



## **Eindimensionale Tabellierung Lage- und Streuungsmaße**

Kurt Holm

Almo Statistik-System

[www.almo-statistik.de](http://www.almo-statistik.de)

[holm@almo-statistik.de](mailto:holm@almo-statistik.de)

[kurt.holm@jku.at](mailto:kurt.holm@jku.at)

Autor: em. Prof. Dr. Kurt Holm, Universität Linz, Österreich

### **Zum Thema Tabellierung siehe auch die Almo-Dokumente**

- 1b Zwei- und drei-dimensionale Tabellierung
- 2 Beliebig-dimensionale Tabellierung
- 27 Mehrfachantworten, Tabellierung

### **Weitere Almo-Dokumente**

Die folgenden Dokumente können alle kostenlos von der Handbuchseite in [www.almo-statistik.de](http://www.almo-statistik.de) heruntergeladen werden

- 0. Arbeiten\_mit\_Almo.PDF (1 MB)
- 1a. Eindimensionale Tabellierung.PDF (2.4 MB)
- 1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
- 2. Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
- 3. Nicht-parametrische Verfahren.PDF (0.9 MB)
- 4. Kanonische Analysen.PDF (1.8 MB)  
Diskriminanzanalyse.PDF (1.8 MB)  
enthält: Kanonische Korrelation, Diskriminanzanalyse, bivariate Korrespondenzanalyse, optimale Skalierung
- 5. Korrelation.PDF (1.4 MB)
- 6. Allgemeine multiple Korrespondenzanalyse.PDF (1.5 MB)
- 7. Allgemeines ordinales Rasch-Modell.PDF (0.6 MB)
- 7a. Wie man mit Almo ein Rasch-Modell rechnet.PDF (0.2 MB)
- 8. Tests auf Mittelwertsdifferenz, t-Test.PDF (1,6 MB)
- 9. Logitanalyse.pdf (1,2MB) enthält Logit- und Probitanalyse
- 10. Koeffizienten der Logitanalyse.PDF (0,06 MB)
- 11. Daten-Fusion.PDF (1,1 MB)
- 12. Daten-Imputation.PDF (1,3 MB)
- 13. ALM Allgemeines Lineares Modell.PDF (2.3 MB)
- 13a. ALM Allgemeines Lineares Modell II.PDF (2.7 MB)
- 14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer, Cox-Regression.PDF (1,5 MB)
- 15. Faktorenanalyse.PDF (1,6 MB)
- 16. Konfirmatorische Faktorenanalyse.PDF (0,3 MB)
- 17. Clusteranalyse.PDF (3 MB)
- 18. Pisa 2012 Almo-Daten und Analyse-Programme.PDF (17 KB)
- 19. Guttman- und Mokken-Skalierung.PFD (0.8 MB)
- 20. Latent Structure Analysis.PDF (1 MB)
- 21. Statistische Algorithmen in C (80 KB)
- 22. Conjoint-Analyse (PDF 0,8 MB)
- 23. Ausreisser entdecken (PDF 170 KB)
- 24. Statistische Datenanalyse Teil I, Data Mining I
- 25. Statistische Datenanalyse Teil II, Data Mining II
- 26. Statistische Datenanalyse Teil III, Arbeiten mit Almo-Datenanalyse-System
- 27. Mehrfachantworten, Tabellierung von Fragen mit Mehrfachantworten
- 28. Metrische multidimensionale Skalierung (MDS) (0,4 MB)
- 29. Metrisches multidimensionales Unfolding (MDU) (0,6 MB)
- 30. Nicht-metrische multidimensionale Skalierung (MDS) (0,4 MB)
- 31. Pfadanalyse als wiederholte Regressionsanalyse (0,7 MB)
- 32. Datei-Operationen mit Almo (1,1 MB)
- 33. Wählerstromanalyse und Wahlhochrechnung (1,6 MB)
- 34. Soziometrie. Auswertung soziometrischer Daten (0,5 MB)

## Inhalt

<b>P4 Eindimensionale Häufigkeitsverteilung .....</b>	<b>4</b>
P4.1 Eingabe mit Maskenprogramm Prog04m1 .....	4
P4.1.1 Erläuterungen zu den Dialogboxen .....	6
P4.5 Eingabe mit selbstgeschriebenem ALMO–Programm mit Optionen .....	7
P4.3 Ergebnisse .....	7
P4.4 Häufigkeitsverteilung mit Optionen .....	12
P4.4.1 Eingabe mit Maskenprogramm Prog04m2.Msk .....	12
P4.4.2 Erläuterungen zu den Dialogboxen .....	15
P4.5 Eingabe mit selbstgeschriebenem ALMO–Programm mit Optionen .....	20
P4.6 Gewichtung .....	20
P4.7 Eingabe von Verteilungen .....	24
P4.7.1 Eingabe mit Maskenprogramm Prog04m3.Msk .....	24
P4.7.2 Erläuterungen zu den Dialogboxen .....	27
P4.8 Erläuterungen zu den Optionen .....	28
P4.8.1 Vertrauensintervalle .....	28
P4.8.2 Lage- und Streuungsmaße (arithmetisches, harmonische, geometrisches Mittel, Standardabweichung, Median, Quartile) .....	28
P4.8.3 Normalverteilungstest, Schiefe und Exzeß .....	29
P4.8.4 Test auf Gleichverteilung .....	38
P4.8.5 Chi-Quadrat-Test für beliebige Verteilungen mit Prog04m4.....	41
P4.8.6 Binomialtest.....	45
P4.8.7 Transformation von Variablen auf Normalverteilung.....	45
P4.10 Auszählung von Variablen mit sehr vielen Ausprägungen .....	46
<b>P5 Lage- und Streuungsmaße .....</b>	<b>47</b>
P5.01 Eingabe in Programm 5 mit Maskenprogramm.....	47
P5.01.1 Erläuterungen zu den Eingabeboxen .....	49
P5.02 Maskenprogramm mit Optionen .....	50
P5.03 Eingabe mit selbstgeschriebenem ALMO–Programm .....	54
P5.04 Optionen für das selbstgeschriebenem ALMO–Programm.....	54
P5.05 Almo-Ausgabe I.....	54
P5.06 Almo-Ausgabe II .....	56
P5.07 Anteilswerte (relative Häufigkeiten) .....	57
Literatur.....	57

## P4 Eindimensionale Häufigkeitsverteilung

Das Programm ermittelt, wie häufig die Ausprägungen von Variablen besetzt sind. Programmiert wurde es von Kurt Holm.

Die Variablen können ganzzahlige Werte aber auch Werte mit Dezimalstellen besitzen.

Almo erstellt Häufigkeitstabellen folgender Art:

```
Variable: Leistung
-----
```

Wert	Fälle	%	% ohne KW	% kumuliert
Kein Wert	2	3.28	-	-
1 gut	15	24.59	25.42	25.42
2 befriedigend	36	59.02	61.02	86.44
3 nicht ausreiche	8	13.11	13.56	100.00
Summe	61	100%		
Summe ohne KW	59		100%	

Almo stellt die ermittelte Häufigkeitsverteilung grafisch dar und berechnet (optional) auch noch:

1. Vertrauensintervalle
2. Mittelwert, Standardabweichung, Standardfehler des Mittelwerts
3. Median, 1. und 3. Quartil, interquartile Differenz, Variabilitätskoeffizient
4. Bei ganzzahligen Variablen: Chi-Quadrat Anpassungstest, für Normalverteilungen oder Gleichverteilung, Schiefe und Exzeß der Normalverteilung (mit deren Signifikanz)
5. Bei Variablen mit Dezimalwerten: Kolmogorov-Smirnov Ein-Stichprobentest zur Überprüfung der Normalverteilung oder Gleichverteilung.
6. Binomialtest
7. Transformation der Daten auf Normalverteilung und zu Prozentrangwerten.

### P4.1 Eingabe mit Maskenprogramm Prog04m1

Klicken Sie auf den Knopf "Verfahren". In der dann erscheinenden Auswahlliste klicken Sie auf "Häufigkeitsverteilung" und danach auf Prog04m1.Msk. ALMO präsentiert Ihnen dann folgendes Maskenprogramm.

Siehe dazu die sehr ausführliche Darstellung in P0 und in der Bedienungsanleitung zu ALMO, Abschnitt 4, insbesondere 4a.

Prog04M1.Msk  
Häufigkeitsverteilung


Das Programm erstellt Häufigkeitsverteilungen folgender Art

Beruf	Fälle	%
1 Arbeiter	16	26.23
2 Angestellter	29	47.54
3 Selbständiger	16	26.23


Grafik: Balkendiagramme  
siehe Handbuch, Teil 3a, Abschnitt P4.1


Programm-Bedienung --->

Vereinbare Variable=


 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert




Datei der Variablennamen



       zeige = Namensdatei in Output zeigen  
leer = nicht zeigen

Freie Namensfelder




 Leere alle Eingabefelder dieser Sub-Box

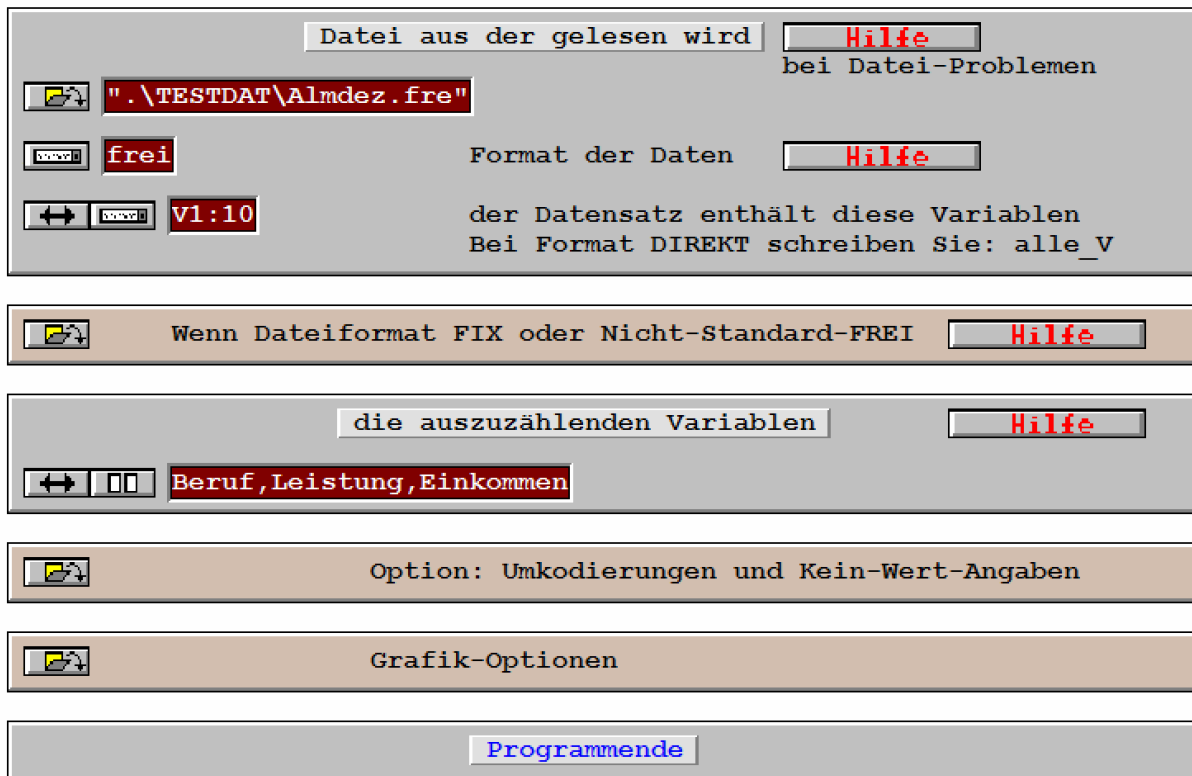
  

erzeuge zusätzliche Namensfelder

Variablennamen in Datei speichern

Eingabefeld leer = nicht speichern



## P4.1.1 Erläuterungen zu den Dialogboxen

### Eingabebox: Speicher für x Variable

Siehe Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.1.

### Eingabebox: Weitere Vereinbarungen

Siehe Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.2.

### Eingabebox: Datei der Variablennamen

Siehe P0.3.

### Eingabebox: Freie Namensfelder

Siehe P0.3.

### Eingabebox: Datei aus der gelesen wird

Siehe P0.4.

### Eingabebox: Die auszuzählenden Variablen



Klicken Sie auf den Knopf mit den zwei kleinen Fenstersymbolen. Almo präsentiert Ihnen dann die Variablen-Auswahl-Box. In ihr klicken Sie auf die Variablen, die ausgezählt werden sollen. Siehe dazu die ausführliche Darstellung in P0.11.

### Eingabebox: Kein\_Wert-Angabe und Umkodierungen

Siehe P0.5.

## **P4.5 Eingabe mit selbstgeschriebenem ALMO-Programm mit Optionen**

Kapitel entfernt. Ist in Handbuch "Teil3a: Grundlegende Verfahren" enthalten.

## **P4.3 Ergebnisse**

Für das Maskenprogramm und das äquivalente selbstgeschriebene Programm liefert ALMO folgende Ausgabe, die wir hier gekürzt (nur für V5 Leistung) wiedergeben:

Eindimensionale Häufigkeitsverteilungen

Zahl der eingelesenen Datensätze: 61

Variable 5 Leistung

	Faelle	%	% kumuliert
1 sehr gut	5	8.20	8.20
2 gut	23	37.70	45.90
3 befriedigend	23	37.70	83.61
4 ausreichend	6	9.84	93.44
5 ungenuegend	4	6.56	100.00
Summe	61	100	

Auf diese Tabelle folgt dann im Output diese Grafik

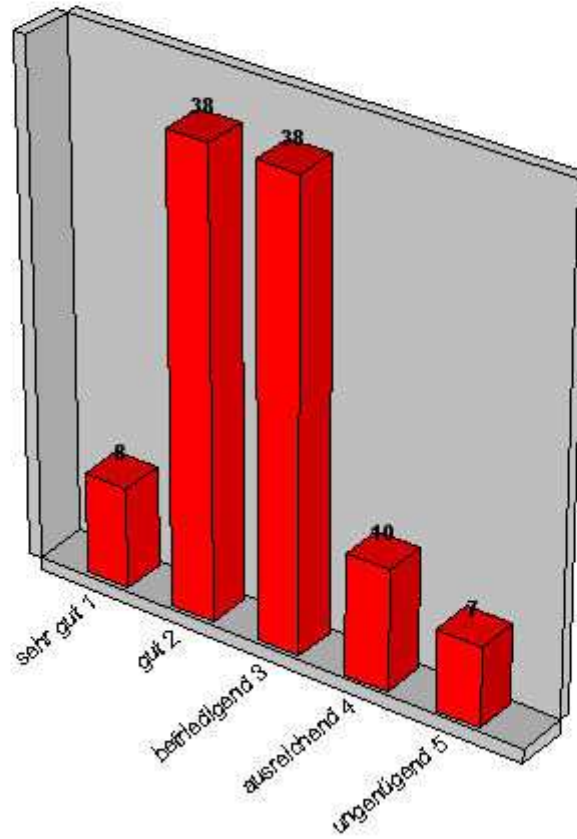
Balkendiagramm  
fuer obige Haeufigkeitsverteilung  
Grafik 01

Hilfe  
Grafik  
1156151802

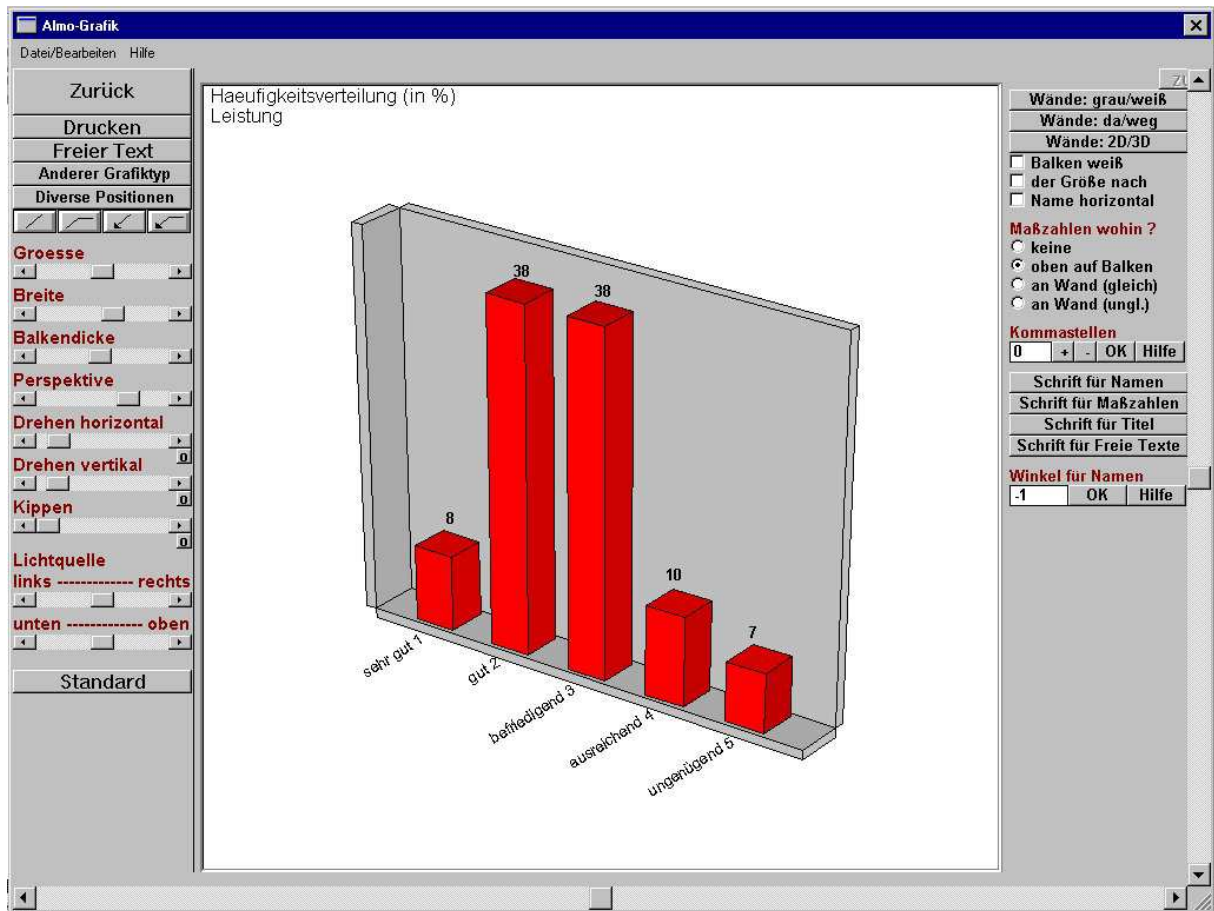
Grafik bearbeiten  
oder neu erzeugen

← Grafik loeschen

Haeufigkeitsverteilung (in %)  
Leistung



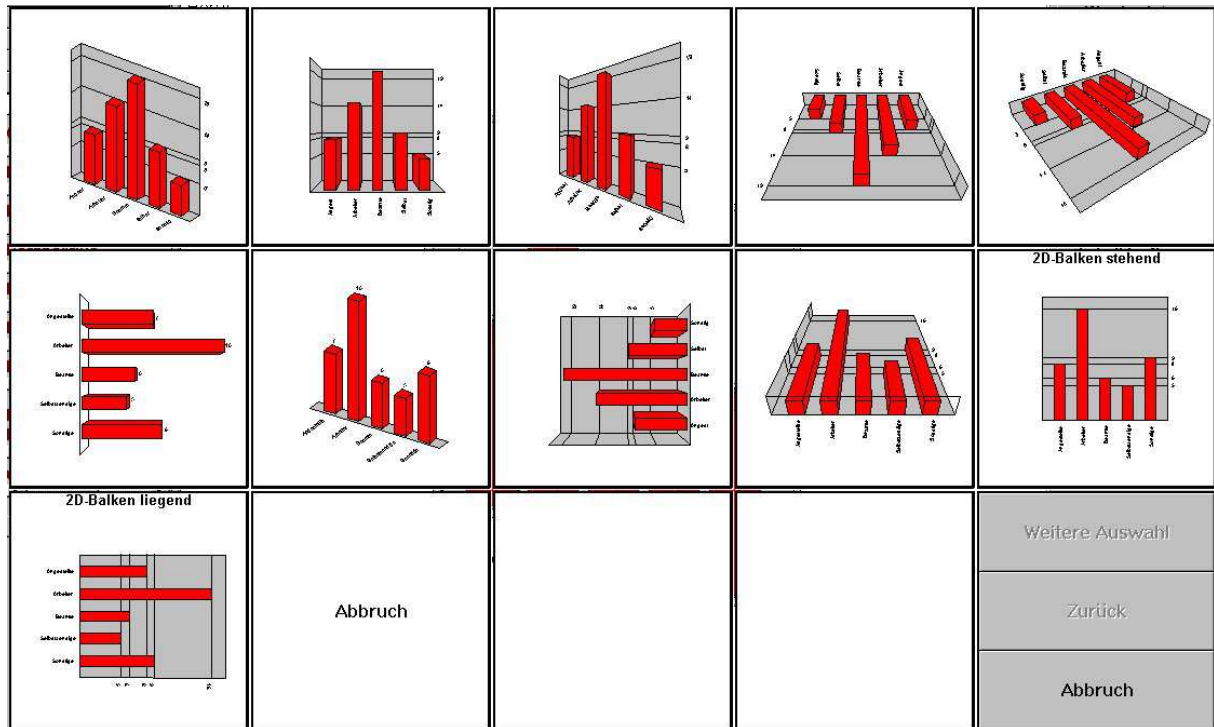
Wenn Sie auf den Knopf "Grafik" klicken dann lädt Almo den Grafikeditor und präsentiert Ihnen das Balkendiagramm, das Sie nunmehr bearbeiten, d.h. in vielfältiger Weise verändern können.



Wir wollen hier die vielfältigen Möglichkeiten des ALMO-Grafik-Editors nicht beschreiben. Eine ausführliche Darstellung ist im Dokument "AnleitungGrafik.pdf" (im Wurzelverzeichnis von Almo) und im Almo-Handbuch, Teil 1: "Bedienungsanleitung" enthalten.

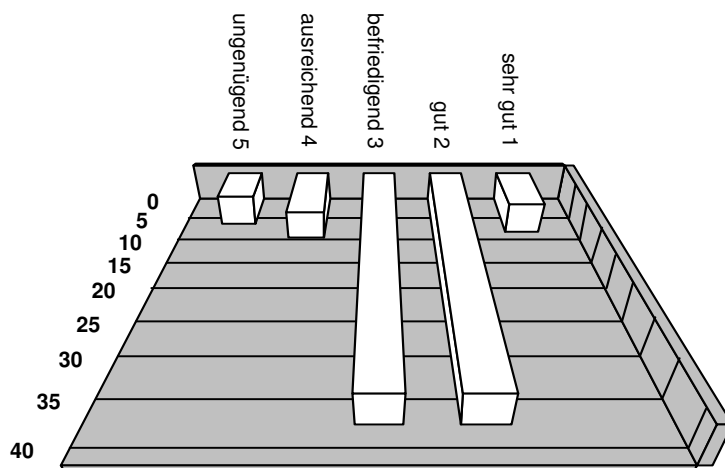
Zwei bedeutsame Gestaltungsmöglichkeiten wollen wir jedoch ausführlicher behandeln.

Klicken Sie auf den Knopf "Diverse Positionen". ALMO bietet Ihnen dann folgende Auswahl an:

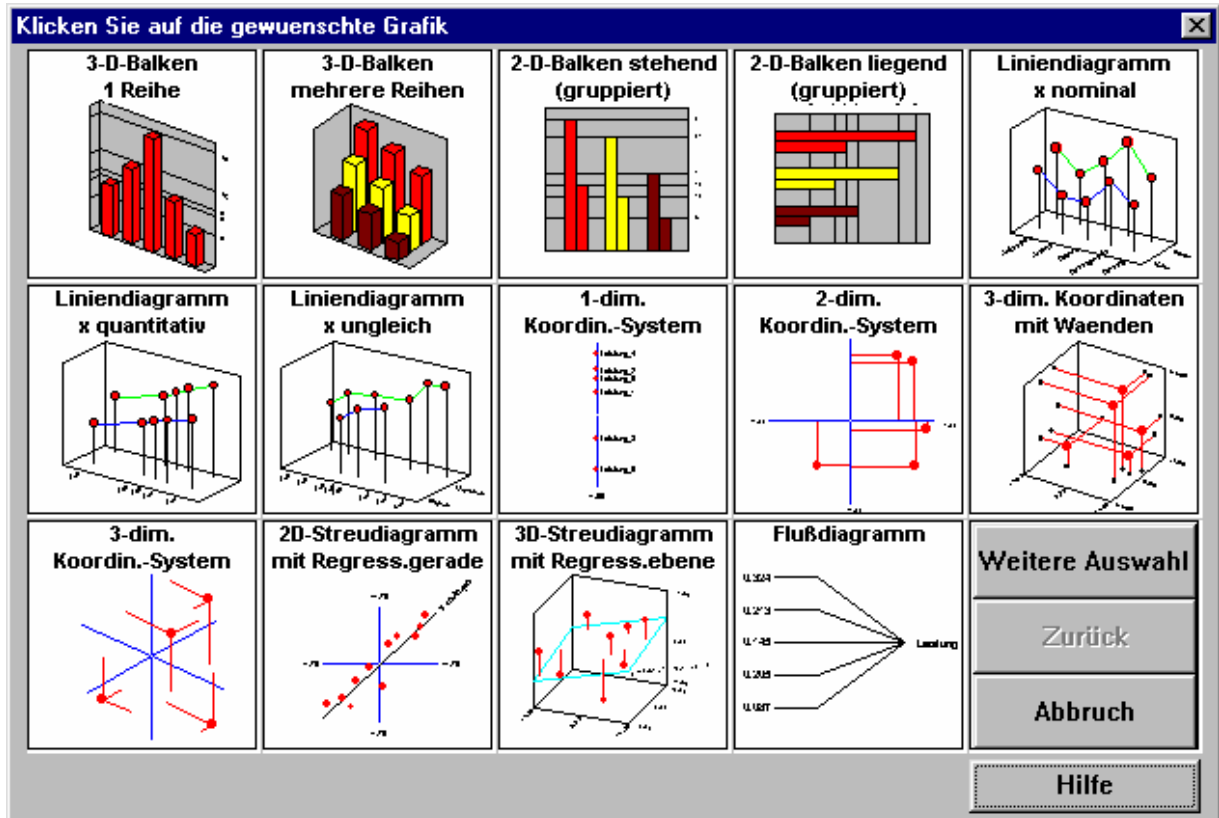


Klicken Sie auf die Position, die Sie wünschen. Also wandelt dann die aktuelle Grafik in diese Position. Also ändert dabei die Einstellung der Schieber 'Grösse', 'Perspektive', 'Drehen horizontal', 'Drehen vertikal' und 'Kippen'. Das bedeutet, dass Sie diese diversen Positionen auch selbst herstellen könnten, wenn Sie diese Schieber verstellen.

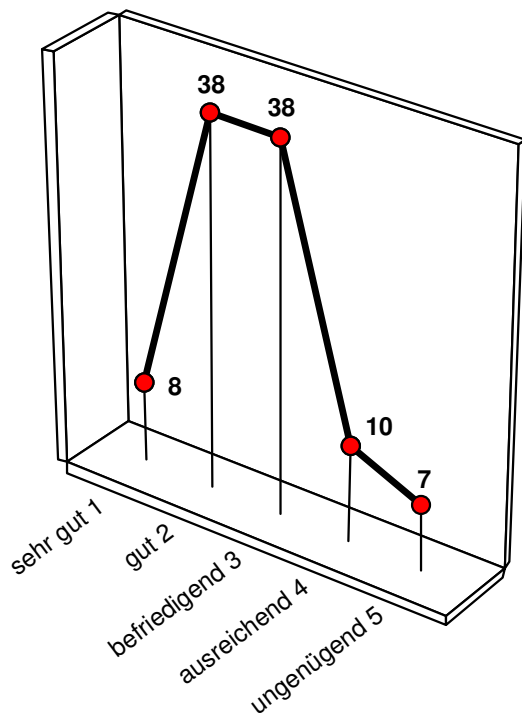
Wenn Sie beispielsweise auf die 4. Grafik in der ersten Reihe klicken, dann erzeugt ALMO folgende Grafik



Sie können auch anstelle eines Balkendiagramms ein Liniendiagramm erzeugen. Klicken Sie auf den Knopf "Anderer Grafiktyp". Sie erhalten dann folgende Auswahl von Grafiktypen präsentiert.



Klicken Sie auf das 5. Bild in der ersten Reihe (Liniendiagramm, x nominal). ALMO wandelt dann das Balkendiagramm in folgendes Liniendiagramm:



Sie können nun auch hier auf den Knopf "Diverse Positionen" klicken und das Liniendiagramm in einer anderen Position darstellen.

#### **P4.4 Häufigkeitsverteilung mit Optionen**

Neben Häufigkeitsverteilungen kann mit Programm 4 optional noch folgendes gerechnet werden.

- 1) Vertrauensintervalle
- 2) Mittelwert, Standardabweichung, Standardfehler des Mittelwerts
- 3) Median, 1. und 3. Quartil, interquartile Differenz, Variabilitätskoeffizient
- 4) Bei ganzzahligen Variablen: Chi-Quadrat Anpassungstest, für Normalverteilungen oder Gleichverteilung, Schiefe und Exzeß der Normalverteilung (mit deren Signifikanz)
- 5) Bei Variablen mit Dezimalwerten: Kolmogorov-Smirnov Ein-Stichprobentest zur Überprüfung der Normalverteilung oder Gleichverteilung.
- 6) Binomialtest
- 7) Transformation der Daten auf Normalverteilung und zu Prozentrangwerten.

##### **P4.4.1 Eingabe mit Maskenprogramm Prog04m2.Msk**

Klicken Sie auf den Knopf **Verfahren**. In der dann erscheinenden Auswahlliste klicken Sie auf **Häufigkeitsverteilung** und danach auf **Prog04m2**. ALMO präsentiert Ihnen dann folgendes Maskenprogramm.

Prog04M2.Msk  
Haeufigkeitsverteilung  
mit Optionen

Das Programm erstellt Häufigkeitsverteilungen folgender Art

Beruf	Fälle	%
1 Arbeiter	16	26.23
2 Angestellter	29	47.54
3 Selbständiger	16	26.23

Dabei besteht die Möglichkeit die Häufigkeitsverteilungen für eine "Gruppierungsvariable" wiederholt zu erzeugen. Beispiel: Die Gruppierungsvariable sei "Schulabschluß" (Volksschule, Gymnasium, Universität). Dann können die Häufigkeitsverteilungen separat ausgezählt werden für Personen mit Volksschulabschluß, Gymnasium, Universität - und dies in einem Programmdurchlauf

Für diese Häufigkeitstabelle können u.a. berechnet werden:

- Vertrauensintervalle
- arithmetisches, harmonisches, geometrisches Mittel
- Median, Quartile, Perzentile
- Standardabweichung
- Schiefe und Exzess der Normalverteilung
- Chi-Quadrat-Anpassungstest für Normalverteilung und Gleichverteilung.
- Kolmogorov-Smirnov-Einstichprobentest für Normalverteilung und Gleichverteilung.
- Binomialtest.
- Transformation der Daten auf Normalverteilung oder zu Prozentrangwerten

Grafik: Balkendiagramme


siehe Handbuch, Teil 3a, Abschnitt P4.4

Programm-Bedienung --->

Hilfe


Speicher fuer x Variable

Vereinbare Variable=

 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Variablennamen

Datei der Variablennamen

 ".\Testdat\Varname2.nam"

**zeige**      zeige = Namensdatei in Output zeigen  
leer = nicht zeigen


Freie Namensfelder

Leere alle Eingabefelder dieser Sub-Box

erzeuge zusätzliche Namensfelder


Variablennamen in Datei speichern

Eingabefeld leer = nicht speichern




Datei aus der gelesen wird

bei Datei-Problemen

 ".\TESTDAT\Almdez.fre"


**frei**      Format der Daten


**V1:10**      der Datensatz enthält diese Variablen  
Bei Format DIREKT schreiben Sie: alle\_v

 Wenn Dateiformat FIX oder Nicht-Standard-FREI

die auszuzählenden Variablen

**V1:8**

 Option: Gruppierungsvariable

 Option: Ein- und Ausschliessen von Untersuchungseinheiten



Die Gruppierungsvariable muß ganzzahlige Werte besitzen und fortlaufend, ohne Lücke, kodiert sein. Ist dies nicht der Fall, dann muss sie nachfolgend auf Ganzzahligkeit und lückenlos fortlaufende Werte umkodiert werden.

### Eingabebox: Ein- und Ausschließen von Untersuchungseinheiten

Siehe Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.7.

### Eingabebox: Kein-Wert-Angabe und Umkodierungen

Siehe dazu die ausführliche Erläuterung im Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.5. und P0.6.

Wird die Optionsbox geöffnet, dann sieht man folgendes:

Es werden 3 Sub-Boxen präsentiert. Nun muss noch die mittlere Sub-Box geöffnet werden. Dies ist ein Ausschnitt aus der geöffneten mittleren Sub-Box:

Bei Leistung und Alter werden Ausprägungen zusammengefasst. Almo gibt z.B. für Leistung folgende Häufigkeitsverteilung aus:

Variable 5 Leistung

Wert	Faelle	%	% kumuliert
1	5	8.20	8.20
2	23	37.70	45.90
3	23	37.70	83.61
4	6	9.84	93.44

5	4	6.56	100.00
Summe	61	100%	

Wird nicht umkodiert, dann gibt Almo aus:

Variable 5 Leistung

-----

Wert	Faelle	%	% kumuliert
1.1	3	4.92	4.92
1.2	2	3.28	8.20
2.1	5	8.20	16.39
2.2	5	8.20	24.59
3.1	3	4.92	29.51
3.2	10	16.39	45.90
4.1	5	8.20	54.10
4.2	10	16.39	70.49
5.1	4	6.56	77.05
5.2	4	6.56	83.61
6.2	2	3.28	86.89
7.1	3	4.92	91.80
7.2	1	1.64	93.44
8.1	2	3.28	96.72
8.2	1	1.64	98.36
9.1	1	1.64	100.00
Summe	61	100%	

### Eingabebox: Gewichtung

Siehe nachfolgenden Abschnitt P4.6

### Eingabebox: Optionen: Verschiedene Tests und Koeffizienten



Optionsbox geöffnet:



Anpassungstest gerechnet werden soll, dann schreiben Sie in das Eingabefeld V3 bzw. Sie klicken auf den Knopf mit den 2 Fenstersymbolen und selektieren dann in der daraufhin erscheinenden Variablen-Auswahl-Box die Variable V3.

*Eingabefelder in der 2. Hälfte der Optionsbox:*

Wir wollen im folgenden auf diese Optionen ausführlich eingehen:

Vertrauensintervall in Abschnitt P4.8.1

Lage- und Streuungsmaße in Abschnitt P4.8.2

Binomialtest in Abschnitt P4.8.5

Transformation auf Normalverteilung in Abschnitt P4.8.6

Kurze Erläuterung zum Binomialtest:

Der Binomialtest kann nur für zweiwertige Variable gerechnet werden. Er überprüft, ob die tatsächliche Verteilung auf die beiden Ausprägungen der Variablen mit einer vermuteten Verteilung übereinstimmt - oder nicht.

Geben Sie im 1. Eingabefeld die Variablen an, für die der Binomialtest gerechnet werden soll. Im 2. Eingabefeld geben Sie den vermuteten Anteilswert für die jeweils 1. Ausprägung der Variablen an.

### **Eingabebox: Grafik-Optionen**

Siehe P0.10.

Zur Erzeugung von Grafiken bietet Ihnen Almo drei Möglichkeiten an:

#### **1. Möglichkeit**

Grafik = Almo;

Es wird eine hochauflösende Grafik erzeugt. In der Ergebnisliste finden Sie einen großen Knopf mit der Aufschrift "Grafik". Wenn Sie auf diesen klicken, dann öffnet Almo das Grafikfenster und erzeugt eine Grafik.

#### **2. und 3. Möglichkeit**

Grafik = Excel;

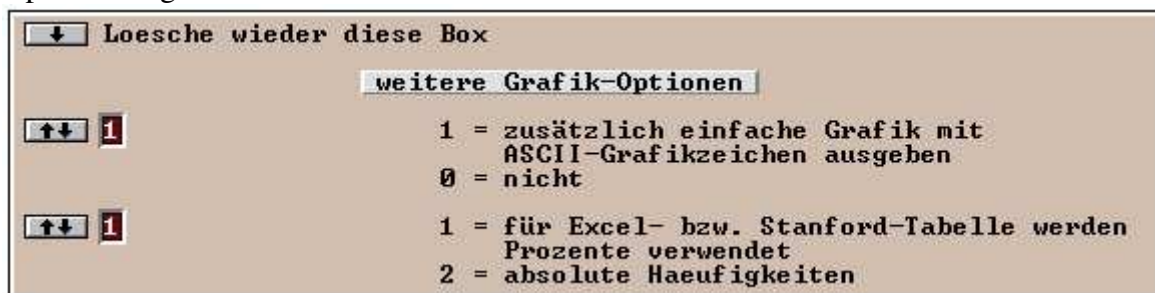
Grafik = Stanford;

ALMO erzeugt Grafik-Tabellen, die als Dateien gespeichert werden. Sie können in Excel bzw. Stanford-Graphics eingelesen werden und dort unmittelbar in eine Graphik Ihrer Wahl gewandelt werden.

### **Eingabebox: weitere Grafik-Optionen**



Optionsbox geöffnet:



*Eingabefeld 1:* Zusätzlich können Sie einfache Balkendiagramme im Textmodus anfordern. Diese können wie Text ohne Probleme ausgedruckt werden.

*Eingabefeld 2:* Für die Excel- und Stanford-Grafik-Tabellen werden entweder die Prozentwerte oder die absoluten Häufigkeiten verwendet.

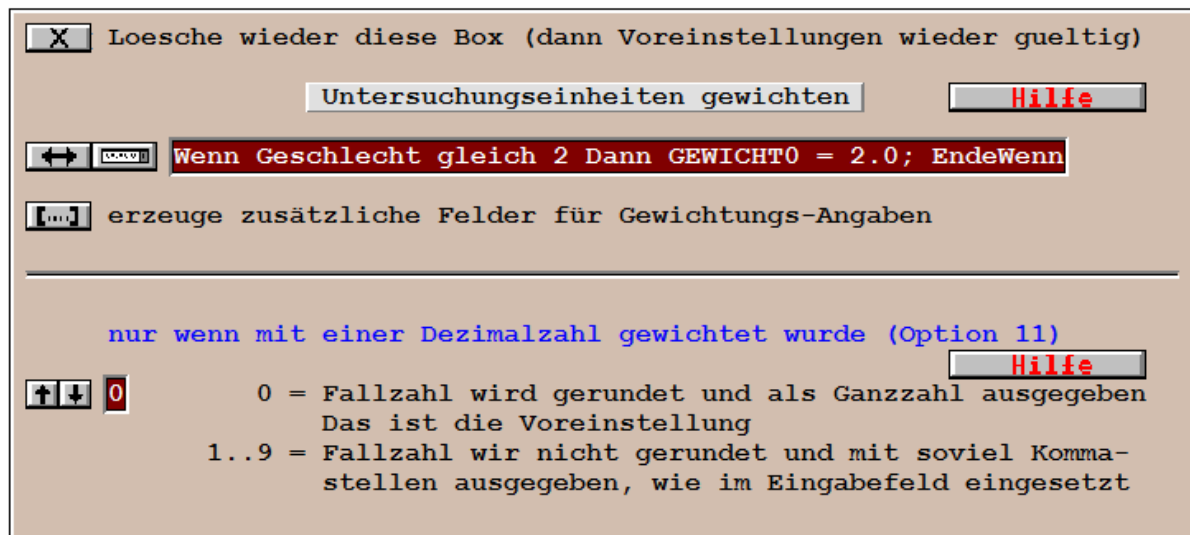
### **P4.5 Eingabe mit selbstgeschriebenem ALMO-Programm mit Optionen**

Kapitel entfernt. Ist in Handbuch "Teil3a: Grundlegende Verfahren" enthalten

### **P4.6 Gewichtung**



Wird die Optionsbox geöffnet, dann sieht man folgendes:



Wenn Sie auf den 2. Knopf vor dem Eingabefeld klicken, dann sehen Sie folgendes:



Wenn Sie auf dem eingeblendeten Text klicken, dann wird dieser in das Eingabefeld eingesetzt. Sie müssen dann die Punkte entsprechend ersetzen.

Als Gewichtungsvariable wird die Almo-Variable

GEWICHT1

verwendet. Hier bei Almo-Programm 4 ausnahmsweise

GEWICHT0.

Wenn Sie keine Gewichtung verwenden wollen, dann lassen Sie die Optionsbox geschlossen

Es gibt 3 Möglichkeiten zu gewichten

### 1. Gewichtung mit einer Wenn-Anweisung

Wenn Sie z.B. in das Eingabefeld schreiben

```
Wenn V1 gleich 1 dann GEWICHT0 = 2.5; EndeWenn  
Wenn V1 gleich 2 dann GEWICHT0 = 3.0; EndeWenn
```

dann wird eine Untersuchungseinheit, die in V1 den Wert 1 besitzt, so behandelt, wie wenn sie aus 2.5 Untersuchungseinheiten bestehen würde und eine Untersuchungseinheit, die in V1 den Wert 2 besitzt, so wie wenn sie aus 3 Untersuchungseinheiten bestehen würde.

#### BEACHTEN:

Dezimalzahlen müssen in Almo mit Punkt als Dezimalzeichen geschrieben werden

V1 könnte beispielsweise das Geschlecht von Untersuchungspersonen sein. Die Ausprägung 1 wäre dann "männlich" und 2 "weiblich".

Das bedeutet an einem Beispiel gezeigt folgendes: Die (ungewichteten) Häufigkeiten der auszählenden Variablen V7,8,9 seien für die Personen, die in V1 den Wert 1 besitzen, folgende

V7	220
V8	80
V9	30

Durch die Gewichtung werden diese Häufigkeiten um den Faktor 2.5 grösser, also zu

V7	550
V8	200
V9	75

Für die Personen, die in V1 den Wert 2 besitzen sind die (ungewichteten) Häufigkeiten folgende:

V7	110
V8	60
V9	40

Durch die Gewichtung werden diese Häufigkeiten um den Faktor 3 grösser, also zu

V7	330
V8	180
V9	120

Zusammen genommen entstehen also diese Häufigkeiten

V7	880
V8	380
V9	195

Dies gilt natürlich nur, wenn V1 ausschliesslich die beiden Ausprägungen 1 und 2 besitzt.

Sollten jedoch auch Untersuchungsobjekte vorhanden sein, die in V1 den Wert 3 oder irgend einen anderen Wert besitzen, dann müssen deren (ungewichtete) Häufigkeiten in V7,8,9 mit hinzuaddiert werden.

## **2. Gewichtung mit einer Variablen**

Gelegentlich befindet sich in der Datei eine Variable, die Gewichtungszahlen enthält - z.B. V48. In diesem Fall schreiben Sie in das Eingabefeld

```
GEWICHT0 = V48;
```

Dann wird eine Untersuchungseinheit in allen auszuzählenden Variablen mit der Zahl gewichtet, die für diese Untersuchungseinheit in V48 gefunden wird.

Bei Stichproben-Untersuchungen sind häufig manche Untergruppen über- oder unterrepräsentiert. In einer speziellen Variablen wird dann jeder Person ein Ausgleichfaktor zugewiesen. Bei der bekannten Pisa-Studie wird z.B. jedem einzelnen befragten Schüler eine Gewichtung zugewiesen, die ihrerseits aus der Multiplikation mehrerer Gewichtungen entsteht. Dadurch sollen Stichproben-Verzerrungen hinsichtlich des Schultyps, des Geschlechts, des Alters usw. ausgeglichen werden.

## **3. Gewichtung mit einer Konstanten**

Wenn Sie wollen, dann können Sie alle Fälle gleich gewichten, was nicht viel Sinn macht. Sie schreiben z.B.

```
GEWICHT0 = 2.5;
```

Dadurch wird jede einzelne Untersuchungseinheit 2.5 Mal in die Analyse aufgenommen.

## **4. Gewichtung mit einer Konstanten zur Hochrechnung**

Mit einer Konstanten zu gewichten macht jedoch Sinn, wenn Sie von der Stichprobe auf die Gesamtheit hochrechnen wollen. Ihre Stichprobe macht z.B. 1/90 -tel der Gesamtheit aus. Dann verwenden Sie

```
Gewicht0=90;
```

## **5. Reihenfolge in der Almo Daten manipuliert**

BEACHTEN: Die Reihenfolge in der Almo Daten manipuliert ist folgende:

1. Schritt: Zuerst werden die in der Optionsbox "Ein- und Ausschliessen von Untersuchungseinheiten" enthaltenen Anweisungen ausgeführt.
2. Schritt: Dann werden die in der Optionsbox "Umkodierungen und Kein-Wert-Angaben" enthaltenen Anweisungen ausgeführt.
3. Schritt: Dann werden die in der Optionsbox "Ausreisser identifizieren" enthaltenen Anweisungen ausgeführt.  
Diese Option wird nicht in allen Programm-Masken angeboten.
4. Schritt: Zum Schluss werden dann die in der Optionsbox "Untersuchungseinheiten gewichten"

enthaltenen Anweisungen ausgeführt.  
Diese Option wird nicht in allen Programm-Masken angeboten.

*Kurz:* Zuerst Ein- und Ausschliessen,  
dann Umkodieren,  
dann Ausreisser identifizieren und behandeln,  
dann gewichten

Beispiel: Werden Variable umkodiert, so geschieht das im Rechenverlauf vor der Gewichtung. Wenn Also die Gewichtungsbedingung einliest und verarbeitet, dann sind die Variablen bereits umkodiert.

### **Ergänzung zur Gewichtung mit der Wenn-Anweisung**

Sehr häufig wird man spezifische Untergruppen gewichten wollen. Sie können z.B. alle Männer doppelt zählen (!!). Sie schreiben dann

```
Wenn Geschlecht gleich 1 dann GEWICHT0 = 2.0; EndeWenn
```

(Geschlecht=1 sei der Code für die Männer)

Oder ein anderes Beispiel:

```
Wenn Beruf gleich 1 dann GEWICHT0 = 0.5 ; EndeWenn
```

```
Wenn Beruf gleich 2 dann GEWICHT0 = 0.75; EndeWenn
```

```
Wenn Beruf gleich 3 dann GEWICHT0 = 1.25; EndeWenn
```

(wobei: 1=Arbeiter, 2=Angestellte, 3=Selbständige)

Sie müssen also zwischen WENN und DANN einen logischen Ausdruck schreiben. Möglich sind z.B. folgende Anweisungen

```
WENN V5 gleich 7 DANN GEWICHT0 = 2.0; EndeWenn
```

```
WENN V6 groesser 8 DANN GEWICHT0 = 2.0; EndeWenn
```

```
WENN V7 kleiner 0.5 DANN GEWICHT0 = 2.0; EndeWenn
```

```
WENN V8 nichtgleich V9 DANN GEWICHT0 = 2.0; EndeWenn
```

```
WENN V10 groessergleich 2.5 DANN GEWICHT0 = 2.0; EndeWenn
```

```
WENN V11 kleingleich 2 DANN GEWICHT0 = 2.0; EndeWenn
```

Anstelle der Worte "gleich", "groesser" usw. sind auch die üblichen mathematischen Symbole möglich, also:

```
= > < ~= >= <=
      |
      |
      nichtgleich
```

Auch UND sowie ODER sind möglich. Beispiel

```
WENN V5 gleich 7 UND V10 groesser V20 DANN
GEWICHT0 = 2.0; EndeWenn
```

```
WENN V6 kleiner V5 ODER V12 nichtgleich 2 DANN
GEWICHT0 = 2.0; EndeWenn
```

Zur WENN...DANN -Anweisung siehe Also-Handbuch, Teil 2, Bedienungsanleitung Abschnitt 29 - 31.

## ***P4.7 Eingabe von Verteilungen***

Gelegentlich kommt es vor, daß man über eine Verteilung schon verfügt, für die man nun einige Koeffizienten berechnen möchte, die im Rahmen von Programm 4 ermittelt werden. Man möchte beispielsweise für eine schon gegebene Verteilung einen Normalverteilungstest durchführen.

### **P4.7.1 Eingabe mit Maskenprogramm Prog04m3.Msk**

Prog04m3.Msk  
Häufigkeitsverteilung  
mit Eingabe einer fertigen Tabellen

Es kann nur 1 Tabelle eingegeben werden !!

Für die fertige Tabelle sollen die Koeffizienten und Test gerechnet werden, die im Rahmen des ALMO-Programms 4 errechenbar sind. Beispielsweise soll folgende Tabelle eingelesen werden

Lebensalter	Häufigkeit
21	4
22	14
24	24
25	32
26	18
27	9

Für diese Häufigkeitstabelle können u.a. berechnet werden:

- Vertrauensintervalle
- arithmetisches, harmonisches, geometrisches Mittel
- Median, Quartile
- Standardabweichung
- Schiefe und Exzess der Normalverteilung
- Chi-Quadrat-Anpassungstest für Normalverteilung und Gleichverteilung.
- Kolmogorov-Smirnov-Einstichprobentest für Normalverteilung und Gleichverteilung.
- Binomialtest.
- Transformation der Daten auf Normalverteilung oder zu Prozentrangwerten

Grafik: Balkendiagramme

Was ist ein Kurzprogramm ? --> Hilfe  
Bedienung --> Hilfe

Name für die auszuzählende Variable U1

1 ↔ Name 1=Alter:21 Jahre,22,23,24,25,26;

Zahl der Zeilen der Tabelle hinter diesem Kurzprogramm

2 ↔ 6

Optionen: Normal- und Gleichverteilungstest

3 ↔ U1      Normalverteilungstest für U1  
Schiefe, Exzess

↔ U1      Gleichverteilungstest für U1

↔ U1      U1 ist ganzzahlig. Almo rechnet dann den  
Chi-Quadrat-Anpassungstest  
Besitzt U1 Dezimalwerte, dann löschen Sie  
den Eintrag durch Klick auf den Raus-Rein-  
Knopf. Almo rechnet dann den Kolmogorov-  
Smirnov-Test



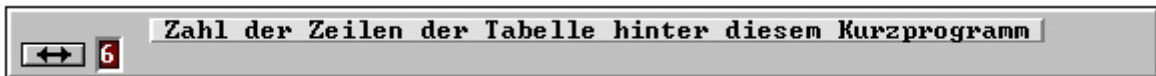
## P4.7.2 Erläuterungen zu den Dialogboxen

**Eingabebox 1:** Name für die auszuzählende Variable V1



Sie können einen Variablennamen und (hinter dem Doppelpunkt) Ausprägungsnamen geben. Wenn Sie dies nicht wollen, dann löschen Sie das Editfeld durch Klick auf den Knopf mit dem doppelköpfigen Pfeil.

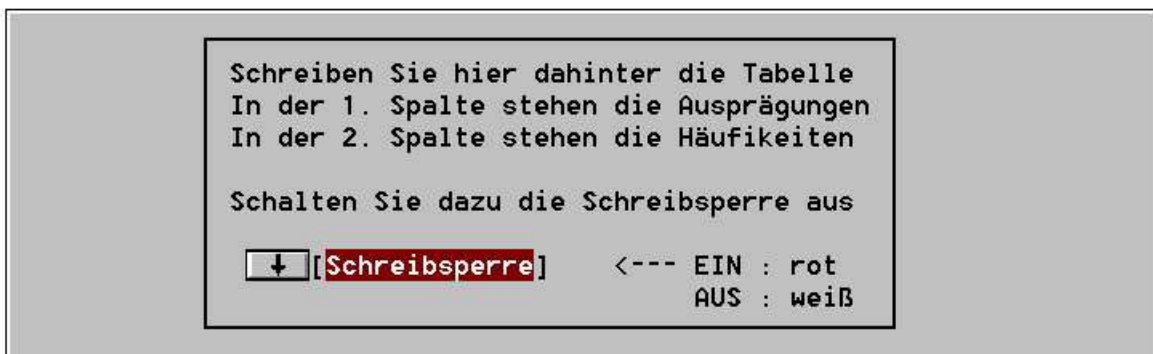
**Eingabebox 2:** Zahl der Zeilen der Tabelle



Geben Sie an wieviel Zeilen die Tabelle besitzt, die eingelesen und analysiert werden soll

**Eingabebox 3 bis 5** sind weitgehend gleich mit denen aus dem Maskenprogramm Prog04m2.Msk, Eingabebox 12.

**Eingabebox 6:** Schreibsperre



```
21 4
22 14
23 24
24 32
25 18
26 9
```

Hinter Box 6 wird die Tabelle geschrieben. Sie besteht prinzipiell aus 2 Spalten. In der 1. Spalte stehen die Ausprägungsnummern der Variablen und in der 2. Spalte die Häufigkeiten. In unserem Beispiel steht in der 1. Spalte das Lebensalter in Jahren. Die Zahlen müssen nicht notwendigerweise fortlaufend sein. In unserem Beispiel fehlen z.B. die 23-Jährigen.

Um hinter Box 6 überhaupt schreiben, löschen oder ändern zu können, müssen Sie durch Klick auf den Knopf mit Pfeil nach unten die Schreibsperre ausschalten.

## P4.8 Erläuterungen zu den Optionen

in den Maskenprogrammen Prog04m2, Prog04m3

### P4.8.1 Vertrauensintervalle

Beim Maskenprogramm Prog04m2 geben Sie, wie in P4.2.2 dargestellt, in der Box "Weitere Optionen" die Variable an, für die Vertrauensintervall ermittelt werden sollen. Dabei ist beim Maskenprogramm ein Konfidenzniveau von 95% voreingestellt.

Für die angegebenen Variable werden die Vertrauensintervalle bei einem Konfidenzniveau von 95% ermittelt. Die Vertrauensintervalle werden berechnet nach der Formel

$$p \pm z * \sqrt{\frac{p * (100 - p)}{n}}$$

p = Anteil der jeweiligen Ausprägung in %

n = Stichprobengröße

z = z-Wert der Normalverteilung (in der Voreinstellung =1.96)

Die obige Formel wird nur verwendet, wenn p\*n bzw. (100-p)\*n (der jeweils kleinere Ausdruck) größer/gleich 1000 ist. Ist diese Bedingung nicht gegeben dann wird das Binomialverteilungs-Modell verwendet. ALMO berechnet dann das exakte Clopper-Pearson-Intervall.

ALMO liefert folgendes Ergebnis:

Variable 3 Beruf

-----	Faelle	%	Vertrauensintervall		
			von	bis	% kumuliert
1 Arbeiter	20	25.00	15.51	34.49	25.00
2 Angestellter	42	52.50	41.56	63.44	77.50
3 Selbstaendiger	18	22.50	13.35	31.65	100.00
Summe	80	100			

### P4.8.2 Lage- und Streuungsmaße (arithmetisches, harmonische, geometrisches Mittel, Standardabweichung, Median, Quartile)

Beim Maskenprogramm Prog04m2 geben Sie, wie in P4.2.2 dargestellt, in der Box "Weitere Optionen" die Variable an, für die Sie Lage- und Streuungsmaße berechnen wollen.

Die Standardabweichung wird als deskriptive Statistik mit n (und nicht mit n-1) dividiert. Der Median und die Quartile werden nach der Methode der linearen Interpolation (siehe Clauß/Ebner, 1974, S.77) und nach der Methode des "mittleren Wertes" errechnet. Die beiden Methoden können zu verschiedenen Ergebnissen führen. Bei der linearen Interpolation wird ein Werteintervall von 1 vorausgesetzt. Dies gilt auch, wenn der Median bzw. die Quartile in einen Bereich fallen, der auf Grund nicht besetzter Ausprägungen scheinbar kleiner als 1 ist.

Bei der Methode des "mittleren Wertes" wird die Fallzahl n multipliziert mit 0.25 (1.Quartil), mit 0.50 (Median), mit 0.75 (3.Quartil). Das Ergebnis sei a. Ist a nicht ganzzahlig, dann ist das Quartil bzw. der Median  $x_{a+1}$ . Ist a ganzzahlig, dann ist das Quartil bzw. der Median die Intervallmitte zwischen  $x_a$  und  $x_{a+1}$ .

$x_a$  = a-ter Wert in der nach Größe sortierten Wertefolge.

$x_{a+1}$  = (a+1)-ter Wert

### P4.8.3 Normalverteilungstest, Schiefe und Exzeß

Beim Maskenprogramm Prog04m2 erfolgt die Eingabe wie in Abschnitt P4.2.2 dargestellt. Berechnet werden

- a. Schiefe der Normalverteilung
- b. Exzeß der Normalverteilung
- c. Chi-Quadrat-Anpassungstest auf Normalverteilung
- d. Kolmogorov-Smirnov Einstichprobentest für Normalverteilung.

#### *P4.8.3.1 Der Chi-Quadrat-Anpassungstest, Schiefe und Exzeß.*

ALMO unterstellt, daß sich die Ausprägungen der Variablen von ihrer Untergrenze bis zu ihrer Obergrenze in Einer-Schritten bewegen. Also z.B. 3 (Untergrenze) 4 5 6 7 8 (Obergrenze). Ist eine "erwartete Häufigkeit" kleiner als 5 dann wird sie mit ihrer Nachbar-Ausprägung zusammengefasst. Für den Chi-Quadrat-Test auf Schiefe und Exzeß werden keine Zusammenfassungen vorgenommen. Der Chi-Quadrat-Anpassungstest in Almo wurde von Gunter Geske programmiert.

Wir wollen folgendes Beispiel aus Peatman: Applied Statistics, 1963, S.257 rechnen. Wir verwenden das Maskenprogramm Prog04m3. Das Programm ist als Beispielsprogramm unter dem Namen „ChiQuad3.Alm“ vorhanden. Es wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme".

Wir haben in diesem Programm-Beispiel folgende bereits ausgezählte Verteilung eingelesen.

Wert	Haeufigkeit
-----	-----
1	3
2	2
3	6
4	21
5	27
6	42
7	35
8	23
9	17
10	6
11	1

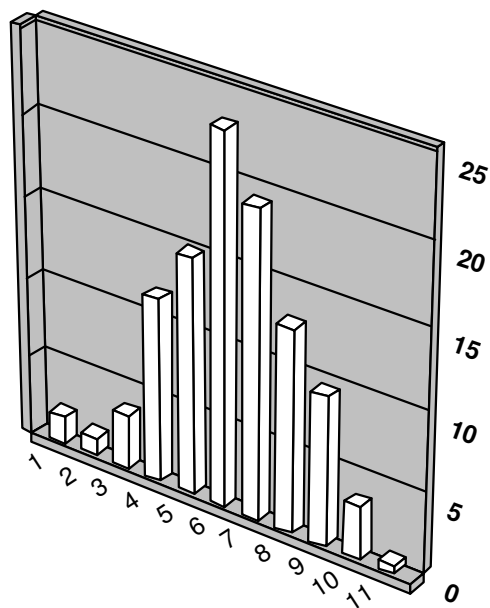
In Abschnitt P4.7 wurde erläutert, wie schon ausgezählte Verteilungen eingelesen werden.

Almo liefert folgende Ergebnisse:

Variable 1

	Faelle	%	% kumuliert
1	3	1.64	1.64
2	2	1.09	2.73
3	6	3.28	6.01
4	21	11.48	17.49
5	27	14.75	32.24
6	42	22.95	55.19
7	35	19.13	74.32
8	23	12.57	86.89
9	17	9.29	96.17
10	6	3.28	99.45
11	1	0.55	100.00
Summe	183	100	

ALMO liefert uns folgende Grafik, aus der optisch ersichtlich wird, daß wahrscheinlich eine Normalverteilung vorliegen wird.



arithmetisches Mittel	6.278689	
Standardabweichung	1.897348	(im Nenner: n)
Standardabweichung	1.902553	(im Nenner: n-1)
Stand.fehler arithm. Mittels	0.140641	
Variabilitaetskoeffizient	30.301759	
(mit 100 multipliziert)		

geometrisches Mittel            5.915587  
 harmonisches Mittel            5.365814

	mittlerer Wert bzw. Intervallmitte	linear interpoliert mit Wertintervall=1
1.Quartil	5.000000	5.009259
Median	6.000000	6.273810
3.Quartil	8.000000	7.554348
Quartilsabstand (mit 2 dividiert)	1.500000	1.272544

Schiefe-Test

-----  
 Schiefe                            -0.180716  
 Standardfehler                    0.181071  
 t-Wert                             0.998038  
 Signifikanz (1-p)\*100            68.043507

\*\*\*\*\* **Erläuterung:** Die Schiefe ist mit 68.3315 nicht signifikant. Der üblicherweise geforderte Wert von 95% wird nicht erreicht. Die Verteilung ist nicht schief

Exzess-Test

-----  
 Exzess (-3.0)                    -0.029349  
 Standardfehler                    0.362143  
 t-Wert                             0.081043  
 Signifikanz (1-p)\*100            6.663211

\*\*\*\*\* **Erläuterung:** Der Exzeß ist nicht signifikant. Die ermittelte Signifikanz liegt weit unterhalb von 95%. Die Verteilung hat eine der Normalverteilung entsprechende Gipfelbreite.

Chi-Quadrat-Test auf Normalverteilung

-----  
 \*\*\*\*\* MITTEILUNG NR. 201  
 Aus ermittelter Haeufigkeitsverteilung wird fuer Test uebernommen:  
 Mittelwert= 6.2787

\*\*\*\*\* MITTEILUNG NR. 201  
 Aus ermittelter Haeufigkeitsverteilung wird fuer Test uebernommen:  
 Standardabweichung= 1.9026

	H a e u f i g k e i t e n	
	erwartete	beobachtete
1	0.817	3
2	3.060	2
3	8.692	6
4	18.730	21
5	30.615	27
6	37.963	42
7	35.712	35
8	25.485	23
9	13.796	17
10	5.666	6
11	1.765	1

Chi-Quadrat                            2.3528

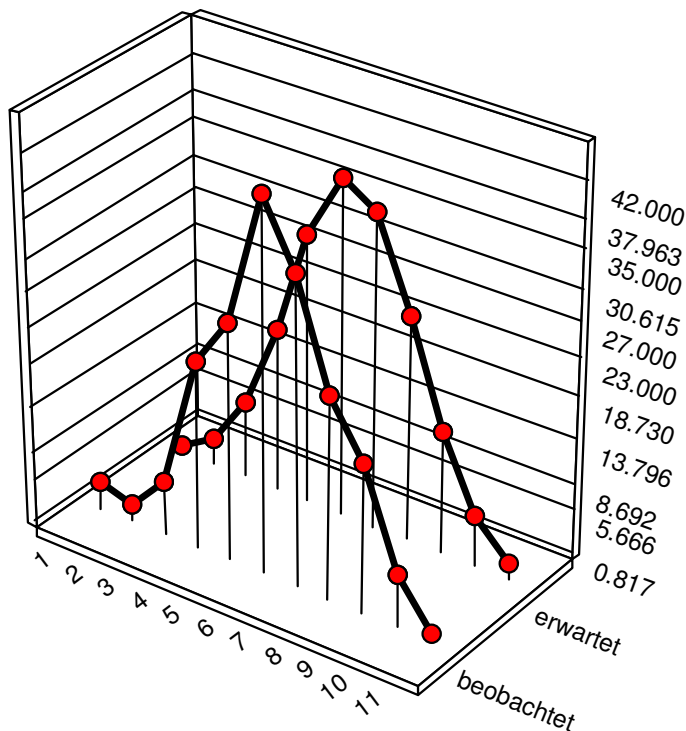
Freiheitsgrade	5
Wahrscheinlichkeit $(1-p) \cdot 100$	
dass beobachtete Verteilung	
nicht Normalverteilung	
entspricht	19.9802 %

Die Normalverteilung kann angenommen werden, wenn obiger Wert kleiner ca. 95 % ist

\*\*\* Durch Zusammenfassungen reduzierte sich die Zahl der Freiheitsgrade um 3.

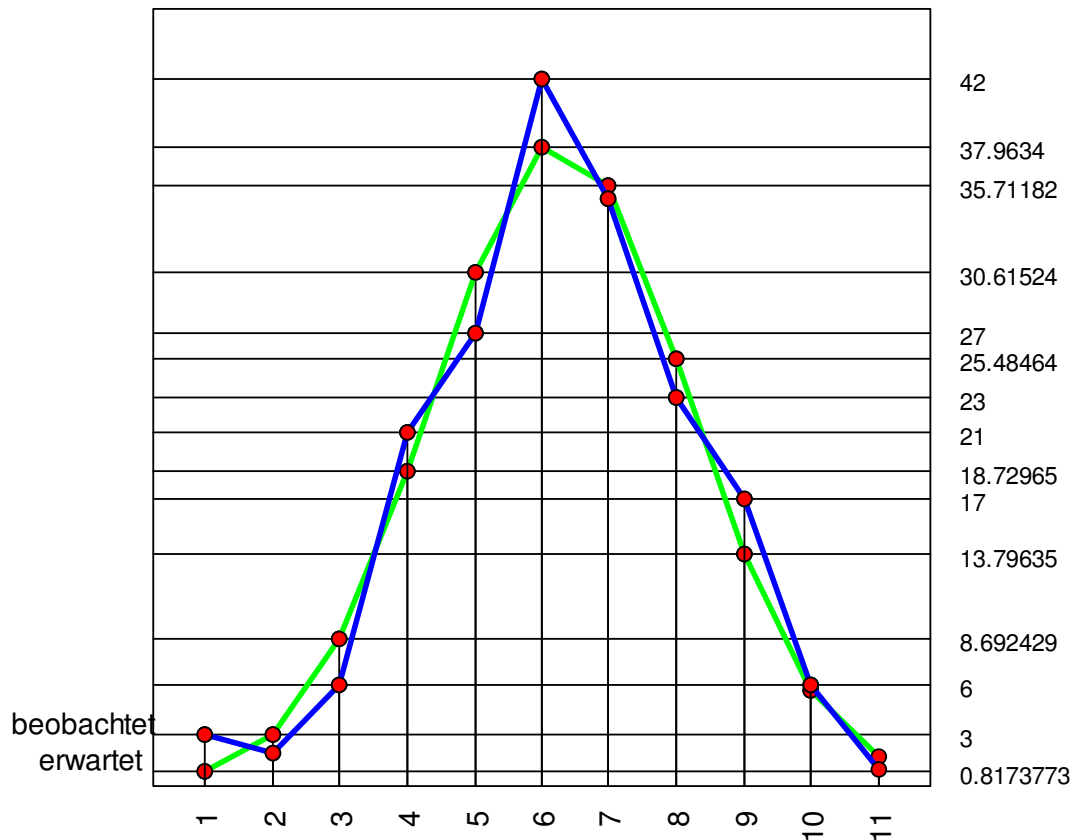
\*\*\*\*\* **Erläuterung:** Die Wahrscheinlichkeit, dass keine Normalverteilung vorliegt, ist mit 19,851% weit unterhalb der üblichen Signifikanzgrenze von 95%. Die Verteilung ist normalverteilt.

ALMO zeichnet folgendes Liniendiagramm:



Das Diagramm kann auch 2-dimensional dargestellt werden, so dass die Abweichung der beobachteten Werte von den erwarteten besser sichtbar wird.

Erwartete und beobachtete  
Häufigkeiten  
grüne Linie: erwartet  
blaue Linie: beobachtet



### Exkurs: Berechnung von Schiefe und Exzeß

Die Schiefe  $S$  ist ein Maß dafür, inwieweit die gefundene Verteilung symmetrisch oder asymmetrisch ist. Die Normalverteilung hat die Schiefe  $S = 0$ . Abweichungen von der Schiefe  $S = 0$  bei ermittelten Häufigkeitsverteilungen können mittels des t-Tests auf ihre Signifikanz überprüft werden. Die Schiefe ist aus der Gleichung

$$S = \frac{\sum x^3}{N \cdot s_x^3}$$

zu bestimmen. Darin bedeutet:

$\sum x^3$  = die Summe der dritten Potenzen aller Abweichungen der Beobachtungswerte vom arithmetischen Mittelwert

$N$  = Anzahl der Befragten

$s_x$  = Standardabweichung der Häufigkeitsverteilung

Die notwendige Testgröße  $t$  wird mittels der Gleichung

$$t = \frac{S}{s_s}$$

mit  $df = N - 1$  Freiheitsgraden berechnet.

Die noch unbekannte Größe  $s_s$  ist der sogenannte Standardfehler der Schiefe  $S$ , der durch

$$s_s = \sqrt{\frac{6}{N}}$$

definiert ist.

Nun kann die gefundene Häufigkeitsverteilung nicht nur symmetrisch sein; bei vollkommener Symmetrie kann der Gipfel der Verteilung unterschiedlich breit oder spitz ausgeprägt auftreten. Das Kriterium für die Gipfform ist der Exzeß  $E$ , der für die ideale Normalverteilung den Wert  $E = + 3,0$  hat. Der Exzeß kann aus der Beziehung

$$E = \frac{\sum x^4}{N \cdot s_x^4}$$

ermittelt werden, die Testgröße mit der Formel

$$t = \frac{E - 3}{s_E}$$

mit  $df = N - 1$  Freiheitsgraden

Die in den Gleichungen auftretenden Größen entsprechen sinngemäß den bei der Schiefe verwendeten. Der Standardfehler des Exzesses wird allerdings mit der Formel bestimmt:

$$s_E = 2\sqrt{\frac{6}{N}}$$

Die Testgröße Chi-Quadrat wird mit folgender Summenformel bestimmt:

$$\text{ChiQuadrat} = \sum \frac{(f_o - f_e)^2}{f_e}$$

mit  $df = NZ - 3$  Freiheitsgraden

$NZ$  = Zahl der Ausprägungen der Variablen

$f_o$  = empirische Häufigkeit je Ausprägung der Variablen

$f_e$  = erwartete Häufigkeit

Die Erwartungswerte  $f_e$  werden üblicherweise durch Integration der Normalverteilung zwischen den Klassengrenzen berechnet. Auf dieses manchmal sehr rechenaufwendige Verfahren kann verzichtet werden, wenn die Beziehung

$$f_e = \frac{i \cdot N}{s_x} \cdot y$$

benutzt wird. Hierin sind noch die Variablen  $i$  und  $y$  zu erläutern:

$i$  = Breite des Klassenintervalls

$y$  = Ordinate der standardisierten Normalverteilung für die Klassenmitte

$y$  wird nach der Formel für die Normalverteilung ermittelt.

$$y = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$$

wobei

$z_i = (x_i - M) / s_x$

$x_i$  = Variablenwert für Intervall

$M$  = Gesamtintervall

### P4.8.3.2 Der Kolmogorov-Smirnov Einstichprobentest (KS-Test) für die Normalverteilung.

Mit ihm wird die Normalverteilungs-Hypothese für Dezimalwert-Variable überprüft. Der KS-Test ist nur für stetige, intervallskalierte Variable brauchbar.

Jeder Messwert (Ausprägung) sollte nur 1 Mal besetzt sein. Auf Grund von Messungenauigkeiten dürfen auch einige Messwerte doppelt oder mehrfach besetzt sein. In diesem Falle ist der KS-Test von höherer Güte als der Chi-Quadrat-Anpassungstest. Vor allem ist er auch noch bei nur wenigen Messwerten durchführbar. Er "bricht zusammen" wenn einzelne Messwerte zu häufig besetzt sind. Die Probleme des KS-Test werden kurz bei Büning/ Trenkler (1978, S. 91) diskutiert. Wir rechnen ein Beispiel aus Büning/ Trenkler (1978, S. 86). Wir verwenden das Maskenprogramm Prog04m3. Das Programm ist als Beispielsprogramm unter dem Namen „Kolmog\_4.Alm“ vorhanden. Es wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme".

Wir haben in diesem Programm-Beispiel eine bereits ausgezählte Verteilung eingelesen.

ALMO liefert folgende Ausgabe (gekürzt):

Variable 1 Benzinverbrauch

-----

	Faelle	%	% kumuliert
11.5	1	10.00	10.00
11.8	1	10.00	20.00
12	1	10.00	30.00
12.4	1	10.00	40.00
12.5	1	10.00	50.00
12.6	1	10.00	60.00
12.8	1	10.00	70.00
12.9	1	10.00	80.00
13	1	10.00	90.00
13.2	1	10.00	100.00
Summe	10	100	
arithmetisches Mittel		12.470000	
Standardabweichung		0.523546	(im Nenner: n)
Standardabweichung		0.551866	(im Nenner: n-1)
Stand.fehler arithm. Mittels		0.174515	
Variabilitaetskoeffizient (mit 100 multipliziert)		4.425545	
geometrisches Mittel		12.458852	
harmonisches Mittel		12.447551	
		mittlerer Wert bzw.Intervallmitte	linear interpoliert mit Wertintervall=1
1.Quartil		12.000000	12.000000
Median		12.550000	13.000000
3.Quartil		12.900000	12.900000
Quartilsabstand (mit 2 dividiert)		0.450000	0.450000
Schiefe-Test			
-----			
Schiefe		-0.393899	
Standardfehler		0.774597	
t-Wert		0.508521	
Signifikanz (1-p)*100		37.864878	

Exzess-Test

Exzess (-3.0)	-1.334472
Standardfehler	1.549193
t-Wert	0.861398
Signifikanz (1-p)*100	58.965375

Kolmogorov-Smirnov-Test auf Normalverteilung

Ausprägung	kumulierte relative Haeufigkeiten			Differenz	
	erwartete	beobachtete linksseitig	beobachtete rechtsseitig	links	rechts
11.5	0.0394	0.1000	0.0000	-0.0606	0.0394
11.8	0.1124	0.2000	0.1000	-0.0876	0.0124
12	0.1972	0.3000	0.2000	-0.1028	-0.0028
12.4	0.4490	0.4000	0.3000	0.0490	0.1490
12.5	0.5228	0.5000	0.4000	0.0228	0.1228
12.6	0.5924	0.6000	0.5000	-0.0076	0.0924
12.8	0.7252	0.7000	0.6000	0.0252	0.1252
12.9	0.7820	0.8000	0.7000	-0.0180	0.0820
13	0.8316	0.9000	0.8000	-0.0684	0.0316
13.2	0.9070	1.0000	0.9000	-0.0930	0.0070

Absolute maximale Differenz 0.1490

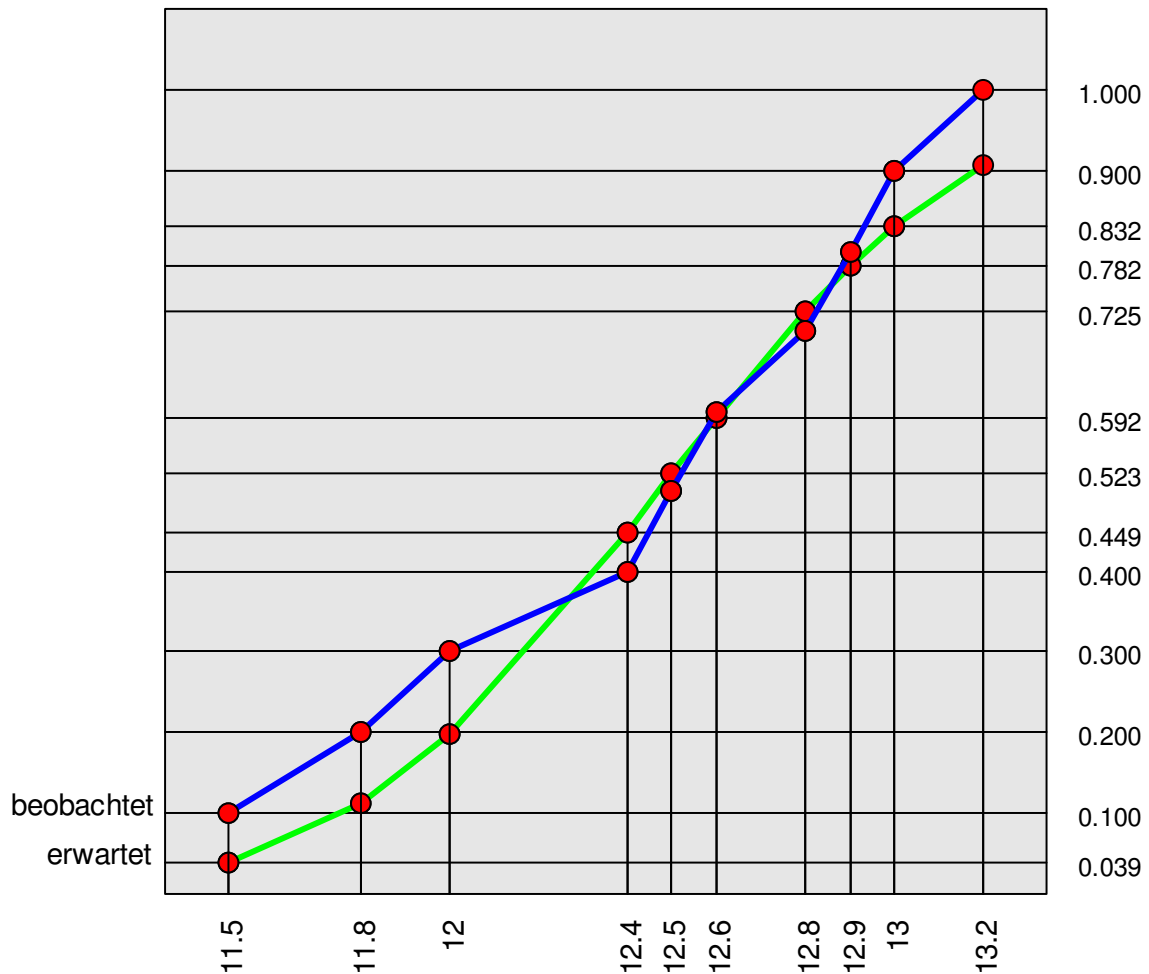
Wahrscheinlichkeit (1-p)\*100  
dass keine Normalverteilung 2.0607 %  
(zweiseitiger Test)

Die Normalverteilung kann  
angenommen werden, wenn obiger  
Wert kleiner ca. 95 % ist

**Ergebnis:** Die Hypothese, dass die empirische Verteilung einer Normalverteilung entspricht, kann also nicht verworfen werden, da die ausgegebene Wahrscheinlichkeit unter der (konventionellen) 95%-Grenze liegt. Die Verteilung ist normalverteilt.

ALMO liefert noch folgendes Liniendiagramm:

Erwartete und beobachtete  
kumulierte Häufigkeiten  
gruene Linie: erwartet  
blaue Linie: beobachtet



### Signifikanz p des Kolmogorov-Smirnov-Einstichproben-Test

In ALMO wird die Irrtumswahrscheinlichkeit p nach folgender Formel ermittelt (siehe dazu B.L. van der Waerden: Mathematische Statistik, 1971, Springer Verlag, S.72)

$$p = 2 \cdot (e^{-2 \cdot n \cdot D^2} - e^{-2.4 \cdot n \cdot D^2} + e^{-2.9 \cdot n \cdot D^2} - e^{-2.16 \cdot n \cdot D^2})$$

n = Zahl der Untersuchungseinheiten

D = max. Differenz

Der Benutzer versuche folgende Simulation mit normalverteilten Zufallszahlen durchzuführen: Das Programm ist als Beispielsprogramm unter dem Namen „Zufall2.Alm“ vorhanden. Dies ist ein Syntax-Programm.

```

#-----#
# Normalverteilte Zufallszahlen erzeugen und #
# mit Kolmogorov-Smirnov Einstichprobentest auf #
# Normalverteilung pruefen #
#-----#

Vereinbare
  Variable=10;
Anfang
Programm=4;
  Quantitative_V = V7; # V7 ist quantitativ und nicht gruppiert #
  Diverse_Werte 7 = 20; # V7 nimmt maximal 20 unterschiedliche Werte an #

  Normalverteilungstest= V7; # fuer V7 Normalverteilungstest.Da V7 quant. #
  # wird Kolmogorov-Smirnov gerechnet #
  Mittelwert 7 = KW; # Mittelwert und Stand.abwg. nicht vorgeben #
  Stand_abwg 7 = KW; # ALMO ermittelt sie aus empirischer Verteilung #
#
Ende_Programmparameter

Von H1=1; # 20 Datensaeetze erzeugen #
  V7(Zufall 0:1; Normal 10,2.5) # V7 bleibt intervallskaliert und #
  Gehe_in_Programm # wird nicht auf Ganzzahl gerundet #
Bis H1=20;

ENDE

```

## P4.8.4 Test auf Gleichverteilung

Beim Maskenprogramm Prog04m2 oder Prog04m3 erfolgt die Eingabe wie in Abschnitt P4.4.1 oder P4.7.1 dargestellt.

Werden die zu testenden Variablen als ganzzahlig angegeben, dann rechnet ALMO den Chi-Quadrat-Anpassungstest. Werden sie als Variable angegeben, die Dezimalwerte besitzen, dann wird der Kolmogorov-Smirnov Einstichprobentest (KS-Test) angewendet. Der KS-Test funktioniert hier, im Unterschied zum Normalverteilungstest nur, wenn die Ausprägungen mehrfach besetzt sind. Ist jede Ausprägung nur 1 Mal besetzt, dann muss man die Variable durch eine Umkodierungsanweisung gruppieren, z.B. so V1(0 Schritt 0.5 bis 10 = I)  
Betrachten wir ein Beispiel:

### P4.8.4.1 Kolmogorov-Smirnov Ein-Stichprobentest auf Gleichverteilung

Wir verwenden das Maskenprogramm Prog04m3.

Beispiel aus Siegel: Nichtparametrische statistische Methoden, 1987, S.49. Das Programm ist als Beispielsprogramm unter dem Namen "Kolmog\_3.alm" vorhanden. Es wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme".

ALMO liefert folgendes Ergebnis;

```

Variable 1
-----

```

Wert	Faelle	%	% kumuliert
2	1	10.00	10.00
4	5	50.00	60.00
5	4	40.00	100.00
Summe	10	100	

## Kolmogorov-Smirnov-Test auf Gleichverteilung

Ausprägung	kumulierte relative Häufigkeiten		Differenz
	erwartete	beobachtete	
1	0.2000	0.0000	0.2000
2	0.4000	0.1000	0.3000
3	0.6000	0.1000	0.5000
4	0.8000	0.6000	0.2000
5	1.0000	1.0000	0.0000

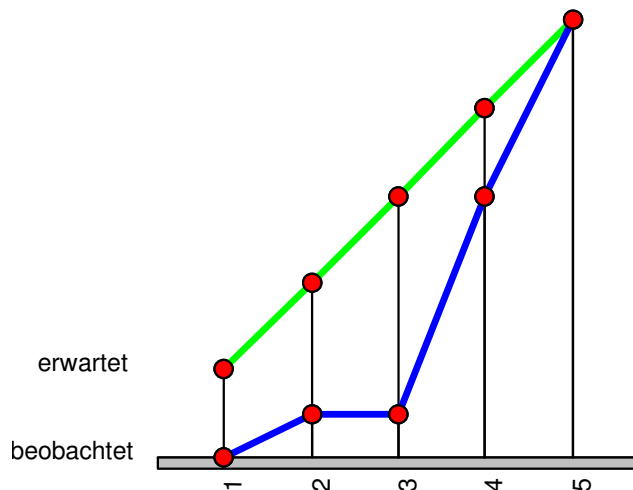
Absolute maximale Differenz 0.5000

Wahrscheinlichkeit  $(1-p) \cdot 100$   
 dass keine Gleichverteilung 98.6524 %  
 (zweiseitiger Test)

Die Gleichverteilung kann angenommen werden, wenn obiger Wert kleiner ca. 95 % ist.

Mit "Option 9=0;" kann die Ausgabe der kumulierten Häufigkeiten verhindert werden.

ALMO liefert noch folgende Grafik der kumulierten erwarteten und beobachteten Häufigkeiten.



Beachte: Die Kurve der kumulierten erwarteten Häufigkeiten ist in unserem Fall eine Gerade mit 45° Steigung. Die Kurve der kumulierten beobachteten Häufigkeiten weicht beträchtlich ab.

#### P4.8.4.2 Chi-Quadrat-Anpassungstest auf Gleichverteilung

Beispiel aus Siegel: Nichtparametrische statistische Methoden, 1987, S.44. Wir verwenden das Maskenprogramm Prog04m3. Das Programm ist als Beispielsprogramm unter dem Namen "Chiquad4.alm" vorhanden. Es wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme".

ALMO liefert folgendes Ergebnis:

Variable 1 positions

Wert	Faelle	%	% kumuliert
1	29	20.14	20.14
2	19	13.19	33.33
3	18	12.50	45.83
4	25	17.36	63.19
5	17	11.81	75.00
6	10	6.94	81.94
7	15	10.42	92.36
8	11	7.64	100.00
Summe	144	100	

Chi-Quadrat-Test auf Gleichverteilung

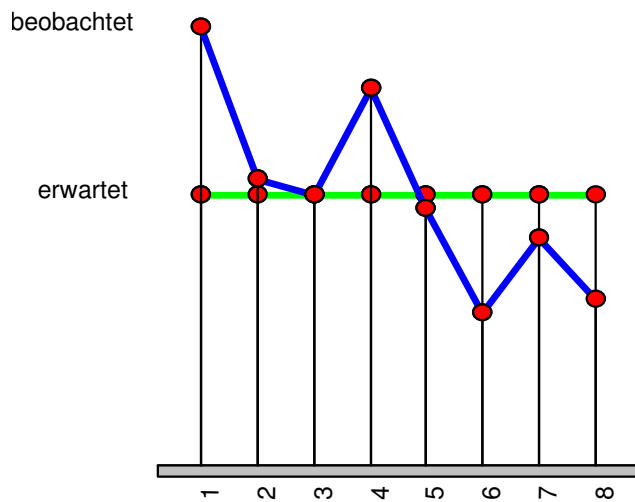
	H a e u f i g k e i t e n	
	erwartete	beobachtete
1	18.000	29
2	18.000	19
3	18.000	18
4	18.000	25
5	18.000	17
6	18.000	10
7	18.000	15
8	18.000	11

Chi-Quadrat 16.3333  
Freiheitsgrade 7

Wahrscheinlichkeit (1-p)\*100  
dass beobachtete Verteilung  
nicht Gleichverteilung  
entspricht 97.7213 %

Die Gleichverteilung kann  
angenommen werden, wenn obiger  
Wert kleiner ca. 95 % ist

ALMO liefert dann noch folgende Grafik der erwarteten und beobachteten Häufigkeiten:



#### P4.8.5 Chi-Quadrat-Test für beliebige Verteilungen mit Prog04m4

Das Maskenprogramm für diesen Test ist folgendes:

**Prog04m4.Msk**  
**Chi-Quadrat-Anpassungstest für beliebige Verteilungen**  
**mit vom Benutzer eingegebenen erwarteten Häufigkeiten**

**Beispiel:**

In einer Stichprobe wurde nachfolgende Altersverteilung angetroffen. Die Altersverteilung in der Grundgesamtheit ist bekannt. Die Werte sind folgende:

Alter	Alters- gruppe	tatsächliche Häufigkeiten also: Häufigkeit in Stichprobe		erwartete Häufigkeiten also: Häufigkeit in Gesamtheit ungerechnet auf n=900
		Prozent in Gesamtheit		
bis 20	1	40	6	54
21-30	2	120	10	90
31-40	3	240	28	252
41-50	4	340	35	315
51-60	5	140	17	153
über 61	6	20	4	36
		900	100 %	900

Die Frage lautet: Sind die Stichproben-Ergebnisse repräsentativ für die Grundgesamtheit.

Beachte: Der Chi-Quadrat-Anpassungstest kann nur mit Häufigkeiten, nicht mit Prozentwerten gerechnet werden. Es müssen also die %-Werte aus der Grundgesamtheit umgerechnet werden in Häufigkeiten. Dabei muss dann für die Grundgesamtheit das gleiche n wie für die Stichprobe angenommen werden

Was ist ein Kurzprogramm ? --> Hilfe  
 Bedienung --> Hilfe

Name für die auszuzählende Variable U1

1 ↔ Name 1=Alter:bis 20,21-30,31-40,41-50,51-60,über 61;

Zahl der Zeilen der Tabelle hinter diesem Programm

2 ↔ 6

Erwartete Häufigkeiten

3 ↔ 54,90,252,315,153,36

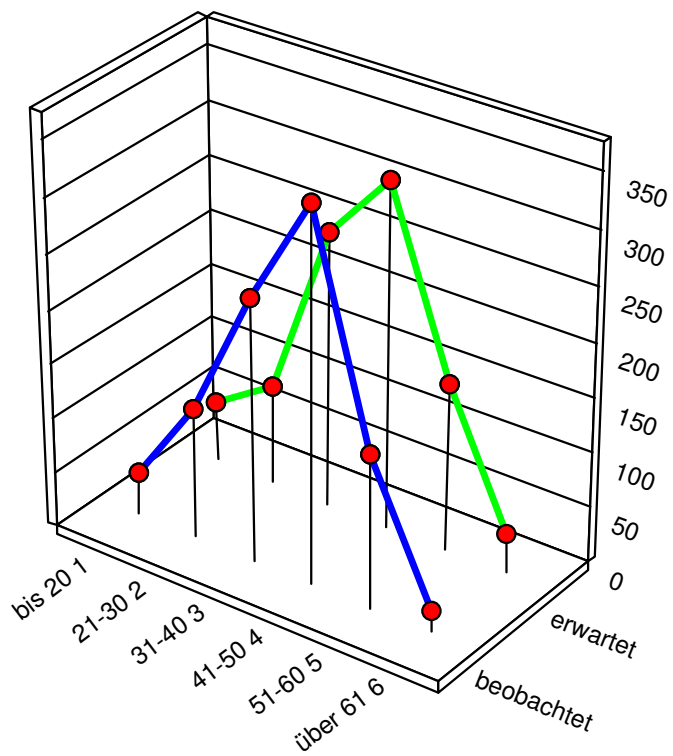
↔ |

↔ |

... erzeuge zusätzliche Felder für weitere erwartete Häufigkeiten



# Erwartete und beobachtete Haeufigkeiten



## P4.8.6 Binomialtest

Beim Maskenprogramm Prog04m2 oder Prog04m3 erfolgt die Eingabe wie in Abschnitt P4.4.1 oder P4.7.1 dargestellt.

Der Binomialtest wird selbstverständlich nur gerechnet, wenn die Variable dichotom ist. Der Benutzer muss also polytome Variable durch eine Umkodierungsanweisung dichotomisieren, z.B. so

V10(1:5=1; 6:10=2)

Wenn  $n \cdot p$ , bzw.  $n \cdot (1-p)$  größer 10, dann rechnet ALMO die Normalverteilungs-Approximation der Binomialverteilung. ( $n$ =Gesamtzahl der Untersuchungseinheiten,  $p$  = der in die Variable G eingegebene Anteilswert der 1. Ausprägung). Der Benutzer bemerkt dies daran, dass ALMO einen z-Wert ausgibt.

Betrachten wir ein Beispiel aus Lienert (1973,S.147). Wir verwenden das Maskenprogramm Prog04m3. Das Programm ist als Beispielsprogramm unter dem Namen "Binomial2.alm" vorhanden. Es wird gefunden durch Klick auf das Menü "Almo/Liste aller Almo-Programme".

ALMO liefert folgendes Ergebnis:

Variable 1 Geburt						
-----						
Wert	Faelle	%	Vertrauensintervall		% kumuliert	
			von	bis		
1 weiblich	3	23.08	8.18	50.26	23.08	
2 maennlich	10	76.92	49.74	91.82	100.00	
-----						
Summe	13	100				

Binomialtest (einseitig)

```

-----
***** MITTEILUNG NR. 201
folgender vom Benutzer eingegebene Wert wird fuer Test verwendet:
vorgegebener Anteilswert= 0.5000

Glieder des Binomialausdrucks
0      0.00012207
1      0.00158691
2      0.00952148
3      0.03491211
-----
p =    0.04614258

```

```

Der beobachtete Anteilswert von      23.0769 %
weicht ab vom
vorgegebenen Anteilswerts von      50.0000 %
mit einer Signifikanz
(1-p)*100 von                        95.3857 %

```

## P4.8.7 Transformation von Variablen auf Standard-Normalverteilung

Beim Maskenprogramm Prog04m2 oder Prog04m3 geben Sie in der Box "Weitere Optionen" die Variablen an, die auf Standardnormalverteilung transformiert werden sollen.

Beispiel: Die Variablen 6 und 8 werden

(1) auf Prozentrangwerte transformiert (siehe dazu auch P10.5) und sie werden

(2) in ihren Werten so transformiert, dass sie standardnormalverteilt sind (mit Mittelwert=0 und Standardabweichung=1).

Dies ist eine nicht-lineare Transformation, die z.B. die Korrelation von V6 und V8 mit anderen Variablen erheblich ändern kann. Der Forscher muss es mit seinem "Gewissen ausmachen", ob er einen solchen gravierenden Eingriff in seine Daten vornehmen will. ALMO liefert beispielsweise folgendes Ergebnis:

	kumulierte Haeufigkeit	Prozentrangwerte	auf Normalvertlg. transform. Werte (z-Werte)
1	10	0.0820	-1.3919
2	19	0.2377	-0.7140
3	29	0.3934	-0.2725
4	40	0.5656	0.1653
5	44	0.6885	0.4916
6	49	0.7623	0.7140
7	58	0.8770	1.1604
8	60	0.9672	1.8403
9	61	0.9918	2.4016

Es besteht nun die Möglichkeit in späteren Programmen durch eine Umkodierungsanweisung die gewonnenen normalverteilten z-Werte zu verwenden, etwa in folgender Weise:

V6 (1= -1.3919; 2= -0.7140; ..... 9=2.4016)

Nach dieser Umkodierung besitzt V6 einen Mittelwert von =0 und eine Standard-abweichung von=1. Zitat aus Lohninger: „Manchmal liest man in der Literatur, dass die z-Werte normalverteilt sind. Das ist falsch - die z-Transformation ändert nichts an der Form der Verteilung, es werden lediglich der Mittelwert und die Standardabweichung verändert. Bildlich gesprochen: Die Verteilung wird entlang der x-Achse verschoben und gedehnt bzw. gestaucht.“

Durch eine zweite (lineare) Umformung ist es möglich, die z-Werte zu ändern, etwa so:

$$V6 = 2 * V6 + 5;$$

Die z-Werte werden hierdurch positiv.

Als zweite Umformung wäre auch die sogenannte T-Transformation möglich. Sie wird erreicht durch

$$V6 = 50 - 10 * V6;$$

Siehe: Lohse, Ludwig, Röhr(1982,S.89)

Siehe Hans Lohninger: [http://www.statistics4u.info/fundstat\\_germ/ee\\_ztransform.html](http://www.statistics4u.info/fundstat_germ/ee_ztransform.html)

## **P4.10 Auszählung von Variablen mit sehr vielen Ausprägungen**

Eine Variable V10 (z.B. das Jahreskommen) habe einen möglichen Wertebereich von 0 bis 32 000.

Es ist sicherlich sinnvoll, die 32 000 Ausprägungen dieser Variablen durch eine Umkodierungsanweisung zu einer vertretbaren Zahl von Ausprägungen zusammenzufassen, z.B. so:

V10 (0 Schritt 4000 bis 32000 = I)

V10 besitzt dann 8 Ausprägungen.

## **P5 Lage- und Streuungsmaße**

Programm 5 ermittelt die Mittelwerte, Standardabweichungen, Untergrenzen und Obergrenzen der Variablen. Bei ordinalen Variablen wird anstelle des Mittelwertes und der Standardabweichung der Median und der halbe Quartilsabstand (bzw. optional die Standardabweichung nach Kendall's tau-b Kalkül) errechnet. Das Programm wurde von Kurt Holm geschrieben. Die Programmteile, die die Standardabweichung der ordinalen Variablen ermitteln, wurden von H. Denz verfasst.

### ***P5.01 Eingabe in Programm 5 mit Maskenprogramm***

**Prog05m1.MSK**  
 Einige wichtige Kennwerte der Variablen ermitteln

Die Kennwerte sind:

1. Werte-Unter- und Obergrenzen der Variablen
2. Zahl der diversen Werte, die die Variablen besitzen
3. Zahl der Kein-Wert-Fälle (fehlende Werte)
4. Mittelwerte der Variablen (auch Median, häufigster Wert)
5. Streuungen der Variablen

Was ist ein Kurzprogramm ? --> Hilfe  
 Bedienung --> Hilfe

1 Hilfe

**Speicher fuer x Variable**

Vereinbare Variable= 20 ;

2 Hilfe

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

3 Hilfe

**Datei der Variablennamen**

"C:\Almo7\Testdat\Uarnamen.nam"

zeige      zeige = Namensdatei in Output zeigen  
 leer = nicht

4 Hilfe

**Freie Namensfelder**

-

erzeuge zusätzliche Namensfelder

5 Hilfe

**Datei aus der gelesen wird** Hilfe

"C:\Almo7\TESTIDAT\Testdat.fre"      bei Datei-Problemen

frei      Format der Daten Hilfe

U1:20      der Datensatz enthält diese Variablen  
 Bei Format DIREKT schreiben Sie: alle\_U

6 Hilfe

**Wenn Dateiformat FIX oder Nicht-Standard-FREI**

7 Hilfe

**zu analysierende Variable**

quantitative Variable

Leistung,Alter,Einkommen,Kinderzahl

---

nominale Variable

U1:4,9:19

---

ordinale Variable

Bewertung

8 Hilfe

Option: Werte-Unter- und Obergrenzen bzw. "Diverse Werte"

9 Hilfe

Option: Umkodierungen und Kein-Wert-Angaben

## P5.01.1 Erläuterungen zu den Eingabeboxen

Eingabebox 1 bis Eingabebox 6 und Eingabebox 9:

Siehe Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.1, P0.3, P0.4, P0.5.

**Eingabebox 7: Zu analysierende Variable**

zu analysierende Variable Hilfe

quantitative Variable

↔ □□ Leistung, Alter, Einkommen, Kinderzahl

---

nominale Variable

↔ □□ U1:4, 9:19

---

ordinale Variable

↔ □□ Bewertung

Almo unterscheidet die 3 Messniveaus: quantitativ, ordinal und nominal und rechnet dafür teilweise unterschiedliche Koeffizienten. Siehe dazu P19.0.1. Sie können aber auch ein und dieselbe Variable als quantitativ und als nominal und als ordinal angeben. Sie erhalten dann für diese Variable die Koeffizienten, die Almo für diese 3 Messniveaus ermittelt.

Sie können die Analyse-Variablen „von Hand“ in die Eingabefelder schreiben oder Sie klicken auf den Knopf mit den 2 kleinen symbolischen Fenstern. Almo öffnet dann die Dialogbox „Variable für Analyse auswählen“. In dieser können Sie die Variable, die in die Eingabefelder eingeschrieben werden sollen per Mausklick selektieren. Siehe die ausführliche Beschreibung dieser Dialogbox in P0.11.

**Eingabebox 8: Unter- und Obergrenzen oder diverse Werte**

↓ Option: Werte-Unter- und Obergrenzen bzw. "Diverse Werte"

Optionsbox geöffnet:

↓ Loesche wieder diese Box

Unter- und Obergrenzen  
bzw.  
"diverse Werte" ermitteln

↑ ↓ 0

0 = für alle Variable  
werden die Unter- und Obergrenzen ermittelt

1 = für die nominalen und ordinalen Variablen  
werden die Unter- und Obergrenzen ermittelt  
für die quantitativen Variablen wird  
die Zahl der diversen Werte ermittelt

2 = für alle Variablen wird  
die Zahl der diversen Werte ermittelt

Beachte: Bei 2 wird für nominale Variable  
zusätzlich der "häufigste Wert" ermittelt

### ***P5.01.1.1 Zum Begriff der „diversen Werte“***

Betrachten wir ein Beispiel: Die quantitative Variable V10 besitzt über alle Datensätze hinweg die Werte

1.2

1.4

2.1

2.9

3.7

Die Zahl der diversen Werte ist dann 5. Diese Zahl der diversen Werte wird dann in die Obergrenze gegeben, also ist  $OG_{10} = 5$ . Die Untergrenze wird auf eins gesetzt, also ist  $UG_{10} = 1$ .

Sie können in das Eingabefeld 0 oder 1 oder 2 eingeben. Almo berechnet dann folgendes:

Wenn Sie **0** eingeben, dann berechnet Almo

- a. Unter- und Obergrenze (= Minimum und Maximum) aller Variablen (der nominalen, ordinalen, quantitativen)
- b. Mittelwerte für die quantitativen Variablen und Median für die ordinalen Variablen
- c. Standardabweichung für die quantitativen Variablen, halber Quartilsabstand für die ordinalen Variablen

Wenn Sie **1** eingeben, dann berechnet Almo

- a. die Unter- und Obergrenzen (= Minimum und Maximum) der nominalen und ordinalen Variablen – nicht jedoch für die quantitativen Variablen
- b. für die quantitativen Variablen wird die Zahl der diversen Werte ermittelt, die diese Variable annehmen
- c. für die quantitativen Variablen wird der Mittelwert und für die ordinalen der Median errechnet
- d. für die quantitativen Variablen wird die Standardabweichung ermittelt.

Diese Ausgabe wird in Abschnitt P5.05 dargestellt.

Wenn Sie **2** eingeben, dann berechnet Almo

- a. die Zahl der diversen Werte für alle Variable (nominale, ordinale, quantitative)
  - b. Mittelwert für quantitative, Median für ordinale und häufigster Wert für nominale Variable
  - c. Standardabweichung für quantitative und halber Quartilsabstand für ordinale Variable.
- Diese Ausgabe wird in P5.06 gezeigt.

### **Eingabebox 6: Kein\_Wert-Angabe und Umkodierung**

Siehe dazu Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.5.

## ***P5.02 Maskenprogramm mit Optionen***





## Eingabebox 9: verschiedene Optionen

↓ verschiedene Optionen

Optionsbox geöffnet:

↓ Loesche wieder diese Box

↑↓ 0 verschiedene Optionen

Unter- und Obergrenzen  
oder  
"diverse Werte" ermitteln

0 = für alle Variable  
werden die Unter- und Obergrenzen ermittelt

1 = für die nominalen und ordinalen Variablen  
werden die Unter- und Obergrenzen ermittelt  
für die quantitativen Variablen wird  
die Zahl der diversen Werte ermittelt

2 = für alle Variablen wird  
die Zahl der diversen Werte ermittelt

Beachte: Bei 2 wird für nominale Variable  
zusätzlich der "häufigste Wert" ermittelt

---

↑↓ 0

Streuungsmaß für ordinale Variable

0 = fuer die ordinalen Variablen wird als  
Streuungsmaß der halbe Quartilsabstand  
berechnet

1 = das Streuungsmass für ordinale Variable  
wird nach dem Kendall'schen tau-b Kalkül  
errechnet

---

↔ -1

-1 = Bei der Berechnung der Standardabwgn.  
soll im Nenner n-1 stehen.  
Editfeld leer = im Nenner steht n

---

↑↓ 4

Kommastellen für Ausgabe

---

↑↓ 2

Ausgabe der Ergebnisse

0 = keine Ergebnisse ausgeben

1 = die Zahl der diversen Werte ermitteln  
aber die diversen Werte nicht einzeln  
ausgeben

2 = die diversen Werte für jede Variable  
einzeln ausgeben

---

↑↓ 0

0 = für die quantitativen Variablen wird  
als Lagemaß der Mittelwert ermittelt

1 = für die quantitativen Variablen wird  
anstelle des Mittelwerts die Summe  
ermittelt u. in die beigeordnete  
Variable M gegeben

Eingabefeld 1: Siehe Erläuterungen zu Box 6 in P5.01.1.

Eingabefeld 2: Wenn Sie 0 eingeben, dann berechnet Almo als Streuungsmaß für die ordinalen Variablen den „halben Quartilsabstand“. Wenn Sie 1 eingeben, dann wird ein Streuungsmaß nach dem Kendall'schen tau-b Kalkül errechnet. Siehe dazu Denz:

Regressionsanalyse mit ordinalen Variablen in Holm (Hrsg.): Die Befragung 5, UTB 435.

*Eingabefeld 3:* Bei der Berechnung der Standardabweichung für quantitative Variable verwendet Almo standardmäßig  $n$  im Nenner. Der Benutzer kann in das 1. Eingabefeld eine beliebige negative oder positive Zahl einsetzen. Eine negative Zahl wird von  $n$  subtrahiert und eine positive zu  $n$  addiert. Die Voreinstellung ist 0, d.h. im Nenner steht  $n$ . Wenn Sie die Standardabweichung dazu verwenden wollen, Variablenwerte zu standardisieren, dann sollten Sie die Standardabweichung mit  $n$  dividieren.

*Eingabefeld 4:* Wenn Sie hier z.B. 4 eingeben, dann werden die Ergebnisse mit 4 Kommastellen ausgegeben.

*Eingabefeld 5:* Art der Ergebnisausgabe

2 = alle Ergebnisse ausgeben. Falls für die quantitativen Variablen die Zahl der diversen Werte ermittelt wurde, dann gibt Almo alle aufgetretenen diversen Werte aus. Dies können sehr viele sein. Mit 1 kann ihre Ausgabe unterdrückt werden.

1 = wie 2, aber die aufgetretenen diversen Werte nicht ausgeben

0 = Ergebnisse nicht ausgeben

*Eingabefeld 6:* 1 = für die quantitativen Variablen wird anstelle des Mittelwerts die Summe ermittelt und in die beigeordnete Variable M gegeben. 0 = der Mittelwert wird ermittelt.

### **P5.03 Eingabe mit selbstgeschriebenen ALMO-Programm**

Kapitel entfernt. Ist in Handbuch "Teil3a: Grundlegende Verfahren" enthalten.

### **P5.04 Optionen für das selbstgeschriebenen ALMO-Programm**

Kapitel entfernt. Ist in Handbuch "Teil3a: Grundlegende Verfahren" enthalten.

### **P5.05 Almo-Ausgabe I**

Almo liefert nachstehende Ausgabe wenn im Maskenprogramm Prog05m1 in Eingabebox 8 oder Prog05m2 in Eingabebox 9 im 1. Eingabefeld eine 1 eingesetzt wurde.

Zahl der eingelesenen Datensätze: 61  
(in Variable "Objekte" uebernommen)

```
***** MITTEILUNG
Die Standardabweichung der quantitativen Variablen
wurde mit n im Nenner gerechnet
```

```
***** MITTEILUNG
Fuer die ordinalen Variablen wird als "Mittelwert" der Median
und als Stand.abwg der halbe Quartilsabstand ausgegeben
```

```
***** WARNUNG
Almo setzt bei der Berechnung des Medians und des
Streuungsmaßes der ordinalen Variablen voraus
dass diese mit 1 beginnend mit Schrittweite 1 und ohne Luecken
kodiert sind. Ist dies nicht der Fall, dann sollten Sie Option 1 = 2
setzen bzw. im Maskenprogramm in der Box "zu analysierende Variable"
im 4. Eingabefeld "2" einsetzen
```

Variable	Mittelwert	Stand.abwg.	Untergrenze	Obergrenze
Nominale Variable				
1 Geschlecht	-	-	1.0000	2.0000
2 Wohnort	-	-	1.0000	2.0000
3 Beruf	-	-	1.0000	3.0000
4 Schulbildung	-	-	1.0000	2.0000
9 Item1	-	-	0.0000	1.0000
10 Item2	-	-	0.0000	1.0000
11 Item3	-	-	0.0000	1.0000
12 Item4	-	-	0.0000	1.0000
13 Item5	-	-	0.0000	1.0000
14 AB	-	-	1.0000	3.0000
15 AC	-	-	1.0000	3.0000
16 AD	-	-	1.0000	3.0000
17 BC	-	-	1.0000	3.0000
18 BD	-	-	1.0000	3.0000
19 CD	-	-	1.0000	3.0000

Ordinale Variable				
20 Bewertung	2.0000 (Median)	0.5000 (Quartilsabstand/2)	1.0000	3.0000

Quantitative Variable				
5 Leistung	3.8852	1.9086	-	Diverse_Werte 9.0000
6 Alter	3.9344	2.2093	-	9.0000
7 Einkommen	3.7049	2.4850	-	9.0000
8 Kinderzahl	4.6557	2.5278	-	9.0000

Aufgetretene Diverse\_Werte fuer V5 Leistung  
1.0000 2.0000 3.0000 4.0000 5.0000 6.0000  
7.0000 8.0000 9.0000

Aufgetretene Diverse\_Werte fuer V6 Alter  
1.0000 2.0000 3.0000 4.0000 5.0000 6.0000  
7.0000 8.0000 9.0000

Aufgetretene Diverse\_Werte fuer V7 Einkommen  
1.0000 2.0000 3.0000 4.0000 5.0000 6.0000  
7.0000 8.0000 9.0000

Aufgetretene Diverse\_Werte fuer V8 Kinderzahl  
1.0000 2.0000 3.0000 4.0000 5.0000 6.0000  
7.0000 8.0000 9.0000

Für nominale Variable wird in der Rubrik Mittelwert und Standardabweichung von ALMO ein Strich ausgegeben. Die beigeordneten Variablen SA 1:4, 9:19 und M 1:4, 9:19 bleiben leer.

Für ordinale Variable, wird in der Rubrik Mittelwert der Median ausgegeben. Als Standardabweichung wird der halbe Quartilsabstand ausgegeben.

Für die quantitative Variable werden alle beigeordnete Variable ermittelt.

Da wir in Eingabefeld 1=1 gesetzt haben, werden anstelle der Obergrenzen, die Zahl der diversen Werte ausgegeben, die die quantitativen Variablen annehmen. Da wir auch bei Option "Ausgabe der Ergebnisse =2" gesetzt haben, werden die diversen Werte auch noch ausgegeben.

## P5.06 Almo-Ausgabe II

Almo liefert folgende Ausgabe, wenn im "selbst geschriebenen" Almo-Programm Option 1=2 gesetzt wurde bzw. wenn im Maskenprogramm Prog05m1 in Box 8 oder Prog05m2 in Box 9 im 1. Eingabefeld eine 2 gesetzt wurde.

Die Ausgabe ist verkürzt

```

***** MITTEILUNG
Die Standardabweichung der quantitativen Variablen
wurde mit n im Nenner gerechnet

***** MITTEILUNG
Fuer die nominalen Variablen wird als "Mittelwert"
der haeufigste Wert ausgegeben

***** MITTEILUNG
Fuer die ordinalen Variablen wird als "Mittelwert" der Median
und als Stand.abwg der halbe Quartilsabstand ausgegeben

```

### \*\*\*\*\* Erläuterung:

Almo teilt zunächst mit, dass unter der Überschrift "Mittelwert" für die nominalen Variablen der "häufigste Wert" und für die ordinalen Variablen der Median ausgegeben wird. Für die ordinalen Variablen wird unter der Überschrift "Stand.abwg" der halbe Quartilsabstand ausgegeben.

In nachfolgender Ergebnistabelle wird jetzt anstelle der Werte-Obergrenze der Variablen deren "Zahl der diversen Werte" ausgegeben. Siehe zu diesem Begriff P5.01,1.1

In die beigeordneten Variablen UG1 bis UG20 wird jeweils 1 eingegeben. In die beigeordneten Variablen OG1 bis OG20 wird jeweils die Zahl der diversen Werte eingegeben.

Variable	Mittelwert	Stand.abwg.	Untergrenze	Diverse Werte	Kein-Wert		
					Faelle abs	in %	
Nominale Variable							
1 Geschlecht	1.0000	-	-	2.0000	0	0.00	
2 Wohnort	2.0000	-	-	2.0000	0	0.00	
3 Beruf	2.0000	-	-	3.0000	0	0.00	
4 Schulbildung	2.0000	-	-	2.0000	0	0.00	
9 Item1	1.0000	-	-	2.0000	0	0.00	
10 Item2	1.0000	-	-	2.0000	0	0.00	
11 Item3	0.0000	-	-	2.0000	0	0.00	
12 Item4	0.0000	-	-	2.0000	0	0.00	
13 Item5	0.0000	-	-	2.0000	0	0.00	
14 AB	3.0000	-	-	3.0000	0	0.00	
15 AC	3.0000	-	-	3.0000	0	0.00	
16 AD	3.0000	-	-	3.0000	0	0.00	
17 BC	3.0000	-	-	3.0000	0	0.00	
18 BD	3.0000	-	-	3.0000	0	0.00	
19 CD	3.0000	-	-	3.0000	0	0.00	

Ordinale Variable

-----						
20 Bewertung	2.0000 (Median)	0.5000 (Quartilsabstand/2)	-	3.0000	3	4.92
Quantitative Variable						
5 Leistung	3.8852	1.9086	-	9.0000	0	0.00
6 Alter	3.9344	2.2093	-	9.0000	0	0.00
7 Einkommen	3.7049	2.4850	-	9.0000	0	0.00
8 Kinderzahl	4.6557	2.5278	-	9.0000	0	0.00

## P5.07 Anteilswerte (relative Häufigkeiten)

Wenn wir eine 0-1 kodierte Variable als quantitative Variable in Programm 5 einführen, dann erhalten wir für sie einen Mittelwert (z.B. 0.35) und eine Standardabweichung (z.B. 0.477). Der Mittelwert entspricht dabei dem Anteilswert  $p$  der Ausprägung 1. D.h. der Anteil der Untersuchungseinheiten, die sich in der Ausprägung 1 befinden, ist 0.35 bzw. 35 %. Die Standardabweichung, die Programm 5 ausgibt, ist gleich der gewohnten Standardabweichung für nominal-dichotome Variable  $\sqrt{p*(1-p)}$ .

Die dargestellte Eigenschaft 0-1 kodierter Variablen können wir also sehr gut dazu verwenden, um die Anteilswerte der Ausprägungen auch einer polytomen nominalen Variablen zu erhalten. Wir lösen diese in 0-1 kodierte Dummy-Variable auf und berechnen für diese die Mittelwerte. Siehe dazu das Programm "Anteilsw.Alm". Man findet dieses Programm durch Klick auf das Menü "Almo / Liste aller Almo-Programme".

In diesem Beispiel-Programm wird die nominale Variable V3 (mit 3 Ausprägungen) in 3 Dummy-Variable aufgelöst. Die für diese errechneten Mittelwerte sind die Anteilswerte der 3 Ausprägungen von V3. In der Umkodierungsbox dieses Programms werden die Dummy-Variable für V3 in folgender Weise gebildet



Variable 3 besitzt 3 Ausprägungen. Wir lösen sie in die 3 0-1 kodierten Dummy-Variable V101 bis 103 auf. Wenn beispielsweise der Mittelwert der Dummy-Variablen 103=0.26 ist, dann bedeutet dies, dass sich die Untersuchungseinheiten mit einem Anteil von 0.26 (=26 %) in der Ausprägung 3 von V3 befinden.

**Beachte:** Nur die Untersuchungseinheiten, die in V3 einen validen Wert besitzen (also nicht KEIN\_WERT) sind, werden zusammen auf = 1.0 (= 100%) gesetzt.

**Beachte** die Voraussetzungen für die DUMMY-Transformation im Almo-Handbuch, Teil 2, Bedienungsanleitung, Abschnitt 16.7.

## Literatur

- Büning, Trenkler:** Nichtparametrische Methoden, Berlin, 1978  
**Clauß G., Ebner H.:** Grundlagen der Statistik, Berlin, 1974  
**Geske, Gunther:** Eindimensionale Grundauszählungen und Normalverteilungs-test, in Holm, K. (Hrsg.) Die Befragung 2, Francke, UTB 373, München 1975  
**Lienert, G.A.:** Verteilungsfreie Methoden in der Biostatistik, Meisenheim am Glan,

1973  
**Lohse, Ludwig, Röhr:** Statistische Verfahren, Berlin, 1982  
**Siegel, Sidney:** Nonparametric statistics, 1956 (Deutsch: Nichtparametrische statistische Methoden 1976)