,1          | 83,3               |
|       | 3,00  | 3         | 16,7    | 16,7          | 100,0              |
|       | Total | 18        | 100,0   | 100,0         |                    |

### Aufgabe 07

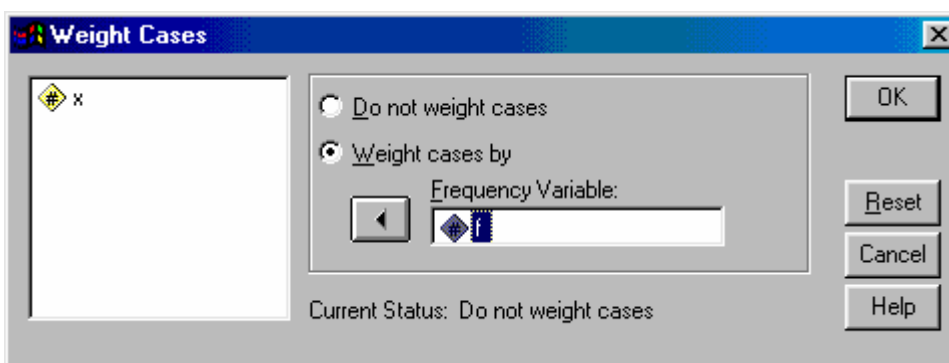
Bei einem Würfel, dessen 6 Oberflächen die Augenzahlen von 1 bis 6 tragen, soll überprüft werden, ob er unverfälscht ist, also beim Würfeln eine Verteilung der Augenzahlen liefert, die mit der Nullhypothese eines Laplace-Ereignisraumes im Einklang steht. Der Würfel wird  $n = 360$  mal geworfen. Man erhält die folgenden Augenzahlen:

|               |    |    |    |    |    |     |
|---------------|----|----|----|----|----|-----|
| Augenzahl $x$ | 1  | 2  | 3  | 4  | 5  | 6   |
| $f$           | 42 | 62 | 58 | 91 | 49 | 58. |

Sind die Abweichungen der beobachteten Häufigkeiten signifikant? ( $\alpha = 5\%$ )

### Lösung

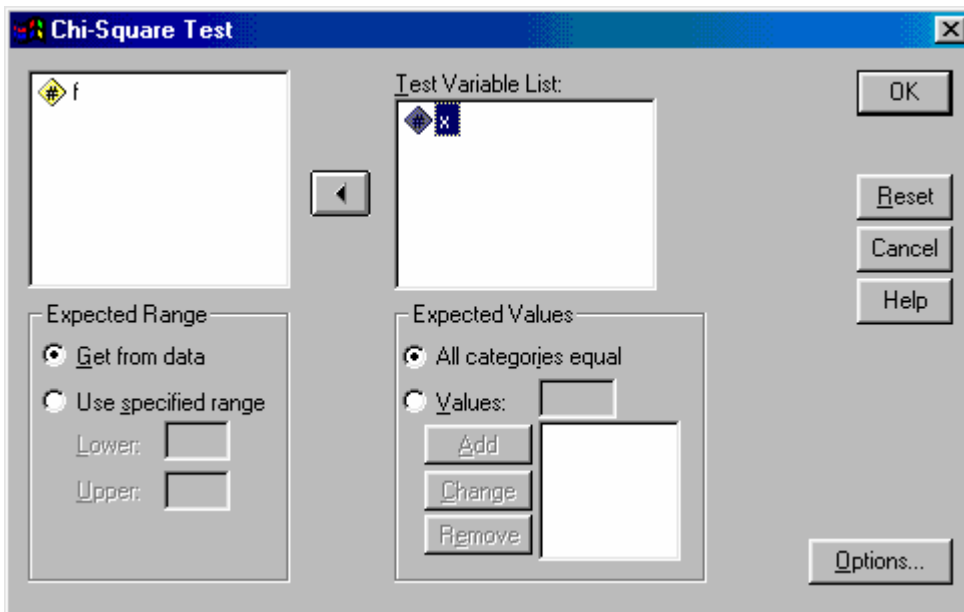
Wie in der Aufgabe 05 haben wir hier keine Urliste, sondern schon eine Häufigkeitsverteilung. Wir geben also die Variable  $x$  und die Variable  $f$  in SPSS ein. Dann verwenden wir wieder die Gewichtungsfunktion. Mit *Data* → *Weight Cases* erhalten wir das Fenster für die Gewichtungsfunktion, das wir ausfüllen, wie es abgebildet ist:



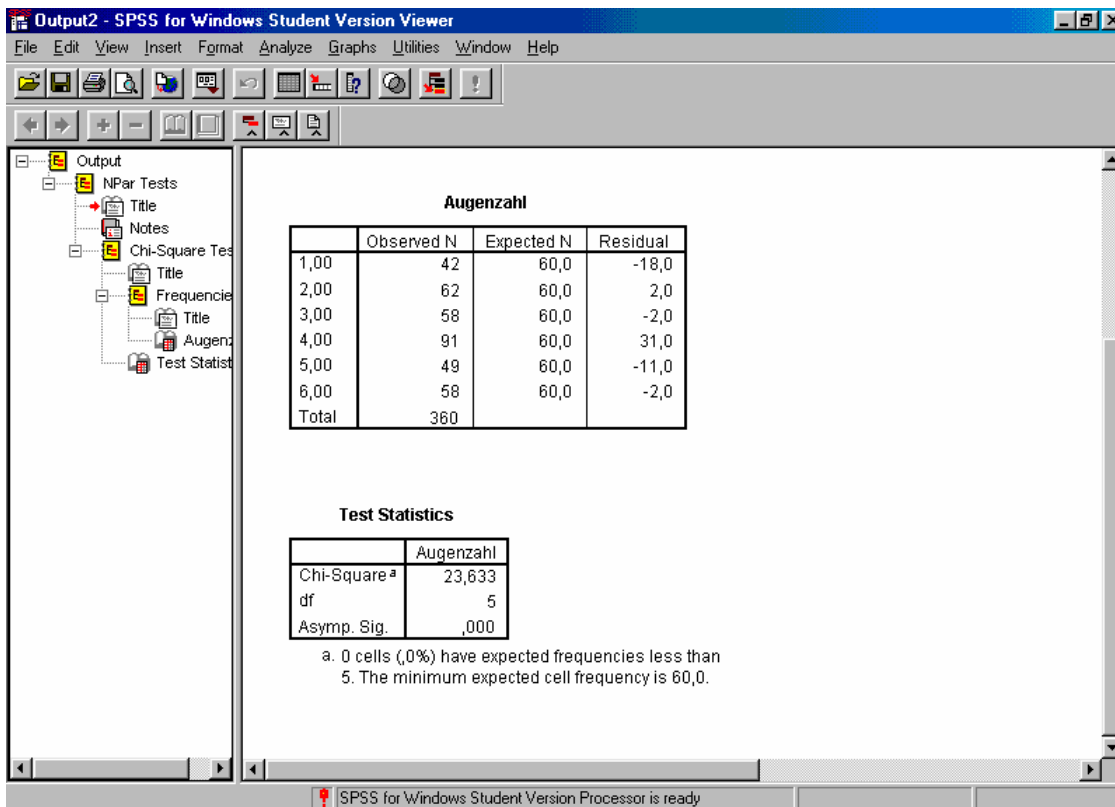
Mit OK kehren wir zum Datenbildschirm zurück, der uns jetzt wieder anzeigt, daß die Gewichtungsfunktion aktiv ist (schwarzer Pfeil). Speichern Sie Ihre Datei unter *Aufgabe07.sav*. Hier ist noch auf eine Besonderheit hinzuweisen. Wenn Sie die Datei nach dem Setzen der Gewichtungsfunktion speichern, wird diese Funktion mit abgespeichert. Sie ist dann beim Laden der Datei zu einem späteren Zeitpunkt von Anfang an aktiv.

|    | x    | f     | var | var | var | var | var | var | var |
|----|------|-------|-----|-----|-----|-----|-----|-----|-----|
| 1  | 1,00 | 42,00 |     |     |     |     |     |     |     |
| 2  | 2,00 | 62,00 |     |     |     |     |     |     |     |
| 3  | 3,00 | 58,00 |     |     |     |     |     |     |     |
| 4  | 4,00 | 91,00 |     |     |     |     |     |     |     |
| 5  | 5,00 | 49,00 |     |     |     |     |     |     |     |
| 6  | 6,00 | 58,00 |     |     |     |     |     |     |     |
| 7  |      |       |     |     |     |     |     |     |     |
| 8  |      |       |     |     |     |     |     |     |     |
| 9  |      |       |     |     |     |     |     |     |     |
| 10 |      |       |     |     |     |     |     |     |     |
| 11 |      |       |     |     |     |     |     |     |     |
| 12 |      |       |     |     |     |     |     |     |     |
| 13 |      |       |     |     |     |     |     |     |     |
| 14 |      |       |     |     |     |     |     |     |     |
| 15 |      |       |     |     |     |     |     |     |     |

Den Chi-quadrat-Test auf Gleichverteilung rufen wir über *Analyze* → *Nonparametric Tests* → *Chi-Square* auf:



Wir bringen die Variable *x* in das Feld *Test Variable List*. Das Feld *Expected Values* enthält die Nullhypothese der Gleichverteilung als Voreinstellung *All categories equal*. Nach dem Anklicken von *Values* könnte man auch erwartete Häufigkeiten eingeben, die man aufgrund einer anderen Nullhypothese anderenorts berechnet hat. Das Fenster *Options* erlaubt uns, die Behandlung fehlender Werte zu steuern. Wir belassen es bei der Voreinstellung und klicken dieses Feld nicht an. Nach *OK* erhalten wir den Ausgabebildschirm:



Er enthält die Verteilung der beobachteten und erwarteten Häufigkeiten sowie deren Differenz. Schließlich wird uns Chi-quadrat, die Zahl seiner Freiheitsgrade und die Überschreitungswahrscheinlichkeit angegeben. Im Beispiel ist Chi-quadrat signifikant auf dem 1-Promille-Niveau, es muß also angenommen werden, daß der Würfel nicht korrekt ist.

### Aufgabe 08

Ein Psychologe hat die Hypothese, daß die Zeit, die Versuchspersonen zum Lösen eines bestimmten Problems benötigen, von der Art der gegebenen Hilfen abhängt. Er untersucht deshalb  $n_1 = 14$  Personen bei der Problemlösung mit konkreten und  $n_2 = 14$  Personen bei der Problemlösung mit abstrakten Hinweisen. Abhängige Variable ist die benötigte Zeit in Minuten. Er erhält die folgenden Resultate:

Hinweise auf Lösungsweg:

|          |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| konkret  | 10 | 16 | 12 | 20 | 15 | 11 | 17 | 19 | 13 | 10 | 17 | 14 | 21 | 16 |
| abstrakt | 17 | 22 | 12 | 14 | 25 | 13 | 15 | 23 | 16 | 27 | 15 | 19 | 24 | 17 |

Läßt sich die Hypothese, konkrete Hilfen seien wirksamer als abstrakte, aufrechterhalten ( $\alpha \leq 5\%$ , zweiseitig)?

(Quelle: Nachholklausur Statistik II Sommersemester 1998 v. 14.12.1998, Aufg. 6, Übungsblatt *stat211.doc*)

### Lösung

Die Aufgabe verlangt den Mittelwertsvergleich zwischen zwei unabhängigen Gruppen. Zeitmessungen haben Intervallskalenniveau. Auch wenn die Verteilungen meistens rechtsschief sind, wird der t-Test üblicherweise als zulässig angesehen. Der Lösungsweg entspricht dem von Aufgabe 03 und soll deshalb nicht in den Details wiedergegeben werden.

Definieren Sie die Variable *x* mit dem Bezeichner *Problemlösezeit* und die Variable *ind* mit dem Bezeichner *Hinweis* sowie den Wertebezeichnern *konkret* (1) und *abstrakt* (2). Ihr Datenbildschirm sieht dann so aus:

|    | x     | ind  | var | var | var | var | var | var | var |
|----|-------|------|-----|-----|-----|-----|-----|-----|-----|
| 8  | 19,00 | 1,00 |     |     |     |     |     |     |     |
| 9  | 13,00 | 1,00 |     |     |     |     |     |     |     |
| 10 | 10,00 | 1,00 |     |     |     |     |     |     |     |
| 11 | 17,00 | 1,00 |     |     |     |     |     |     |     |
| 12 | 14,00 | 1,00 |     |     |     |     |     |     |     |
| 13 | 21,00 | 1,00 |     |     |     |     |     |     |     |
| 14 | 16,00 | 1,00 |     |     |     |     |     |     |     |
| 15 | 17,00 | 2,00 |     |     |     |     |     |     |     |
| 16 | 22,00 | 2,00 |     |     |     |     |     |     |     |
| 17 | 12,00 | 2,00 |     |     |     |     |     |     |     |
| 18 | 14,00 | 2,00 |     |     |     |     |     |     |     |
| 19 | 25,00 | 2,00 |     |     |     |     |     |     |     |
| 20 | 13,00 | 2,00 |     |     |     |     |     |     |     |
| 21 | 15,00 | 2,00 |     |     |     |     |     |     |     |
| 22 | 23,00 | 2,00 |     |     |     |     |     |     |     |

Speichern Sie Ihre Daten unter *Aufgabe08.sav* ab.

Über *Analyze* → *Compare Means* → *Independent-Samples T Test* erhalten Sie das Steuerfenster für den gewünschten t-Test. Nachdem Sie es ausgefüllt und mit *OK* verlassen haben, erhalten Sie:

Group Statistics

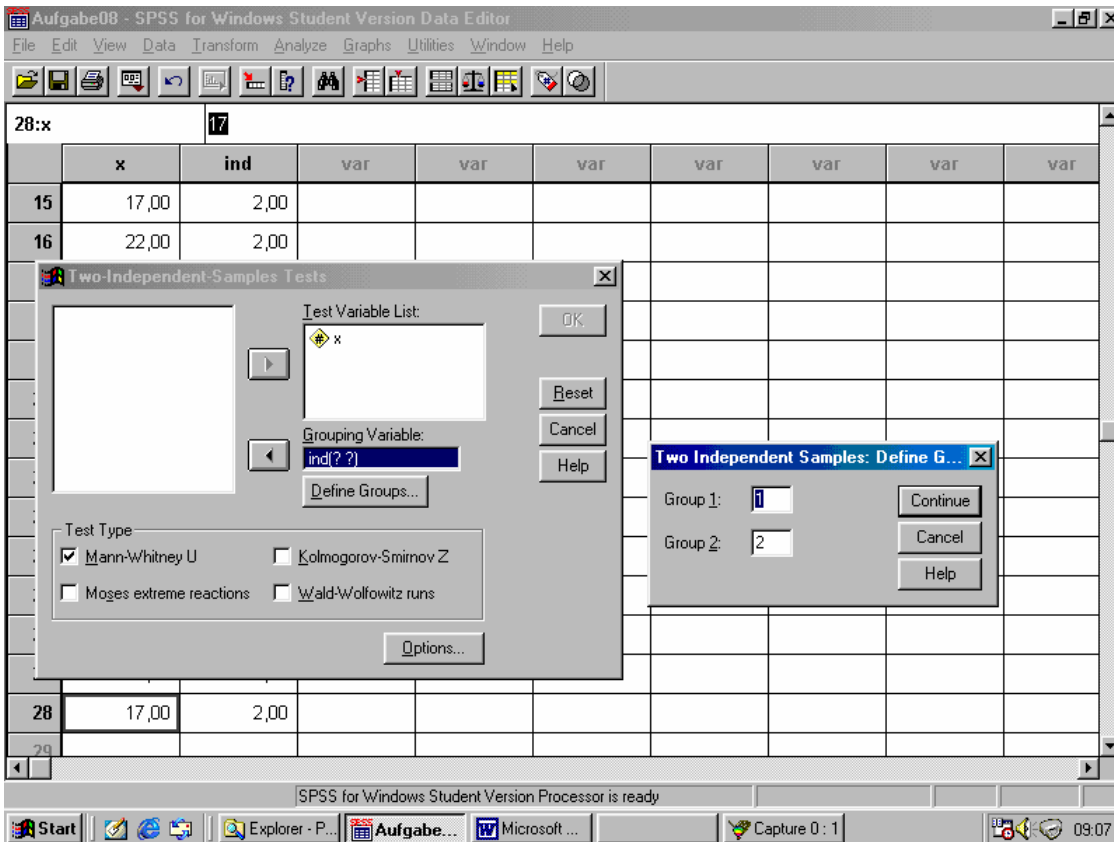
|                 | Hinweis  | N  | Mean    | Std. Deviation | Std. Error Mean |
|-----------------|----------|----|---------|----------------|-----------------|
| Problemlösezeit | konkret  | 14 | 15,0714 | 3,5834         | ,9577           |
|                 | abstrakt | 14 | 18,5000 | 4,8477         | 1,2956          |

Independent Samples Test

|                 |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                       |   |        |
|-----------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|--------|
|                 |                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |        |
|                 |                             |   |      |                              |        |                 |                 |                       | Lower                                     | Upper  |
| Problemlösezeit | Equal variances assumed     | 2,395                                   | ,134 | -2,128                       | 26     | ,043            | -3,4286         | 1,6111                | -6,7403                                   | -,1168 |
|                 | Equal variances not assumed |   |      | -2,128                       | 23,940 | ,044            | -3,4286         | 1,6111                | -6,7542                                   | -,1029 |

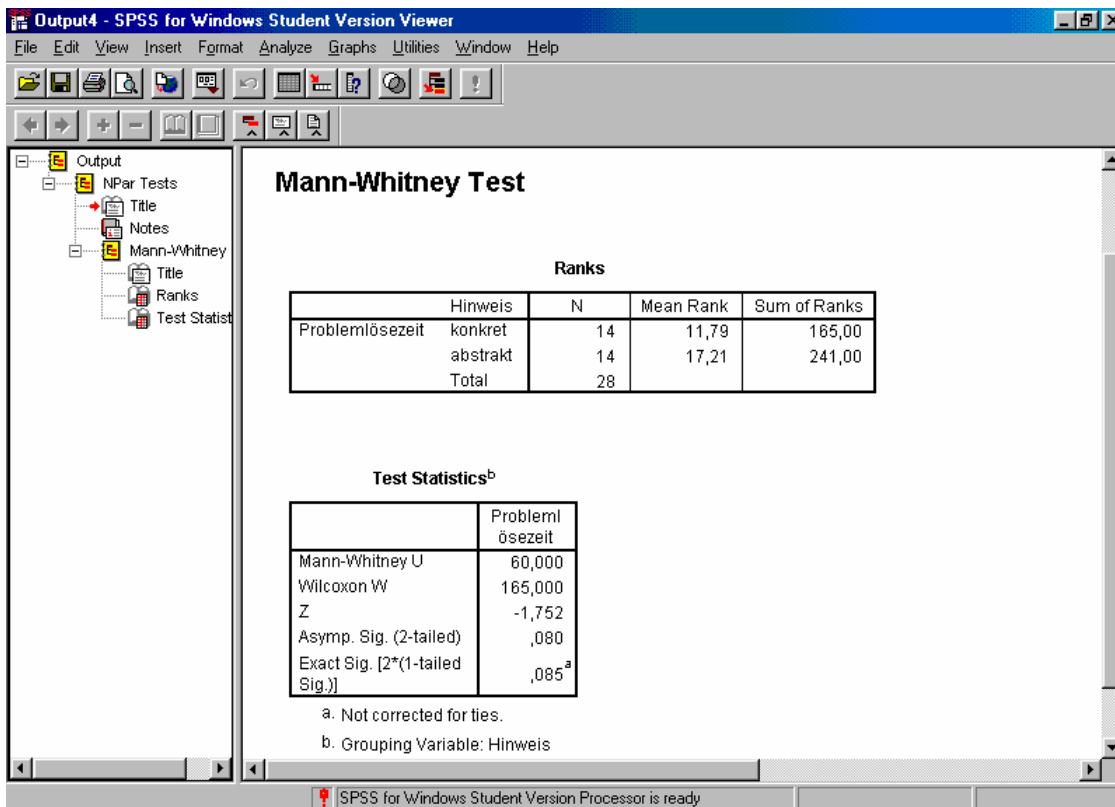
Die Ergebnisse besagen, daß wir die Varianzen als homogen annehmen können. Der Mittelwertsunterschied in der Problemlösezeit für konkrete minus abstrakte Hilfen beträgt -3,429 Minuten und ist auf 5%-Niveau zweiseitig signifikant. Konkrete Hilfen sind also wirksamer als abstrakte.

Es könnte sein, daß Sie Zweifel daran bekommen, ob die Voraussetzungen für den t-Test hier wirklich erfüllt sind. Bei Nichterfüllung wäre, wie Sie in der Vorlesung Statistik II hören, der U-Test nach Mann-Whitney anzuwenden. Das ist in SPSS sehr einfach. Sie erhalten über *Analyze* → *Nonparametric Tests* → *2 Independent Samples* das passende Steuerfenster:



Sie müssen es in der gleichen Weise ausfüllen, wie beim t-Test. Im Feld *Test Type* wählen Sie, wenn es nicht schon eingestellt ist, *Mann-Whitney U*. Die anderen hier noch angebotenen Tests werden in der Vorlesung nicht behandelt.

Nach *Continue* und *OK* erhalten Sie:



Unter *Asymp. Sig. (2-tailed)* sehen Sie, daß der Unterschied zwischen beiden Stichproben in diesem Test die zweiseitige Signifikanzschranke für  $\alpha = 5\%$  nicht erreicht. Aufgrund dieses Tests dürfen wir also nicht auf eine unterschiedliche Wirkung von konkreten und abstrakten Hilfen in diesem Problemlöseversuch schließen.

Für die Praxis heißt das, daß man natürlich nicht einfach beide Tests, t-Test und U-Test, machen und sich, falls die Ergebnisse voneinander abweichen, dasjenige aussuchen kann, das einem am besten paßt. Im vorliegenden Beispiel müßte gründlich nachgedacht werden, ob die Voraussetzungen des t-Tests als erfüllt angesehen werden können. In diesem Fall ist dann auch dessen Ergebnis zu interpretieren. Kommt man zum Resultat, daß die Voraussetzungen des t-Tests als nicht erfüllt gelten müssen, darf nur noch der U-Test interpretiert werden.

### Aufgabe 09

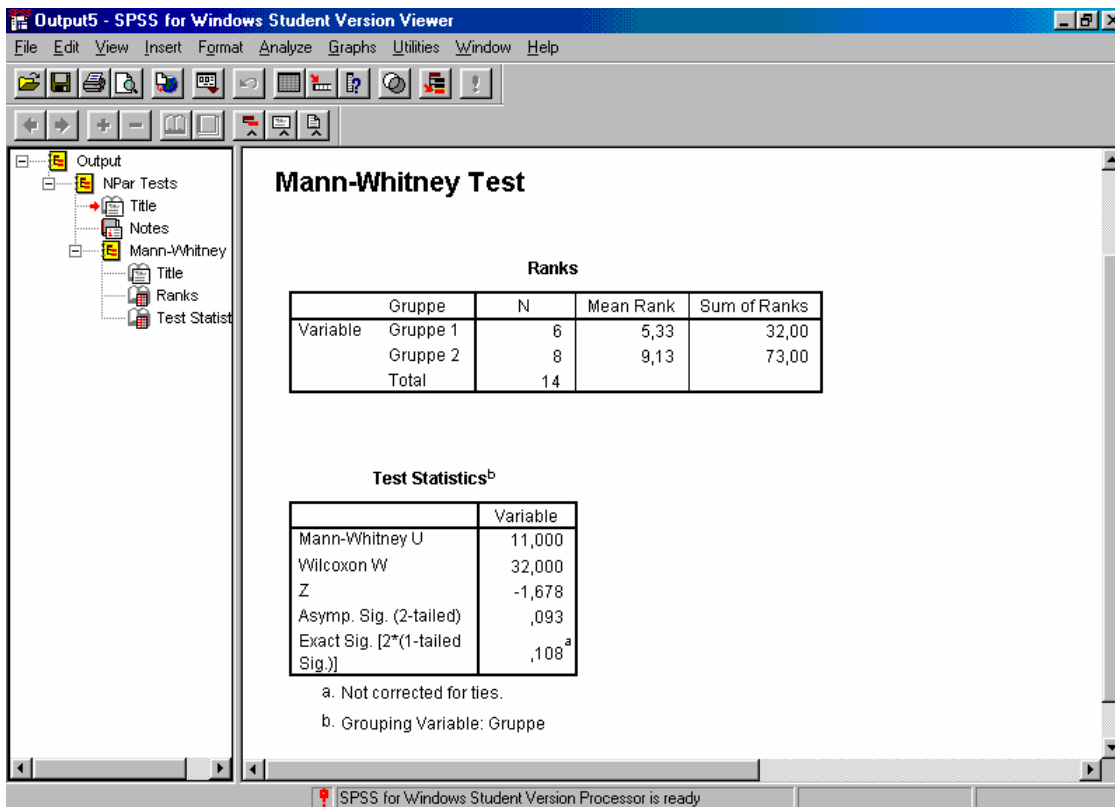
Diese Aufgabe dient zur Wiederholung des in Aufgabe 08 bereits behandelten U-Tests. Gegeben sei die Urliste:

Gruppe 1    8   13    2   18    6   17  
 Gruppe 2    25   7   15   10   23   20   22   16.

Prüfen Sie den Unterschied in der Zentraltendenz zwischen beiden Gruppen mit einem nichtparametrischen Verfahren, dem U-Test, auf Signifikanz ( $\alpha = 5\%$  zweiseitig).

### Lösung

Die Lösung folgt den in der Aufgabe 08 im einzelnen dargestellten Schritten. Speichern Sie Ihre Rohdatei unter *Aufgabe09.sav* ab. Mit *Analyze* → *Nonparametric Tests* → *2 Independent Samples* erhalten Sie das Steuerfenster für den U-Test. Füllen Sie es einschließlich *Grouping Variable* und *Define Groups* aus und verlassen Sie es mit *OK*. Sie erhalten:



Auch hier wird, wie die Tabelle *Test Statistics* zeigt, die gesetzte Signifikanzschranke nicht erreicht, auf einen Unterschied zwischen beiden Gruppen kann also nicht geschlossen werden. Außer *U* enthält die Tabelle auch noch die Prüfgröße *W*. Sie gehört zu einer etwas abweichenden Form dieses Tests, die unter dem Namen *Wilcoxon-White* bekannt ist, in der Vorlesung aber nicht behandelt wird

### Aufgabe 10

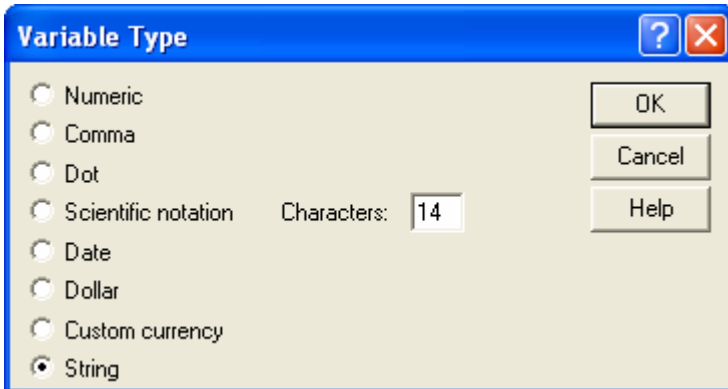
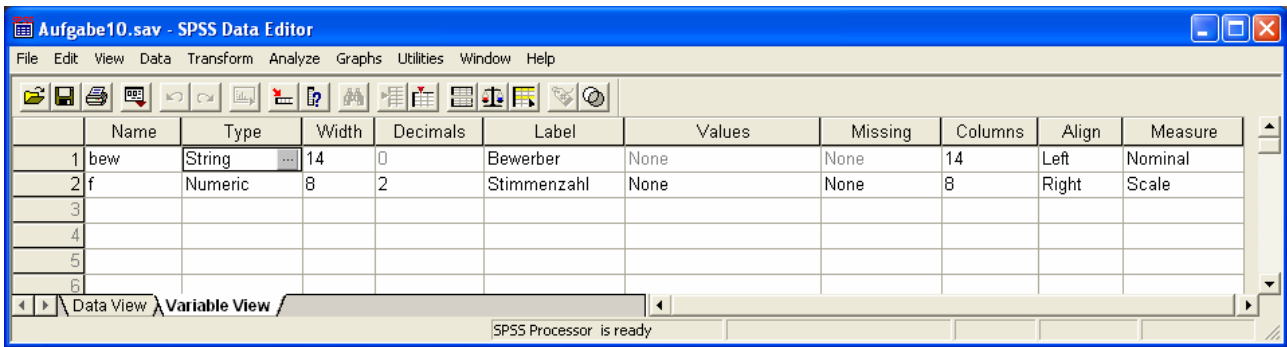
In dieser Aufgabe holen wir ein Detail zur deskriptiven Statistik nach: Das Kreisdiagramm. Beim ersten Durchgang der Oberbürgermeisterwahl in Tübingen am 8. November 1998 erhielten die einzelnen Bewerber die folgenden Stimmzahlen:

| Bewerber             | Stimmzahl |
|----------------------|-----------|
| Freisler             | 4366      |
| Hasenclever          | 7817      |
| Klink                | 4908      |
| Russ-Scherer         | 7410      |
| Restliche 5 zusammen | 3413      |

Stellen Sie das Wahlergebnis als Kreisdiagramm graphisch dar.

### Lösung

Die Merkmalsvariable nennen wir *bew* mit dem Bezeichner *Bewerber*. Sie hat jetzt den Typ *String* mit *14 Characters*, da der längste Name 12 Buchstaben hat. Das Variablendefinitionsfenster füllen Sie wie folgt aus:

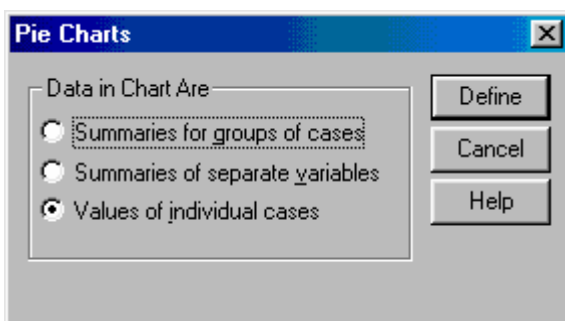


Der Variablentyp wird jetzt auf *String* mit 14 Stellen gesetzt.

Nach *OK* setzt SPSS *Measure* auf *Nominal*. Nach Rückkehr in *Data View* können Sie die Bewerbernamen in die Spalte *bew* eintragen. Die Variable *f* müssen Sie in *Variable View* mit dem Bezeichner *Stimmzahl* definieren und dann in *Data View* die Stimmzahlen eingeben. Obwohl Sie hier wieder eine Häufigkeitsverteilung und keine Urliste eingegeben haben, müssen Sie die Gewichtungsfunktion über *Data* → *Weight Cases* **nicht** setzen, weil SPSS bei der Zeichnung des Kreisdiagrammes die Werte der Variablen *f* automatisch richtig als Häufigkeiten interpretiert. Speichern Sie die Daten in *Aufgabe10.sav* ab.

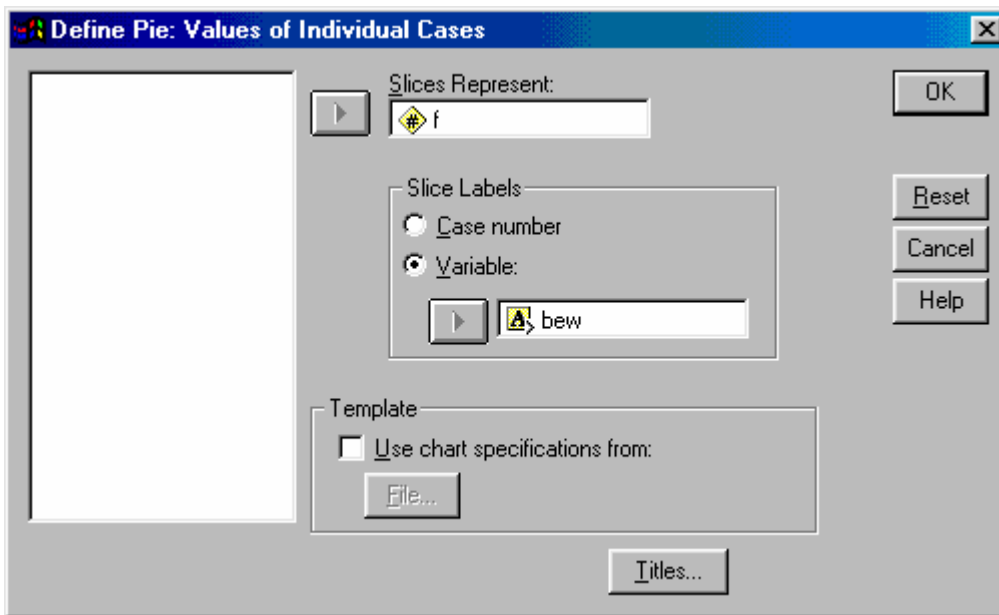
(Wollen Sie jedoch die Häufigkeitsverteilung der Wählerstimmen mit *Analyze* → *Descriptive Statistics* → *Frequencies* bestimmen, müssen Sie vorher die Gewichtungsfunktion mit der Variablen *f* aktivieren.)

Für das Kreisdiagramm rufen Sie *Graphs* → *Pie* auf. Im folgenden Fenster



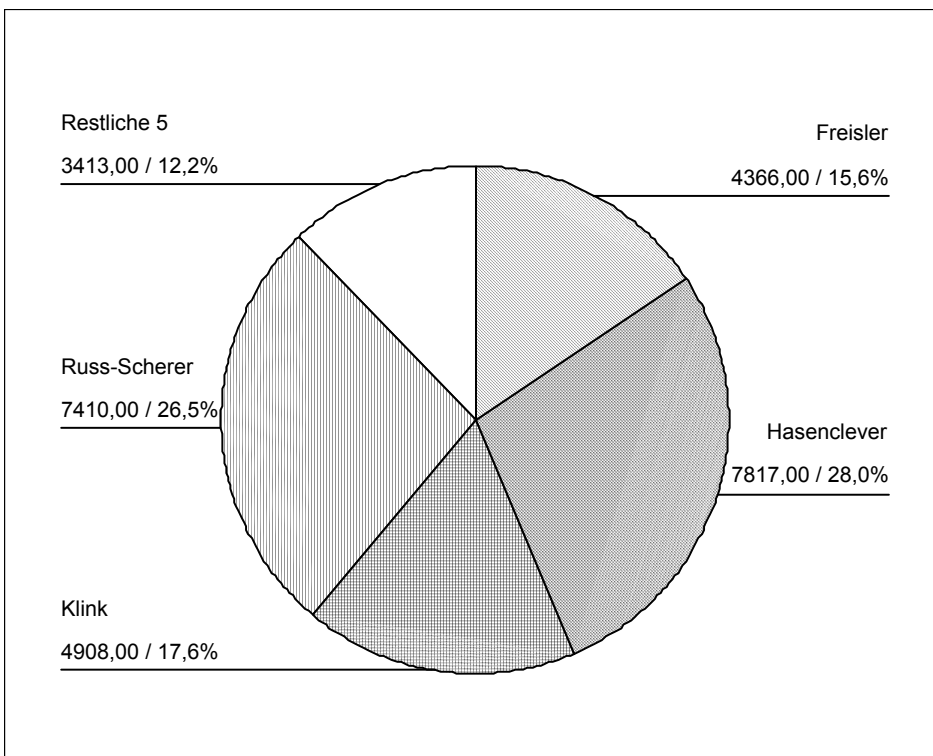
klicken Sie *Values of individual cases* und dann *Define* an. Sie erhalten ein neues Fenster,





das Sie wie abgebildet ausfüllen. *Slices Represent* erhält den Variablennamen *f*, da die Sektoren ja die Häufigkeiten wiedergeben sollen. Unter *Slice Labels* klicken Sie *Variable* an und fügen den Variablennamen *bew* ein. Damit werden die Bewerbernamen, die in der Variablen *bew* enthalten sind, zu Bezeichnern der einzelnen Sektoren im Kreisdiagramm.

Mit *OK* starten Sie die Zeichnung. Mit ein wenig Nacharbeit im *Chart Editor* (Zeichnung im Ausgabebildschirm doppelt anklicken) sieht sie etwa so aus:



In diesem Beispiel habe ich die Farbfüllungen, die SPSS zunächst liefert, durch schwarz-weiße Füllmuster ersetzt, da dieses Skriptum überwiegend schwarz-weiß gedruckt wird. Vor allem bei Kreisdiagrammen kann man sich meistens nicht darauf verlassen, daß die Farben einer Bildschirmdarstellung im Schwarz-weiß-Druck als brauchbare, gut zu unterscheidende Grautöne herauskommen. Wie Sie weiter sehen, habe ich im *Chart Editor* sowohl den Ausdruck der absoluten Stimmzahlen als auch den der Stimmenprozente ausgewählt.

## Aufgabe 11

Gegeben ist die folgende Urliste von Maßzahlen, die in vier Gruppen,  $B_1$  bis  $B_4$ , vorliegen. Jede Maßzahl ist das Resultat einer Messung an einer Person; die vier Personengruppen mögen unter vier verschiedenen Versuchsbedingungen ("Treatments") untersucht worden sein. Es soll geprüft werden, ob die Unterschiede zwischen den Mittelwerten der vier Gruppen interpretiert, also auf die vier verschiedenen Versuchsbedingungen zurückgeführt werden dürfen (generelle Alternativhypothese  $H_1$ ), oder ob angenommen werden muß, daß sie nur auf Zufall beruhen (generelle Nullhypothese,  $H_0$ ). Als Signifikanzniveau wird  $\alpha = 5\%$  gewählt.

|       |   |   |   |    |   |   |    |   |    |   |   |   |   |
|-------|---|---|---|----|---|---|----|---|----|---|---|---|---|
| $B_1$ | 2 | 5 | 5 | 3  | 8 | 4 | 7  | 4 | 6  | 5 | 6 | 5 | 5 |
| $B_2$ | 6 | 5 | 4 | 5  | 2 | 4 | 4  | 5 | 3  | 4 | 2 |   |   |
| $B_3$ | 4 | 0 | 1 | -1 | 3 | 0 | 2  | 1 | 2  | 2 |   |   |   |
| $B_4$ | 3 | 2 | 6 | -1 | 1 | 1 | -2 | 4 | -1 | 0 | 0 | 1 |   |

(Quelle: Übung Statistik II, Übungsbeispiele zur ein-, zwei- und dreifaktoriellen Varianzanalyse, Datei *stat212.doc*, Aufgabe 1)

## Lösung

Öffnen Sie SPSS mit dem Dateneingabebildschirm. Definieren Sie die erste Variable mit dem Namen  $x$  und dem Variablenbezeichner *Abhängige Variable*. Geben Sie dann die insgesamt 46 Maßzahlen der Urliste untereinander ein. Vergessen Sie dabei nicht die vereinzelt vorkommenden negativen Vorzeichen.

Definieren Sie jetzt die zweite Variable als Gruppierungsvariable mit dem Namen  $grp$  und dem Bezeichner *Gruppierung*. Definieren Sie auch die Wertebezeichner  $1 = "B1"$ ,  $2 = "B2"$ ,  $3 = "B3"$  und  $4 = "B4"$ . (Wenn Sie nicht mehr wissen, wie das geht, schauen Sie den Bildschirm auf S. 31 zu Aufgabe 03 an.) Geben Sie jetzt unter  $grp$  für die ersten 13 Fälle 1, die nächsten 11 Fälle 2, die nächsten 10 Fälle 3 und die letzten 12 Fälle 4 ein. Speichern Sie Ihre Daten in der Datei *Aufgabe11.sav* ab. Der tiefste Teil Ihres Datenbildschirms muß nach dem Abspeichern wie folgt aussehen:

Aufgabe11 - SPSS for Windows Student Version Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1:x

|    | x     | grp  | var | var | var | var | var | var | var |
|----|-------|------|-----|-----|-----|-----|-----|-----|-----|
| 34 | 2,00  | 3,00 |     |     |     |     |     |     |     |
| 35 | 3,00  | 4,00 |     |     |     |     |     |     |     |
| 36 | 2,00  | 4,00 |     |     |     |     |     |     |     |
| 37 | 6,00  | 4,00 |     |     |     |     |     |     |     |
| 38 | -1,00 | 4,00 |     |     |     |     |     |     |     |
| 39 | 1,00  | 4,00 |     |     |     |     |     |     |     |
| 40 | 1,00  | 4,00 |     |     |     |     |     |     |     |
| 41 | -2,00 | 4,00 |     |     |     |     |     |     |     |
| 42 | 4,00  | 4,00 |     |     |     |     |     |     |     |
| 43 | -1,00 | 4,00 |     |     |     |     |     |     |     |
| 44 | ,00   | 4,00 |     |     |     |     |     |     |     |
| 45 | ,00   | 4,00 |     |     |     |     |     |     |     |
| 46 | 1,00  | 4,00 |     |     |     |     |     |     |     |
| 47 |       |      |     |     |     |     |     |     |     |
| 48 |       |      |     |     |     |     |     |     |     |

SPSS for Windows Student Version Processor is ready

Mit *Analyze* → *General Linear Model* → *Univariate* öffnen Sie das Auswertungsfenster für univariate Varianzanalysen. Da *x* unsere abhängige Variable ist, wird sie nach *Dependent Variable* gebracht. Die Gruppierungsvariable *grp* fassen wir als festen Faktor auf und bringen sie nach *Fixed Factor(s)*. Von den Schaltfeldern rechts im Anwendungsfenster klicken wir zunächst *Plots* an, um SPSS zum Erstellen einer graphischen Darstellung der Mittelwerte zu veranlassen. Wir bringen *grp* zunächst in das Feld *Horizontal Axis* und dann durch Anklicken von *Add* in das Feld für *Plots*. Der Bildschirm sieht jetzt so aus:

Aufgabe11 - SPSS for Windows Student Version Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1:x

Univariate

Dependent Variable: x

Fixed Factor(s): grp

Random Factor(s):

Covariate(s):

WLS Weight:

Model...  
Contrasts...  
Plots...  
Post Hoc...  
Save...  
Options...

Univariate: Profile Plots

Factors: grp

Horizontal Axis:

Separate Lines:

Separate Plots:

Plots: Add Change Remove

grp

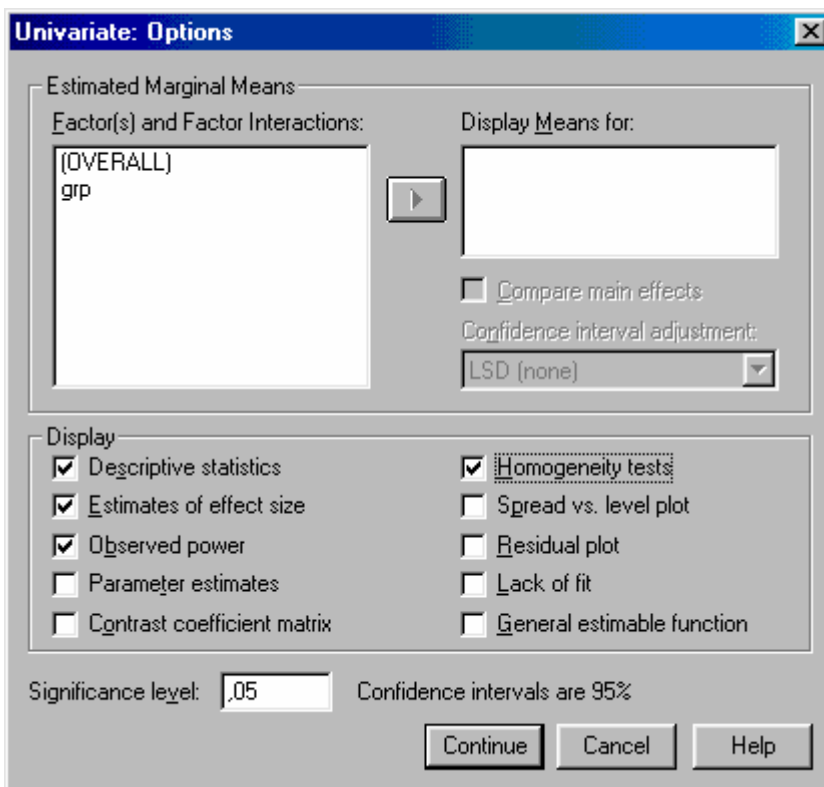
Continue  
Cancel  
Help

OK Reset Can

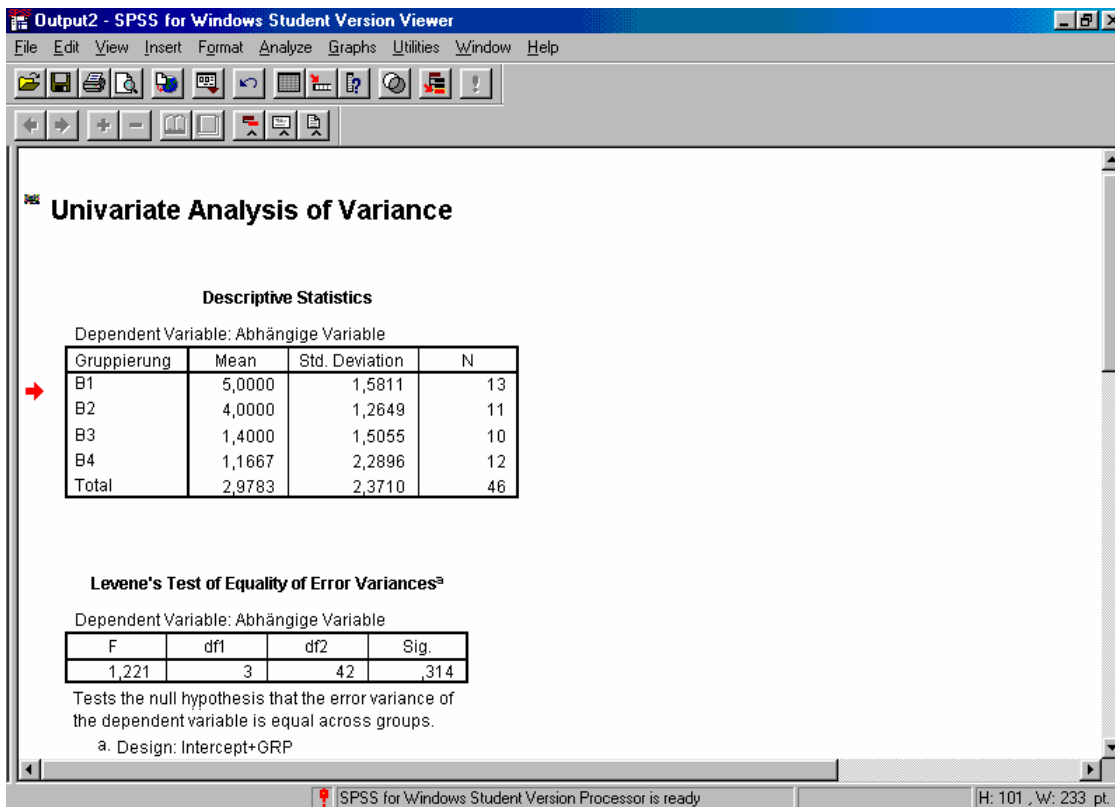
SPSS for Windows Student Version Processor is ready

Start Explorer - Uc133-w95-2 (D:) Explorer - AbbSPSS Aufgabe11 - SPSS fo... 17:38

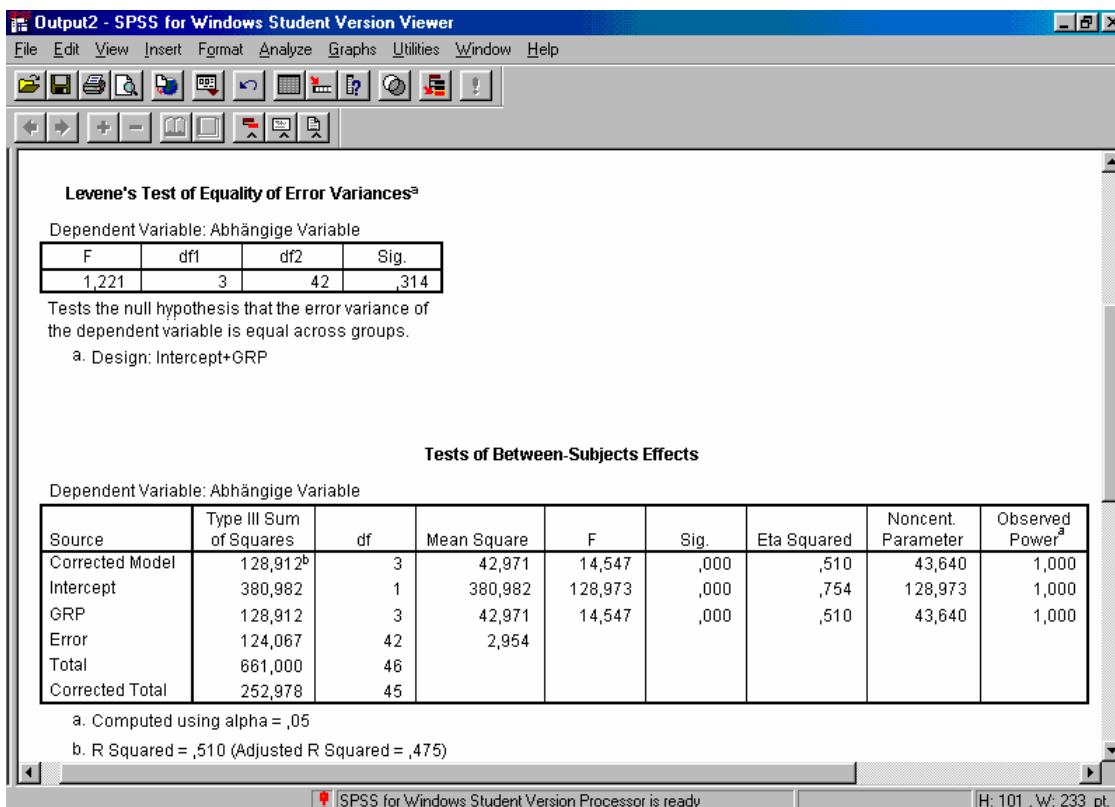
Mit *Continue* kehren wir zum Auswertungsfenster für univariate Varianzanalysen zurück. Mit *Options* öffnen wir jetzt noch das Optionenfenster:



Wir klicken die vier in der Abbildung mit Haken versehenen Optionen an und verschaffen uns damit die wichtigsten deskriptiven Statistiken, also vor allem die einzelnen Mittelwerte, eine Schätzung der gefundenen Effektgröße, die Teststärke und den Präliminartest auf Homogenität der Varianzen innerhalb der einzelnen Gruppen. Auch dieses Fenster schließen wir mit *Continue*. Von dem angegebenen Signifikanzniveau macht SPSS bei den Vertrauensbereichen Gebrauch; bei den Signifikanztests der Varianzanalyse liefert es uns jeweils die Überschreitungswahrscheinlichkeit, die wir dann selbst mit dem Signifikanzniveau vergleichen müssen (Methode I in der Vorlesung). Mit *OK* starten wir nun die Varianzanalyse. Der Ausgabebildschirm zeigt uns zunächst die wichtigsten Statistiken für die einzelnen Gruppen, Mittelwerte, Standardabweichungen und Anzahlen der Fälle, sowie Mittelwert, Standardabweichung und Anzahl der Fälle *total*. Darunter findet sich das Ergebnis des Levene-Tests, eines modifizierten F-Tests für Varianzhomogenität. Die Überschreitungswahrscheinlichkeit ist hier 31,4 %, wir können also davon ausgehen, daß diese Voraussetzung der Varianzanalyse erfüllt ist.



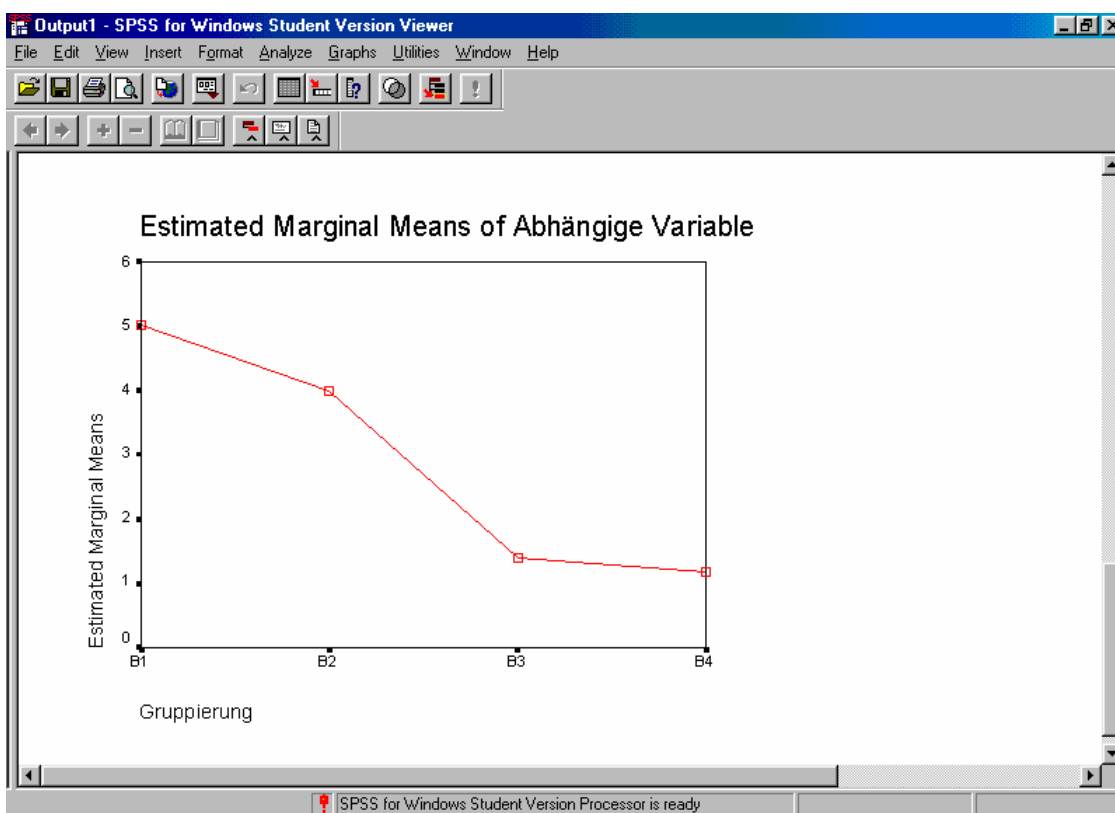
Weiter unten zeigt der Ausgabebildschirm das Ergebnis der Varianzanalyse:



Die Zeilen *Corrected Model*, *Intercept* und *Total* können wir hier außer Betracht lassen. Die in der Vorlesung behandelten Quadratsummen finden wir in der Spalte *Type III Sum of Squares*. Dabei bedeuten *GRP* = *zwischen den Gruppen*, also Faktor B, *Error* = *innerhalb der Gruppen* und *Corrected Total* = *Total* nach unserer Terminologie. Die zugehörigen Varianzschätzungen stehen in der Spalte *Mean Square*, der relevante F-Wert in der Spalte *F*. Wie sich unschwer überprüfen läßt, ist  $F = s^2_{\text{zwischen}}/s^2_{\text{innerhalb}} = 14,547$ . In der Spalte *Sig.* steht die Überschreitungswahrscheinlichkeit,

die hier kleiner ist als 0,0005. Die generelle Nullhypothese wird also verworfen, die generelle Alternativhypothese angenommen: Die verschiedenen Versuchsbedingungen haben eine unterschiedliche Wirkung gehabt. Unter *Eta Squared* ( $\eta^2$ ) wird der Quotient aus der Quadratsumme zwischen den Gruppen und der Quadratsumme total als Maß der Varianzaufklärung nach Bortz (1999, 5. Aufl., Gl. (7.21)) angegeben. Den Zusammenhang mit der Effektgröße  $\varepsilon$  gibt Bortz in Gl. (7.29), die Definition von  $\varepsilon$  bei Varianzanalysen in Gl. (7.27) und die Berechnung von  $\varepsilon$  aus  $\eta^2$  in Gl. (7.29) an. In unserem Beispiel werden durch die Versuchsbedingungen 51,0 % der Varianz aufgeklärt, Gl. 7.29 liefert  $\varepsilon = 1,020$ , was einen "starken" Effekt darstellt. Die Spalte *Power* schließlich liefert die Macht des Tests,  $1 - \beta$ , die Wahrscheinlichkeit, eine in Wirklichkeit geltende Alternativhypothese auch durch ein signifikantes Resultat zu entdecken. SPSS gibt hier den Wert 1,000 aus. Die Mittelwertsunterschiede sind also an der Standardabweichung innerhalb der Gruppen gemessen so groß, daß eine Entscheidung des F-Tests zugunsten der generellen  $H_1$  praktisch sicher ist.

Danach zeigt der Ausgabebildschirm die graphische Darstellung der Mittelwerte:

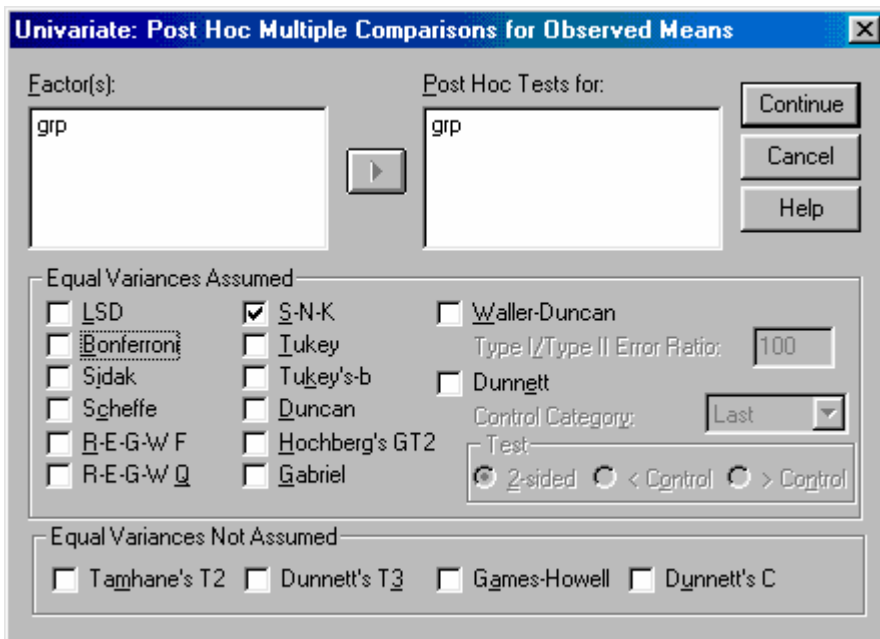


Die Abbildung kann im *Chart Editor* nachbearbeitet werden, wenn man sie doppelt anklickt, wie es in Aufgabe 01 näher beschrieben wird.

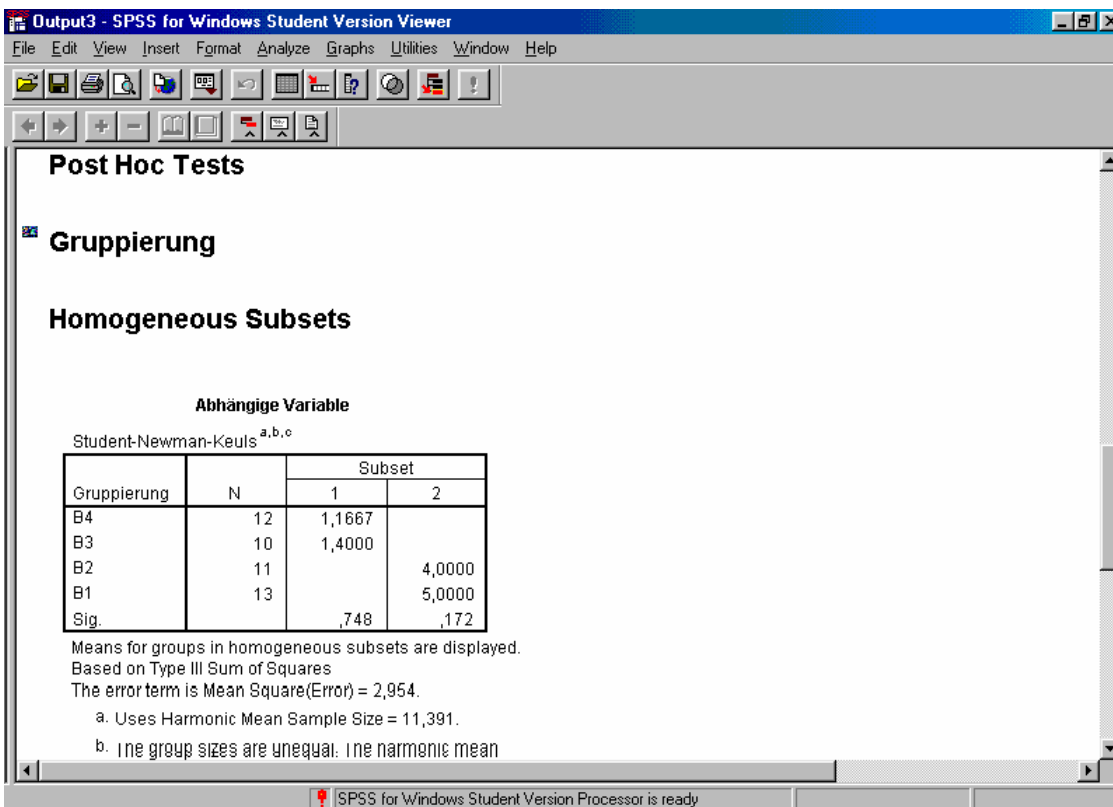
In der Graphik fällt auf, daß die Mittelwerte von B<sub>1</sub> und B<sub>2</sub> sowie von B<sub>3</sub> und B<sub>4</sub> jeweils recht nahe beisammen liegen, während sich zwischen den Mittelwerten von B<sub>2</sub> und B<sub>3</sub> eine größere Differenz zeigt. Das regt zur Prüfung spezieller, aposteriorischer Nullhypothesen an. Über die hier bestehenden Möglichkeiten gibt es umfangreiche Spezialliteratur; eine Zusammenfassung findet sich in Kirk (1982, 2. Aufl., S. 90-133). Eine Zusammenstellung von neun wichtigen Verfahren enthält die Vervielfältigung zur Vorlesung *Tests für Einzelvergleiche zur Varianzanalyse*, Datei *anovst.doc*. Wir wählen für dieses Beispiel den häufig angewandten Newman-Keuls-Test, der alle kombinatorisch möglichen ganzen (also nicht gebrochenen) Mittelwertsdifferenzen auf Signifikanz prüft und dabei Teilmengen von untereinander nicht signifikant verschiedenen Mittelwerten bildet.

Wir gehen zurück zum Dateneingabebildschirm und starten erneut mit *Analyse* → *General Linear Model* → *Univariate* den Auswertungsbildschirm für die Varianzanalyse. In den Schaltfeldern auf der rechten Seite klicken wir *Post Hoc* an. Nachdem wir unseren Faktornamen, hier *grp*, in das

Fenster *Post Hoc Tests for* gebracht haben, wird die Auswahl von 14 verschiedenen Tests für *Equal Variances Assumed* aktiv. Wir markieren nur *S-N-K*, die Abkürzung für *Student-Newman-Keuls*, wie die nächste Abbildung zeigt.



Mit *Continue* → *OK* starten wir die Varianzanalyse. (Die anderen Einstellungen des Auswertungsbildschirms für die Varianzanalyse sind noch vom letzten Lauf vorhanden. Sollten Sie mit dem Newman-Keuls-Test neu in das Beispiel eingestiegen sein, müssen Sie diese Einstellungen jetzt setzen, wie es oben angegeben wurde.) Der Ausgabebildschirm enthält danach zusätzlich zur Varianzanalyse eine Tabelle für den Newman-Keuls-Test:



Die Tabelle ist so zu lesen, daß der Test zwei Untergruppen von Mittelwerten gefunden hat, die sich jeweils nicht signifikant voneinander unterscheiden. Jeder Mittelwert der einen Untergruppe unterscheidet sich hingegen signifikant von jedem Mittelwert der anderen Untergruppe. Beim Blick

auf die graphische Darstellung der vier Mittelwerte erweist sich dieses Resultat als sehr plausibel und leicht interpretierbar.

### Aufgabe 12

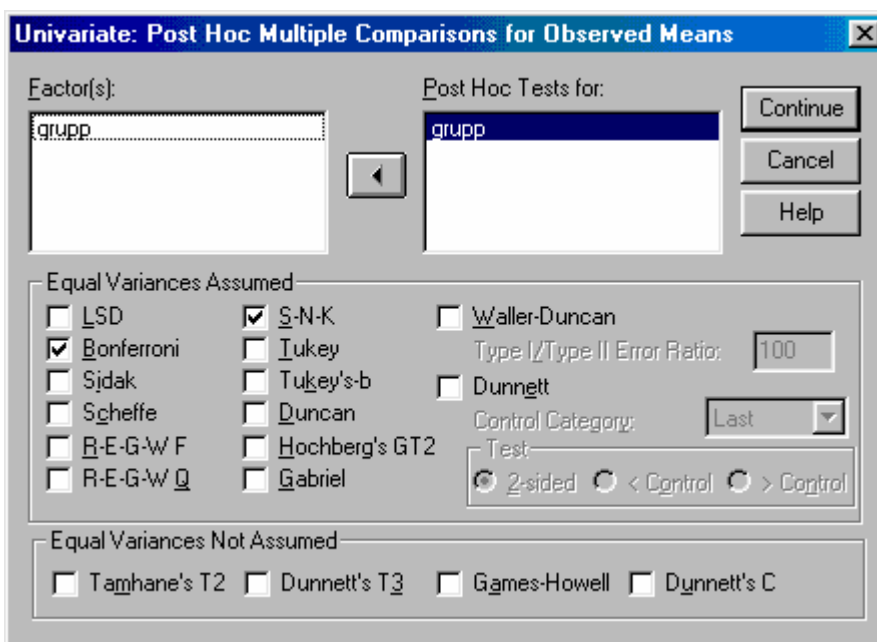
Diese Aufgabe ist eine reine Wiederholungsaufgabe zu Aufgabe 11. Sie entstammt ebenfalls den Übungsbeispielen zur Varianzanalyse *stat212.doc* und hat dort die Nummer 2. Die Urliste lautet:

B<sub>1</sub> 5 4 7 3 6 5 6 4 5 5 6 4 5 6 5  
B<sub>2</sub> 11 9 10 11 13 10 12 12 10 11 13 11 12 12 11  
B<sub>3</sub> 5 9 6 7 5 6 8 7 6 5 7 6 8 5 6.

Prüfen Sie den globalen Unterschied zwischen allen Mittelwert mit einer Varianzanalyse auf Signifikanz. Prüfen Sie zusätzlich alle einzelnen ganzen Mittelwertsdifferenzen mit dem Bonferroni- und dem Newman-Keuls-Test.

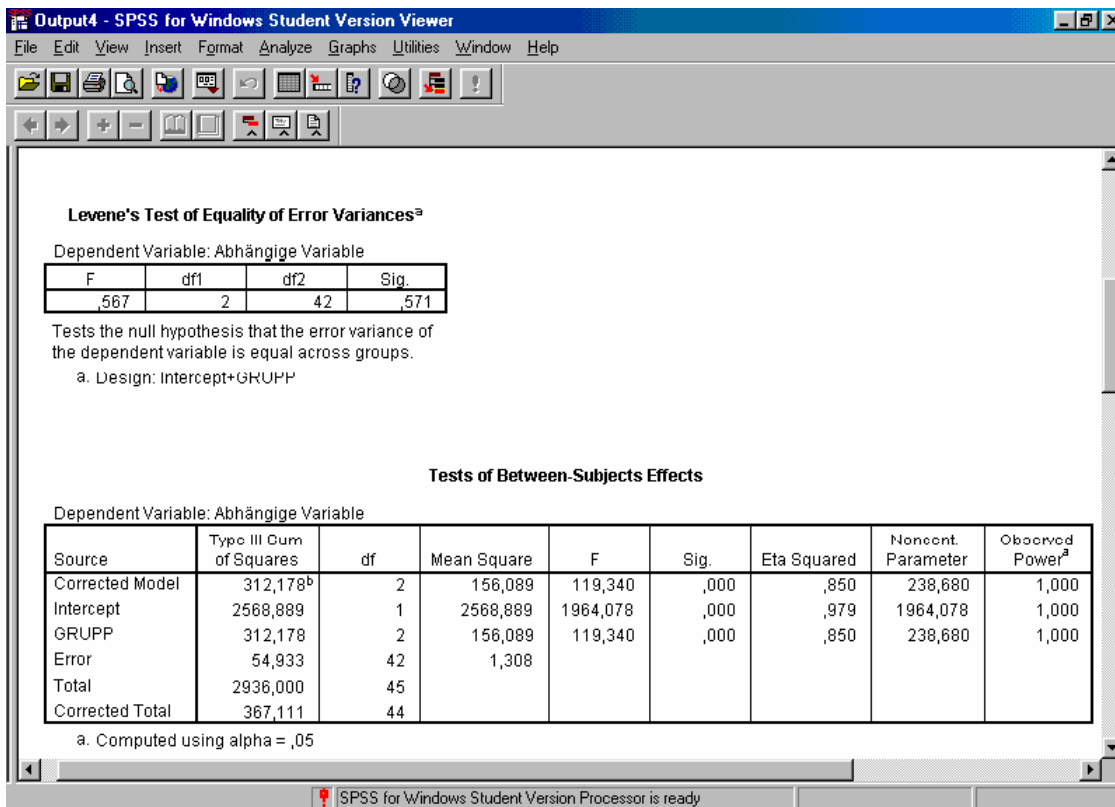
### Lösung

Definieren Sie wie in Aufgabe 11 die abhängige Variable  $x$  mit dem Bezeichner *Abhängige Variable*. Die Gruppierungsvariable möge hier *grupp* mit dem Bezeichner *Gruppierung* und den Wertebezeichnern  $1 = B_1$ ,  $2 = B_2$  und  $3 = B_3$  heißen. Geben Sie die Daten ein und speichern Sie die Datendatei unter *Aufgabe12.sav*. Holen Sie sich mit *Analyze* → *General Linear Model* → *Univariate* den Auswertungsbildschirm für die Varianzanalyse. Füllen Sie die Unterbildschirme *Plots* und *Options* so aus wie in Aufgabe 11. Klicken Sie im Unterbildschirm *Post Hoc Bonferroni* und *S-N-K* an:

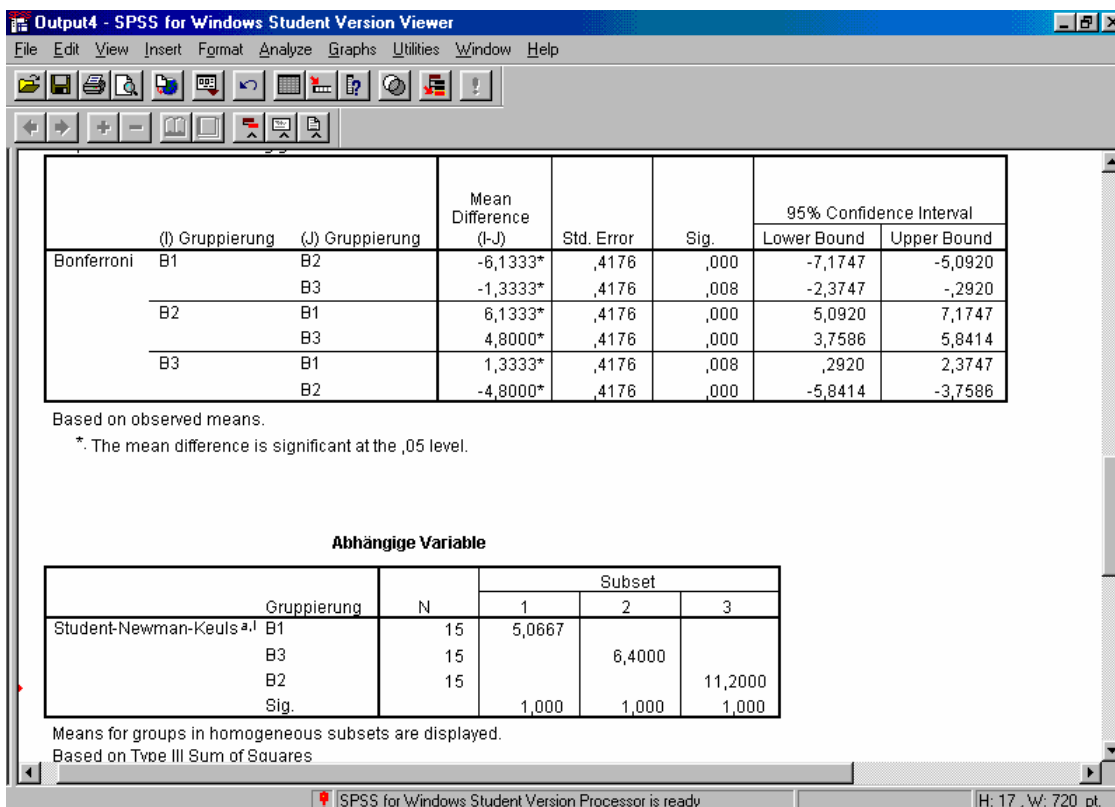


Starten Sie nun die Varianzanalyse mit *Continue* → *OK*. Das Resultat hat das übliche Format:





Die Varianzen innerhalb der Gruppen sind homogen, diese Voraussetzung der Varianzanalyse ist also erfüllt. Die generelle Nullhypothese wird mit einem  $\alpha \leq 0,0005$  verworfen, das Resultat ist also signifikant. Erklärter Varianzanteil und Macht des Tests sind sehr hoch. Die Ergebnisse von Bonferroni- und Newman-Keuls-Test zeigt das nächste Bild:



Im Bonferroni-Test werden alle Mittelwertsdifferenzen berechnet und auf Signifikanz geprüft. Die Basis ist der t-Test mit der sogenannten Bonferroni-Korrektur für die Tatsache, daß die einzelnen Mittelwertvergleiche zu einer Grundmenge von Mittelwerten gehören und daher nicht voneinander

unabhängig sind. In der Spalte *Sig.* wird wie üblich die Überschreitungswahrscheinlichkeit angegeben. Alle Mittelwertsdifferenzen zeigen sich dabei auf mindestens 1-Prozent-Niveau signifikant. Der Newman-Keuls-Test erkennt diese zwischen allen einzelnen Mittelwerten bestehenden Signifikanzen ebenfalls und macht daher jeden der drei Mittelwerte zu einer eigenen Teilmenge.

### Aufgabe 13

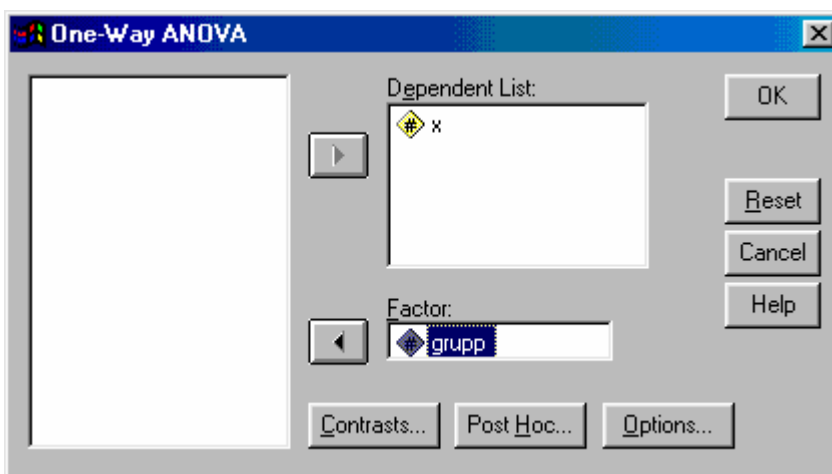
Auch hier handelt es sich um ein weiteres Zahlenbeispiel zur einfaktoriellem Varianzanalyse. Es ist mit der Nummer 3 in den Übungsbeispielen zur ein-, zwei- und dreifaktoriellem Varianzanalyse, Datei *stat212.doc*, enthalten. Die Urliste lautet:

|                |   |   |   |   |   |    |
|----------------|---|---|---|---|---|----|
| B <sub>1</sub> | 7 | 2 | 1 | 4 | 8 | 5  |
| B <sub>2</sub> | 9 | 7 | 8 | 4 | 5 | 10 |
| B <sub>3</sub> | 3 | 1 | 6 | 4 | 2 | .  |

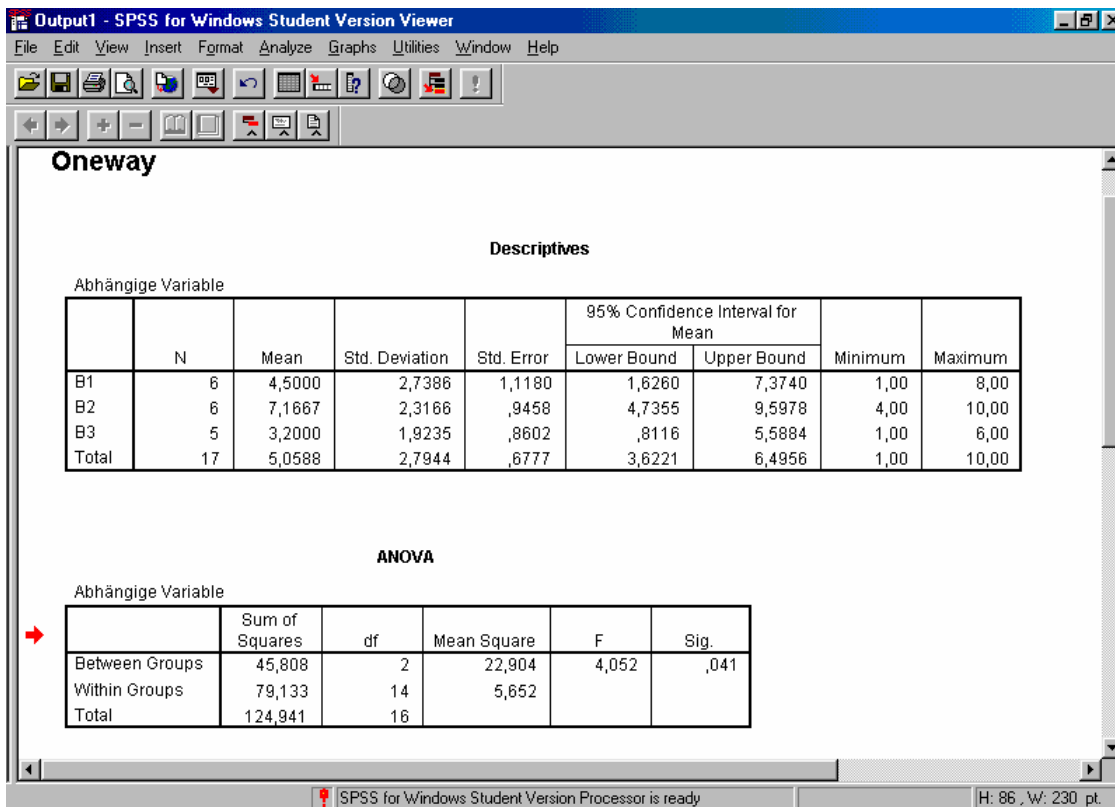
Berechnen Sie die Varianzanalyse, nachdem Sie die Urliste unter *Aufgabe13.sav* abgespeichert haben.

### Lösung

Die einfaktoriellem Varianzanalyse findet sich auch noch an einer anderen Stelle in SPSS. Man kann sie über *Analyse* → *Compare Means* → *One-Way-ANOVA* aufrufen. Das wollen wir in diesem Beispiel tun. Dabei erhält man einen anderen Eingabebildschirm, der nur für die einfaktoriellem Varianzanalyse gedacht ist:



Werden in *Dependent List* mehrere abhängige Variablen eingegeben, so rechnet SPSS für jede von ihnen eine eigene Varianzanalyse. Die Untermenüs *Post Hoc* und *Options* entsprechen im wesentlichen denen bei der Varianzanalyse innerhalb von *General Linear Model*. Die graphische Darstellung wird jetzt unter *Options* angefordert. (Das Untermenü *Contrasts*, das auch im Auswertungsfenster von *General Linear Model* enthalten ist, berechnet sogenannte Trendanalysen, die im Rahmen dieses Skriptums nicht behandelt werden. Literatur: Bortz, 1999, 5. Aufl., Kap. 7.4, S. 265-273.) Innerhalb geringfügig anderer Tabellen hat das Ergebnis die gewohnte Form:



Im Beispiel ist die Überschreitungswahrscheinlichkeit 4,1 %, die generelle Nullhypothese wird also bei  $\alpha = 5\%$  verworfen. Die Voraussetzung der Varianzhomogenität ist erfüllt, die Tabelle für den Levene-Test wurde hier aber nicht wiedergegeben.

### Aufgabe 14

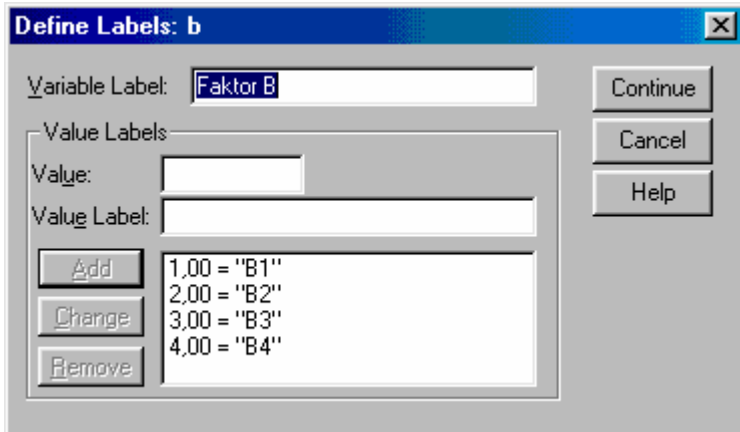
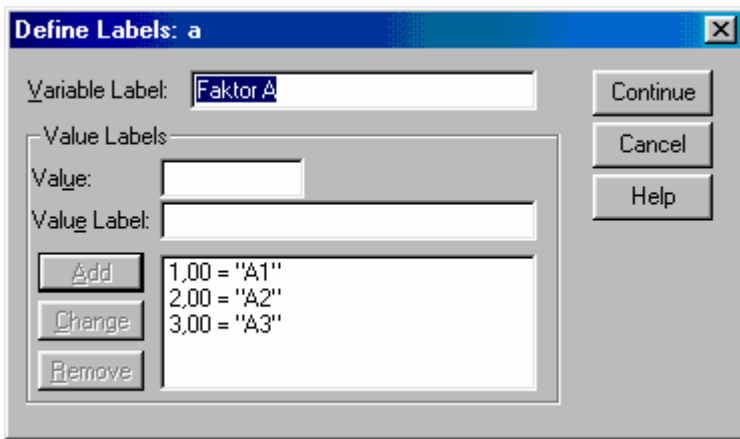
Gegeben ist die folgende Urliste, bei der jede Maßzahl  $x_{hij}$  das Ergebnis einer Messung an der  $h$ -ten Person in der Faktorstufenkombination  $A_i \times B_j$  darstellt. Die Faktoren A und B sind zwei miteinander gekreuzte unabhängige Variablen. Das Beispiel ist mit der Nummer 4 in den Übungsbeispielen zur ... Varianzanalyse, Datei *stat212.doc*, enthalten.

|                | B <sub>1</sub> | B <sub>2</sub> | B <sub>3</sub> | B <sub>4</sub> |
|----------------|----------------|----------------|----------------|----------------|
| A <sub>1</sub> | 5 4 3 6        | 5 5 4 6        | 5 6 8 5        | 4 6 6 4        |
| A <sub>2</sub> | 2 7 3 4        | 4 5 4 5        | 4 6 5 3        | 2 3 5 4        |
| A <sub>3</sub> | 7 4 2 5        | 3 5 2 6        | 6 5 7 4        | 5 4 6 3        |

Prüfen Sie die generellen Nullhypothesen hinsichtlich der Faktoren A und B und deren Wechselwirkung.

### Lösung

Da jede Maßzahl in zwei Dimensionen, A und B, kategorisiert ist, benötigen wir jetzt zwei Gruppierungsvariablen. Wir nennen sie  $a$  und  $b$ . Wir wählen Variablenbezeichner und Wertebezeichner gemäß den beiden folgenden Abbildungen:

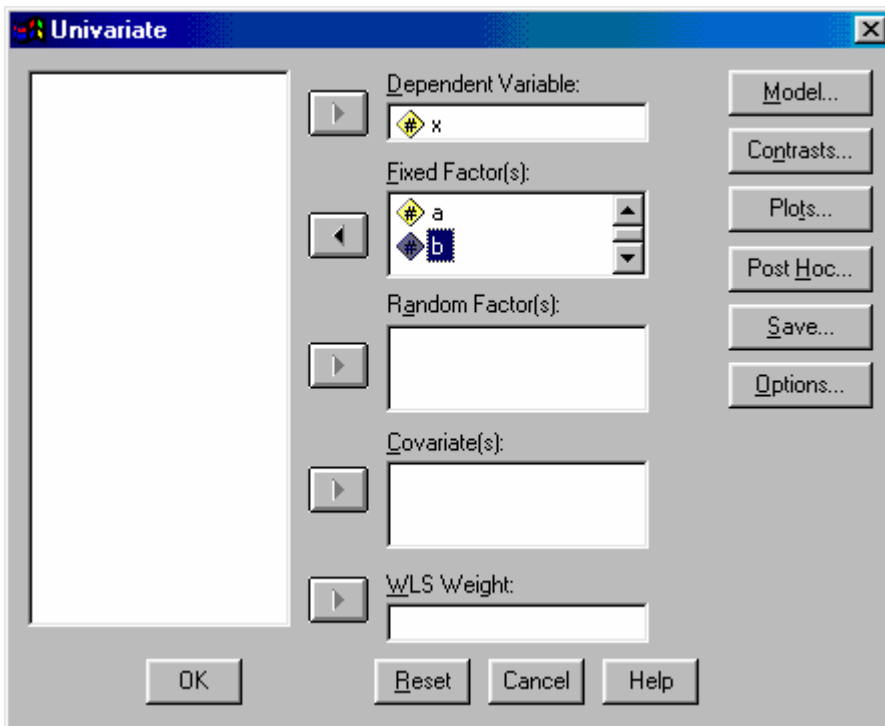


Die beiden Gruppierungsvariablen geben wir dann so ein, daß jede Maßzahl nach A und nach B richtig klassifiziert wird. Für die ersten 14 Maßzahlen sieht das so aus:

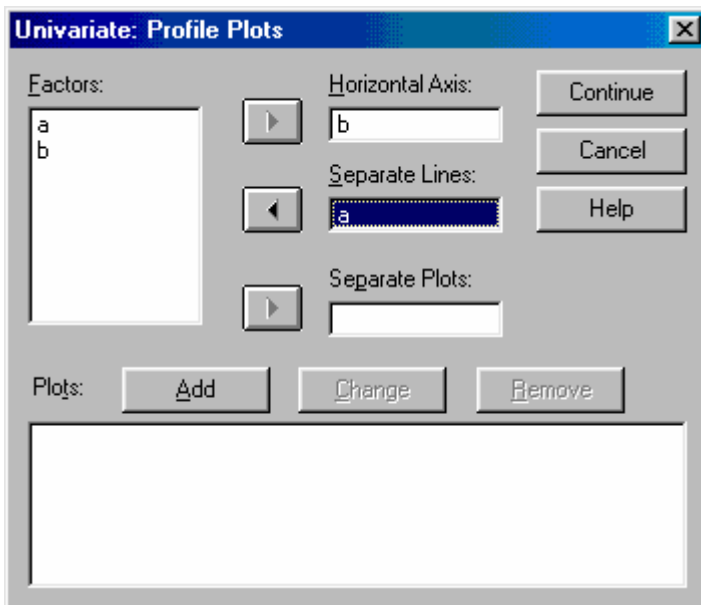
|    | x    | a    | b    | var | var | var | var | var | var |
|----|------|------|------|-----|-----|-----|-----|-----|-----|
| 1  | 5,00 | 1,00 | 1,00 |     |     |     |     |     |     |
| 2  | 4,00 | 1,00 | 1,00 |     |     |     |     |     |     |
| 3  | 3,00 | 1,00 | 1,00 |     |     |     |     |     |     |
| 4  | 6,00 | 1,00 | 1,00 |     |     |     |     |     |     |
| 5  | 5,00 | 1,00 | 2,00 |     |     |     |     |     |     |
| 6  | 5,00 | 1,00 | 2,00 |     |     |     |     |     |     |
| 7  | 4,00 | 1,00 | 2,00 |     |     |     |     |     |     |
| 8  | 6,00 | 1,00 | 2,00 |     |     |     |     |     |     |
| 9  | 5,00 | 1,00 | 3,00 |     |     |     |     |     |     |
| 10 | 6,00 | 1,00 | 3,00 |     |     |     |     |     |     |
| 11 | 8,00 | 1,00 | 3,00 |     |     |     |     |     |     |
| 12 | 5,00 | 1,00 | 3,00 |     |     |     |     |     |     |
| 13 | 4,00 | 1,00 | 4,00 |     |     |     |     |     |     |
| 14 | 6,00 | 1,00 | 4,00 |     |     |     |     |     |     |
| 15 | 6,00 | 1,00 | 4,00 |     |     |     |     |     |     |

(Machen Sie sich sorgfältig klar, daß Sie die Entstehung der Zahlen unter *a* und *b* auch wirklich verstanden haben!) Speichern Sie die Daten unter *Aufgabe14.sav* ab.

Mit *Analyze* → *General Linear Model* → *Univariate* holen wir den Auswertungsbildschirm, in dem wir jetzt *x* wieder als *Dependent Variable*, *a* und *b* beide als *Fixed Factor(s)* eingeben:



Mit *Plots* holen wir uns den Bildschirm für die Festlegung der Graphik. Im Beispiel ordnen wir *b* der Abszisse, *a* den verschiedenen Kurvenzügen zu:



Bevor wir diesen Bildschirm mit *Continue* verlassen können, müssen wir unsere Wahl noch mit *Add* in das untere Fenster *Plots* bringen. Nach der Rückkehr in den Auswertungsbildschirm wählen wir noch über *Options* den Präliminartest auf Varianzenhomogenität, die deskriptiven Statistiken sowie Größen wie Effektstärke und Macht aus. Nach *OK* zeigt SPSS den Ausgabebildschirm. Einer Tabelle mit den Mittelwerten, Standardabweichungen und Gruppengrößen für alle Zellen, Zeilen und Spalten

Output2 - SPSS for Windows Student Version Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: Abhängige Variable

| Faktor A | Faktor B | Mean   | Std. Deviation | N  |
|----------|----------|--------|----------------|----|
| A1       | B1       | 4,5000 | 1,2910         | 4  |
|          | B2       | 5,0000 | ,8165          | 4  |
|          | B3       | 6,0000 | 1,4142         | 4  |
|          | B4       | 5,0000 | 1,1547         | 4  |
|          | Total    | 5,1250 | 1,2042         | 16 |
| A2       | B1       | 4,0000 | 2,1602         | 4  |
|          | B2       | 4,5000 | ,5774          | 4  |
|          | B3       | 4,5000 | 1,2910         | 4  |
|          | B4       | 3,5000 | 1,2910         | 4  |
|          | Total    | 4,1250 | 1,3601         | 16 |
| A3       | B1       | 4,5000 | 2,0817         | 4  |
|          | B2       | 4,0000 | 1,8257         | 4  |
|          | B3       | 5,5000 | 1,2910         | 4  |
|          | B4       | 4,5000 | 1,2910         | 4  |
|          | Total    | 4,6250 | 1,5864         | 16 |
| Total    | B1       | 4,3333 | 1,7233         | 12 |
|          | B2       | 4,5000 | 1,1677         | 12 |
|          | B3       | 5,3333 | 1,3707         | 12 |
|          | B4       | 4,3333 | 1,3027         | 12 |
|          | Total    | 4,6250 | 1,4236         | 48 |

SPSS for Windows Student Version Processor is ready

folgen die Tabellen für den Präliminartest und die varianzanalytische Auswertung:

Output2 - SPSS for Windows Student Version Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: Abhängige Variable

| F    | df1 | df2 | Sig. |
|------|-----|-----|------|
| ,894 | 11  | 36  | ,555 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+A+B+A \* B

Tests of Between-Subjects Effects

Dependent Variable: Abhängige Variable

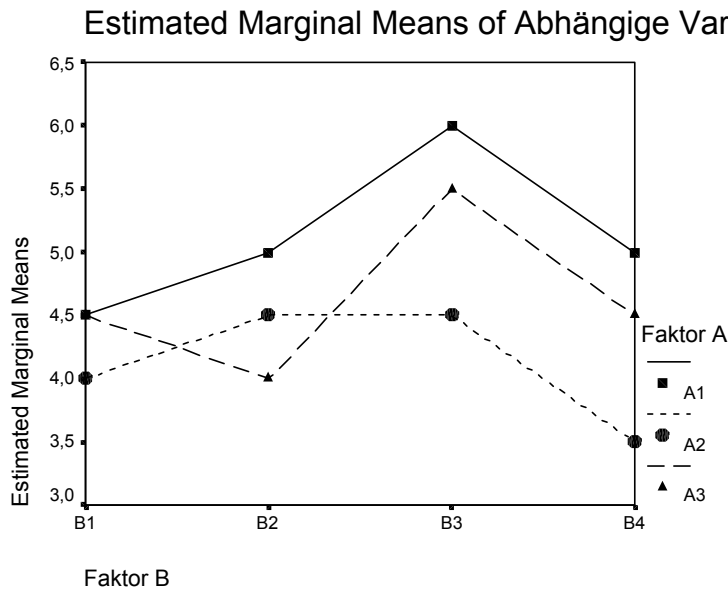
| Source          | Type III Sum of Squares | df | Mean Square | F       | Sig. |
|-----------------|-------------------------|----|-------------|---------|------|
| Corrected Model | 20,250 <sup>a</sup>     | 11 | 1,841       | ,884    | ,564 |
| Intercept       | 1026,750                | 1  | 1026,750    | 492,840 | ,000 |
| A               | 8,000                   | 2  | 4,000       | 1,920   | ,161 |
| B               | 8,250                   | 3  | 2,750       | 1,320   | ,283 |
| A * B           | 4,000                   | 6  | ,667        | ,320    | ,922 |
| Error           | 75,000                  | 36 | 2,083       |         |      |
| Total           | 1122,000                | 48 |             |         |      |
| Corrected Total | 95,250                  | 47 |             |         |      |

SPSS for Windows Student Version Processor is ready

Der Präliminartest zeigt, daß die Voraussetzung der Varianzhomogenität wiederum erfüllt ist. Die Überschreitungswahrscheinlichkeiten der Spalte *Sig.* sind für die Haupteffekte A und B sowie die Wechselwirkung A\*B größer als 5 %, wir haben also keine Signifikanz bekommen. Alle Nullhypothesen werden beibehalten. Die Zeilen *Intercept* und *Total* lassen wir außer Betracht; die Zeile *Corrected Model* enthält die Prüfung der generellen Nullhypothese in Bezug auf alle Zellen ohne Rücksicht auf deren Gruppierung nach den Faktoren A und B. Man würde sie auch

bekommen, wenn man den Datensatz mit durchnummerierten Zellen einer einfaktoriellen Varianzanalyse unterziehen würde (Prüfen Sie das einmal nach!).

Nach einer geringfügigen Nacharbeit im Chart Editor sieht die Graphik folgendermaßen aus:



Der nicht parallele Verlauf der drei Kurvenzüge spricht für eine Wechselwirkung, deren Betrag jedoch, wie die Varianzanalyse gezeigt hat, die Signifikanzgrenze nicht erreicht. Die - zur Zeichnung hinzuzudenkenden - Mittelwerte aus den drei Punkten für jedes  $B_j$  deuten einen Haupteffekt B an, der jedoch ebenfalls die Signifikanzgrenze verfehlt hat. Schließlich entspricht die unterschiedliche durchschnittliche Höhe der drei Kurvenzüge für einen Haupteffekt A. Da alle drei Effekte unterhalb der Signifikanzschranke liegen, dürfen sie inhaltlich nicht interpretiert werden. Einzelvergleiche erübrigen sich bei dieser Sachlage. Für die Haupteffekte A und B sowie die Wechselwirkung  $A*B$  ergaben sich Werte für  $\eta^2$  von 0,096; 0,099 und 0,051; also geringe Effektgrößen. Für die Macht der drei Tests resultierten die Werte  $1 - \beta$  von 0,372; 0,321 und 0,126. Falls also in Wirklichkeit Alternativhypothesen mit den hier geschätzten Mittelwerten zutreffen sollten, haben die hier vorliegenden geringen Fallzahlen zu einer zu hohen Wahrscheinlichkeit des  $\beta$ -Fehlers geführt. (Es handelt sich um ein Übungsbeispiel mit deutlich weniger Fällen als in der Praxis üblich.)

### Aufgabe 15

Gegeben ist die folgende Urliste, bei der jede Maßzahl  $x_{hijk}$  das Ergebnis einer Messung an der  $h$ -ten Person in der Faktorstufenkombination  $A_i \times B_j \times C_k$  darstellt. Den Daten liegt also ein Versuchsplan von drei miteinander gekreuzten Faktoren A, B und C zugrunde. Die Zellenbesetzung ist  $l = 4$ . Das Beispiel ist mit der Nummer 5 in den *Übungsbeispielen zur ... Varianzanalyse*, Datei *stat212.doc*, enthalten.

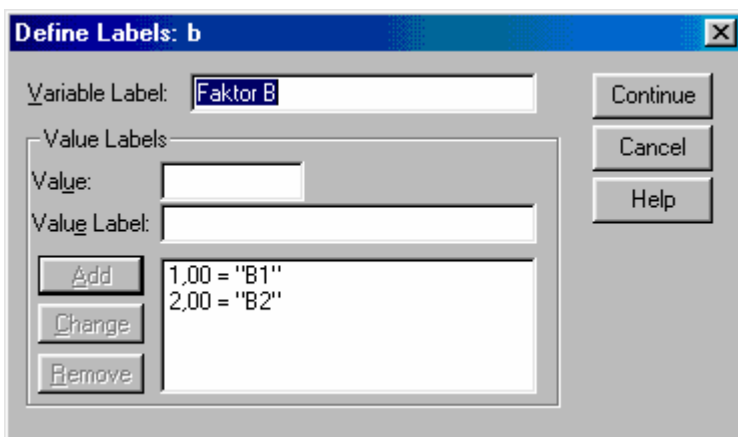
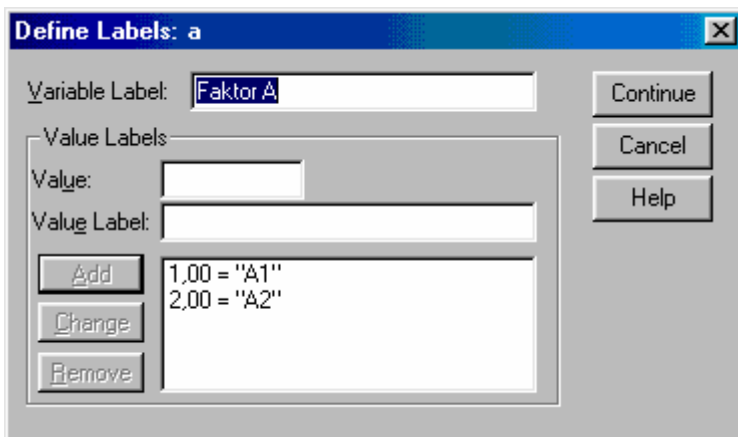
|                | C <sub>1</sub> |   |   |   |                |   |   |   | C <sub>2</sub> |   |   |   |                |    |   |    |
|----------------|----------------|---|---|---|----------------|---|---|---|----------------|---|---|---|----------------|----|---|----|
|                | B <sub>1</sub> |   |   |   | B <sub>2</sub> |   |   |   | B <sub>1</sub> |   |   |   | B <sub>2</sub> |    |   |    |
| A <sub>1</sub> | 3              | 6 | 3 | 3 | 7              | 8 | 7 | 6 | 4              | 5 | 4 | 3 | 7              | 8  | 9 | 8  |
| A <sub>2</sub> | 1              | 2 | 2 | 2 | 5              | 6 | 5 | 6 | 2              | 3 | 4 | 3 | 10             | 10 | 9 | 11 |

(nach Kirk, 1968, S. 218)

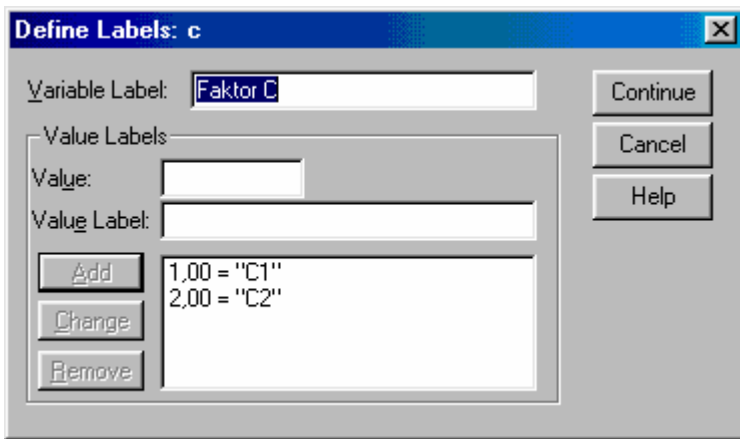
Überprüfen Sie die generellen Nullhypothesen hinsichtlich der Haupteffekte A, B und C sowie der Wechselwirkungen A\*B, A\*C, B\*C und A\*B\*C.

### Lösung

Da jede Maßzahl in drei Dimensionen, A, B und C, klassifiziert ist, benötigen wir jetzt drei Gruppierungsvariablen. Wir wählen Variablen- und Wertebezeichner gemäß den folgenden Abbildungen:



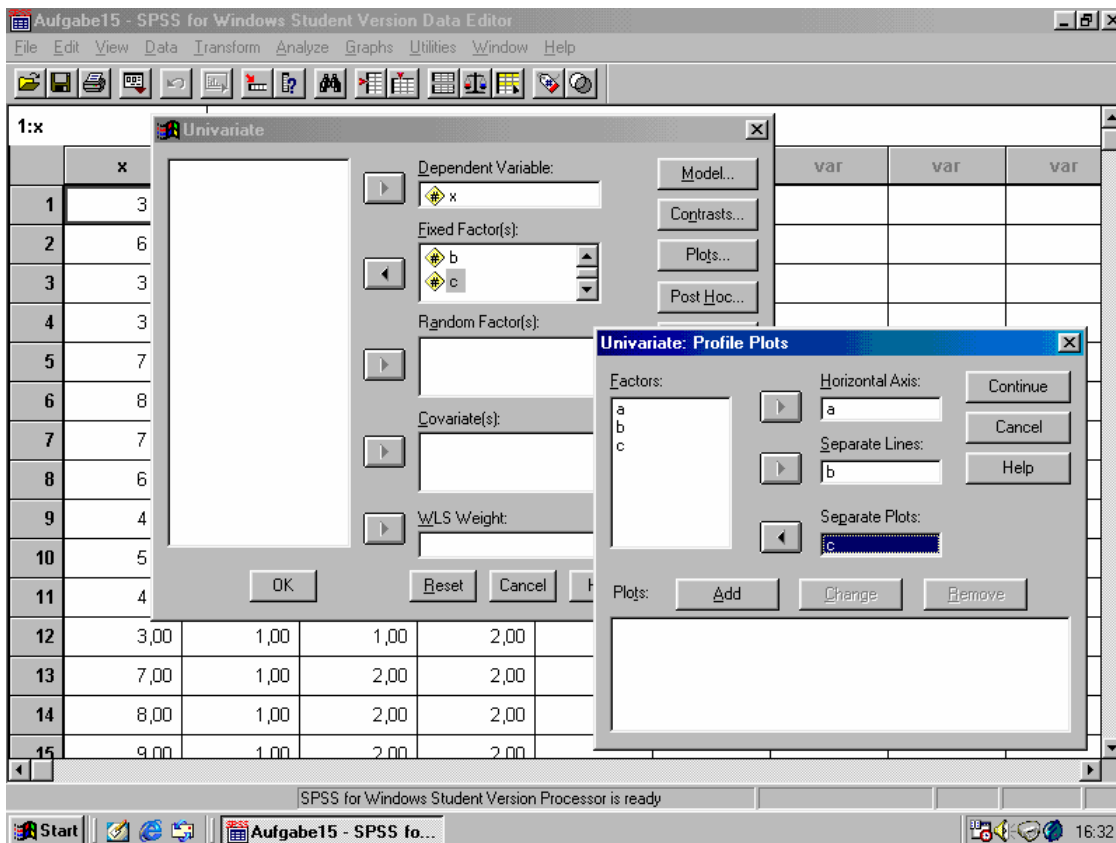




Die drei Gruppierungsvariablen geben wir dann so ein, daß jede Maßzahl  $x$  nach A, B und C richtig klassifiziert wird. Für die ersten 14 Maßzahlen sieht das so aus:

|    | x    | a    | b    | c    | var | var | var | var | var |
|----|------|------|------|------|-----|-----|-----|-----|-----|
| 1  | 3,00 | 1,00 | 1,00 | 1,00 |     |     |     |     |     |
| 2  | 6,00 | 1,00 | 1,00 | 1,00 |     |     |     |     |     |
| 3  | 3,00 | 1,00 | 1,00 | 1,00 |     |     |     |     |     |
| 4  | 3,00 | 1,00 | 1,00 | 1,00 |     |     |     |     |     |
| 5  | 7,00 | 1,00 | 2,00 | 1,00 |     |     |     |     |     |
| 6  | 8,00 | 1,00 | 2,00 | 1,00 |     |     |     |     |     |
| 7  | 7,00 | 1,00 | 2,00 | 1,00 |     |     |     |     |     |
| 8  | 6,00 | 1,00 | 2,00 | 1,00 |     |     |     |     |     |
| 9  | 4,00 | 1,00 | 1,00 | 2,00 |     |     |     |     |     |
| 10 | 5,00 | 1,00 | 1,00 | 2,00 |     |     |     |     |     |
| 11 | 4,00 | 1,00 | 1,00 | 2,00 |     |     |     |     |     |
| 12 | 3,00 | 1,00 | 1,00 | 2,00 |     |     |     |     |     |
| 13 | 7,00 | 1,00 | 2,00 | 2,00 |     |     |     |     |     |
| 14 | 8,00 | 1,00 | 2,00 | 2,00 |     |     |     |     |     |
| 15 | 9,00 | 1,00 | 2,00 | 2,00 |     |     |     |     |     |

Speichern Sie Ihre Eingaben unter *Aufgabe15.sav* ab. Den Auswertungsbildschirm für die Varianzanalyse öffnen wir mit *Analyze* → *General Linear Model* → *Univariate*. Hier geben wir  $x$  als *Dependent Variable*,  $a$ ,  $b$  und  $c$  als *Fixed Factor(s)* ein. Nach Drücken von *Plots* definieren wir die Abbildungen:



Mit *Add* legen wir die zwei Abbildungen definitiv fest, mit *Continue* verlassen wir das Definitionsfenster für die Abbildungen. Nach der Rückkehr in den Auswertungsbildschirm wählen wir noch über *Options* den Präliminartest auf Varianzenhomogenität, die deskriptiven Statistiken sowie Größen wie Effektstärke und Macht aus. Nach *OK* zeigt SPSS den Ausgabebildschirm. Die deskriptiven Statistiken lauten:

Descriptive Statistics  
 Dependent Variable: Abhängige Variable

| Faktor A | Faktor B | Faktor C | Mean    | Std. Deviation | N  |
|----------|----------|----------|---------|----------------|----|
| A1       | B1       | C1       | 3,7500  | 1,5000         | 4  |
|          |          | C2       | 4,0000  | ,8165          | 4  |
|          | Total    | 3,8750   | 1,1260  | 8              |    |
|          | B2       | C1       | 7,0000  | ,8165          | 4  |
|          |          | C2       | 8,0000  | ,8165          | 4  |
|          | Total    | 7,5000   | ,9258   | 8              |    |
|          | Total    | C1       | 5,3750  | 2,0659         | 8  |
|          |          | C2       | 6,0000  | 2,2678         | 8  |
|          | Total    | 5,6875   | 2,1203  | 16             |    |
| A2       | B1       | C1       | 1,7500  | ,5000          | 4  |
|          |          | C2       | 3,0000  | ,8165          | 4  |
|          | Total    | 2,3750   | ,9161   | 8              |    |
|          | B2       | C1       | 5,5000  | ,5774          | 4  |
|          |          | C2       | 10,0000 | ,8165          | 4  |
|          | Total    | 7,7500   | 2,4928  | 8              |    |
|          | Total    | C1       | 3,6250  | 2,0659         | 8  |
|          |          | C2       | 6,5000  | 3,8173         | 8  |
|          | Total    | 5,0625   | 3,3160  | 16             |    |
| Total    | B1       | C1       | 2,7500  | 1,4880         | 8  |
|          |          | C2       | 3,5000  | ,9258          | 8  |
|          | Total    | 3,1250   | 1,2583  | 16             |    |
|          | B2       | C1       | 6,2500  | 1,0351         | 8  |
|          |          | C2       | 9,0000  | 1,3093         | 8  |
|          | Total    | 7,6250   | 1,8212  | 16             |    |
|          | Total    | C1       | 4,5000  | 2,1909         | 16 |
|          |          | C2       | 6,2500  | 3,0441         | 16 |
|          | Total    | 5,3750   | 2,7562  | 32             |    |

Der Levene-Test auf Varianzhomogenität ist mit  $F(7, 24) = 0,748$  nicht signifikant. Für die Varianzanalyse erhalten wir die Tafel:

Output1 - SPSS for Windows Student Version Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Tests of Between-Subjects Effects

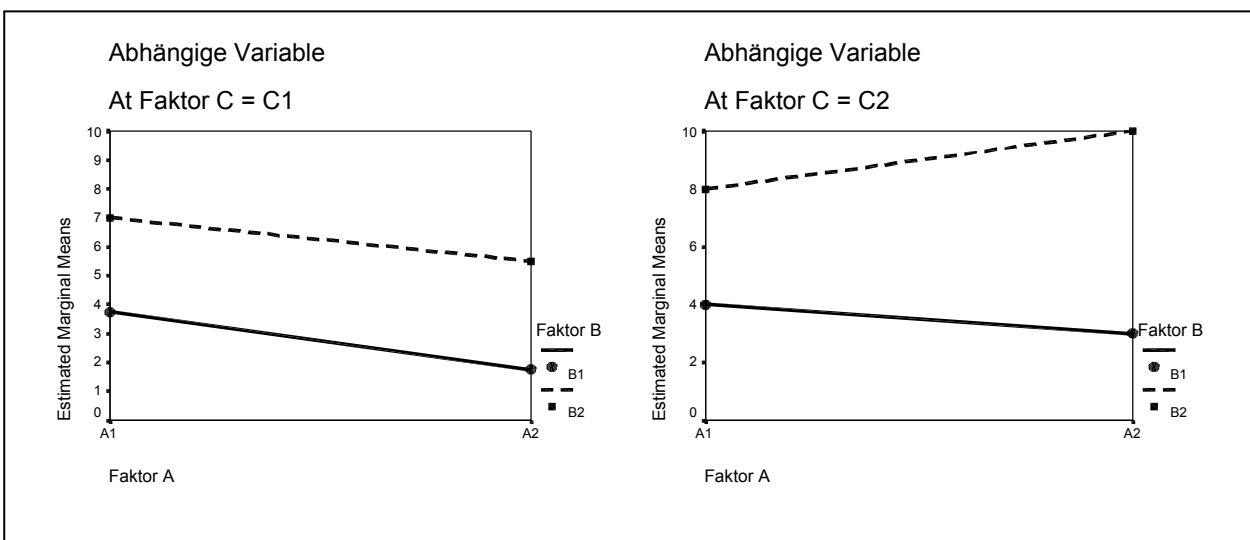
Dependent Variable: Abhängige Variable

| Source          | Type III Sum of Squares | df | Mean Square | F        | Sig. | Eta Squared | Noncent. Parameter | Observed Power <sup>a</sup> |
|-----------------|-------------------------|----|-------------|----------|------|-------------|--------------------|-----------------------------|
| Corrected Model | 217,000 <sup>b</sup>    | 7  | 31,000      | 40,216   | ,000 | ,921        | 281,514            | 1,000                       |
| Intercept       | 924,500                 | 1  | 924,500     | 1199,351 | ,000 | ,980        | 1199,351           | 1,000                       |
| A               | 3,125                   | 1  | 3,125       | 4,054    | ,055 | ,145        | 4,054              | ,489                        |
| B               | 162,000                 | 1  | 162,000     | 210,162  | ,000 | ,898        | 210,162            | 1,000                       |
| C               | 24,500                  | 1  | 24,500      | 31,784   | ,000 | ,570        | 31,784             | 1,000                       |
| A * B           | 6,125                   | 1  | 6,125       | 7,946    | ,010 | ,249        | 7,946              | ,772                        |
| A * C           | 10,125                  | 1  | 10,125      | 13,135   | ,001 | ,354        | 13,135             | ,935                        |
| B * C           | 8,000                   | 1  | 8,000       | 10,378   | ,004 | ,302        | 10,378             | ,871                        |
| A * B * C       | 3,125                   | 1  | 3,125       | 4,054    | ,055 | ,145        | 4,054              | ,489                        |
| Error           | 18,500                  | 24 | ,771        |          |      |             |                    |                             |
| Total           | 1160,000                | 32 |             |          |      |             |                    |                             |
| Corrected Total | 235,500                 | 31 |             |          |      |             |                    |                             |

a. Computed using alpha = ,05  
b. R Squared = ,921 (Adjusted R Squared = ,899)

SPSS for Windows Student Version Processor is ready H: 430 , W: 335 pt.

Nur der Haupteffekt A und die Dreifachwechselwirkung A\*B\*C verfehlen die Signifikanzschranke. In diesen beiden Fällen ist die Macht des Tests kleiner als 0,5. Alle anderen Haupteffekte und Wechselwirkungen sind auf dem 1 %-Niveau signifikant. Die Mittelwerte sind in den beiden folgenden Abbildungen enthalten:



Das Bild für C<sub>2</sub> zeigt eine Wechselwirkung zwischen A und B, die in Bild C<sub>1</sub> fehlt. Das bedeutet eine Dreifachwechselwirkung, die jedoch, wie die Varianzanalyse zeigt, die Signifikanzschranke nicht erreicht. Denkt man sich die einander entsprechenden Punkte der beiden Bilder gemittelt, so bleibt eine Wechselwirkung A\*B erhalten. Die Unterschiede zwischen A<sub>1</sub> und A<sub>2</sub> sind auf den beiden Stufen von C verschieden, was eine Wechselwirkung A\*C bedeutet. Die unterschiedlichen Abstände zwischen der durchgezogenen und der gestrichelten Geraden in beiden Bildern bedeuten

eine Wechselwirkung B\*C (Bitte machen Sie sich klar, warum das so ist, und lesen Sie erst dann weiter!). Denkt man sich alle vier Punkte für A<sub>1</sub> links in beiden Bildern und alle vier Punkte für A<sub>2</sub> rechts in beiden Bildern gemittelt, so bleibt nur ein geringer Unterschied, der dem nicht signifikanten Haupteffekt A entspricht. Der signifikante Haupteffekt B entspricht dem beträchtlichen und in beiden Bildern in die gleiche Richtung gehenden Abstand zwischen der durchgezogenen und der gestrichelten Linie. Denkt man sich schließlich je einen Mittelwert über alle vier Punkte der linken und der rechten Abbildung, so bedeutet deren Unterschied den ebenfalls signifikanten Haupteffekt von C.

Vor allem für die signifikanten Effekte erhalten wir auch recht beträchtliche Effektgrößen  $\eta^2$ . Angesichts der Signifikanzen bei den hier vorliegenden, sehr kleinen Zellenbesetzungen ist das nicht verwunderlich. Es handelt sich um ein konstruiertes Demonstrationsbeispiel, das schon bei kleinen Zellenbesetzungen viele Signifikanzen erreichen sollte.

### Aufgabe 16

Gegeben ist die folgende Urliste mit den Faktoren A, B und C. Sie gehört also zu einem dreifaktoriellen Versuchsplan. Es fällt jedoch auf, daß die einzelnen Zellen nur eine Maßzahl  $x_{ijk}$  enthalten. In diesem Plan sind die einzelnen Stufen des Faktors A, A<sub>i</sub>, einzelne Personen. Folglich bedeuten die Daten, daß jede Person *i* in jeder Faktorstufenkombination B<sub>j</sub>\*C<sub>k</sub> untersucht wurde. Das ist typisch für einen Versuchsplan mit Meßwiederholungen, im vorliegenden Fall auf den zwei gekreuzten Faktoren B und C. Solche Versuchspläne haben in der Psychologie beträchtliche Bedeutung.

|       |     |   |   |   |   |   |   |   |
|-------|-----|---|---|---|---|---|---|---|
| C:    | k = | 1 | 1 | 1 | 2 | 2 | 2 |   |
| B:    | j = | 1 | 2 | 3 | 1 | 2 | 3 |   |
| ----- |     |   |   |   |   |   |   |   |
| A:    | i = | 1 | 4 | 6 | 7 | 9 | 6 | 7 |
|       |     | 2 | 7 | 2 | 5 | 5 | 3 | 4 |
|       |     | 3 | 5 | 8 | 9 | 7 | 7 | 5 |
|       |     | 4 | 1 | 3 | 8 | 8 | 2 | 9 |
|       |     | 5 | 6 | 4 | 3 | 6 | 4 | 6 |
|       |     | 6 | 2 | 4 | 2 | 9 | 8 | 8 |

Prüfen Sie die Haupteffekte B und C sowie die Wechselwirkung B\*C auf Signifikanz.

### Lösung

Leider sind Varianzanalysen mit Meßwiederholungen in der Studentenversion 9.0 von SPSS nicht enthalten. Da solche Fragestellungen in der Psychologie aber immer wieder vorkommen, wird für diese Lösung, abweichend von den anderen Beispielen dieses Skriptums, die Vollversion 10.0.5 verwendet.

Nach der Variablendefinition und der Eingabe der Daten sieht der Datenbildschirm wie folgt aus:

|    | b1c1 | b2c1 | b3c1 | b1c2 | b2c2 | b3c2 | var | var | var | var | var | var | var |
|----|------|------|------|------|------|------|-----|-----|-----|-----|-----|-----|-----|
| 1  | 4,00 | 6,00 | 7,00 | 9,00 | 6,00 | 7,00 |     |     |     |     |     |     |     |
| 2  | 7,00 | 2,00 | 5,00 | 5,00 | 3,00 | 4,00 |     |     |     |     |     |     |     |
| 3  | 5,00 | 8,00 | 9,00 | 7,00 | 7,00 | 5,00 |     |     |     |     |     |     |     |
| 4  | 1,00 | 3,00 | 8,00 | 8,00 | 2,00 | 9,00 |     |     |     |     |     |     |     |
| 5  | 6,00 | 4,00 | 3,00 | 6,00 | 4,00 | 6,00 |     |     |     |     |     |     |     |
| 6  | 2,00 | 4,00 | 2,00 | 9,00 | 8,00 | 8,00 |     |     |     |     |     |     |     |
| 7  |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 8  |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 9  |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 10 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 11 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 12 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 13 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 14 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 15 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 16 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 17 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 18 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 19 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 20 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 21 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 22 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 23 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 24 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 25 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 26 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 27 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 28 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 29 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 30 |      |      |      |      |      |      |     |     |     |     |     |     |     |
| 31 |      |      |      |      |      |      |     |     |     |     |     |     |     |

Nach der Grundregel für SPSS, wonach Daten, die zur gleichen Person gehören, stets einen "Fall" bilden und daher in die gleiche Zeile eingegeben werden, stehen jetzt jeweils sechs Zahlen nebeneinander. Die Urliste in SPSS sieht genauso aus wie auf dem Aufgabenblatt. Die Variablenamen wurden mit Hilfe der Faktorennamen gebildet: b1c1, b2c1 usw., und zwar in deren Reihenfolge in den Kopfzeilen im Aufgabenblatt.

Wenn Sie am Anfang das am weitesten links stehende *var* angeklickt haben, um die Variablen zu definieren, werden Sie bemerkt haben, daß das Variablendefinitionsfenster in der Version 10 von SPSS umgestaltet wurde. Es sieht jetzt so aus:

Aufgabe16.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

|    | Name | Type    | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|----|------|---------|-------|----------|-------|--------|---------|---------|-------|---------|
| 2  | b2c1 | Numeric | 8     | 2        | B2 C1 | None   | None    | 8       | Right | Scale   |
| 3  | b3c1 | Numeric | 8     | 2        | B3 C1 | None   | None    | 8       | Right | Scale   |
| 4  | b1c2 | Numeric | 8     | 2        | B1 C2 | None   | None    | 8       | Right | Scale   |
| 5  | b2c2 | Numeric | 8     | 2        | B2 C2 | None   | None    | 8       | Right | Scale   |
| 6  | b3c2 | Numeric | 8     | 2        | B3 C2 | None   | None    | 8       | Right | Scale   |
| 7  |      |         |       |          |       |        |         |         |       |         |
| 8  |      |         |       |          |       |        |         |         |       |         |
| 9  |      |         |       |          |       |        |         |         |       |         |
| 10 |      |         |       |          |       |        |         |         |       |         |
| 11 |      |         |       |          |       |        |         |         |       |         |
| 12 |      |         |       |          |       |        |         |         |       |         |
| 13 |      |         |       |          |       |        |         |         |       |         |
| 14 |      |         |       |          |       |        |         |         |       |         |
| 15 |      |         |       |          |       |        |         |         |       |         |
| 16 |      |         |       |          |       |        |         |         |       |         |
| 17 |      |         |       |          |       |        |         |         |       |         |
| 18 |      |         |       |          |       |        |         |         |       |         |
| 19 |      |         |       |          |       |        |         |         |       |         |
| 20 |      |         |       |          |       |        |         |         |       |         |
| 21 |      |         |       |          |       |        |         |         |       |         |
| 22 |      |         |       |          |       |        |         |         |       |         |
| 23 |      |         |       |          |       |        |         |         |       |         |
| 24 |      |         |       |          |       |        |         |         |       |         |
| 25 |      |         |       |          |       |        |         |         |       |         |
| 26 |      |         |       |          |       |        |         |         |       |         |
| 27 |      |         |       |          |       |        |         |         |       |         |
| 28 |      |         |       |          |       |        |         |         |       |         |
| 29 |      |         |       |          |       |        |         |         |       |         |
| 30 |      |         |       |          |       |        |         |         |       |         |
| 31 |      |         |       |          |       |        |         |         |       |         |
| 32 |      |         |       |          |       |        |         |         |       |         |
| 33 |      |         |       |          |       |        |         |         |       |         |

Data View Variable View

SPSS Processor is ready

Der wesentliche Vorteil ist, daß Sie die Variablenbezeichner und alle Eigenschaften für alle Variablen auf einen Blick sehen können. Die Felder für die Wertebezeichner können Sie für die einzelnen Variablen als Pop-up-Menu unter *Values* aufrufen. Zwischen diesem Fenster und dem Eingabefenster können Sie wechseln, indem Sie unten links wahlweise *Data View* oder *Variable View* anklicken. Wie Sie unter *Label* sehen, wurden auch Variablenbezeichner nach dem gleichen Prinzip wie die Variablennamen gebildet.

Die Befehle für die Varianzanalyse mit Meßwiederholungen lassen sich nach dem Anklicken von *Analyze* → *General Linear Model* → *Repeated Measures* eingeben. Danach erhalten Sie folgendes Fenster:

Repeated Measures Define Factor(s)

Within-Subject Factor Name: c

Number of Levels: 2

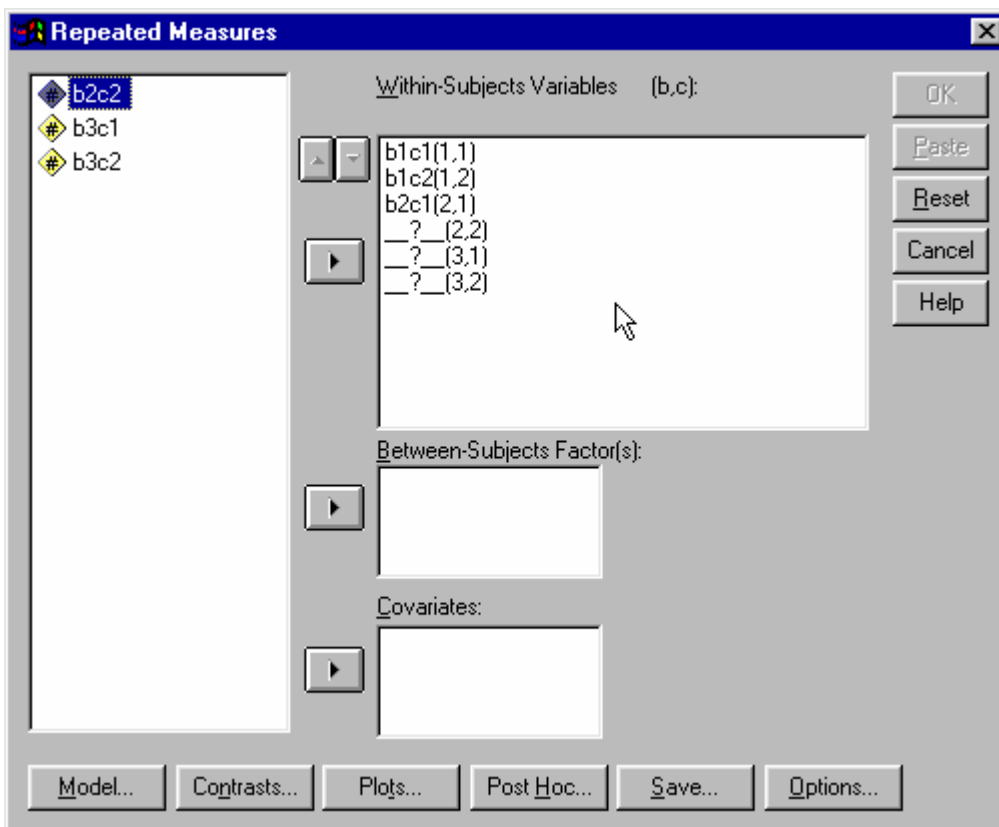
Add Change Remove

b(3)

Define Reset Cancel Help Measure >>

Da die Faktoren und Faktorstufen jetzt nicht mehr zwischen den Fällen (= Zeilen) variieren, können sie auch nicht mehr mit Gruppierungsvariablen eingegeben werden. Stattdessen erwartet SPSS neue, über die Spalten laufende Definitionen der Faktoren mit Meßwiederholungen. Unter *Within-Subject Factor Name* geben wir zunächst *b*, unter *Number of Levels* 3 ein, wie es unserem Beispiel entspricht. Mit *Add* bringen wir diese Eingaben in das untere Fenster und setzen die Eingaben mit *c*

und 2 fort. Im Prinzip ist es gleichgültig, ob wir diese Eingaben mit *b* oder mit *c* beginnen, obwohl dies unterschiedliche Auswirkungen hat. Mit *Define* holen wir uns ein weiteres Definitionsfenster:



Hier müssen wir nun unsere Variablen *b1c1*, *b1c2*, usw. den neuen Variablen *b* und *c* richtig zuordnen. Beachten Sie, daß im Fenster *Within Subjects Variables* die Variable, die wir zuerst eingegeben haben, also *b*, am weitesten links steht und über die Zeilen hinweg am langsamsten "läuft". Wenn Ihr Bildschirm genau so aussieht wie hier, auf der linken Seite also *b2c2* aktiviert ist, dann müssen sie dies mit dem Transferfeld nach rechts in *?(2,2)* schaffen. (Nehmen Sie sich die Zeit sich zu vergewissern, daß Sie das vollständig verstanden haben!) Die übrigen Schaltfelder dieses Bildschirms gleichen denen der bisherigen Varianzanalysen ohne Meßwiederholungen. Insbesondere sollten Sie unter *Plots* eine Zeichnung der Mittelwerte und unter *Options* die wichtigsten Statistiken anfordern. Die entsprechenden Fenster sollen hier nicht wiedergegeben werden. Das Feld *Between Subjects Factor(s)* gibt Ihnen die Möglichkeit, weitere Faktoren ohne Meßwiederholungen, dafür mit zugehörigen Gruppierungsvariablen in der Ihnen schon bekannten Weise einzugeben. Das vorliegende Beispiel enthält keine solchen Faktoren. Mit *OK* starten Sie schließlich die Berechnung. Sie erhalten den folgenden Ausgabebildschirm:

Output1 - SPSS Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Output

- General Linear Model
  - Title
  - Notes
  - Within-Subjects Factors
  - Descriptive Statistics
  - Mauchly's Test of Sphericity
  - Tests of Within-Subjects Effects
  - Tests of Within-Subjects Contrasts
  - Tests of Between-Subjects Effect
  - Profile Plots
    - Title
    - B \* C

| B | C | Dependent Variable |
|---|---|--------------------|
| 1 | 1 | B1C1               |
|   | 2 | B1C2               |
| 2 | 1 | B2C1               |
|   | 2 | B2C2               |
| 3 | 1 | B3C1               |
|   | 2 | B3C2               |

**Descriptive Statistics**

|            | Mean   | Std. Deviation | N |
|------------|--------|----------------|---|
| B1C1 B1 C1 | 4,1667 | 2,3166         | 6 |
| B1C2 B1 C2 | 7,3333 | 1,6330         | 6 |
| B2C1 B2 C1 | 4,5000 | 2,1679         | 6 |
| B2C2 B2 C2 | 5,0000 | 2,3664         | 6 |
| B3C1 B3 C1 | 5,6667 | 2,8048         | 6 |
| B3C2 B3 C2 | 6,5000 | 1,8708         | 6 |

**Mauchly's Test of Sphericity<sup>b</sup>**

Measure: MEASURE\_1

| Within Subjects Effect | Mauchly's W | Approx. Chi-Square | df | Sig. | Epsilon <sup>a</sup> |             |             |
|------------------------|-------------|--------------------|----|------|----------------------|-------------|-------------|
|                        |             |                    |    |      | Greenhouse e-Geisser | Huynh-Feldt | Lower-bound |
| B                      | ,858        | ,612               | 2  | ,737 | ,876                 | 1,000       | ,500        |
| C                      | 1,000       | ,000               | 0  | .    | 1,000                | 1,000       | 1,000       |
| B * C                  | ,744        | 1,183              | 2  | ,553 | ,796                 | 1,000       | ,500        |

Tests the null hypothesis that the error covariance matrix of the nonnormalized transformed dependent variables is proportional to an identity matrix.

SPSS Processor is ready | H: 293 , W: 612 pt

In einer ersten Tafel werden Ihnen die Zuordnung Ihrer Variablen zu den Faktoren mit Meßwiederholungen angezeigt. Offensichtlich haben wir dabei keinen Fehler gemacht. Eine zweite Tafel liefert die Mittelwerte, Standardabweichungen und Gruppengrößen der  $B_j * C_k$ -Zellen. Die dritte Tafel schließlich enthält das Ergebnis von *Mauchly's Test of Sphericity*. Letzteres ist eine wichtige Voraussetzung für die Interpretierbarkeit der Ergebnisse und entspricht der Varianzhomogenität bei Varianzanalysen ohne Meßwiederholungen. In der Spalte *Sig.* sehen Sie, daß die Voraussetzung erfüllt ist (wie bei der Varianzhomogenität bedeutet hier ein nicht signifikantes Ergebnis die Erfüllung der Voraussetzung).

Die Ergebnisse der Varianzanalyse zeigt die nächste Tabelle im Ausgabebildschirm:



Output1 - SPSS Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Tests of Within-Subjects Effects

Measure: MEASURE\_1

| Source     |                    | Type III Sum of Squares | df     | Mean Square | F     | Sig. | Eta Squared | Noncent. Parameter | Observed Power <sup>a</sup> |
|------------|--------------------|-------------------------|--------|-------------|-------|------|-------------|--------------------|-----------------------------|
| B          | Sphericity Assumed | 11,556                  | 2      | 5,778       | 1,235 | ,332 | ,198        | 2,470              | ,210                        |
|            | Greenhouse-Geisser | 11,556                  | 1,752  | 6,597       | 1,235 | ,330 | ,198        | 2,164              | ,196                        |
|            | Huynh-Feldt        | 11,556                  | 2,000  | 5,778       | 1,235 | ,332 | ,198        | 2,470              | ,210                        |
|            | Lower-bound        | 11,556                  | 1,000  | 11,556      | 1,235 | ,317 | ,198        | 1,235              | ,149                        |
| Error(B)   | Sphericity Assumed | 46,778                  | 10     | 4,678       |       |      |             |                    |                             |
|            | Greenhouse-Geisser | 46,778                  | 8,758  | 5,341       |       |      |             |                    |                             |
|            | Huynh-Feldt        | 46,778                  | 10,000 | 4,678       |       |      |             |                    |                             |
|            | Lower-bound        | 46,778                  | 5,000  | 9,356       |       |      |             |                    |                             |
| C          | Sphericity Assumed | 20,250                  | 1      | 20,250      | 2,306 | ,189 | ,316        | 2,306              | ,236                        |
|            | Greenhouse-Geisser | 20,250                  | 1,000  | 20,250      | 2,306 | ,189 | ,316        | 2,306              | ,236                        |
|            | Huynh-Feldt        | 20,250                  | 1,000  | 20,250      | 2,306 | ,189 | ,316        | 2,306              | ,236                        |
|            | Lower-bound        | 20,250                  | 1,000  | 20,250      | 2,306 | ,189 | ,316        | 2,306              | ,236                        |
| Error(C)   | Sphericity Assumed | 43,917                  | 5      | 8,783       |       |      |             |                    |                             |
|            | Greenhouse-Geisser | 43,917                  | 5,000  | 8,783       |       |      |             |                    |                             |
|            | Huynh-Feldt        | 43,917                  | 5,000  | 8,783       |       |      |             |                    |                             |
|            | Lower-bound        | 43,917                  | 5,000  | 8,783       |       |      |             |                    |                             |
| B * C      | Sphericity Assumed | 12,667                  | 2      | 6,333       | 2,135 | ,169 | ,299        | 4,270              | ,337                        |
|            | Greenhouse-Geisser | 12,667                  | 1,592  | 7,955       | 2,135 | ,183 | ,299        | 3,399              | ,293                        |
|            | Huynh-Feldt        | 12,667                  | 2,000  | 6,333       | 2,135 | ,169 | ,299        | 4,270              | ,337                        |
|            | Lower-bound        | 12,667                  | 1,000  | 12,667      | 2,135 | ,204 | ,299        | 2,135              | ,222                        |
| Error(B*C) | Sphericity Assumed | 29,667                  | 10     | 2,967       |       |      |             |                    |                             |
|            | Greenhouse-Geisser | 29,667                  | 7,962  | 3,726       |       |      |             |                    |                             |
|            | Huynh-Feldt        | 29,667                  | 10,000 | 2,967       |       |      |             |                    |                             |
|            | Lower-bound        | 29,667                  | 5,000  | 5,933       |       |      |             |                    |                             |

a. Computed using alpha = ,05

SPSS Processor is ready | H: 293, W: 612 pt

Für die beiden Faktoren B und C und für die Wechselwirkung B\*C werden jeweils vier Zeilen ausgedruckt. Die erste davon enthält die üblichen Angaben, Quadratsumme, Zahl der Freiheitsgrade, Varianzschätzung gleich mittleres Quadrat, gegebenenfalls F-Wert und Überschreitungswahrscheinlichkeit für die Erfüllung der Voraussetzung *Sphericity*. Die zwei folgenden Zeilen enthalten Berechnungen nach Greenhouse-Geisser und nach Huynh-Feldt für den Fall einer Verletzung der Voraussetzung. Die Zeile *Lower-bound* bezieht sich auf den möglichen ungünstigsten Fall. Da in unserem Beispiel die Voraussetzung erfüllt ist, interessiert nur die jeweils erste Zeile. Wie die Überschreitungswahrscheinlichkeiten zeigen, wurde die Signifikanzgrenze nirgends erreicht. Das liegt natürlich an dem kleinen Stichprobenumfang dieses Beispiels. Das Effektmaß  $\eta^2$  erreicht trotz Nichtsignifikanz noch ganz passable Werte nahezu bis 0,3, was ebenfalls mit dem geringen Stichprobenumfang zusammenhängt. Die Macht  $1 - \beta$  bleibt bei allen drei Signifikanztests deutlich unter 0,5.

Die Graphik zeigt wie gewohnt die sechs Zellenmittelwerte. Aus Platzgründen soll sie hier nicht wiedergegeben werden. Der Ausgabebildschirm zu dieser Aufgabe zeigt noch weitere Tabellen, die hier aber außer Betracht bleiben können.

**Anmerkung:** In den Dateien **Aufgabe17.sav**, **Aufgabe18.sav** und **Aufgabe19.sav** befinden sich die Daten zu den Erklärungsbeispielen in der Vorlesung zur einfaktoriellen Varianzanalyse, zum Newman-Keuls-Test und zur zweifaktoriellen Varianzanalyse (Glaser, 1978, S. 168 ff.). Diese Beispiele enthalten gegenüber den Aufgaben 11 bis 16 dieses Scriptums nichts Neues und werden daher hier nicht behandelt. Hier wird unter den Aufgabennummern 17, 18 und 19 die Vektoren- und Matrizenrechnung mit EXCEL dargestellt, zu der keine Datendateien vorgegeben werden.

## Aufgabe 17

Gegeben sind die drei Vektoren

$\mathbf{a} = (7 \ 4 \ 3 \ 5 \ 8 \ 9 \ 4)$ ,  $\mathbf{b} = (2 \ 1 \ 8 \ 4 \ 3 \ 1 \ 5)$  und  $\mathbf{c} = (6 \ 4 \ 1 \ 7 \ 3 \ 5 \ 8)$ .

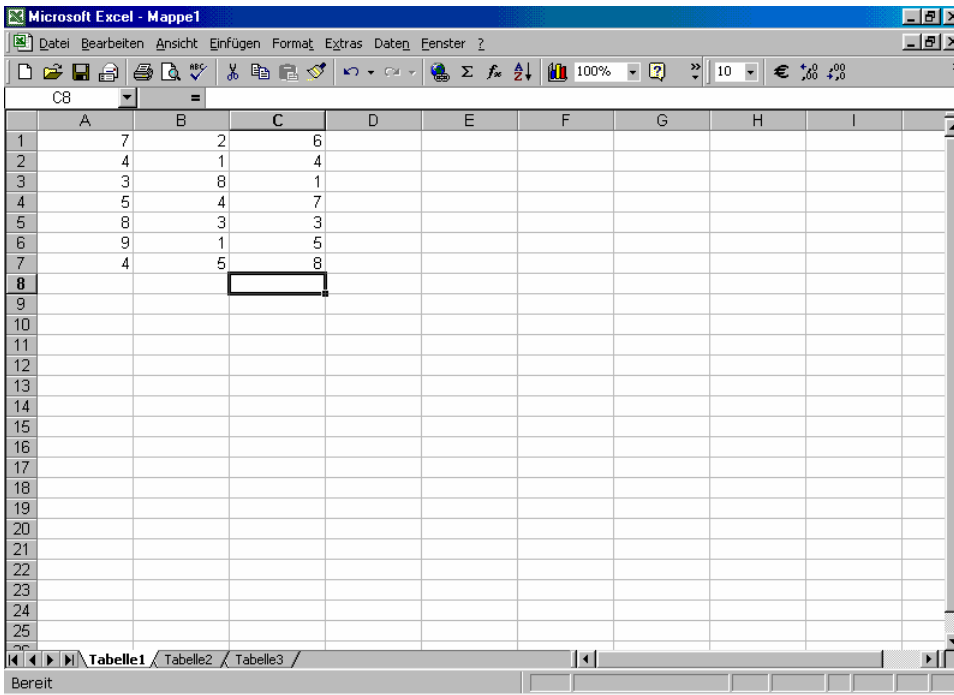
- 17.1 Wie lautet der Vektor  $\mathbf{d} = \mathbf{a} + \mathbf{b}$ ?
- 17.2 Wie lautet der Vektor  $\mathbf{e} = \mathbf{a} + \mathbf{b} + \mathbf{c}$ ?
- 17.3 Wie lautet der Vektor  $\mathbf{f} = \mathbf{a} + \mathbf{b} - \mathbf{c}$ ?
- 17.4 Wie lautet der Vektor  $\mathbf{g} = \mathbf{a} - \mathbf{b} - \mathbf{c}$ ?
- 17.5 Berechnen Sie den Abstand der Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ .
- 17.6 Berechnen Sie die Distanz zwischen den Vektoren  $\mathbf{e}$  und  $\mathbf{f}$ .
- 17.7 Berechnen Sie den Betrag des Vektors  $\mathbf{a}$ .
- 17.8 Berechnen Sie den Betrag des Vektors  $\mathbf{g}$ .
- 17.9 Bilden Sie die Produkte  $\mathbf{h} = 5\mathbf{a}$  und  $\mathbf{i} = 5\mathbf{b}$ .
- 17.10 Berechnen Sie den Abstand zwischen den Vektoren  $\mathbf{h}$  und  $\mathbf{i}$ .
- 17.11 Bilden Sie die Produkte  $(\mathbf{ab})\mathbf{c}$  und  $\mathbf{a}(\mathbf{bc})$ .

(Quelle: Übungsbeispiele zur Faktorenanalyse, Datei *Stat3\_01.mcd*, Aufgabe 1)

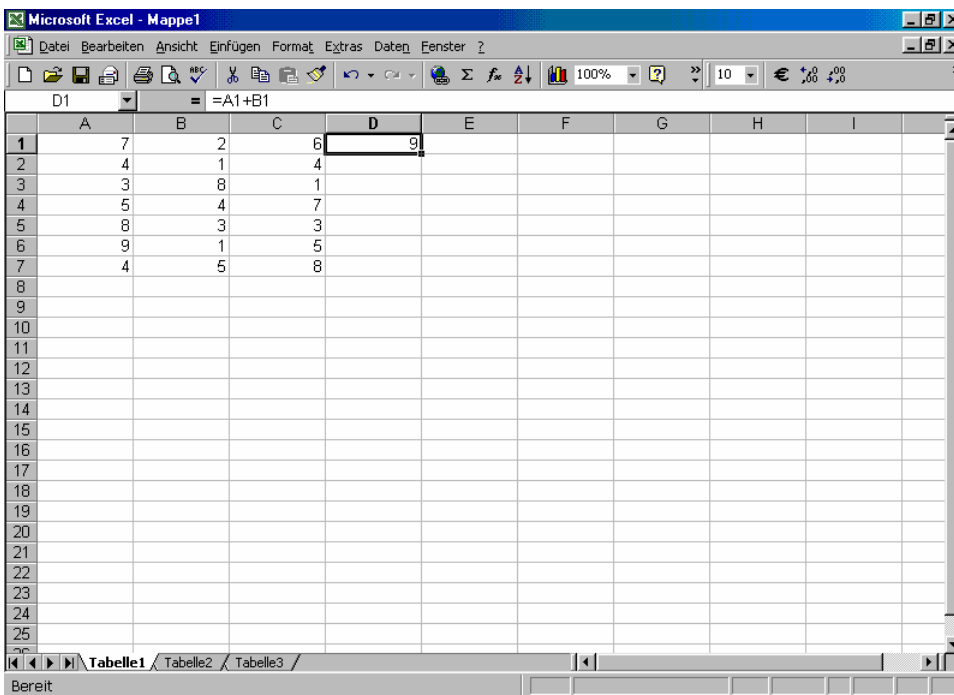
## Lösung

Wir bearbeiten diese Aufgabe mit dem Rechenblatt von EXCEL 2000. Nach dem Öffnen des Programmes erhalten wir ein Rechenblatt, das dem Datenbildschirm von SPSS nicht unähnlich ist. Die Zeilen sind numeriert, die Spalten mit Buchstaben gekennzeichnet. Eine einzelne Zelle wird mit dem zugehörigen Buchstaben und der zugehörigen Zahl adressiert: z. B. *C17*. Einen Vektor kennzeichnet man mit der ersten und der letzten Zelle, verbunden mit einem Doppelpunkt: z. B. für einen Spaltenvektor *B3:B10*, oder für einen Zeilenvektor *C5:AB5*. (Die 26. Spalte heißt *Z*, die 27. Spalte *AA* usw. Es werden also auch zwei- und mehrstellige Buchstaben "zahlen" zur Numerierung der Spalten verwendet.) Für die Bezeichnung von Matrixelementen  $A_{ij}$  gilt im allgemeinen die Regel, daß zuerst die Zeilenkennung, dann die Spaltenkennung genannt wird.  $A_{ij}$  bedeutet also den Skalar in der *i*-ten Zeile und der *j*-ten Spalte. Leider verstößt EXCEL durch die Nennung des Buchstabens vor der Zahl gegen diese Konvention. Das ist gewöhnungsbedürftig. Es läßt sich in dem Fenster ändern, das sich nach *Extras* → *Optionen* → *Allgemein* öffnet, indem man dort *ZISI-Bezugsart* anklickt. Das soll hier jedoch unterbleiben, da üblicherweise mit der Standardeinstellung gearbeitet wird.

Wir tragen die drei Vektoren  $\mathbf{a}$ ,  $\mathbf{b}$  und  $\mathbf{c}$  als Spaltenvektoren ein. Das sieht dann so aus:



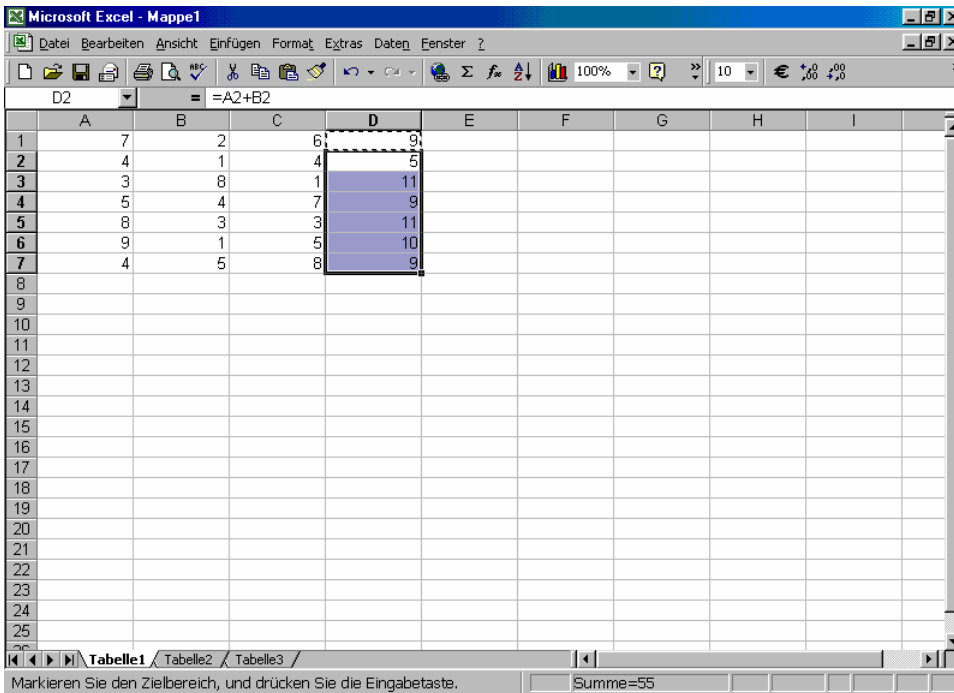
17.1 Den Vektor  $\mathbf{d}$  bilden wir als Spaltenvektor unter  $D$ . Dazu klicken wir zunächst die Zelle  $D1$  an. Danach klicken wir die lange Eingabezeile oberhalb der Buchstabenzeile an. Links vom Gleichheitszeichen erscheinen ein rotes Kreuz und ein grüner Haken. Als erstes Zeichen geben wir das Gleichheitszeichen ein. Es bedeutet, daß die nachfolgende Eingabe weder als Ziffern-, noch als Texteingabe, sondern als Gleichung interpretiert wird, deren Wert zu berechnen ist. Wir schließen diese Eingabe mit dem Anklicken des grünen Hakens ab und erhalten den folgenden Bildschirm:



Im Adressenfeld oben links steht jetzt  $D1$ , weil wir diese Zelle vor dem Ausfüllen der Eingabezeile aktiviert haben. In der Eingabezeile steht die Gleichung  $=A1+B1$ , und in der Zelle  $D1$  steht das Zahlenergebnis  $9$ . Daran sehen Sie das Grundprinzip der Funktionsweise von EXCEL: In die einzelnen Zellen können Gleichungen eingegeben werden, die Berechnungen mit den Werten anderer Zellen ausführen. Deshalb heißen solche Programme *Tabellenkalkulation*.

Bis hierher haben wir nur den ersten Skalar des gesuchten Vektors  $\mathbf{d}$ . Zur Berechnung der anderen Skalare des Vektors nutzen wir eine weitere Funktion von EXCEL. Klicken Sie bei noch aktivierter

Zelle  $D1$  Bearbeiten  $\rightarrow$  Kopieren an. Zur Anzeige, daß sich der Inhalt von  $D1$  jetzt in der Zwischenablage befindet, beginnt der Rand der Zelle zu flimmern. Klicken Sie dann  $D2$  an und ziehen Sie die Aktivierung bis  $D7$ . Drücken Sie danach *Bearbeiten*  $\rightarrow$  Einfügen. Ihr Bildschirm sieht jetzt so aus:



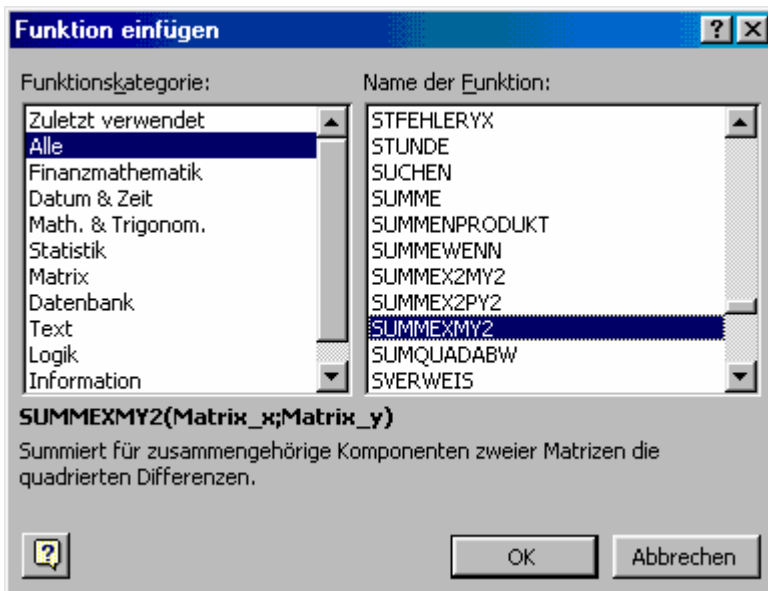
Wie Sie sehen, steht in  $D1:D7$  der gesuchte Vektor  $\mathbf{d}$ . Was hat EXCEL gemacht? Aktivieren Sie zur Erklärung die Zelle  $D6$ , in der der Skalar  $10$  steht, und betrachten Sie den Inhalt der Eingabezeile. Hier steht jetzt  $=A6+B6$ . Beim Kopieren des Inhalts der Zelle  $D1$  in die Zellen  $D2:D7$  hat also EXCEL die Zeilennummern automatisch so abgeändert, daß immer die zugehörige Zeile der Spaltenvektoren  $\mathbf{a}$  und  $\mathbf{b}$  addiert wird. Das gilt ganz generell für EXCEL: Beim Kopieren von Gleichungen in andere Zeilen oder Spalten werden die Adressen der Operanden so umgerechnet, daß entsprechende Matrizenoperationen, im einfachsten Fall Addition und Subtraktion, entstehen.

17.2 Auf die gleiche Art und Weise erzeugen wir den Vektor  $\mathbf{e}$  in  $E1:E7$ .

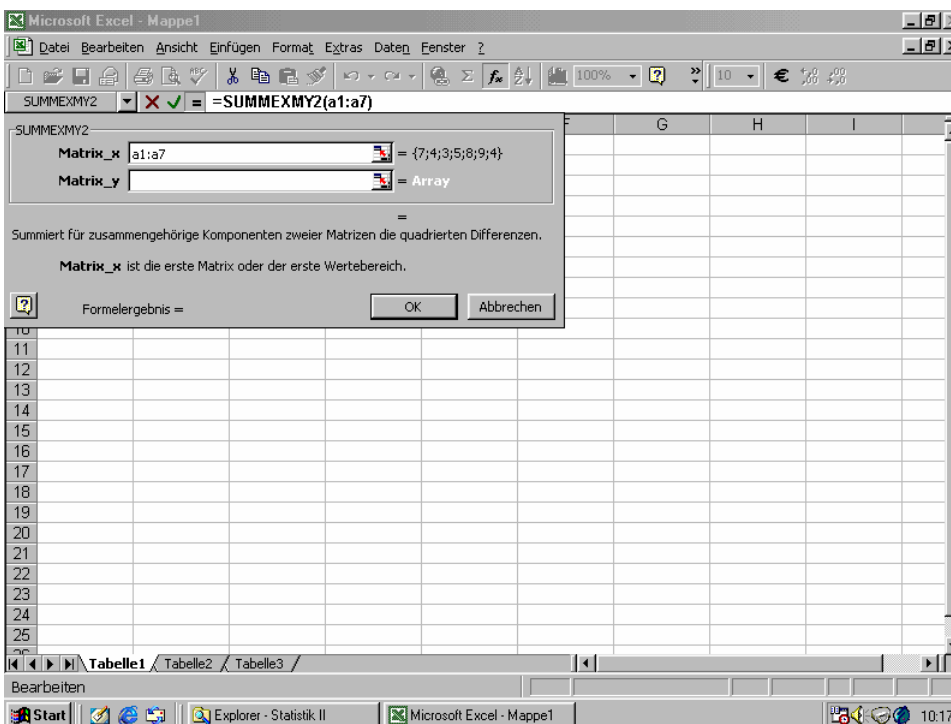
17.3 Auf die gleiche Art und Weise erzeugen wir den Vektor  $\mathbf{f}$  in  $F1:F7$ .

17.4 Auf die gleiche Art und Weise erzeugen wir den Vektor  $\mathbf{g}$  in  $G1:G7$ . Die Resultate sind in einer der nächsten Abbildungen enthalten.

17.5 Der Abstand zweier Vektoren ist die Quadratwurzel aus der Summe der quadrierten Differenzen einander entsprechender Skalare beider Vektoren (Vorlesung, Gl. (107)). EXCEL verfügt über eine Bibliothek von Funktionen, die wir hier ausnutzen können. Die Summe der quadrierten Differenzen wollen wir in die Zelle  $A9$  eintragen. Aktivieren Sie deshalb diese Zelle durch Anklicken. Ein Auswahlmenü für die Bibliotheksfunktionen erhalten Sie danach durch Anklicken des Icons  $fx$  in der oberen Symbolleiste:



Wählen Sie hier auf der linken Seite *Alle* und gehen Sie dann auf der rechten Seite im Alphabet zu *Summe...* Die Bedeutung jeder einzelnen Funktion wird Ihnen als kurzer Text angezeigt, wenn Sie sie anklicken. Im Bild sehen Sie, daß die Funktion *SUMMEXMY2* liefert, was wir jetzt brauchen. Durch *OK* können Sie sie in die vorher aktivierte Zelle *A9* eintragen. Zunächst erhalten Sie den Bildschirm:



Hier können Sie hinter *Matrix\_x* den Vektor **a**, also *a1:a7*, und hinter *Matrix\_y* den Vektor **b**, also *b1:b7*, eintragen. Die beiden Vektoren werden Ihnen dann noch zur Kontrolle angezeigt (im Bild nur Vektor **a**). Nach *OK* sehen Sie in der Zelle *A9* den Zahlenwert *150* und in der Eingabezeile die berechnete Funktion *SUMMEXMY2(A1:A7;B1:B7)*. EXCEL erlaubt also bei der Eingabe für die Adressen auch Kleinbuchstaben, wandelt sie für seine Anzeigen und Ausgaben aber in Großbuchstaben um.

Auf die gleiche Art und Weise tragen wir jetzt die Quadratwurzel aus *A9* in *A10* ein. Die zugehörige Funktion heißt *WURZEL()*. In englischsprachigen Programmen sollten Sie nach *SQRT*, der Abkürzung für *square root*, suchen. Der Zahlenwert lautet auf *12,2474487*.

17.6 In der gleichen Weise berechnen Sie die Summe der quadrierten Differenzen der einzelnen Komponenten von **e** und **f** in der Zelle E9, ihre Quadratwurzel in der Zelle E10. Die Zahlenwerte lauten 800 und 28,2842712.

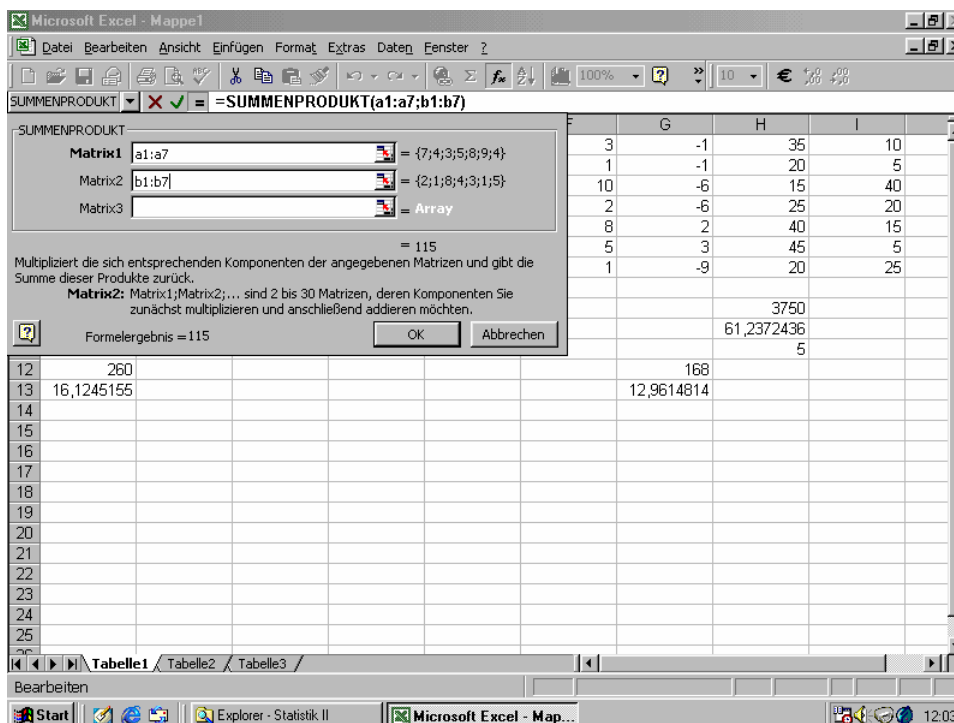
17.7 Der Betrag eines Vektors ist gleich der Quadratwurzel aus der Summe seiner quadrierten Komponenten. Auch hier gehen wir in zwei Schritten vor; die Summe der quadrierten Komponenten berechnen wir als Skalarprodukt des Vektors **a** mit sich selbst mit der Funktion *SUMMENPRODUKT* in der Zelle A12. Hier müssen wir den Vektor in der Form *a1:a7* zweimal eingeben. Den Betrag rechnen wir mit der Funktion *WURZEL* in der Zelle A13. Die Zahlen lauten 260 und 16,1245155.

17.8 Den Betrag des Vektors **g** rechnen wir in der gleichen Weise in den Zellen G12 und G13. Das geht am einfachsten, indem wir die Zelle A12 in G12 und die Zelle A13 in G13 kopieren. Die Zahlen lauten 168 und 12,9614814.

17.9 Klicken Sie die Zelle H1 an und tragen Sie in die Eingabezeile  $=5*a1$  ein. Kopieren Sie diese Zelle dann in die Zwischenablage. Aktivieren Sie jetzt die Zellen H1:I7 und übertragen Sie den Inhalt der Zwischenablage mit *Einfügen*. Die gesuchten Vektoren **h** und **i** erscheinen in H1:H7 und I1:I7.

17.10 Aktivieren Sie die Zellen A9:A10. Klicken Sie *Bearbeiten* → *Kopieren*. Aktivieren Sie jetzt die Zellen H9:H10 und klicken Sie *Bearbeiten* → *Einfügen*. In H9 erscheint die Summe der quadrierten Differenzen von **h** und **i**, in H10 die Quadratwurzel daraus. Die Zahlenwerte sind 3750 und 61,2372436. (Bitte machen Sie sich genau klar, warum das funktioniert!) Geben Sie jetzt in die Zelle H11 ein:  $=h10/a10$ . Als Ergebnis erhalten Sie 5. Warum?

17.11 Zunächst berechnen wir in Zelle B9 das Skalarprodukt **ab** mit der Funktion *SUMMENPRODUKT* (Die Arbeit mit dem Auswahlfenster für die Funktionen können Sie sich erleichtern, wenn Sie auf der linken Seite *Zuletzt benutzt* aktivieren):



Danach kopieren wir die Zelle B9 in die Zelle C9. Diese enthält nun das Skalarprodukt **bc** (warum?). Die Zahlen lauten 115 und 106. Für das Produkt **(ab)c** müssen wir nun einen Vektor erzeugen, bei dem die Skalare von **c** mit der Konstanten **ab** multipliziert werden. Zu diesem Zweck geben wir in die Zelle B11 ein:  $=c1*\$b\$9$ . Diese Eingabe kopieren wir über die Zwischenablage in die Zellen B12 bis B17. Im Bereich B11:B17 erhalten wir nun den gesuchten Vektor **(ab)c**.

Die Dollarzeichen in der Eingabe für Zelle *B11* bedeuten *absolute Adressierung*. In der Zelle *B11* wurde aufgrund der Eingabe der Inhalt von *C1* mit dem Inhalt von *B9* multipliziert. Hätten wir nun *B11* wie bisher ausgefüllt, dann wäre beim Kopieren, beispielsweise nach *B15*, dort die Multiplikation von *C5* mit *B13* eingetragen worden, die in unserem Beispiel keinen Sinn, in vielen Fällen auch einen Fehler, ergibt. Beim Kopieren müssen also die Ziffern der C-Adressen angepaßt, die Adresse *B9* hingegen konstantgehalten werden. Das ist die Funktion der absoluten Adressierung *\$b\$9*. Das vor eine Zeilen- oder Spaltenangabe oder vor beide gesetzte Dollarzeichen besagt also, daß die jeweils folgende Adresse beim Kopieren der Formel in eine andere Zelle unverändert erhalten bleiben soll.

Für den Vektor **a(bc)** tragen wir in die Zelle *C11* ein:  $=a1*\$c\$9$ . Nach dem Kopieren des Zelleninhaltes in die Zellen *C12* bis *C17* erhalten wir den gesuchten Vektor in *C11:C17*.

Mit allen diesen Berechnungen zeigt der Bildschirm schließlich:

|    | A          | B   | C   | D  | E          | F  | G          | H          | I  |
|----|------------|-----|-----|----|------------|----|------------|------------|----|
| 1  | 7          | 2   | 6   | 9  | 15         | 3  | -1         | 35         | 10 |
| 2  | 4          | 1   | 4   | 5  | 9          | 1  | -1         | 20         | 5  |
| 3  | 3          | 8   | 1   | 11 | 12         | 10 | -6         | 15         | 40 |
| 4  | 5          | 4   | 7   | 9  | 16         | 2  | -6         | 25         | 20 |
| 5  | 8          | 3   | 3   | 11 | 14         | 8  | 2          | 40         | 15 |
| 6  | 9          | 1   | 5   | 10 | 15         | 5  | 3          | 45         | 5  |
| 7  | 4          | 5   | 8   | 9  | 17         | 1  | -9         | 20         | 25 |
| 8  |            |     |     |    |            |    |            |            |    |
| 9  | 150        | 115 | 106 |    | 800        |    |            | 3750       |    |
| 10 | 12,2474487 |     |     |    | 28,2842712 |    |            | 61,2372436 |    |
| 11 |            | 690 | 742 |    |            |    |            | 5          |    |
| 12 | 260        | 460 | 424 |    |            |    | 168        |            |    |
| 13 | 16,1245155 | 115 | 318 |    |            |    | 12,9614814 |            |    |
| 14 |            | 805 | 530 |    |            |    |            |            |    |
| 15 |            | 345 | 848 |    |            |    |            |            |    |
| 16 |            | 575 | 954 |    |            |    |            |            |    |
| 17 |            | 920 | 424 |    |            |    |            |            |    |
| 18 |            |     |     |    |            |    |            |            |    |
| 19 |            |     |     |    |            |    |            |            |    |
| 20 |            |     |     |    |            |    |            |            |    |
| 21 |            |     |     |    |            |    |            |            |    |
| 22 |            |     |     |    |            |    |            |            |    |
| 23 |            |     |     |    |            |    |            |            |    |
| 24 |            |     |     |    |            |    |            |            |    |
| 25 |            |     |     |    |            |    |            |            |    |

Ihre Eingaben und Resultate können Sie nun noch als EXCEL-"Mappe" in die Datei *Aufgabe17.xls* speichern.

Denken Sie daran, daß wir mit der Eingabe der Gleichungen Berechnungen programmiert haben. Wenn Sie bei den Daten *A1:A7*, *B1:B7* und *C1:C7* auch nur eine einzige Zahl ändern, werden alle Berechnungen mit diesen Vektoren sofort aktualisiert. Ein solches Tabellenkalkulationsprogramm ist also ein enorm leistungsfähiges Hilfsmittel für alltägliche, vor allem auch kaufmännische, Berechnungsaufgaben. Wir befassen uns hier nur mit einem kleinen Ausschnitt seiner möglichen Verwendungen, der Vektoren- und Matrizenrechnung.

## Aufgabe 18

Gegeben sind die beiden Matrizen

$$A = \begin{bmatrix} 2 & 3 & 1 & 4 \\ 1 & 7 & 4 & 2 \\ 2 & 4 & 5 & 1 \\ 3 & 2 & 4 & 5 \end{bmatrix} \quad \text{und} \quad B = \begin{bmatrix} 1 & 5 & 8 & 4 \\ 2 & 3 & 6 & 5 \\ 1 & 4 & 3 & 2 \\ 5 & 3 & 1 & 6 \end{bmatrix}.$$

18.1 Berechnen Sie das Produkt  $\mathbf{AB}$  und das Produkt  $\mathbf{BA}$ .

18.2 Mit welcher Matrix  $\mathbf{X}$  muß  $\mathbf{A}$  postmultipliziert werden, damit im resultierenden Produkt  $\mathbf{C} = \mathbf{AX}$  der erste Spaltenvektor gleich dem Produkt des Skalars 5 mit dem ersten Spaltenvektor von  $\mathbf{A}$  wird? Es soll also  $\mathbf{c}_1 = 5\mathbf{a}_1$  werden. Bestimmen Sie  $\mathbf{X}$  weiterhin so, daß  $\mathbf{c}_2 = 3\mathbf{a}_2$ ,  $\mathbf{c}_3 = 7\mathbf{a}_3$  und  $\mathbf{c}_4 = 2\mathbf{a}_4$ .

18.3 Berechnen Sie die Transponierte des Produktes  $\mathbf{AB}$  und vergleichen Sie es mit den Produkten der transponierten Ausgangsmatrizen  $\mathbf{A'B'}$  und  $\mathbf{B'A'}$ .

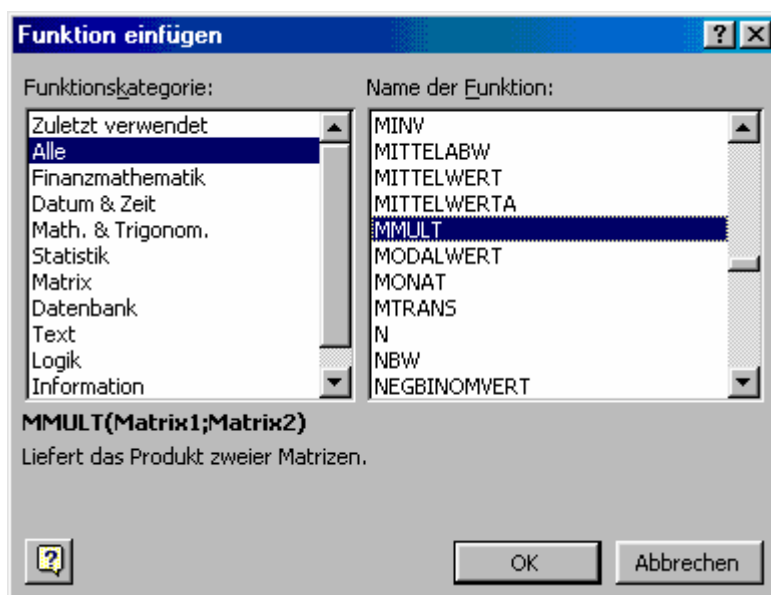
18.4 Berechnen Sie die Beträge der Zeilenvektoren der Matrix  $\mathbf{A}$  und die Winkel, die diese Zeilenvektoren miteinander bilden.

18.5 Lassen sich die Kosinusfunktionen der in 18.4 berechneten Winkel als Korrelationskoeffizienten zwischen den Zeilenvektoren interpretieren, wenn die Zeilenvektoren als psychologische Variablen aufgefaßt werden?

(Quelle: Übungsbeispiele zur Faktorenanalyse, Datei *Stat3\_01.mcd*, Aufgabe 4)

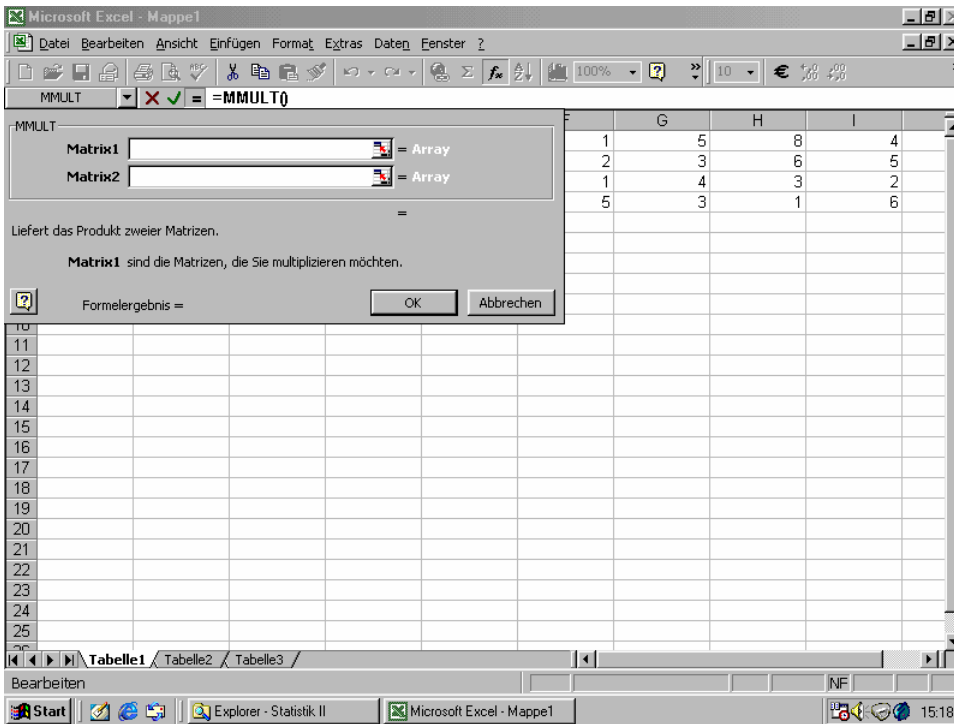
## Lösung

18.1 Öffnen Sie EXCEL 2000 und tragen Sie  $\mathbf{A}$  in  $A1:D4$  und  $\mathbf{B}$  in  $F1:I4$  ein. Das Produkt  $\mathbf{AB}$  existiert; es hat 4 Zeilen und 4 Spalten. Wir sehen dafür den Bereich  $A6:D9$  vor. Markieren Sie diesen Bereich. Klicken Sie jetzt in der oberen Symbolleiste  $f_x$  an und suchen Sie im nachfolgenden Fenster die Funktion  $\mathbf{MMULT}$  für die Matrizenmultiplikation:

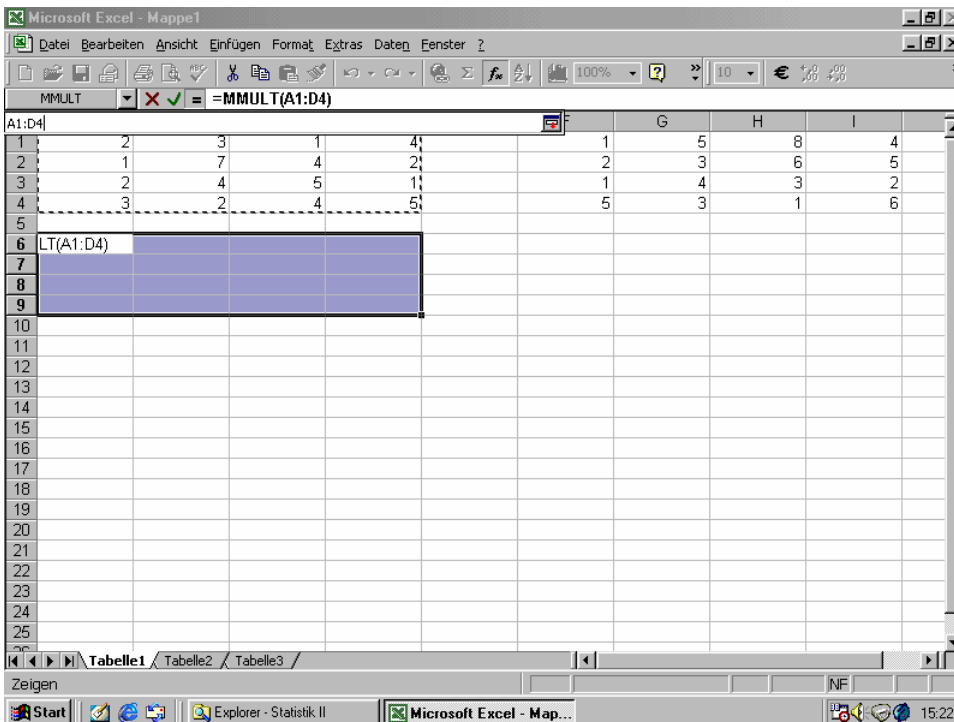


Mit  $\mathbf{OK}$  erhalten Sie:

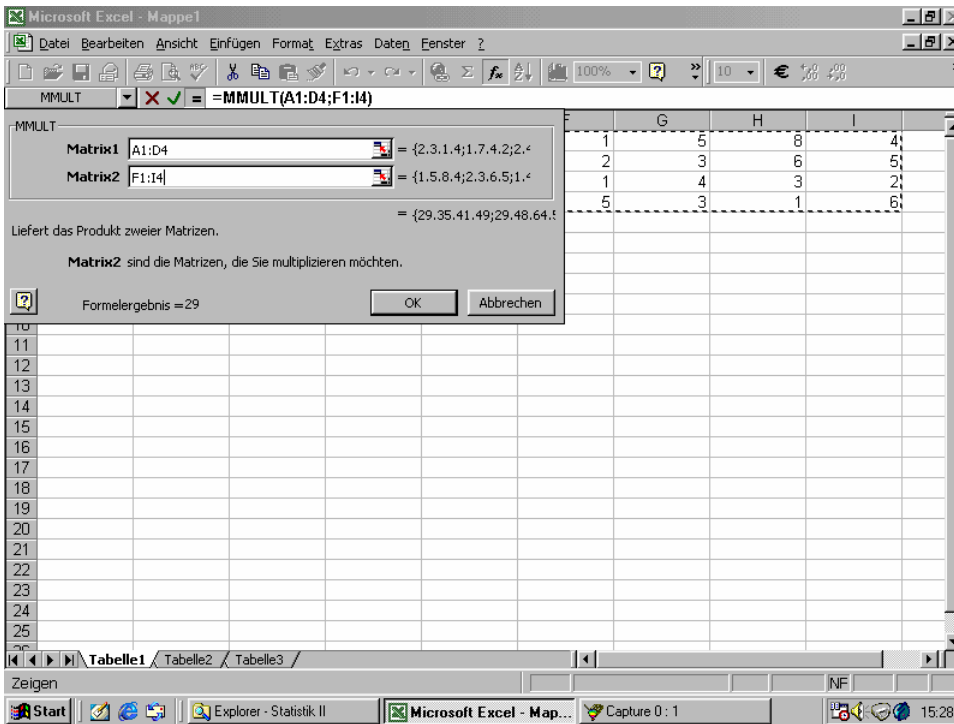




Hinter *Matrix1* und *Matrix2* können Sie jetzt *a1:d4* und *f1:i4* eintragen, was der Vorgehensweise in der vorangegangenen Aufgabe entspricht. Sie können aber auch den kleinen roten Pfeil vor *= Array* anklicken, und zwar zunächst für *Matrix1*. Im Datenbildschirm erhalten Sie eine Adressenzeile für die erste Matrix, die die Buchstabenzeile für die Spalten verdeckt:

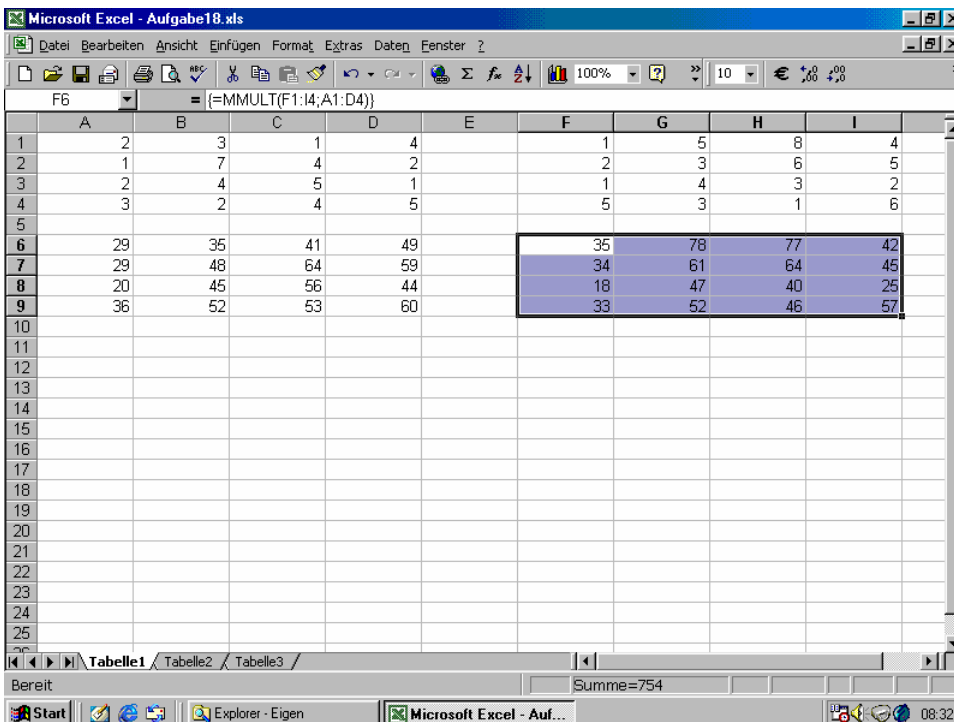


Jetzt können Sie die Matrix **A** durch Ziehen mit der Maus von der linken oberen Zelle *A1* bis zur rechten unteren Zelle *D4* eingeben; wie Sie sehen, wird dabei automatisch die Adresse *A1:D4* in die Adressenzeile eingesetzt. Die Matrix selbst erhält einen flimmernden Rand. Durch Anklicken des kleinen roten Pfeils am rechten Rand der Adressenzeile kehren Sie in den vorangehenden Bildschirm zurück; die Matrix **A** ist jetzt als erster Operand eingegeben. Nachdem Sie mit der Matrix **B** ebenso verfahren sind, erhalten Sie den Bildschirm:



Bevor Sie diesen mit *OK* verlassen, müssen Sie die Tasten *Control* und *Shift* (deutsch *Strg* und  $\uparrow$ ) drücken und bis zum Anklicken von *OK* gedrückt halten. Das Ergebnis ist das Matrizenprodukt **AB** im Bereich *A6:D9*. Wenn Sie *Control* und *Shift* vergessen, erzeugt EXCEL nur ein Element der Produktmatrix in der Zelle *A6*, jedoch keine komplette Matrix. Das ist softwareergonomisch nicht besonders geglückt; schlimmer ist, daß diese Information ist in den meisten Handbüchern und Hilfefunktionen von EXCEL gut versteckt wurde.

Für das Produkt **BA** aktivieren wir den Bereich *F6:I9* und wiederholen das Ganze sinngemäß. Mit beiden Matrizenprodukten sieht unser Bildschirm dann so aus:



18.2 Wenn wir eine Matrix mit spaltenweise konstanten Skalaren multiplizieren wollen, müssen wir sie mit einer Diagonalmatrix postmultiplizieren, die diese Skalare in den Diagonalzellen enthält. Diese Matrix **C** tragen wir in den Bereich *A11:D14* ein; die Zahlenwerte liefert der Aufgabentext.

Wir multiplizieren **AC** aus; für das Resultat wählen wir den Bereich *F11:I14*. Das Ergebnis zeigt der Bildschirm:

|    | A  | B  | C  | D  | E | F  | G  | H  | I  |
|----|----|----|----|----|---|----|----|----|----|
| 1  | 2  | 3  | 1  | 4  |   | 1  | 5  | 8  | 4  |
| 2  | 1  | 7  | 4  | 2  |   | 2  | 3  | 6  | 5  |
| 3  | 2  | 4  | 5  | 1  |   | 1  | 4  | 3  | 2  |
| 4  | 3  | 2  | 4  | 5  |   | 5  | 3  | 1  | 6  |
| 5  |    |    |    |    |   |    |    |    |    |
| 6  | 29 | 35 | 41 | 49 |   | 35 | 78 | 77 | 42 |
| 7  | 29 | 48 | 64 | 59 |   | 34 | 61 | 64 | 45 |
| 8  | 20 | 45 | 56 | 44 |   | 18 | 47 | 40 | 25 |
| 9  | 36 | 52 | 53 | 60 |   | 33 | 52 | 46 | 57 |
| 10 |    |    |    |    |   |    |    |    |    |
| 11 | 5  | 0  | 0  | 0  |   | 10 | 9  | 7  | 8  |
| 12 | 0  | 3  | 0  | 0  |   | 5  | 21 | 28 | 4  |
| 13 | 0  | 0  | 7  | 0  |   | 10 | 12 | 35 | 2  |
| 14 | 0  | 0  | 0  | 2  |   | 15 | 6  | 28 | 10 |
| 15 |    |    |    |    |   |    |    |    |    |
| 16 |    |    |    |    |   |    |    |    |    |
| 17 |    |    |    |    |   |    |    |    |    |
| 18 |    |    |    |    |   |    |    |    |    |
| 19 |    |    |    |    |   |    |    |    |    |
| 20 |    |    |    |    |   |    |    |    |    |
| 21 |    |    |    |    |   |    |    |    |    |
| 22 |    |    |    |    |   |    |    |    |    |
| 23 |    |    |    |    |   |    |    |    |    |
| 24 |    |    |    |    |   |    |    |    |    |
| 25 |    |    |    |    |   |    |    |    |    |

18.3 Mit der Funktion *MTRANS* erzeugen wir **A'** im Bereich *A16:D19* und **B'** im Bereich *F16:I19*. Danach erzeugen wir **A'B'** im Bereich *A21:D24* und **B'A'** im Bereich *F21:I24*. **AB** liegt aufgrund der Aufgabe 18.1 im Bereich *A6:D9* vor; wir erzeugen **(AB)'** im Bereich *A26:D29*. Das Ergebnis zeigt der Bildschirm:

|    | A             | B  | C  | D  | E | F  | G  | H  | I  |  |
|----|---------------|----|----|----|---|----|----|----|----|--|
| 15 | Aufgabe 18.3: |    |    |    |   |    |    |    |    |  |
| 16 | 2             | 1  | 2  | 3  |   | 1  | 2  | 1  | 5  |  |
| 17 | 3             | 7  | 4  | 2  |   | 5  | 3  | 4  | 3  |  |
| 18 | 1             | 4  | 5  | 4  |   | 8  | 6  | 3  | 1  |  |
| 19 | 4             | 2  | 1  | 5  |   | 4  | 5  | 2  | 6  |  |
| 20 |               |    |    |    |   |    |    |    |    |  |
| 21 | 35            | 34 | 18 | 33 |   | 29 | 29 | 20 | 36 |  |
| 22 | 78            | 61 | 47 | 52 |   | 35 | 48 | 45 | 52 |  |
| 23 | 77            | 64 | 40 | 46 |   | 41 | 64 | 56 | 53 |  |
| 24 | 42            | 45 | 25 | 57 |   | 49 | 59 | 44 | 60 |  |
| 25 |               |    |    |    |   |    |    |    |    |  |
| 26 | 29            | 29 | 20 | 36 |   |    |    |    |    |  |
| 27 | 35            | 48 | 45 | 52 |   |    |    |    |    |  |
| 28 | 41            | 64 | 56 | 53 |   |    |    |    |    |  |
| 29 | 49            | 59 | 44 | 60 |   |    |    |    |    |  |
| 30 |               |    |    |    |   |    |    |    |    |  |
| 31 |               |    |    |    |   |    |    |    |    |  |
| 32 |               |    |    |    |   |    |    |    |    |  |
| 33 |               |    |    |    |   |    |    |    |    |  |
| 34 |               |    |    |    |   |    |    |    |    |  |
| 35 |               |    |    |    |   |    |    |    |    |  |
| 36 |               |    |    |    |   |    |    |    |    |  |
| 37 |               |    |    |    |   |    |    |    |    |  |
| 38 |               |    |    |    |   |    |    |    |    |  |
| 39 |               |    |    |    |   |    |    |    |    |  |

Durch Vergleichen stellen wir fest, daß **(AB)' = B'A'**. Das ist ein Satz der Matrizenrechnung: Die Transponierte eines Matrizenproduktes ist gleich dem Produkt der transponierten **und vertauschten** Matrizen. Das bedeutet in der Regel **(AB)' ≠ A'B'**.

18.4 Für diese Aufgabe nehmen wir der Übersicht halber ein neues Tabellenblatt, *Tabelle2*, in das wir den Bereich *A1:D4* hineinkopieren. Dazu aktivieren wir zunächst *A1:D4* durch Ziehen mit der

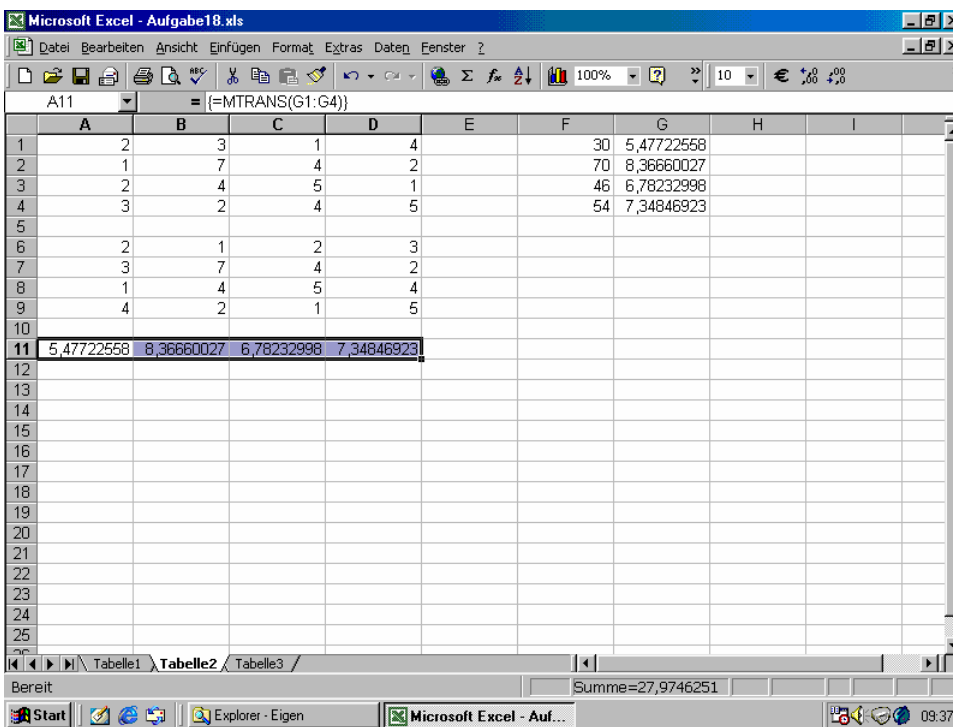
Maus. Danach klicken wir *Bearbeiten* → *Kopieren* an. Durch Anklicken von *Tabelle2* am unteren Fensterrand wählen wir diese Tabelle aus. Sie erscheint, leer und mit aktiver Zelle *A1*. Durch Anklicken von *Bearbeiten* → *Einfügen* kopieren wir die Matrix **A** in dieses Blatt.

Für die Beträge der Zeilenvektoren erzeugen wir zunächst deren Skalarprodukte mit sich selbst im Bereich *F1:F4*. Dazu geben wir in *F1* die Funktion `=SUMMENPRODUKT(A1:D1;A1:D1)` ein (vgl. Lösung 17.7 und 17.8). Danach kopieren wir diese Gleichung in *F2:F4*. Schließlich berechnen wir in *G1:G4* die gesuchten Beträge als Quadratwurzeln aus *F1:F4*. Die vier Beträge bilden jetzt einen Spaltenvektor, den wir **g** nennen wollen.

Den Winkel  $\theta$  zwischen zwei Vektoren **x** und **y** erhalten wir mit Gleichung (113) aus der

Vorlesung: 
$$\vartheta = \cos^{-1} \left( \frac{x^* y}{|x|^* |y|} \right).$$

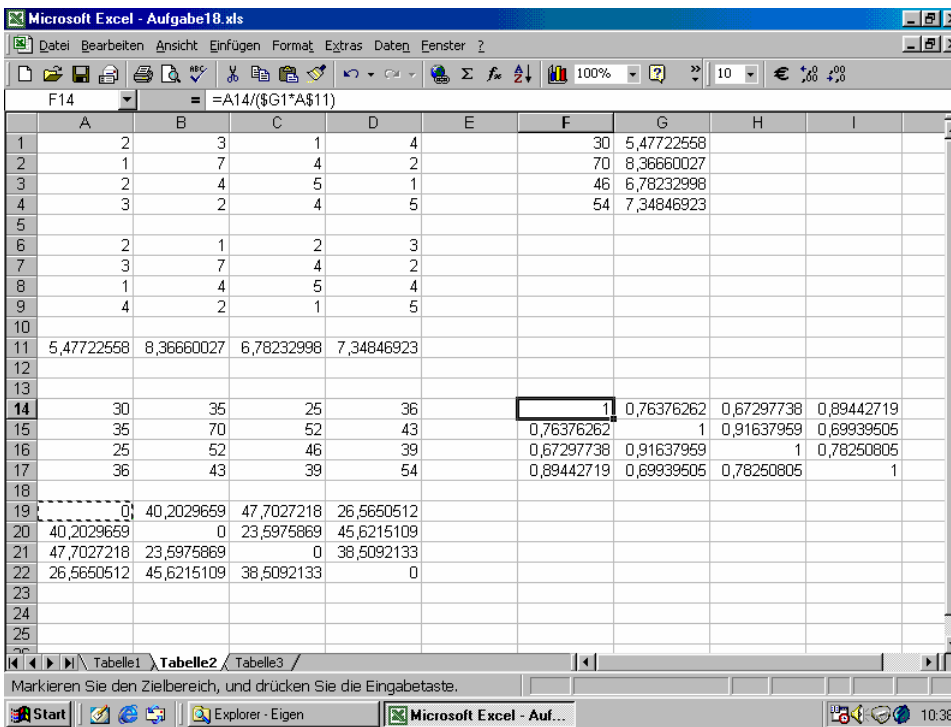
Um uns diese Rechnung zu erleichtern, erzeugen wir zunächst **A'** im Bereich *A6:D9* und die Transponierte von **g** als Zeilenvektor **g'** = *A11:D11*. Das Ergebnis zeigt der Bildschirm:



Jetzt verschaffen wir uns die Klammerausdrücke der Gleichung (113). Dazu berechnen wir zunächst in dem Bereich *A14:D17* das Matrizenprodukt **AA'**. Es enthält sämtliche Zähler des Klammerausdruckes von (113). Machen Sie sich klar, warum das so ist! Nun tragen wir in *F14* die Gleichung `=A14/($G1*A$11)` ein. Damit erreichen wir die Berechnung des gesamten Bruches für den Winkel zwischen dem ersten Spaltenvektor von **A** und sich selbst (ein an und für sich uninteressanter Grenzfall, den wir aber formal in der Berechnung mitführen). Der Zahlenwert zeigt sofort 1;  $\cos^{-1}(1) = 0$ , dieser Winkel beträgt also trivialerweise 0 Grad. Wenn wir jetzt die Zelle *F14* in den gesamten Bereich *F14:I17* kopieren, bekommen wir sämtliche Klammerausdrücke dieses Beispiels. Auch hier: durchdenken Sie das Beispiel gründlich und machen Sie sich klar, warum wir auf diesem Wege zum Ziel kommen. Wiederholen Sie insbesondere die Bedeutung der  $\$$ -Zeichen aus der Lösung 17.11.

Zum Schluß müssen wir noch die Winkel berechnen. Dazu tragen wir in das Feld *A19* die Gleichung `=ARCCOS(F14)*180/PI()` ein.  $ARCCOS(x)$  ist die Funktion, die in EXCEL  $\cos^{-1}(x)$  berechnet. Sie liefert ein Ergebnis im Bogenmaß, das noch in Grad umgerechnet werden muß. Das geschieht durch Multiplikation mit 180 und Division durch  $\pi$ . Nach dem Kopieren von *A19* in den

gesamten Bereich  $A19:D22$  erhalten wir die gesuchten Winkel in Matrixform. Der Bildschirm zeigt sich wie folgt:



Der Bereich  $A19:D22$  ist also so zu lesen, daß der Winkel zwischen dem zweiten und dem dritten Zeilenvektor von  $A$   $23,5975869$  Grad beträgt usw.

18.5 Die Kosinusfunktionen in  $F14:I17$  der gefundenen Winkel lassen sich nicht als Korrelationen zugehöriger Variablen interpretieren, weil es sich bei den Zeilenvektoren nicht um Abweichungsmaßzahlen handelt, deren Mittelwert also ungleich null ist.

Sie können Ihre Arbeit in der Datei *Aufgabe18.xls* speichern. Die beiden Tabellen gehören zu einer *Mappe*, die von EXCEL in einer einzigen Datei abgespeichert wird.

### Aufgabe 19

Gegeben sind die beiden Matrizen

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 4 & 2 & 3 \\ 1 & 4 & 2 \end{bmatrix} \text{ und } B = \begin{bmatrix} -8 & 14 & -1 \\ -5 & -3 & 17 \\ 14 & -1 & -10 \end{bmatrix} * \frac{1}{47}.$$

Bilden Sie das Produkt  $AB$ . Interpretieren Sie das Ergebnis.

(Quelle: Übungsbeispiele zur Faktorenanalyse, Datei *Stat3\_01.mcd*, Aufgabe 6)

### Lösung

Öffnen Sie EXCEL und tragen Sie  $A$  in  $A1:C3$ , den ganzzahligen Teil von  $B$  in  $E1:G3$  ein. Tragen Sie dann in  $E6 = E1/47$  ein. Kopieren Sie  $E6$  in  $E6:G9$ . Der Bereich  $E6:G9$  enthält jetzt die Matrix  $B$ . Berechnen Sie mittels *MMULT* das Matrizenprodukt  $AB$  im Bereich  $A6:C8$ . (Die Matrizenmultiplikation wird detailliert in Lösung 18.1 gezeigt.) Erzeugen Sie jetzt noch im Bereich  $A11:C13$  das Produkt aus  $A$  und dem ganzzahligen Teil von  $B$ . Das Ergebnis zeigt der Bildschirm:

|    | A  | B            | C          | D | E           | F           | G           | H | I |
|----|----|--------------|------------|---|-------------|-------------|-------------|---|---|
| 1  | 1  | 3            | 5          |   | -8          | 14          | -1          |   |   |
| 2  | 4  | 2            | 3          |   | -5          | -3          | 17          |   |   |
| 3  | 1  | 4            | 2          |   | 14          | -1          | -10         |   |   |
| 4  |    |              |            |   |             |             |             |   |   |
| 5  |    |              |            |   |             |             |             |   |   |
| 6  | 1  | -1,38778E-17 | 2,2204E-16 |   | -0,17021277 | 0,29787234  | -0,0212766  |   |   |
| 7  | 0  |              | 1          |   | -0,10638298 | -0,06382979 | 0,36170213  |   |   |
| 8  | 0  | 6,93889E-18  | 1          |   | 0,29787234  | -0,0212766  | -0,21276596 |   |   |
| 9  |    |              |            |   |             |             |             |   |   |
| 10 |    |              |            |   |             |             |             |   |   |
| 11 | 47 | 0            | 0          |   |             |             |             |   |   |
| 12 | 0  | 47           | 0          |   |             |             |             |   |   |
| 13 | 0  | 0            | 47         |   |             |             |             |   |   |
| 14 |    |              |            |   |             |             |             |   |   |
| 15 |    |              |            |   |             |             |             |   |   |
| 16 |    |              |            |   |             |             |             |   |   |
| 17 |    |              |            |   |             |             |             |   |   |
| 18 |    |              |            |   |             |             |             |   |   |
| 19 |    |              |            |   |             |             |             |   |   |
| 20 |    |              |            |   |             |             |             |   |   |
| 21 |    |              |            |   |             |             |             |   |   |
| 22 |    |              |            |   |             |             |             |   |   |
| 23 |    |              |            |   |             |             |             |   |   |
| 24 |    |              |            |   |             |             |             |   |   |
| 25 |    |              |            |   |             |             |             |   |   |

Das Matrizenprodukt ergibt die Identitätsmatrix, die in den Diagonalzellen 1, in den Nichtdiagonalzellen 0 enthält. Die drei Zahlen in Exponentendarstellung mit  $E-16$  bis  $E-18$  weichen nur um die Rundungsgenauigkeit des Rechners auf der 16. bis 18. Stelle rechts vom Komma von 0 ab. Die Matrix  $B$  ist also die Invertierte von  $A$ ,  $B = A^{-1}$ . Das Produkt von  $A$  mit dem ganzzahligen Teil von  $B$  zeigt eine Diagonalmatrix mit 47 in den Diagonalzellen. Deshalb wurde auch die Darstellung von  $B$  in der Form *ganzzahlige Matrix multipliziert mit 1/ganze Zahl* gewählt. Diese Zahl, im vorliegenden Beispiel 47, hat eine besondere Bedeutung bei der Matrizeninversion. Sie wird als *Determinante* von  $A$  bezeichnet. Matrizen, deren Determinante den Wert null hat, sind nicht invertierbar, weil die Berechnung dann auf eine Division durch null hinauslaufen würde. Sie werden als *singulär* bezeichnet.

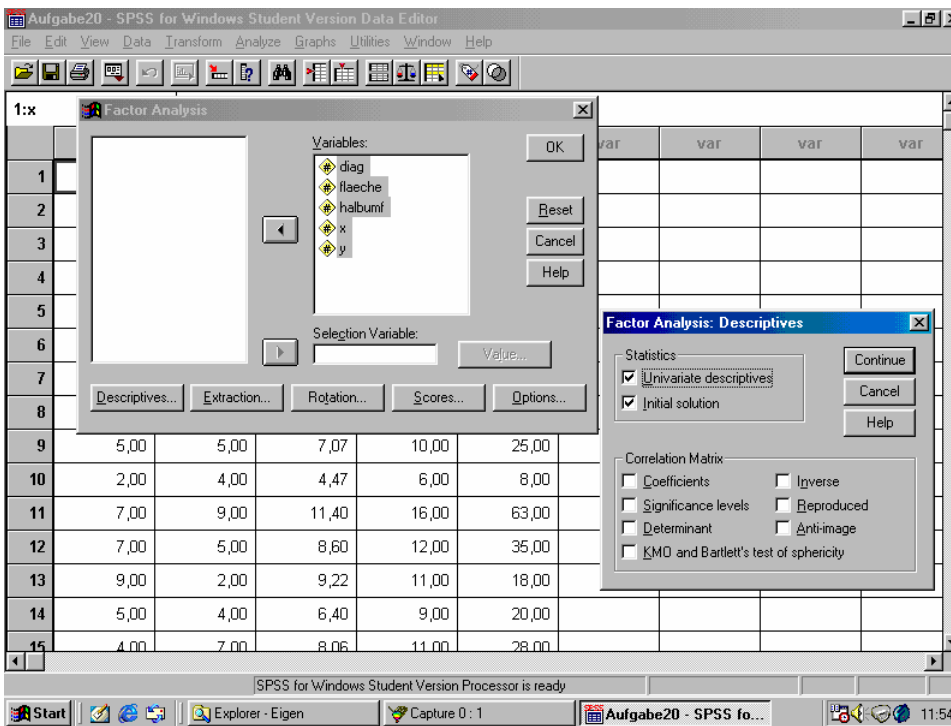
Sie können das Ergebnis dieser Aufgabe unter *Aufgabe19.xls* speichern.

## Aufgabe 20

Öffnen Sie die Datei *Aufgabe20.sav*. Sie enthält für 5 Variablen  $x$ ,  $y$ ,  $diag$ ,  $halbumf$  und  $flaeche$  50 Fälle. Unterziehen Sie diese Daten einer Faktorenanalyse, wobei 2 Faktoren extrahiert werden sollen. Lassen Sie sich das Scree-Plot und die Varimax-rotierte Lösung graphisch darstellen. Interpretieren Sie das Ergebnis.

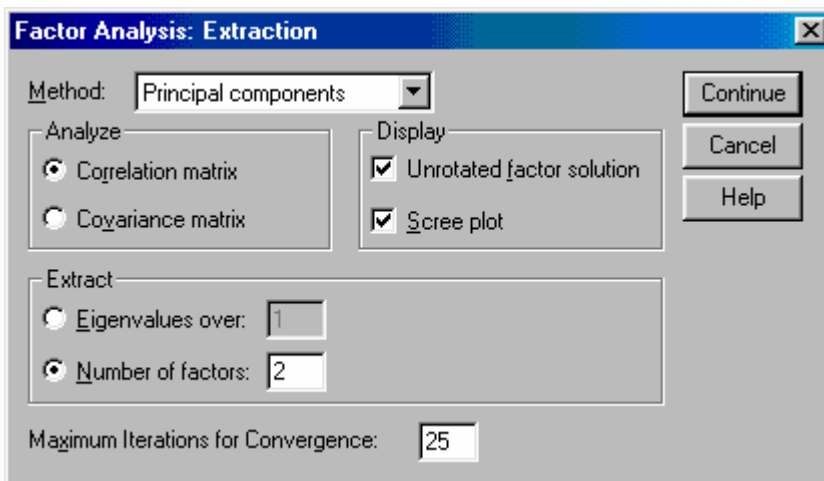
## Lösung

Mit *Analyse* → *Data Reduction* → *Factor* erhalten Sie das Definitionsfenster für die Faktorenanalyse. Bringen Sie zunächst die fünf Variablen in das rechte Fenster *Variables*. Danach müssen Sie sich durch die fünf Schaltfelder von *Descriptives* bis *Options* hindurcharbeiten. *Descriptives* liefert das folgende Fenster:

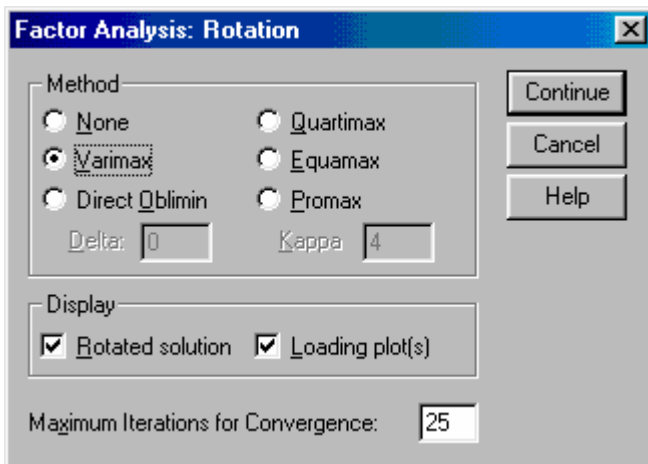


Außer *Initial solution*, das voreingestellt ist, sollten Sie hier *Univariate descriptives* aktivieren. Darüber hinaus können Sie hier eine Reihe von Informationen und Transformationen der Korrelationsmatrix anfordern. Wir lassen das jetzt außer Betracht; bei einer großen Zahl von Variablen erzeugen Sie hier unter Umständen eine sehr umfangreiche Ausgabe, die Sie eine erhebliche Rechnerzeit, eine sehr verminderte Übersicht und beim Druckbefehl einen gewaltigen Papierverbrauch kosten kann.

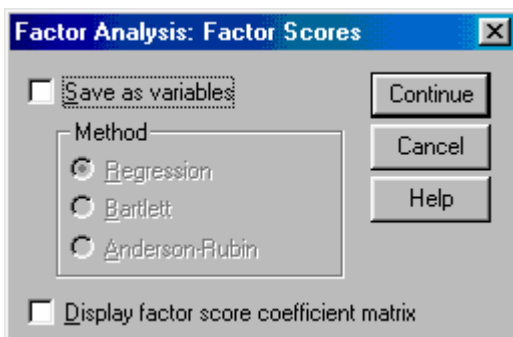
Nach der Rückkehr mit *Continue* öffnen wir *Extraction*:



Die Voreinstellungen *Method: Principal components*, *Analyze Correlation matrix* und *Display Unrotated factor solution* können Sie beibehalten. Aktivieren Sie darüber hinaus noch *Display Scree plot* und *Extract Number of factors: 2*, wie es die Abbildung zeigt. Nach der Rückkehr mit *Continue* öffnen Sie *Rotation*:

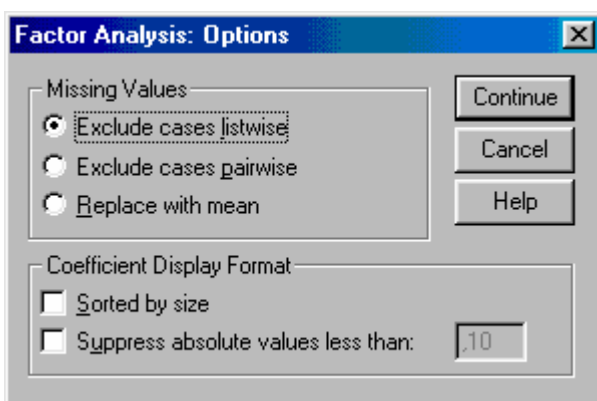


Ändern Sie die Voreinstellungen so ab, wie es das Bild zeigt. Wählen Sie also *Method Varimax*, die in der Praxis weitaus häufigste Einstellung, und fordern Sie mit *Loading plot(s)* eine Abbildung an. Nach der Rückkehr über *Continue* schauen Sie sich *Scores* an:



Hier können Sie die Berechnung der Faktorwertematrix anfordern, indem Sie *Save as variables* anklicken. Danach wird *Method* aktiv, wo *Regression* die übliche Wahl ist. Die Formulierung *Save as variables* sagt Ihnen, in welcher Form SPSS das Ergebnis liefert. Die Faktorenwertematrix enthält die Personen und die Faktoren als Eingänge. SPSS nutzt die Zeilen der Datenmatrix auch als Personeneingang für die Faktorenwertematrix; die Faktoren werden als neue Variablen in die Datenmatrix eingeführt. Wenn Sie also Faktorenwerte berechnen lassen, finden Sie diese nach dem Lauf von SPSS als zusätzliche Variablen in Ihrem Dateneingabebebildschirm vor. Unter *factor score coefficient matrix* versteht man die Matrix **B** der multiplen Regressionskoeffizienten der Faktoren auf die standardisierten Variablen, die Sie sich durch Anklicken von SPSS erzeugen lassen können. Es gilt  $\mathbf{P} = \mathbf{B}'\mathbf{Z}$  und  $\mathbf{B}' = \mathbf{A}'\mathbf{R}^{-1}$ .

Wir bleiben in diesem Fenster bei den Voreinstellungen. Nach *Continue* aktivieren wir *Options*:



Hier können wir die Behandlung fehlender Daten und die spätere Darstellung der Faktorenmatrix steuern. *Exclude cases listwise* bedeutet, daß jede Person, bei der auch nur eine Maßzahl fehlt, ganz aus der Faktorenanalyse ausgeschlossen wird. Das ist eine sehr rigorose, wenn auch statistisch



sichere Vorgehensweise, die man sich nur leisten kann, wenn insgesamt eine große Zahl von Daten vorliegt. *Exclude cases pairwise* bedeutet, daß bei jeder einzelnen Korrelation zwischen zwei Variablen nur genau die Personen ausgeschlossen werden, bei denen lokal wenigstens ein Wert fehlt. Solche Personen bleiben in anderen Teilen der Analyse, wo sie Daten geliefert haben, erhalten. Diese Methode hat den Vorteil, daß man weniger Personen für die Auswertung verliert. Sie hat aber den Nachteil, daß statistische Verzerrungen entstehen können, wenn die Datenausfälle von den Variablen abhängig sind, was sich in der Praxis nur schwer feststellen läßt. *Replace with mean* schließlich füllt fehlende Werte mit den Mittelwerten der jeweiligen Variablen auf. Damit verliert man keine Personen wegen fehlender Daten für die Analyse, mindert aber vorhandene Kovarianzen und Korrelationen, so daß die Faktorenanalyse mit einer kleinen Anzahl von Faktoren weniger Gesamtvarianz erklärt. Man verliert also bei diesem Verfahren Kommunalität.

Der untere Teil des Fensters erlaubt es, die Faktorenladungen, die hier einfach *Coefficients* heißen, in der Ausgabe nach der Größe zu ordnen und kleine Werte nicht auszudrucken. Das kann bei einer großen Zahl von Variablen die Übersicht verbessern.

Da wir keine *missing data* in diesem Beispiel haben, bleiben wir auch in diesem Fenster bei den Voreinstellungen. Nach *Continue* fordern wir mit *OK* die Faktorenanalyse an.

Der Ausgabebildschirm zeigt uns zunächst einige deskriptive Statistiken und die anfänglichen Schätzungen sowie späteren Werte der Kommunalitäten:

**Factor Analysis**

**Descriptive Statistics**

|               | Mean    | Std. Deviation | Analysis N |
|---------------|---------|----------------|------------|
| Diagonale     | 7,5193  | 2,6536         | 50         |
| Fläche        | 25,0400 | 19,8679        | 50         |
| Halber Umfang | 10,0000 | 3,7143         | 50         |
| Länge         | 5,0000  | 2,4578         | 50         |
| Breite        | 5,0000  | 2,7701         | 50         |

**Communalities**

|               | Initial | Extraction |
|---------------|---------|------------|
| Diagonale     | 1,000   | ,953       |
| Fläche        | 1,000   | ,909       |
| Halber Umfang | 1,000   | ,997       |
| Länge         | 1,000   | ,998       |
| Breite        | 1,000   | ,999       |

Extraction Method: Principal Component Analysis.

Die hohen Kommunalitäten, alle nahe 1,000, die niedrigste bei 0,909, zeigen an, daß die gefundenen Faktoren bei den einzelnen Variablen einen sehr großen Varianzanteil erklären.

Die nächste Tabelle gibt uns die wichtigsten Grundinformationen über die gerechnete Faktorenanalyse:

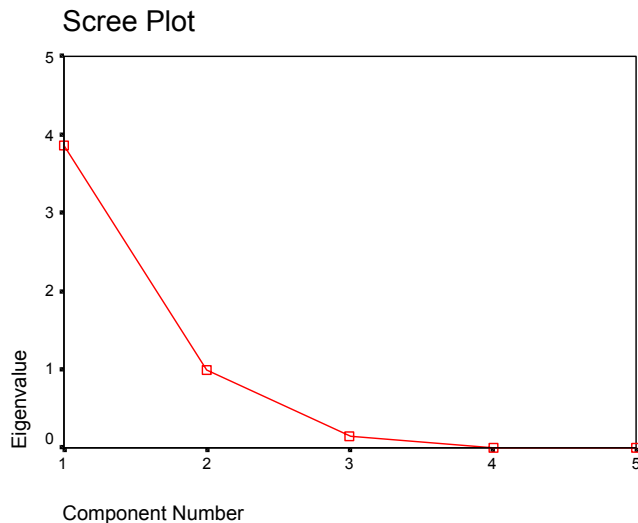
### Total Variance Explained

| Component | Initial Eigenvalues |               |              | Extraction Sums of Squared Loadings |               |              | Rotation Sums of Squared Loadings |               |              |
|-----------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|-----------------------------------|---------------|--------------|
|           | Total               | % of Variance | Cumulative % | Total                               | % of Variance | Cumulative % | Total                             | % of Variance | Cumulative % |
| 1         | 3,862               | 77,231        | 77,231       | 3,862                               | 77,231        | 77,231       | 2,814                             | 56,284        | 56,284       |
| 2         | ,994                | 19,890        | 97,121       | ,994                                | 19,890        | 97,121       | 2,042                             | 40,837        | 97,121       |
| 3         | ,139                | 2,774         | 99,895       |                                     |               |              |                                   |               |              |
| 4         | 5,262E-03           | ,105          | 100,000      |                                     |               |              |                                   |               |              |
| 5         | 9,875E-17           | 1,975E-15     | 100,000      |                                     |               |              |                                   |               |              |

Extraction Method: Principal Component Analysis.

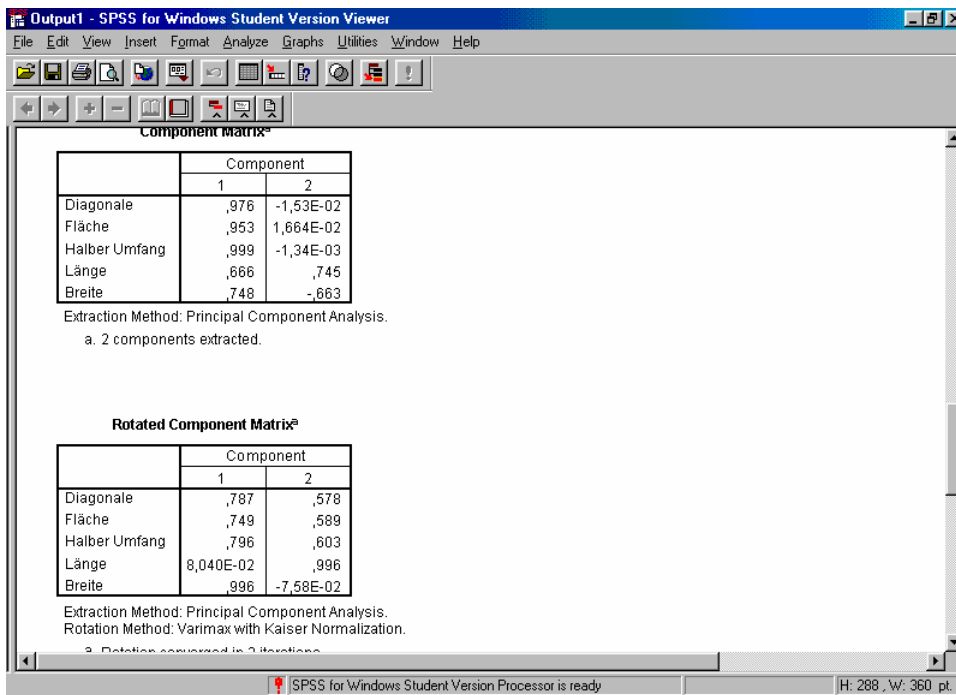
Der erste Eigenwert beträgt 3,862, der zweite 0,994. Die restlichen drei Eigenwerte sind zu vernachlässigen. Vor der Rotation erklärt der erste Faktor 77,231 % der Varianz, der zweite 19,890 %. Die Rotation verteilt die Varianzanteile besser auf beide Faktoren: Jetzt erklärt der erste Faktor 56,284 %, der zweite 40,837 % der Gesamtvarianz. Die beiden Faktoren erklären zusammen 97,121 % der Varianz, ein so hoher Anteil, wie er in der psychologischen Praxis kaum vorkommt.

Das Scree-Plot veranschaulicht die Größe der Eigenwerte:

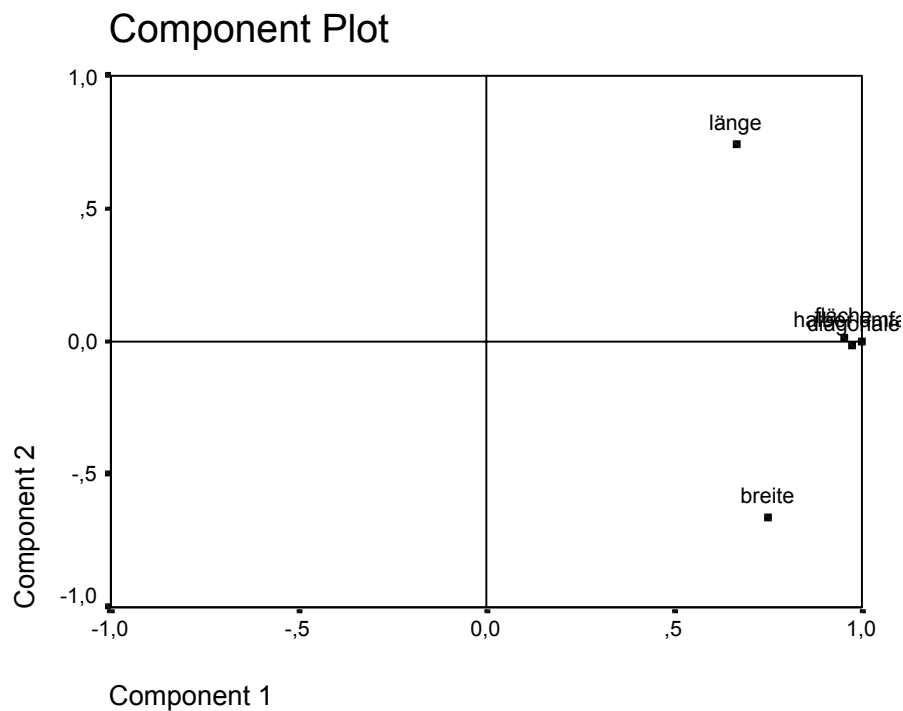


Die Entscheidung, zwei Faktoren zu extrahieren, ist angesichts dieses Diagrammes gerechtfertigt.

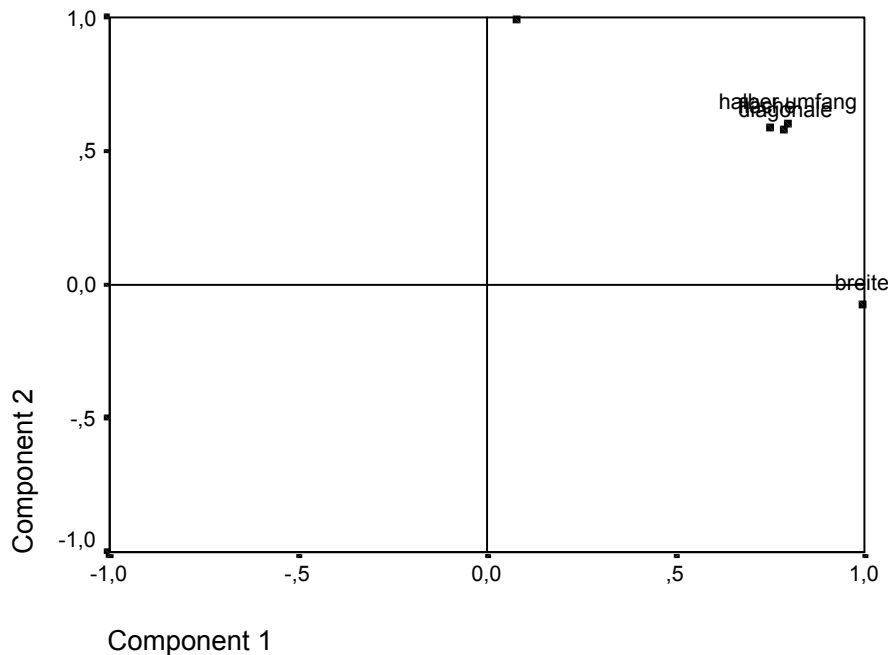
Als nächstes zeigt der Ausgabebildschirm die Faktorenmatrix vor und nach der Rotation:



Die Ausgabe endet mit einer graphischen Darstellung der rotierten Faktorenmatrix. Wenn wir unter *Rotation* die Voreinstellung *None* beibehalten, zeichnet SPSS die unrotierte Faktorenmatrix. Beide Graphiken sehen wie folgt aus, zunächst die unrotierte, dann die rotierte Faktorenmatrix:



## Component Plot in Rotated Space



Die Faktoren bilden die Koordinatenachsen, die Variablen sind als Punkte und nicht als Vektorpfeile eingetragen. Es ist deutlich zu sehen, daß die Extraktion einen ersten Faktor liefert, der praktisch perfekt mit den drei Variablen *diag*, *halbumfang* und *flaeche* zusammenfällt. Der zweite Faktor erklärt einen mittleren Teil der Varianz der Variablen *x* (Länge) und *y* (Breite). Nach der Rotation entspricht der Abszissenfaktor praktisch perfekt der Variablen *y* (Breite), der Ordinatenfaktor der Variablen *x* (Länge). Die drei Variablen *diag*, *halbumfang* und *flaeche* werden jetzt zu je etwa gleichen Teilen von beiden Faktoren erklärt.

Das führt zum letzten Schritt einer Faktorenanalyse, der Interpretation der Faktoren. Dieses Beispiel basiert nicht auf einer psychologischen Messung, sondern ist nach einer Idee von Thurstone konstruiert. Die Variablen *x* und *y* sind je eine Zufallsfolge der gleichverteilten Zahlen 1 ... 9. Sie müssen deshalb die Korrelation null aufweisen, also orthogonal sein. Zur Konstruktion der drei weiteren Variablen wurde jeder Wert von *x* als die Länge, jeder Wert von *y* als die Breite eines Rechtecks aufgefaßt. Die Größen *Diagonale*, *halber Umfang* und *Fläche* wurden dann nach den Regeln der Geometrie berechnet und als Werte in die Variablen *diag*, *halbumfang* und *flaeche* eingesetzt. In diesem Beispiel ist also jeder Fall, jede Zeile in der Datenmatrix, ein Rechteck und nicht, wie sonst üblich, eine Person. Die Idee Thurstones war es, daß die Faktorenanalyse in einem solchen konstruierten Beispiel genau die Struktur sichtbar machen sollte, die in die Daten hineingesteckt worden war: zwei orthogonale Variablen und drei von diesen jeweils etwa gleichstark abhängige Variablen. Wie das Beispiel zeigt, ist das vollständig gelungen. Vor der Rotation kann man den Abszissenfaktor als *Größe*, den Ordinatenfaktor als *Form* (lang, breit) des Rechtecks bezeichnen. Nach der Rotation werden die "hineingesteckten" Variablen *x* (Länge) und *y* (Breite) wieder als Faktoren reproduziert. Folglich sind die Faktoren jetzt auch als *Länge* und *Breite* zu bezeichnen. Die geometrisch abgeleiteten Eigenschaften der Rechtecke, Diagonale, halber Umfang und Fläche, korrelieren jetzt auch wieder mit beiden Faktoren jeweils etwa gleich hoch.

Die Graphiken, die SPSS für die Faktorenanalyse liefert, sind, auch wenn man sie im *Chart Editor* nachbearbeitet, nicht sehr gut. Dazu gehört auch die Kleinschreibung der *Variable labels* und deren Übereinanderdruck bei eng benachbarten Variablenpunkten. Zur Abhilfe kann man die Faktorenmatrix, rotiert oder unrotiert, über die Zwischenablage in ein Tabellenblatt von EXCEL kopieren. Fügt man danach noch zwischen die Spalten, die die Faktoren wiedergeben, mit Nullen gefüllte Spalten ein, erhält man:

Microsoft Excel - Mappe1

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster ?

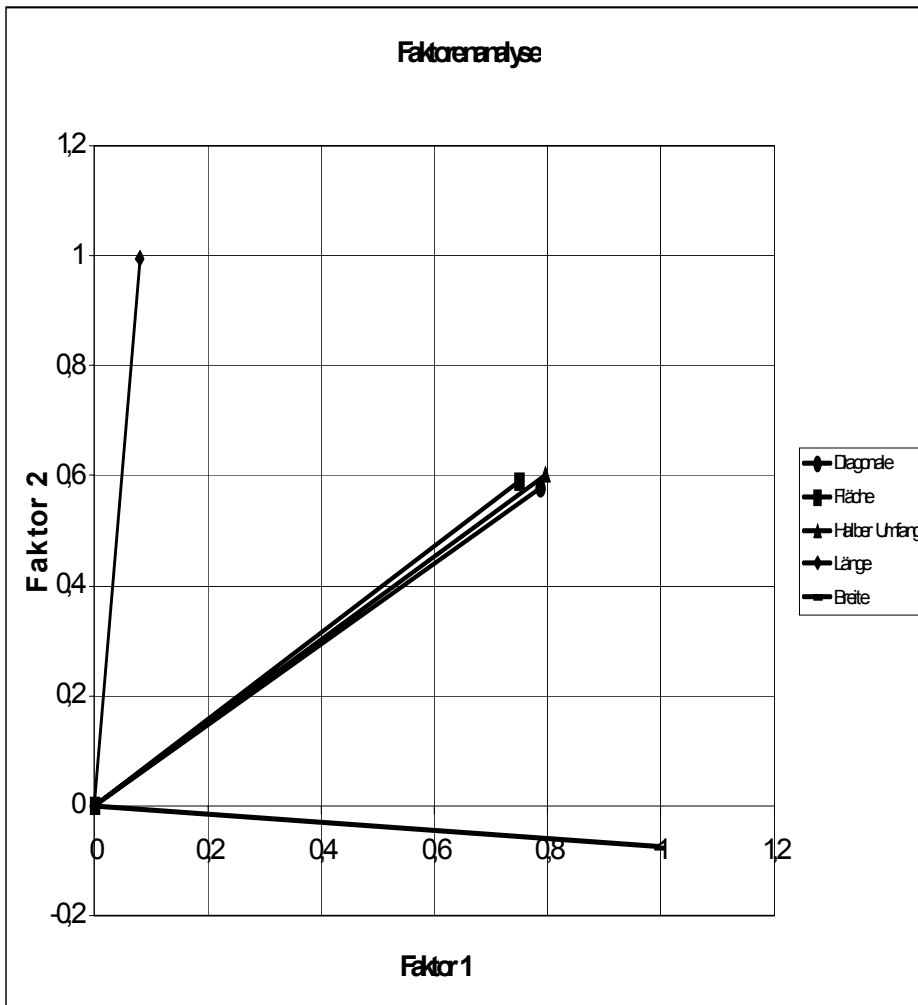
E9

|    | A   | B          | C | D           | E | F | G | H | I |
|----|---|------------|---|-------------|---|---|---|---|---|
| 1  | Rotated Component Matrix  |            |   |             |   |   |   |   |   |
| 2  |   | Component  |   |             |   |   |   |   |   |
| 3  |   | 1          |   | 2           |   |   |   |   |   |
| 4  | Diagonale   | 0,7867215  | 0 | 0,57769232  | 0 |   |   |   |   |
| 5  | Fläche  | 0,74939192 | 0 | 0,58942562  | 0 |   |   |   |   |
| 6  | Halber Umfan  | 0,79637811 | 0 | 0,60250074  | 0 |   |   |   |   |
| 7  | Länge   | 0,08040146 | 0 | 0,99596391  | 0 |   |   |   |   |
| 8  | Breite  | 0,99648453 | 0 | -0,07581935 | 0 |   |   |   |   |
| 9  | Extraction Method: Principal Component Analysis. <input type="checkbox"/> Rotation Method: Varimax with Kaiser Normalization. |            |   |             |   |   |   |   |   |
| 10 | a Rotation converged in 3 iterations.   |            |   |             |   |   |   |   |   |
| 11 |   |            |   |             |   |   |   |   |   |
| 12 |   |            |   |             |   |   |   |   |   |
| 13 |   |            |   |             |   |   |   |   |   |
| 14 |   |            |   |             |   |   |   |   |   |
| 15 |   |            |   |             |   |   |   |   |   |
| 16 |   |            |   |             |   |   |   |   |   |
| 17 |   |            |   |             |   |   |   |   |   |
| 18 |   |            |   |             |   |   |   |   |   |
| 19 |   |            |   |             |   |   |   |   |   |
| 20 |   |            |   |             |   |   |   |   |   |
| 21 |   |            |   |             |   |   |   |   |   |
| 22 |   |            |   |             |   |   |   |   |   |
| 23 |   |            |   |             |   |   |   |   |   |
| 24 |   |            |   |             |   |   |   |   |   |
| 25 |   |            |   |             |   |   |   |   |   |

Tabelle1 / Tabelle2 / Tabelle3 /

Bereit

Durch Anklicken des bunten Histogramms in der oberen Symbolleiste kann man den *Diagramm-Assistenten* von EXCEL starten. Hier ist dann *Punkt (X,Y) Punkte mit Linien* auszuwählen. Bei entsprechender Definition der *Reihen* und mit etwas graphischer Nacharbeit läßt sich die folgende Graphik erzeugen:



Hier sind die Variablenvektoren zwar nicht als Pfeile, aber doch immerhin als Linien, die im Koordinatenanfang beginnen, zu sehen. Wenn man die Graphik quadratisch skaliert, kann man auch erreichen, daß die Winkel zwischen den Variablenvektoren dem Arcuscosinus der Korrelationen entsprechen. Im vorliegenden Fall ging das aber auf Kosten des Schriftbildes. - Diese EXCEL-Mappe ist in der Datei *Aufgabe20.xls* enthalten.

## Aufgabe 21

Das semantische Differential ist eine klassische Methode der Sozialpsychologie, die auf Osgood, Suci & Tannenbaum (1957) zurückgeht. Ertel hat es unter der Bezeichnung *Eindrucksdifferential* im deutschen Sprachraum untersucht. Eine von ihm (Ertel, 1964; 1965a; 1965b) entwickelte und standardisierte Variante läßt sich leicht im Hörsaalversuch replizieren. Das zugehörige Material sieht wie folgt aus (*Beurteilungsbogen zum Eindrucksdifferential*, Datei *sdbogen.doc*). Den befragten Personen wird eine Liste von "Personentypen" vorgegeben. Sie werden instruiert, sich jeweils eine entsprechende Person möglichst anschaulich und genau vorzustellen, sie auf dem *Beurteilungsbogen* (nächste Seite) in die Lücke unter *Ein/eine ..... macht auf mich den Eindruck* einzutragen und dann für jedes der folgenden 30 Adjektive die Stärke des Eindruckes anzukreuzen. Fordert man von jeder befragten Person die Beurteilung aller 20 "Personentypen", wird der Versuch recht aufwendig, weil jede Person 20 Bögen auszufüllen, also  $20 \cdot 30 = 600$  Urteile abzugeben hat. Fordert man andererseits nur die Beurteilung eines einzigen "Personentypen", ist zwar der Versuch nur kurz, die erhaltene Datenmenge aber auch nur klein. Ein Kompromiß zwischen diesen beiden Extremen entsteht, wenn jede befragte Person eine Teilmenge von beispielsweise 5 "Personentypen" beurteilt. Die Datei *Aufgabe21.sav* enthält das Ergebnis eines entsprechenden Hörsaalversuches in der Statistikübung in Tübingen. Es basiert auf 142 ausgefüllten Beurteilungsbögen.

### Zu beurteilende "Personentypen"

1. Gentleman
2. Kleines Kind
3. Gemütsmensch
4. Held
5. Verbrecher
6. Göttliche Autorität
7. Sanguiniker
8. Ideale Mutter
9. Clown
10. Verführerischer Vamp
11. Professor
12. Fanatiker
13. Melancholiker
14. Gebrechliche alte Frau
15. Vorbildliche Führerpersönlichkeit
16. Nettes junges Mädchen
17. Feigling
18. Phlegmatiker
19. Tyrann
20. Idealist

## Beurteilungsbogen zum Eindrucksdifferential

Ein/eine ..... macht auf mich den Eindruck

|                           | 1 niemals | 2 sehr selten | 3 hin und wieder | 4 häufig | 5 sehr häufig | 6 immer |
|---------------------------|-----------|---------------|------------------|----------|---------------|---------|
| 1 unterlegen wirkend      |           |               |                  |          |               |         |
| 2 einengend               |           |               |                  |          |               |         |
| 3 Spannung verursachend   |           |               |                  |          |               |         |
| 4 Entspannung vermittelnd |           |               |                  |          |               |         |
| 5 vorsichtig stimmend     |           |               |                  |          |               |         |
| 6 beeinflussbar           |           |               |                  |          |               |         |
| 7 Ruhe verbreitend        |           |               |                  |          |               |         |
| 8 bedrohlich              |           |               |                  |          |               |         |
| 9 anregend                |           |               |                  |          |               |         |
| 10 nachgiebig             |           |               |                  |          |               |         |
| 11 unangenehm             |           |               |                  |          |               |         |
| 12 angenehm               |           |               |                  |          |               |         |
| 13 Wärme verbreitend      |           |               |                  |          |               |         |
| 14 kraftlos wirkend       |           |               |                  |          |               |         |
| 15 Bewunderung erregend   |           |               |                  |          |               |         |
| 16 Unruhe verbreitend     |           |               |                  |          |               |         |
| 17 langweilig             |           |               |                  |          |               |         |
| 18 Verachtung erregend    |           |               |                  |          |               |         |
| 19 friedlich              |           |               |                  |          |               |         |
| 20 Erregung verursachend  |           |               |                  |          |               |         |
| 21 Kälte verbreitend      |           |               |                  |          |               |         |
| 22 anziehend              |           |               |                  |          |               |         |
| 23 unnachgiebig           |           |               |                  |          |               |         |
| 24 dominierend            |           |               |                  |          |               |         |
| 25 glücklich stimmend     |           |               |                  |          |               |         |
| 26 kraftvoll wirkend      |           |               |                  |          |               |         |
| 27 ungefährlich           |           |               |                  |          |               |         |
| 28 belastend              |           |               |                  |          |               |         |
| 29 Erregung vermindern    |           |               |                  |          |               |         |
| 30 abstoßend              |           |               |                  |          |               |         |



Berechnen Sie die Faktorenanalyse für die 30 Adjektive dieser Form des Eindrucksdifferentials. Wählen Sie die Extraktion von 3 Faktoren und die Varimax-Rotation. Interpretieren Sie das Ergebnis.

#### Literatur

- Ertel, S. (1964). Die emotionale Natur des „semantischen“ Raumes. *Psychologische Forschung*, 28, 1-32.
- Ertel, S. (1965a). Standardisierung eines Eindrucksdifferentials. *Zeitschrift für experimentelle und angewandte Psychologie*, 12, 22-58.
- Ertel, S. (1965b). Weitere Untersuchungen zur Standardisierung eines Eindrucksdifferentials. *Zeitschrift für experimentelle und angewandte Psychologie*, 12, 177-208.
- Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL, U.S.A.: University of Illinois Press.

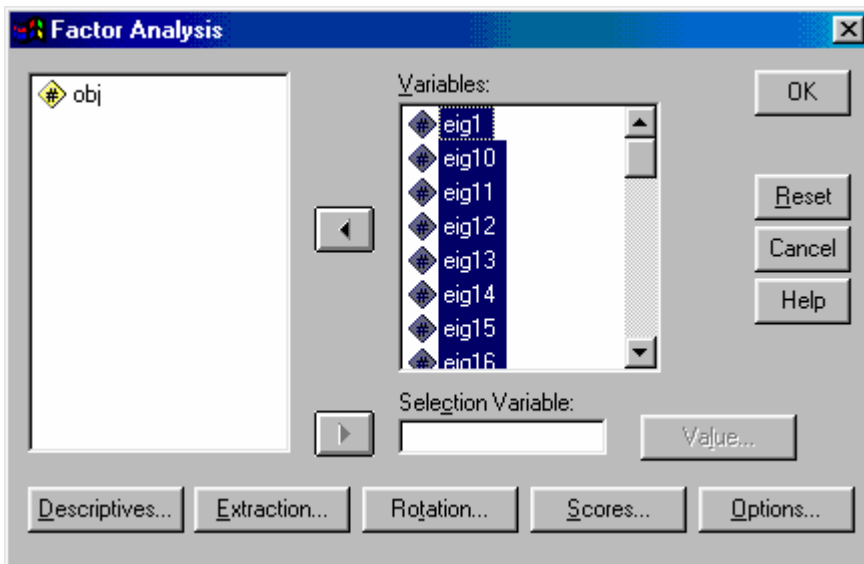
#### Lösung

Laden Sie *Aufgabe21.sav* in das Dateneingabefenster von SPSS. Sie erhalten folgenden Bildschirm:

| obj | eig1 | eig2               | eig3 | eig4 | eig5 | eig6 | eig7 | eig8 |
|-----|------|--------------------|------|------|------|------|------|------|
| 1   | 1,00 | unterlegen wirkend | 3,00 | 1,00 | 3,00 | 3,00 | 2,00 | 4,00 |
| 2   | 1,00 | 2,00               | 1,00 | 1,00 | 3,00 | 1,00 | 5,00 | 3,00 |
| 3   | 1,00 | 3,00               | 4,00 | 4,00 | 3,00 | 4,00 | 2,00 | 3,00 |
| 4   | 1,00 | 1,00               | 1,00 | 2,00 | 4,00 | 3,00 | 2,00 | 5,00 |
| 5   | 1,00 | 2,00               | 2,00 | 2,00 | 4,00 | 3,00 | 3,00 | 4,00 |
| 6   | 1,00 | 2,00               | 4,00 | 2,00 | 5,00 | 3,00 | 2,00 | 5,00 |
| 7   | 1,00 | 1,00               | 5,00 | 5,00 | 5,00 | 3,00 | 3,00 | 6,00 |
| 8   | 2,00 | 3,00               | 3,00 | 3,00 | 2,00 | 4,00 | 5,00 | 2,00 |
| 9   | 2,00 | 3,00               | 1,00 | 3,00 | 3,00 | 3,00 | 5,00 | 2,00 |
| 10  | 2,00 | 4,00               | 3,00 | 3,00 | 2,00 | 2,00 | 5,00 | 2,00 |
| 11  | 2,00 | 3,00               | 1,00 | 4,00 | 3,00 | 4,00 | 5,00 | 2,00 |
| 12  | 2,00 | 3,00               | 2,00 | 3,00 | 3,00 | 3,00 | 4,00 | 4,00 |
| 13  | 2,00 | 6,00               | 4,00 | 3,00 | 2,00 | 4,00 | 6,00 | 2,00 |
| 14  | 2,00 | 3,00               | 1,00 | 4,00 | 4,00 | 3,00 | 5,00 | 4,00 |
| 15  | 2,00 | 4,00               | 5,00 | 5,00 | 3,00 | 5,00 | 5,00 | 2,00 |

Die Variable *obj* enthält die Nummer des beurteilten "Personentyps" gemäß der oben wiedergegebenen Liste. "Personentyp" 1, *Gentleman*, wurde also von 7 Personen beurteilt. Die Nummern der Adjektive auf dem Beurteilungsbogen wurden für die Bildung der Variablennamen *eig1*, *eig2* ... *eig30* verwendet, die Adjektive selbst wurden als Variablenbezeichner eingegeben.

Holen Sie sich mit *Analyze* → *Data Reduction* → *Factor* das Definitionsfenster für die Faktorenanalyse. Bringen Sie die Variablen *eig1* ... *eig30* in das *Variables*-Fenster. Das sieht dann so aus:



Arbeiten Sie jetzt die Schaltfelder von *Descriptives* bis *Options* genau so durch, wie es bei der Lösung zur Aufgabe 20 gezeigt wurde. Nur die folgenden Eingaben sind jetzt etwas anders zu machen: Im Fenster *Extraction* setzen Sie *Number of factors* = 3 und im Fenster *Options* setzen Sie *Missing Values* auf *Exclude cases pairwise*. Mit *OK* starten Sie die Faktorenanalyse.

Das wichtigste Ergebnis ist erst einmal die Tabelle *Total Variance Explained*. Da 30 Variablen einbezogen sind, hat sie auch 30 Zeilen. Aus Platzgründen werden hier nur die ersten 5 wiedergegeben.

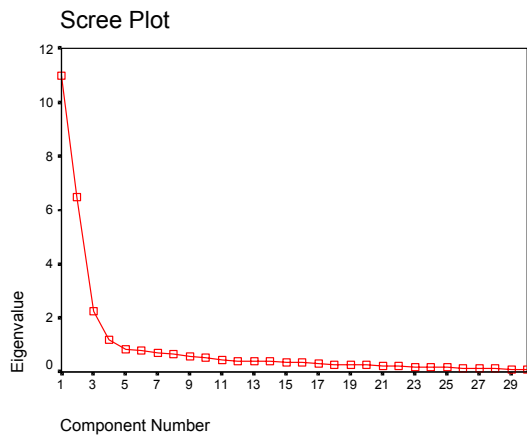
Total Variance Explained

| Component | Initial Eigenvalues |               |              | Extraction Sums of Squared Loadings |               |              | Rotation Sums of Squared Loadings |               |              |
|-----------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|-----------------------------------|---------------|--------------|
|           | Total               | % of Variance | Cumulative % | Total                               | % of Variance | Cumulative % | Total                             | % of Variance | Cumulative % |
| 1         | 10,990              | 36,633        | 36,633       | 10,990                              | 36,633        | 36,633       | 8,947                             | 29,823        | 29,823       |
| 2         | 6,477               | 21,588        | 58,222       | 6,477                               | 21,588        | 58,222       | 5,544                             | 18,479        | 48,302       |
| 3         | 2,237               | 7,457         | 65,679       | 2,237                               | 7,457         | 65,679       | 5,213                             | 17,377        | 65,679       |
| 4         | 1,193               | 3,976         | 69,655       |                                     |               |              |                                   |               |              |
| 5         | ,826                | 2,753         | 72,408       |                                     |               |              |                                   |               |              |

Extraction Method: Principal Component Analysis.

Die Eigenwerte betragen 10,990, 6,477 und 2,237 für die ersten drei Faktoren. Die zugehörigen Varianzanteile gehen von 36,633 % über 21,588 % auf 7,457 %. Die gesamte erklärte Varianz beträgt 65,679 %, ein nicht überwältigendes, aber praktisch brauchbares Ergebnis. Die Rotation gleicht die Verteilung der Varianzanteile auf die drei Faktoren etwas aus: Es resultieren 29,823 %, 18,479 % und 17,377 %.

Das *Scree-Plot* sieht wie folgt aus:



Die Entscheidung, drei Faktoren zu extrahieren, ist also gerechtfertigt. Das Kriterium, alle Faktoren mit Eigenwerten  $\geq 1$  zu berücksichtigen, hätte hier zu einem vierten Faktor geführt, der aber nur noch 3,976 % der Varianz erklärt hätte.

Das entscheidende Ergebnis ist die rotierte Faktorenmatrix (Leider mußte ich wieder die Erfahrung machen, daß WORD 97 diese Tabelle nicht auf allen Druckern korrekt ausdrückt. Zum Teil werden Zeilen und Trennungslinien zwischen den Zeilen verdoppelt. Mit WORD 2000 hat es bei mir stets funktioniert, WG.):

Rotated Component Matrix

|                         | Component    |              |              |
|-------------------------|--------------|--------------|--------------|
|                         | 1            | 2            | 3            |
| unterlegen wirkend      | ,197         | <b>-,805</b> | ,132         |
| nachgiebig              | -,239        | <b>-,625</b> | -,295        |
| unangenehm              | <b>,850</b>  | ,058         | ,301         |
| angenehm                | <b>-,851</b> | ,096         | -,272        |
| Wärme verbreitend       | <b>-,759</b> | -,063        | -,301        |
| kraftlos wirkend        | ,328         | <b>-,785</b> | -,168        |
| Bewunderung erregend    | -,462        | <b>,649</b>  | -,060        |
| Unruhe verbreitend      | ,128         | ,045         | <b>,869</b>  |
| langweilig              | <b>,581</b>  | -,440        | -,443        |
| Verachtung erregend     | <b>,754</b>  | ,017         | ,219         |
| friedlich               | -,418        | -,301        | <b>-,642</b> |
| einengend               | <b>,602</b>  | ,250         | ,245         |
| Erregung verursachend   | ,091         | ,414         | <b>,739</b>  |
| Kälte verbreitend       | <b>,738</b>  | ,253         | ,247         |
| anziehend               | <b>-,705</b> | ,265         | ,042         |
| unnachgiebig            | ,326         | <b>,650</b>  | ,290         |
| dominierend             | ,157         | <b>,773</b>  | ,337         |
| glücklich stimmend      | <b>-,869</b> | ,152         | -,122        |
| kraftvoll wirkend       | -,254        | <b>,772</b>  | ,159         |
| ungefährlich            | <b>-,538</b> | -,483        | -,374        |
| belastend               | <b>,764</b>  | -,164        | ,293         |
| Erregung vermindern     | -,229        | -,056        | <b>-,740</b> |
| Spannung verursachend   | ,381         | ,338         | <b>,614</b>  |
| abstoßend               | <b>,813</b>  | -,002        | ,197         |
| Entspannung vermittelnd | -,515        | -,010        | <b>-,546</b> |
| vorsichtig stimmend     | <b>,325</b>  | -,080        | ,208         |
| beeinflußbar            | -,059        | <b>-,690</b> | ,155         |
| Ruhe verbreitend        | -,266        | ,084         | <b>-,821</b> |
| bedrohlich              | <b>,659</b>  | ,416         | ,407         |
| anregend                | <b>-,660</b> | ,343         | ,337         |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 7 iterations.

Zur Verbesserung der Übersicht wurden die Exponentendarstellungen der kleinen Faktorenladungen, die SPSS ausgibt, ohne Rundung beseitigt. Für die Interpretation wurde von

jeder Variablen die höchste Ladung, also die höchste Ladung jeder **Zeile**, fettgedruckt. Um sich die semantische Bedeutung eines Faktors klarzumachen, sucht man die fettgedruckten Ladungen der einzelnen **Spalten** auf und formuliert einen Begriff, der die Gemeinsamkeiten der zugehörigen Variablen möglichst gut trifft. Man beginnt dabei mit den höchsten Ladungen. Die Variablen mit den höchsten Ladungen auf einem Faktor und möglichst niedrigen Ladungen auf allen anderen Faktoren werden auch *Markiervariablen* genannt. Für Spalte 1 lauten diese *glücklich stimmend, angenehm, unangenehm, abstoßend* und *anziehend*. Bemerkenswert sind auch *Wärme verbreitend* und *Kälte verbreitend*. Schon Osgood hat einen ersten Faktor mit dieser Bedeutung gefunden. Er nannte ihn *Evaluation, Bewertung*. Dieser Faktor gibt also die wertende Stellungnahme zu einem Objekt an. Sie hängt stark mit dem Gefühl von Lust und Unlust zusammen, wie die Adjektive zeigen. Bemerkenswert ist, daß die Adjektive *warm* und *kalt* als Evaluationsvariablen erscheinen. In unserem Beispiel ist die Polung des Faktors negativ, der semantisch positive Pol der zusammengehörigen Adjektivpaare drückt sich in einer negativen Ladung aus.

Für den zweiten Faktor ergeben sich die Markiervariablen *unterlegen wirkend, kraftlos wirkend* (mit negativem Vorzeichen) und *dominierend, kraftvoll wirkend* (mit positivem Vorzeichen). Auch diesen Faktor fand schon Osgood; er nannte ihn *Potency, Macht*. Diese Bezeichnung gibt das Gemeinsame der Markiervariablen sicherlich treffend wieder.

Die Markiervariablen des dritten Faktors lauten *Ruhe verbreitend, Unruhe verbreitend, Erregung vermindern* und *Erregung verursachend*. Nach Osgood heißt dieser Faktor *Activity, Aktivität*. Er drückt das Gegensatzpaar von *Ruhe* und *Bewegung, Passivität* und *Aktivität* aus. In unserem Beispiel entspricht der aktive, unruhige Bedeutungspol dem positiven Vorzeichen der Faktorenladung.

Die graphische Darstellung, die SPSS hier liefert, ist so unbefriedigend, daß sie nicht wiedergegeben werden soll. Die beste Abhilfe besteht darin, die rotierte Faktorenmatrix in ein EXCEL-Datenblatt zu kopieren, dann Spalten mit Nullen einzufügen und eine Graphik zu erzeugen, wie sie in Aufgabe 20 gezeigt wurde. Auch Handzeichnungen mit Vektorpfeilen in zweidimensionalen Koordinatensystemen aus Faktor 1 und Faktor 2 bzw. Faktor 1 und Faktor 3, wie sie in der Vorlesung als Folien gezeigt wurden, sind nicht die schlechteste Lösung.

## Aufgabe 22

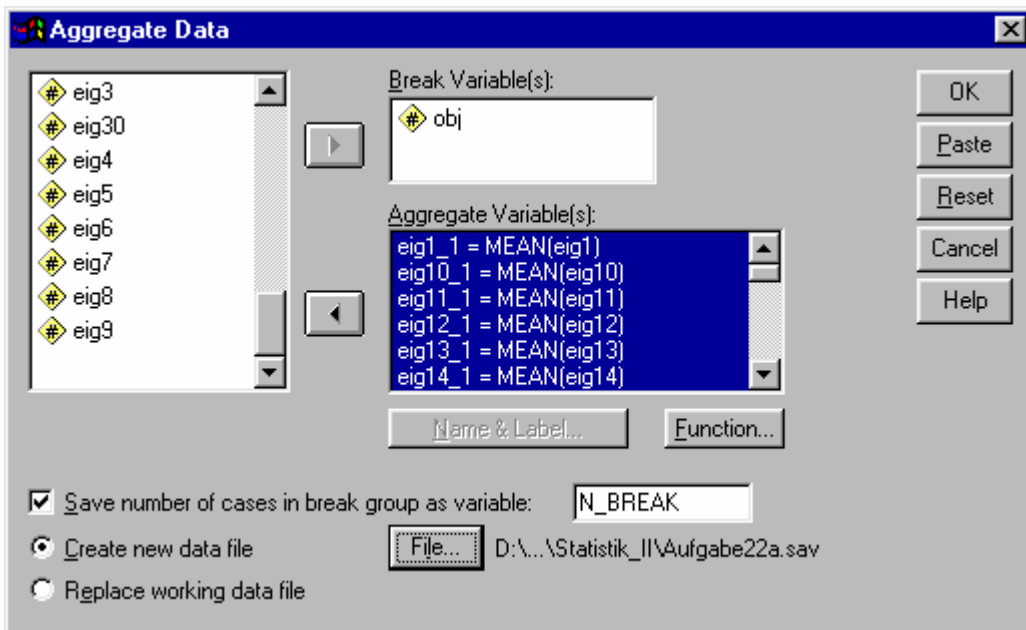
An der Auswertung, die in Aufgabe 21 gezeigt wurde, kann man kritisieren, daß jede Zeile die Beurteilung eines Objektes durch eine Person darstellte, wobei immer mehrere Personen ein Objekt beurteilt haben. Während also die Varianz der Spalten der Datenmatrix wie gewünscht nur auf den Adjektiven als Variablen beruht, beruht die Varianz zwischen den Zeilen sowohl auf Unterschieden zwischen den beurteilten Objekten als auch zwischen den urteilenden Personen. Durch die Bildung arithmetischer Mittelwerte für die Personen innerhalb jedes beurteilten Objektes läßt sich eine reine Objekte x Eigenschaften-Matrix erzeugen.

Entsprechend lautet diese Aufgabe: erzeugen Sie eine Objekte x Eigenschaften-Matrix, die in den Zellen die Mittelwerte über die Personen enthält, und faktorenanalisieren Sie sie nach dem Beispiel von Aufgabe 21.

## Lösung

Die gesuchten Mittelwerte finden wir in SPSS mit der Funktion *Aggregate*. Wie viele besonders leistungsfähige und nützliche Funktionen ist sie in der Studentenversion von SPSS nicht enthalten. Sie wird deshalb hier in der Vollversion 10.0.5 gezeigt.

Öffnen Sie also die Vollversion von SPSS auf einem der Rechner, auf denen sie vorhanden ist. Laden Sie die Datei *Aufgabe21.sav*. Mit *Data* → *Aggregate* erhalten das Fenster:



Füllen Sie es aus, wie es die Abbildung zeigt. Machen Sie *obj* zur *Break Variable(s)* und bringen Sie die Variablen *eig1 ... eig30* in das Fenster *Aggregate Variable(s)*. In diesem Fenster wird angezeigt, was SPSS tun wird: Zu jeder Variablen *eigx* berechnet es eine neue Variable *eigx\_1*, die den Mittelwert *MEAN(eigx)* aller Zeilen enthält, für die *obj* den gleichen Wert hat. Das ist genau die Bedeutung von *Break Variable*. Das Ergebnis wird aufgrund der Voreinstellung *Create new data file* in eine neue Datei geschrieben, der wir den Namen *Aufgabe22a.sav* geben. Mit *Save number of cases in break group as variable N\_BREAK* fordern wir die Information an, wie viele Personen jeden "Personentyp" beurteilt haben.

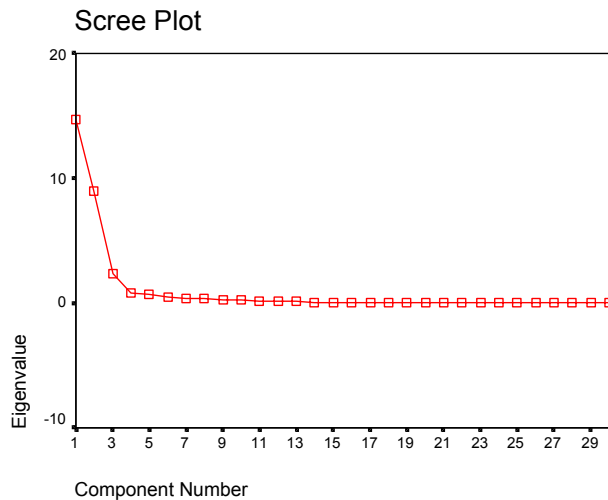
Nach *OK* erhalten wir die gewünschte "Personentypen" x Adjektive-Matrix in der Datei *Aufgabe22a.sav*. Dieses Zwischenresultat ist auch in der Datei *Aufgabe22.sav* enthalten, die Sie in meinem Verzeichnis ... \user0147 \Statistik\_II finden. Damit können Sie jetzt die Faktorenanalyse auch in der Studentenversion von SPSS rechnen. Bitte tun Sie es mit den gleichen Einstellungen wie in Aufgabe 21.

Die Tabelle *Total Variance Explained* lautet, wieder auf die ersten fünf Zeilen gekürzt:

| Component | Initial Eigenvalues |               |              | Extraction Sums of Squared Loadings |               |              | Rotation Sums of Squared Loadings |               |              |
|-----------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|-----------------------------------|---------------|--------------|
|           | Total               | % of Variance | Cumulative % | Total                               | % of Variance | Cumulative % | Total                             | % of Variance | Cumulative % |
| 1         | 14,735              | 49,117        | 49,117       | 14,735                              | 49,117        | 49,117       | 11,569                            | 38,563        | 38,563       |
| 2         | 8,926               | 29,754        | 78,871       | 8,926                               | 29,754        | 78,871       | 7,346                             | 24,487        | 63,050       |
| 3         | 2,376               | 7,922         | 86,793       | 2,376                               | 7,922         | 86,793       | 7,123                             | 23,743        | 86,793       |
| 4         | ,833                | 2,778         | 89,570       |                                     |               |              |                                   |               |              |
| 5         | ,725                | 2,416         | 91,987       |                                     |               |              |                                   |               |              |

Die Faktorenanalyse erklärt jetzt insgesamt 86,793 % der Varianz, ein nach den Maßstäben der Praxis sehr hoher Wert. Nach der Rotation wird die erklärte Varianz auf die drei Faktoren im Verhältnis 38,563 %, 24,487 % und 23,743 % verteilt. Ein möglicher vierter Faktor hätte nur noch einen Eigenwert von 0,833, was 2,778 % der Varianz entsprechen würde.

Das *Scree-Plot* ergibt sich wie folgt:



Die Entscheidung zur Extraktion von 3 Faktoren ist also gerechtfertigt.

Die rotierte Faktorenmatrix schließlich lautet:

Rotated Component Matrix

|                                | Component    |              |              |
|--------------------------------|--------------|--------------|--------------|
|                                | 1            | 2            | 3            |
| EIG1_1 unterlegen wirkend      | ,213         | <b>-,897</b> | ,068         |
| EIG10_1 nachgiebig             | -,249        | <b>-,719</b> | -,488        |
| EIG11_1 unangenehm             | <b>,913</b>  | ,050         | ,318         |
| EIG12_1 angenehm               | <b>-,923</b> | ,143         | -,273        |
| EIG13_1 Wärme verbreitend      | <b>-,871</b> | -,086        | -,280        |
| EIG14_1 kraftlos wirkend       | ,325         | <b>-,861</b> | -,253        |
| EIG15_1 Bewunderung erregend   | -,554        | <b>,767</b>  | ,045         |
| EIG16_1 Unruhe verbreitend     | ,141         | ,111         | <b>,957</b>  |
| EIG17_1 langweilig             | ,512         | -,539        | <b>-,564</b> |
| EIG18_1 Verachtung erregend    | <b>,870</b>  | -,006        | ,252         |
| EIG19_1 friedlich              | -,425        | -,343        | <b>-,796</b> |
| EIG2_1 einengend               | <b>,800</b>  | ,353         | ,309         |
| EIG20_1 Erregung verursachend  | ,166         | ,509         | <b>,788</b>  |
| EIG21_1 Kälte verbreitend      | <b>,843</b>  | ,289         | ,271         |
| EIG22_1 anziehend              | <b>-,808</b> | ,321         | ,049         |
| EIG23_1 unnachgiebig           | ,381         | <b>,673</b>  | ,504         |
| EIG24_1 dominierend            | ,222         | <b>,811</b>  | ,455         |
| EIG25_1 glücklich stimmend     | <b>-,935</b> | ,211         | -,097        |
| EIG26_1 kraftvoll wirkend      | -,251        | <b>,920</b>  | ,209         |
| EIG27_1 ungefährlich           | <b>-,618</b> | -,542        | -,478        |
| EIG28_1 belastend              | <b>,848</b>  | -,236        | ,230         |
| EIG29_1 Erregung vermindern    | -,358        | -,130        | <b>-,873</b> |
| EIG3_1 Spannung verursachend   | ,447         | ,371         | <b>,756</b>  |
| EIG30_1 abstoßend              | <b>,914</b>  | -,059        | ,320         |
| EIG4_1 Entspannung vermittelnd | <b>-,682</b> | -,052        | -,599        |
| EIG5_1 vorsichtig stimmend     | <b>,564</b>  | -,122        | ,291         |
| EIG6_1 beeinflussbar           | -,067        | <b>-,844</b> | ,108         |
| EIG7_1 Ruhe verbreitend        | -,365        | ,140         | <b>-,870</b> |
| EIG8_1 bedrohlich              | <b>,694</b>  | ,469         | ,472         |
| EIG9_1 anregend                | <b>-,746</b> | ,441         | ,414         |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.  
a. Rotation converged in 6 iterations.

Wir haben im wesentlichen das gleiche Muster von Variablen und Faktoren erhalten wie in Aufgabe 21. Wegen des insgesamt höheren erklärten Varianzanteils sind die Faktorenladungen nahezu alle deutlich höher als vorher. Die Faktorenladungen der Markiertvariablen überschreiten großenteils den Zahlenwert von 0,9. Faktor 1 ist wieder *Evaluation* mit den Markiertvariablen *angenehm*, *unangenehm*, *anziehend*, *abstoßend*, *glücklich stimmend*, *Wärme verbreitend* und *Kälte verbreitend*.

Faktor 2 ist *Potency* mit *kraftvoll wirkend, kraftlos wirkend, dominierend* und *unterlegen wirkend*. Faktor 3 schließlich ist wieder *Activity* mit *Unruhe verbreitend, Ruhe verbreitend, Erregung verursachend, Erregung vermindern*.

Auch hier ist die dreidimensionale Abbildung in SPSS nicht viel wert, so daß sie nicht wiedergegeben werden soll. Versuchen Sie, mit EXCEL eine brauchbare Graphik zustandezubringen!

### Aufgabe 23

In der Bevölkerungsbefragung Sternenfels (Glaser, Glaser & Kuder, 2000) waren auch die folgenden 10 Items zu Computern und Computernetzen enthalten:

1. Computer erhöhen den Streß bei der Arbeit
2. Vor Computern habe ich Angst
3. Computer machen die Arbeit anregender und interessanter
4. Computer machen das Arbeitsleben unpersönlicher
5. Durch vermehrten Einsatz von Computern gehen Arbeitsplätze verloren
6. Computer stellen auch für das Privatleben und die Freizeit eine Bereicherung dar
7. Durch die Arbeit am Computer nehmen die gesundheitlichen Probleme zu
8. Nur durch massiven Einsatz von Computern bleibt die Wirtschaft konkurrenzfähig
9. Weltweit vernetzte Computer führen zur Verlagerung von immer mehr Arbeitsplätzen in Billiglohnländer
10. Vernetzte Computer können die Standortnachteile des ländlichen Raums wesentlich verringern.

Für die Antworten wurden die Kategorien *Stimmt überhaupt nicht (1), Stimmt eher nicht (2), Teils-teils bzw. keine Meinung (3), Stimmt eher (4)* und *Stimmt voll und ganz (5)* vorgegeben. Die Daten von 937 Befragten sind in den Variablen *comp1 ... comp10* der Datei *Aufgabe23.sav* enthalten.

Berechnen Sie die Faktorenanalyse über diese Daten mit 3 extrahierten Faktoren und Varimax-Rotation. Behandeln Sie dabei die (hier nicht seltenen) *Missing data* mit *Exclude cases listwise*.

Literatur

Glaser, W. R., Glaser, M. O. & Kuder, T. (2000). Bevölkerungsbefragung zum Innovationszentrum Fabrik Schweitzer und zu Telearbeit in Sternenfels. In W. R. Glaser (Hrsg.). *Telezentren - Zukunft oder schon Vergangenheit?* (S. 7-98). Stuttgart: Kohlhammer.

### Lösung

Öffnen Sie SPSS und laden Sie die Datei *Aufgabe23.sav* in den Eingabebildschirm. Aktivieren Sie mit *Analyse → Data Reduction → Factor* das Fenster der Faktorenanalyse. Bringen Sie die Variablen *comp1 ... comp10* in das *Variables*-Fenster. Bearbeiten Sie die Schaltflächen von *Descriptives* bis *Options* wie in den vorausgegangenen Beispielen, setzen Sie dabei die Angaben zu dieser Aufgabe ein. Fordern Sie mit *OK* die Faktorenanalyse an.

Die Angaben über die Aufteilung der Gesamtvarianz lauten (ungekürzt):

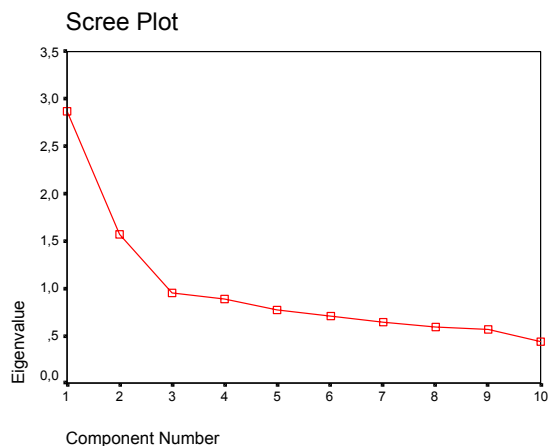
Total Variance Explained

| Component | Initial Eigenvalues |               |              | Extraction Sums of Squared Loadings |               |              | Rotation Sums of Squared Loadings |               |              |
|-----------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|-----------------------------------|---------------|--------------|
|           | Total               | % of Variance | Cumulative % | Total                               | % of Variance | Cumulative % | Total                             | % of Variance | Cumulative % |
| 1         | 2,873               | 28,733        | 28,733       | 2,873                               | 28,733        | 28,733       | 2,564                             | 25,639        | 25,639       |
| 2         | 1,567               | 15,668        | 44,401       | 1,567                               | 15,668        | 44,401       | 1,430                             | 14,305        | 39,944       |
| 3         | ,950                | 9,497         | 53,897       | ,950                                | 9,497         | 53,897       | 1,395                             | 13,954        | 53,897       |
| 4         | ,888                | 8,881         | 62,778       |                                     |               |              |                                   |               |              |
| 5         | ,774                | 7,736         | 70,514       |                                     |               |              |                                   |               |              |
| 6         | ,706                | 7,059         | 77,573       |                                     |               |              |                                   |               |              |
| 7         | ,642                | 6,425         | 83,997       |                                     |               |              |                                   |               |              |
| 8         | ,588                | 5,881         | 89,878       |                                     |               |              |                                   |               |              |
| 9         | ,570                | 5,704         | 95,582       |                                     |               |              |                                   |               |              |
| 10        | ,442                | 4,418         | 100,000      |                                     |               |              |                                   |               |              |

Extraction Method: Principal Component Analysis.

Die Faktorenanalyse erklärt 53,897 % der Gesamtvarianz, ein nicht sehr hoher, in der Praxis aber durchaus noch akzeptierter Wert. Nach der Rotation wird dieser Teil im Verhältnis 25,639 %, 14,305 % und 13,954 % auf die drei Faktoren aufgeteilt.

Wir erhalten folgendes *Scree-Plot*:



Die Entscheidung, drei Faktoren zu extrahieren, erscheint gerechtfertigt, wenn auch möglicherweise nur zwei Faktoren sinnvoll wären.

Die rotierte Faktorenmatrix lautet wie folgt:

Rotated Component Matrix

|   | Component   |             |             |
|---|-------------|-------------|-------------|
|   | 1           | 2           | 3           |
| Computer erhöhen Arbeitsstreß                   | <b>,575</b> | -,372       | -,152       |
| Computer = Vorteil für ländlichen Raum          | -,022       | ,079        | <b>,826</b> |
| Vor Computern habe ich Angst                    | <b>,538</b> | -,186       | -,185       |
| Computer machen Arbeit interessanter            | -,051       | <b>,780</b> | ,088        |
| Computer machen Arbeitsleben unpersönlicher     | <b>,721</b> | -,260       | ,151        |
| Durch Computer Verlust von Arbeitsplätzen       | <b>,735</b> | ,030        | -,036       |
| Computer sind Bereicherung für Privatleben      | -,144       | <b>,663</b> | ,214        |
| Computer bedeuten mehr Gesundheitsprobleme      | <b>,600</b> | -,164       | ,248        |
| Mit Computern bleibt Wirtschaft konkurrenzfähig | -,042       | ,241        | <b>,691</b> |
| Computer = Verlagerung in Billiglohnländer      | <b>,704</b> | ,223        | -,200       |

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.



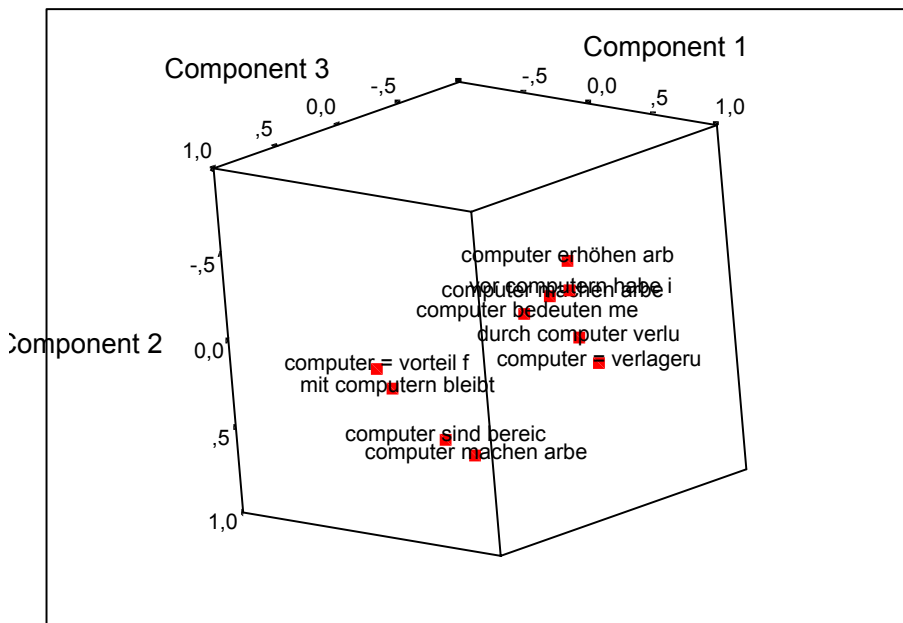
a Rotation converged in 6 iterations.

Die Markiertvariablen sind wiederum fettgedruckt. *Faktor 1* bedeutet offensichtlich *negative Auswirkungen* von Computern und Netzen in ihrem vollen Umfang. *Faktor 2*, wie in diesen Beispielen immer, ist dazu orthogonal und bedeutet offenbar zwei charakteristische, positive Folgen der Computernutzung: *interessantere Arbeit* und *Bereicherung des Privatlebens*. Aus der Orthogonalität folgt, daß diese Auswirkungen nicht als das "Gegenteil" der negativen Wirkungen, sondern als eine davon unabhängige, eigene Dimension gesehen werden. Der *Faktor 3* schließlich kennzeichnet zwei andere positive Wirkungen: *wirtschaftliche Vorteile* und *Kompensation von Standortnachteilen des ländlichen Raumes*.

Zusammenfassend bedeutet dieses Ergebnis also, daß die Befragten drei orthogonale Dimensionen empfanden, eine generelle, negative mit den bekannten Nachteilen und Beeinträchtigungen durch die Computernutzung und zwei jeweils spezielle positive, einmal die wirtschaftlichen Vorteile und zum zweiten die Vorteile für die Qualität von Berufsarbeit und Privatleben.

Die Zeichnung, die SPSS hier liefert, ist wiederum nicht besonders gut. Sie soll dennoch wiedergegeben werden. Nach einer gewissen Rotation im *Chart Editor* entsteht das folgende Bild:

### Component Plot in Rotated Space



Es läßt die drei Gruppen von Variablen, einmal generell negativ und zweimal speziell positiv, als Punkthäufungen erkennen.

Damit sind Sie am Ende dieses Kurses über die für Psychologen wichtigsten statistischen Auswertungen mit SPSS angelangt. Ich hoffe, daß Sie eine gute Basis für weitere, eigene Arbeiten mit SPSS besitzen, wenn Sie dieses Skriptum gründlich durchgearbeitet haben.

## Inhalt

|  |    |
|--|----|
| <b>Aufgabe 01</b> SPSS starten - Variable mit Typ, Namen und Bezeichner definieren -<br>Univariate Urliste eingeben - Datendatei speichern - Die wichtigsten univariaten<br>Statistiken berechnen - Häufigkeitsverteilung und kumulierte Häufigkeitsverteilung<br>aufstellen - Der Ausgabebildschirm - Graphische Darstellung - Der Chart Editor -<br>Maßzahlklassenmitten und Maßzahlklassenbreite - Kumulierte Häufigkeitsverteilung<br>graphisch.....                         | 3  |
| <b>Aufgabe 02</b> Bivariate Urliste eingeben - Variablen definieren - Randverteilungen<br>tabellarisch und graphisch univariat auswerten - Tabellen und Graphiken in ein<br>WORD-Dokument übernehmen - Korrelationen berechnen - Regressionsgeraden<br>berechnen - Bivariate Häufigkeitsverteilung mit Regressionsgerade graphisch<br>darstellen - Bivariate Häufigkeitsverteilung mit <i>Crosstabs</i> tabellarisch darstellen und<br>in EXCEL graphisch weiterverarbeiten..... | 17 |
| <b>Aufgabe 03</b> t-Test für zwei unabhängige Gruppen - Urliste eingeben - Indikatorvariable<br>definieren und eingeben - t-Test rechnen und interpretieren.....   | 29 |
| <b>Aufgabe 04</b> t-Test für abhängige und unabhängige Gruppen in einem komplexen Beispiel -<br>Die Funktion <i>Select Cases</i> .....   | 34 |
| <b>Aufgabe 05</b> Textaufgabe zur Vierfeldertafel. Die Eingabe einer Vierfeldertafel in SPSS -<br>Die Funktion <i>Weight Cases</i> - Deskriptive Statistiken, Korrelationen und<br>Signifikanztests an der Vierfeldertafel .....   | 40 |
| <b>Aufgabe 06</b> Tabellarische und rechnerische Auswertung einer univariaten Urliste (wie in<br>Aufgabe 01) - Maßzahlklassenzusammenfassung und Maßzahlentransformation mit<br>der <i>Recode</i> -Funktion.....   | 46 |
| <b>Aufgabe 07</b> Chi-quadrat-Test auf Gleichverteilung - Dateneingabe - Nutzung der Funktion<br><i>Weight Cases</i> - Berechnung des Tests.....   | 49 |
| <b>Aufgabe 08</b> Vergleich zweier unabhängiger Gruppen hinsichtlich der Zentraltendenz (wie<br>Aufgabe 03) - t-Test und U-Test (Mann-Whitney) im Vergleich .....  | 51 |
| <b>Aufgabe 09</b> U-Test nach Mann-Whitney .....   | 54 |
| <b>Aufgabe 10</b> Darstellung einer Häufigkeitsverteilung mit einer nominalen Variablen als<br>Kreisdiagramm - Dateneingabe bei einer Variablen mit Nominalniveau .....  | 55 |
| <b>Aufgabe 11</b> Einfaktorielle Varianzanalyse - Dateneingabe und Gruppierungsvariable -<br>Graphische Darstellung der Zellenmittelwerte - Newman-Keuls-Test.....   | 58 |
| <b>Aufgabe 12</b> Einfaktorielle Varianzanalyse - Newman-Keuls- und Bonferroni-Test.....   | 64 |
| <b>Aufgabe 13</b> Einfaktorielle Varianzanalyse, aufgerufen über <i>Compare Means</i> .....  | 66 |
| <b>Aufgabe 14</b> Zweifaktorielle Varianzanalyse - Zwei Gruppierungsvariablen - Graphik der<br>Resultate .....   | 67 |
| <b>Aufgabe 15</b> Dreifaktorielle Varianzanalyse.....  | 72 |
| <b>Aufgabe 16</b> Zweifaktorielle Varianzanalyse mit Meßwiederholung auf beiden Faktoren .....   | 76 |
| <b>Aufgabe 17</b> Vektorrechnung mit EXCEL - Summen und Differenzen - Distanz und Betrag<br>- Multiplikation mit einem Skalar und Skalarprodukt.....   | 82 |
| <b>Aufgabe 18</b> Matrizenrechnung mit EXCEL - Matrizenmultiplikation - Transposition -<br>Beträge von Zeilenvektoren - Winkel zwischen Zeilenvektoren .....   | 88 |
| <b>Aufgabe 19</b> Matrizenrechnung mit EXCEL - Matrixinversion - Identitätsmatrix.....   | 93 |
| <b>Aufgabe 20</b> Faktorenanalyse - Demonstrationsbeispiel Rechtecke .....   | 94 |

|  |     |
|--|-----|
| <b>Aufgabe 21</b> Faktorenanalyse - Semantisches Differential .....  | 103 |
| <b>Aufgabe 22</b> Faktorenanalyse - Die Funktion <i>Aggregate</i> - Semantisches Differential nach<br>Aggregierung der Daten ..... | 108 |
| <b>Aufgabe 23</b> Faktorenanalyse - Einstellungen zu Computern und Netzen in der<br>Bevölkerungsbefragung Sternenfels.....         | 111 |
| <b>Inhalt</b> .....  | 114 |
| <b>Glossar</b> zur Arbeit mit SPSS deutsch und englisch .....  | 116 |

**Glossar zur Arbeit mit SPSS deutsch und englisch**  
**Zum Skriptum Statistik II, Sommersemester 2003, Prof. Dr. W. R. Glaser**

| <b>Deutsch</b>                         | <b>Englisch</b>        |
|--|------------------------|
| (Faktoren-)Werte                       | Scores                 |
| Abbrechen                              | Cancel                 |
| Abhängige Variablen                    | Dependent List         |
| Absteigende Werte                      | Descending values      |
| Aggregieren                            | Aggregate              |
| Allgemeines lineares Modell            | General Linear Model   |
| Analysieren                            | Analyze                |
| Anpassung                              | Fit                    |
| Ansicht                                | View                   |
| Aufsteigende Werte                     | Ascending values       |
| Ausgabedatei                           | Output File            |
| Balken                                 | Bar                    |
| Bearbeiten                             | Edit                   |
| Beenden                                | Exit                   |
| Berechnen                              | Compute                |
| Berichte                               | Reports                |
| Binär logistisch                       | Binary Logistic        |
| Binomial                               | Binomial               |
| Bivariat                               | Bivariate              |
| Blatt (= Radius im Sonnenblumendiagr.) | Petal                  |
| Boxplot                                | Boxplot                |
| Break-Variable                         | Break Variable         |
| Chi-Quadrat                            | Chi-square             |
| Clusterzentrenanalyse                  | K-Means Cluster        |
| Datei                                  | File                   |
| Daten                                  | Data                   |
| Daten in Zwischenspeicher              | Cache Data             |
| Datenbank öffnen                       | Open Database          |
| Datendatei                             | Data File              |
| Datenlexikon zuweisen                  | Apply Data Dictionary  |
| Deskriptive Statistiken                | Descriptive Statistics |
| Deskriptive Statistiken                | Descriptives           |
| Dezimalstellen                         | Decimals               |
| Diagramme                              | Charts, Plots          |
| Dimensionsreduktion                    | Data Reduction         |
| Diskriminanzanalyse                    | Discriminant           |
| Distanzen                              | Distances              |
| Einfaktorielle ANOVA                   | One-Way ANOVA          |
| Einfügen                               | Paste                  |
| Einseitig                              | One-tailed             |
| Explorative Datenanalyse               | Explore                |
| Extraktion                             | Extraction             |
| Extras                                 | Utilities              |
| Faktorenanalyse                        | Factor                 |
| Fälle auswählen                        | Select Cases           |
| Fälle gewichten                        | Weight Cases           |

|  |                         |
|--|-------------------------|
| Fälle zusammenfassen                     | Case Summaries          |
| Fehlende Werte                           | Missing (values)        |
| Fehlerbalken                             | Error Bar               |
| Fenster                                  | Window                  |
| Filtervariable                           | Filter Variable         |
| Fläche                                   | Area                    |
| Format                                   | Format                  |
| Gewichtungsschätzung                     | Weight Estimation       |
| Grafiken                                 | Graphs                  |
| Gruppenvariable (= Gruppierungsvariable) | Grouping Variable       |
| Häufigkeiten                             | Frequencies             |
| Hierarchische Cluster                    | Hierarchical Cluster    |
| Hilfe                                    | Help                    |
| Hinzufügen                               | Add                     |
| Histogramm                               | Histogram               |
| Hoch-Tief                                | High-Low                |
| Info über Datendatei                     | Display Data Info       |
| K-S bei einer Stichprobe                 | 1-Sample K-S            |
| K unabhängige Stichproben                | K Independent Samples   |
| K verbundene Stichproben                 | K Related Samples       |
| Klassifizieren                           | Classify                |
| Komponenten (= Faktorenladungen)         | Components              |
| Korrelation                              | Correlate               |
| Korrespondenzanalyse                     | Correspondence Analysis |
| Kreis                                    | Pie                     |
| Kreuztabellen                            | Crosstabs               |
| Kurtosis                                 | Kurtosis                |
| Kurvenanpassung                          | Curve Estimation        |
| Label (= Beschriftung, Bezeichnung)      | Label                   |
| Linear                                   | Linear                  |
| Linie                                    | Line                    |
| Lowess (Name einer Anpassungstechnik)    | Lowess                  |
| Median                                   | Median                  |
| Meßniveau (= Skalenniveau)               | Measure(ment)           |
| Meßwiederholungen                        | Repeated Measures       |
| Metrisch (= Intervall oder Verhältnis)   | Scale                   |
| Mittelwert                               | Mean                    |
| Mittelwerte                              | Means                   |
| Mittelwerte vergleichen                  | Compare Means           |
| Modalwert                                | Mode                    |
| Multinomial logistisch                   | Multinomial Logistic    |
| Multivariat                              | Multivariate            |
| Neu                                      | New                     |
| Nichtlinear                              | Nonlinear               |
| Nichtparametrische Tests                 | Nonparametric Tests     |
| Nominal                                  | Nominal                 |
| Öffnen                                   | Open                    |
| OK                                       | OK                      |
| Optimale Skalierung                      | Optimal Scaling         |
| Optionen                                 | Options                 |

|  |                            |
|--|----------------------------|
| Ordinal                                | Ordinal                    |
| Pareto                                 | Pareto                     |
| Partiell                               | Partial                    |
| Probit                                 | Probit                     |
| Quartile                               | Quartiles                  |
| Regelkarten                            | Control                    |
| Regression                             | Regression                 |
| Rotation                               | Rotation                   |
| Schicht                                | Layer                      |
| Schiefe                                | Skewness                   |
| Screeplot                              | Scree-Plot                 |
| Sequenzen                              | Runs                       |
| Sonnenblume                            | Sunflower                  |
| Sortieren nach                         | Order by                   |
| Spalte                                 | Column                     |
| Spannweite                             | Range                      |
| Speichern                              | Save                       |
| Speichern unter                        | Save As                    |
| Spherizität                            | Sphericity                 |
| Standardabweichung                     | Std. deviation             |
| Statistik                              | Statistics                 |
| Std.-Fehler (des Mittelwertes)         | S.E. mean                  |
| Streudiagramm                          | Scatter                    |
| Summe                                  | Sum                        |
| Syntaxdatei                            | Syntax File                |
| T-Test bei einer Stichprobe            | One-Sample T Test          |
| T-Test bei gepaarten Stichproben       | Paired-Samples T Test      |
| T-Test bei unabhängigen Stichproben    | Independent-Samples T Test |
| Textdaten einlesen                     | Read Text Data             |
| Transformieren                         | Transform                  |
| Trennwert                              | Cut Point                  |
| Umcodieren                             | Recode                     |
| Unabhängige Variablen                  | Independent List           |
| Univariat                              | Univariate                 |
| Variablenlabel (= Variablenbezeichner) | Variable Label             |
| Varianz-Komponenten                    | Variance Components        |
| Varianz                                | Variance                   |
| Werte (= Faktorenwerte)                | Scores                     |
| Weiter                                 | Continue                   |
| Wertelabel (= Wertebezeichner)         | Value Label                |
| Zeile                                  | Row                        |
| Zeitreihen                             | Time Series                |
| Zelle                                  | Cell                       |
| Zuletzt verwendete Dateien             | Recently used files        |
| Zuletzt verwendete Daten               | Recently used data         |
| Zurücksetzen                           | Reset                      |
| Zwei unabhängige Stichproben           | 2 Independent Samples      |
| Zwei verbundene Stichproben            | 2 Related Samples          |
| Zweiseitig                             | Two-tailed                 |
| Zweistufige kleinste Quadrate          | 2-Stage Least Squares      |

| <b>Englisch</b>         | <b>Deutsch</b>                   |
|-------------------------|----------------------------------|
| 1-Sample K-S            | K-S bei einer Stichprobe         |
| 2-Stage Least Squares   | Zweistufige kleinste Quadrate    |
| 2 Independent Samples   | Zwei unabhängige Stichproben     |
| 2 Related Samples       | Zwei verbundene Stichproben      |
| Add                     | Hinzufügen                       |
| Aggregate               | Aggregieren                      |
| Analyze                 | Analysieren                      |
| Apply Data Dictionary   | Datenlexikon zuweisen            |
| Area                    | Fläche                           |
| Ascending values        | Aufsteigende Werte               |
| Bar                     | Balken                           |
| Binary Logistic         | Binär logistisch                 |
| Binomial                | Binomial                         |
| Bivariate               | Bivariat                         |
| Boxplot                 | Boxplot                          |
| Break Variable          | Break-Variable                   |
| Cache Data              | Daten in Zwischenspeicher        |
| Cancel                  | Abbrechen                        |
| Case Summaries          | Fälle zusammenfassen             |
| Cell                    | Zelle                            |
| Charts                  | Diagramme                        |
| Chi-square              | Chi-Quadrat                      |
| Classify                | Klassifizieren                   |
| Column                  | Spalte                           |
| Compare Means           | Mittelwerte vergleichen          |
| Components              | Komponenten (= Faktorenladungen) |
| Compute                 | Berechnen                        |
| Continue                | Weiter                           |
| Control                 | Regelkarten                      |
| Correlate               | Korrelation                      |
| Correspondence Analysis | Korrespondenzanalyse             |
| Crosstabs               | Kreuztabellen                    |
| Curve Estimation        | Kurvenanpassung                  |
| Cut Point               | Trennwert                        |
| Data                    | Daten                            |
| Data File               | Datendatei                       |
| Data Reduction          | Dimensionsreduktion              |
| Decimals                | Dezimalstellen                   |
| Dependent List          | Abhängige Variablen              |
| Descending values       | Absteigende Werte                |
| Descriptive Statistics  | Deskriptive Statistiken          |
| Descriptives            | Deskriptive Statistiken          |
| Discriminant            | Diskriminanzanalyse              |
| Display Data Info       | Info über Datendatei             |
| Distances               | Distanzen                        |
| Edit                    | Bearbeiten                       |

|                            |  |
|----------------------------|--|
| Error Bar                  | Fehlerbalken                             |
| Exit                       | Beenden                                  |
| Explore                    | Explorative Datenanalyse                 |
| Extraction                 | Extraktion                               |
| Factor                     | Faktorenanalyse                          |
| File                       | Datei                                    |
| Filter Variable            | Filtervariable                           |
| Fit                        | Anpassung                                |
| Format                     | Format                                   |
| Frequencies                | Häufigkeiten                             |
| General Linear Model       | Allgemeines lineares Modell              |
| Graphs                     | Grafiken                                 |
| Grouping Variable          | Gruppenvariable (= Gruppierungsvariable) |
| Help                       | Hilfe                                    |
| Hierarchical Cluster       | Hierarchische Cluster                    |
| High-Low                   | Hoch-Tief                                |
| Histogram                  | Histogramm                               |
| Independent-Samples T Test | T-Test bei unabhängigen Stichproben      |
| Independent List           | Unabhängige Variablen                    |
| K-Means Cluster            | Clusterzentrenanalyse                    |
| K Independent Samples      | K unabhängige Stichproben                |
| K Related Samples          | K verbundene Stichproben                 |
| Kurtosis                   | Kurtosis                                 |
| Label                      | Label, Beschriftung, Bezeichnung         |
| Layer                      | Schicht                                  |
| Line                       | Linie                                    |
| Linear                     | Linear                                   |
| Lowess                     | Lowess (Name einer Anpassungstechnik)    |
| Mean                       | Mittelwert                               |
| Means                      | Mittelwerte                              |
| Measure                    | Maßzahl                                  |
| Measure(ment)              | Meßniveau (= Skalenniveau)               |
| Median                     | Median                                   |
| Missing (values)           | Fehlende Werte                           |
| Mode                       | Modalwert                                |
| Multinomial Logistic       | Multinomial logistisch                   |
| Multivariate               | Multivariat                              |
| New                        | Neu                                      |
| Nominal                    | Nominal                                  |
| Nonlinear                  | Nichtlinear                              |
| Nonparametric Tests        | Nichtparametrische Tests                 |
| OK                         | OK                                       |
| One-Sample T Test          | T-Test bei einer Stichprobe              |
| One-tailed                 | Einseitig                                |
| One-Way ANOVA              | Einfaktorielle ANOVA                     |
| Open                       | Öffnen                                   |
| Open Database              | Datenbank öffnen                         |
| Optimal Scaling            | Optimale Skalierung                      |
| Options                    | Optionen                                 |
| Order by                   | Sortieren nach                           |



|                       |  |
|-----------------------|--|
| Ordinal               | Ordinal                                |
| Output File           | Ausgabedatei                           |
| Paired-Samples T Test | T-Test bei gepaarten Stichproben       |
| Pareto                | Pareto                                 |
| Partial               | Partiell                               |
| Paste                 | Einfügen                               |
| Petal                 | Blatt (= Radius im Sonnenblumendiagr.) |
| Plot                  | Diagramm                               |
| Pie                   | Kreis                                  |
| Probit                | Probit                                 |
| Quartiles             | Quartile                               |
| Range                 | Spannweite                             |
| Read Text Data        | Textdaten einlesen                     |
| Recently used data    | Zuletzt verwendete Daten               |
| Recently used files   | Zuletzt verwendete Dateien             |
| Recode                | Umcodieren                             |
| Regression            | Regression                             |
| Repeated Measures     | Meßwiederholungen                      |
| Reports               | Berichte                               |
| Reset                 | Zurücksetzen                           |
| Rotation              | Rotation                               |
| Row                   | Zeile                                  |
| Runs                  | Sequenzen                              |
| S.E. mean             | Std.-Fehler (des Mittelwertes)         |
| Save                  | Speichern                              |
| Save As               | Speichern unter                        |
| Scale                 | Metrisch (= Intervall oder Verhältnis) |
| Scatter               | Streudiagramm                          |
| Scores                | (Faktoren-)Werte                       |
| Scree-Plot            | Screeplot                              |
| Select Cases          | Fälle auswählen                        |
| Skewness              | Schiefe                                |
| Sphericity            | Sphärizität                            |
| Statistics            | Statistik                              |
| Std. deviation        | Standardabweichung                     |
| Sum                   | Summe                                  |
| Sunflower             | Sonnenblume                            |
| Syntax File           | Syntaxdatei                            |
| Time Series           | Zeitreihen                             |
| Transform             | Transformieren                         |
| Two-tailed            | Zweiseitig                             |
| Univariate            | Univariat                              |
| Utilities             | Extras                                 |
| Value Label           | Wertelabel (= Wertebezeichner)         |
| Variable Label        | Variablenlabel (= Variablenbezeichner) |
| Variance              | Varianz                                |
| Variance Components   | Varianz-Komponenten                    |
| View                  | Ansicht                                |
| Weight Cases          | Fälle gewichten                        |
| Weight Estimation     | Gewichtungsschätzung                   |

| Window   | Fenster |
|--|---------|
| <p><b>Anmerkung:</b> Die statistischen Begriffe, die in der englischen Version von SPSS verwendet werden, sind weitgehend korrekt. Die Übersetzungen in den deutschen Versionen von SPSS sind gelegentlich irreführend, schlechtes Deutsch oder in der Statistik ungebräuchlich. In diesen Fällen wurden die richtigen Bezeichnungen oder Erläuterungen in Klammern hinzugefügt.</p> |         |

