

Anwendersoftwareklasse 2

DASYS – Handbuch

Stand: 08.08.86

© 1986

Selbstverlag GMD

Alle Rechte vorbehalten.

Insbesondere ist die Überführung in maschinenlesbare Form, sowie das Speichern in Informationssystemen, auch auszugsweise, nur mit schriftlicher Genehmigung der GMD gestattet.

Herausgeber:

Gesellschaft für Mathematik und Datenverarbeitung mbH

Postfach 1240, Schloß Birlinghoven
D – 5205 Sankt Augustin 1
Telefon(02241) 14 – 1, Telex 8 89 469 gmd d
Telefax(02241) 14 28 89, BTX *43900#
Teletex 2627 – 224135 = GMDVV

Softwareklasse 2 (Anwendersoftware)

Regelmäßige Wartung

Autoren:

Dr. Bendisch, H. von dem Bussche

Texterstellung:

Dieser Text wurde mit der EUMEL – Textverarbeitung erstellt und aufbereitet und mit dem Agfa Laserdrucksystem P400 gedruckt.

Umschlaggestaltung:

Hannelotte Wecken

Hinweis:

Diese Dokumentation wurde mit größtmöglicher Sorgfalt erstellt. Dennoch wird für die Korrektheit und Vollständigkeit der gemachten Angaben keine Gewähr übernommen. Bei vermuteten Fehlern der Software oder der Dokumentation bitten wir um baldige Meldung, damit eine Korrektur möglichst rasch erfolgen kann. Anregungen und Kritik sind jederzeit willkommen.

1. Vorwort

DASYS, ein Datenauswertungssystem, dient zur Auswertung von Datenmaterial mit statistischen Methoden.

DASYS stellt ein in sich geschlossenes System dar, das keine Programmierkenntnisse voraussetzt. Es wird daher empfohlen, daß man einen DASYS-HG (für Single-User-Systeme) bzw. eine DASYS-Task (für Multi-User-Systeme) für Statistikanwender ohne EUMEL-Kenntnisse verfügbar macht. Der erste Teil dieses Heftes ist als Benutzerhandbuch für das komplette System ausgelegt, der zweite Teil richtet sich an Programmierer, die einzelne Statistikmodule verwenden wollen und enthält die Paketschnittstellen der einzelnen Programm-Ebenen.

Inhaltsverzeichnis

Teil I

Benutzerhandbuch:

1.	Vorwort	1
2.	Beschreibung der Kommandos	3
2.1.	Dateneingabe	4
2.2.	Einfache statistische Kenngrößen	6
2.3.	Eindimensionale Häufigkeitsauszählung	6
2.3.1.	Häufigkeitsauszählung bei stetigem Merkmal	7
2.3.2.	Häufigkeitsauszählung bei diskretem Merkmal	9
2.4.	Korrelationsmatrix	10
2.5.	Lineare Regression	10
2.6.	Verarbeitung der Ausgaben	11
2.7.	Unterbrechen	12
3.	Elan – Kommando – Monitor	12
4.	Verhalten bei Fehlern	12
	Anhang: Beispielergebnisse	13

Teil II

Beschreibung für den Programmierer:

1.	Einführung	25
2.	Verwaltung	27
3.	Elementare Statistik	29
3.1.	Grundfunktionen eines Merkmals	29
3.2.	Funktionen von zwei Merkmalen	30
3.3.	Häufigkeitsauszählungen eines Merkmals	30
3.4.	Schnittstelle des Pakets "elementare statistik"	31
4.	Statistik Funktionen	35
5.	Dialog	38
5.1.	Schnittstelle des Pakets "dialog"	39
	Anhang	43

2. Beschreibung der Kommandos

Das Kapitel 2 dieses Handbuches beschreibt das Verhalten einzelner DASYS-Funktionen ausgehend vom DASYS-Monitor. Zu jeder Funktion gibt es ein Kapitel, in dem kurz die Funktion beschrieben wird. Beispiele und Hinweise zur Ausgabe ergänzen die Kapitel.

Anmerkung : Wie man den DASYS-Monitor erreicht, ist von der jeweiligen EUMEL-Installation abhängig. Bei Multi-User-Systemen beginnt man eine DASYS-Sitzung einfach durch :

<SV> - Taste (BREAK oder CONTROL b).

Dann meldet sich Eumel mit :

"gib supervisor kommando".

Darauf gibt man ein:

continue ("dasys") <CR>.

Es erscheint das erste Menu. Dies gilt allerdings nur, wenn DASYS in dem System wie folgt eingerichtet wurde :

- begin ("DASYS") (* Einrichten einer Manager-Task *)
- archive ("DASYS") (* Anmelden des DASYS - Archivs *)
- fetch all (archive) (* Laden der DASYS - Software *)
(* evtl. auch vom zweiten Archive *)
- insert ("dialog")
- global manager (* Task zum Manager erklären *)
- begin ("dasys", "DASYS") (* Arbeitstask einrichten *)
- fetch("generator") (* Generator - Programm holen *)
- run (* Generator - Programm starten *)

Nach etwa 30-40 Minuten erscheint das Monitor-Menü. In den Parametereingabemenüs bekommt man nach ESC ESC die Kommandos Verarbeitung (q) und Abbrechen (h) angeboten.

2.1. Dateneingabe

Daten, die von DASYS ausgewertet werden sollen, müssen in Eumel-Test-Files vorliegen. Um die Daten, die eine solche Datei enthält, auszuwerten, sind diese zunächst von der Textform im EUMEL-File in einen Datenraum reeller Zahlen umzuwandeln. Dieser Datenraum kann für maximal 10 Merkmale bis zu je 1000 Werte aufnehmen.

Das Kommando "Dateneingabe" hat folgenden Parameter :

(1) <dateiname>

Hier ist der Name einer Eumel-Datei einzutragen, die die Daten enthält, mit denen gerechnet werden soll. Diese Datei muß in der eigenen Task existieren (ggf. vorher mit 'fetch' laden).

Nach Anwahl einer EUMEL-Datei kann man entweder die Daten, die diese Datei enthält, mit Hilfe des Editors verändern oder die Daten einlesen. Auf diese eingelesenen Daten beziehen sich alle späteren Berechnungen.

Für DASYS-Eingabedateien ist folgendes Format einzuhalten :

1. Die erste Zeile der Datei enthält die Merkmalsnamen. Ein Merkmalsname beginnt mit einem Buchstaben auf den beliebig viele Buchstaben oder Ziffern folgen dürfen (Da man bei der Angabe eines Merkmalsnamen im Dialog jedoch nur 20 Zeichen eingeben kann, sollte man sich auf diese Länge bei den Namen beschränken).
2. DASYS kann maximal 10 Merkmale verwalten. Möchte man mit mehr als 10 Merkmalen arbeiten, so kann man *alle* Merkmale in eine Datei eingeben und markiert dann die Merkmale, die vorerst nicht eingelesen werden sollen. Es dürfen dann maximal 10 unmarkierte Merkmale übrigbleiben. Ein Merkmal wird markiert, indem man an den Merkmalsnamen in der ersten Zeile ein Minuszeichen anhängt.
3. Die Merkmalsnamen in der ersten Zeile werden durch Kommata voneinander getrennt.
4. Ab der zweiten Zeile werden die Merkmalswerte eingegeben. Es sind auch Daten für die Merkmale einzugeben, die nicht eingelesen werden sollen.
5. Die Merkmalswerte sind durch ein Komma und/oder Blanks zu trennen.

Beispiel einer Eingabedatei :

lfdnr	grz	abs	laen	hub	dmes	modu	teza
1	, 29.4	, 0.5	, 130	, 210	, 140	, 6.9	, 10
2	, 47.2	, 0.6	, 130	, 210	, 130	, 5.85	, 10
3	, 48.7	, 0.6	, 130	, 210	, 130	, 5.85	, 10
4	, 59.5	, 0.8	, 150	, 250	, 110	, 2.55	, 12
5	, 54.8	, 0.6	, 150	, 250	, 110	, 3.85	, 12
6	, 62.6	, 0.7	, 150	, 240	, 110	, 4.5	, 12
7	, 69.9	, 0.5	, 240	, 340	, 110	, 2.75	, 12
8	, 41.7	, 0.5	, 124	, 210	, 60	, 1.00	, 12
9	, 43.9	, 0.8	, 90	, 170	, 80	, 2.75	, 12
10	, 22.7	, 0.4	, 62	, 145	, 60	, 2.75	, 12
11	, 40.8	, 0.5	, 90	, 180	, 80	, 1.25	, 12
12	, 37.3	, 0.6	, 62	, 150	, 60	, 1.25	, 12
13	, 42.1	, 0.6	, 90	, 175	, 80	, 2	, 13
14	, 83.0	, 0.8	, 260	, 350	, 95	, 3.6	, 13
15	, 63.3	, 0.7	, 150	, 230	, 110	, 3.85	, 15
16	, 79.3	, 0.7	, 185	, 270	, 125	, 4.50	, 15
17	, 125.3	, 1.6	, 150	, 240	, 115	, 4.3	, 15
18	, 102.0	, 0.9	, 150	, 250	, 115	, 4.3	, 15
19	, 130.6	, 1.6	, 185	, 260	, 120	, 4.5	, 15
20	, 144.2	, 4.2	, 105	, 190	, 100	, 2.5	, 12
21	, 119.7	, 2.0	, 210	, 300	, 100	, 2.75	, 12
22	, 137.1	, 2.3	, 240	, 340	, 110	, 4.0	, 12
23	, 71.4	, 1.6	, 130	, 220	, 140	, 6.9	, 10

Diese Datei enthält 8 Merkmale, das Merkmal "lfdnr" soll vorerst nicht bearbeitet werden. Die Kommata zwischen den Merkmalswerten könnte man auch weglassen.

Die Daten sind entnommen aus :

John, B. : Statistische Verfahren für technische Messreihen, Hanser Verlag,
München 1979.

2.2. Einfache statistische Kenngrößen

Das Kommando "Einfache statistische Kenngrößen" dient der Berechnung von Minimum, Maximum, Spannweite, Median, Summe, arithmetischem Mittel, Standardabweichung und Varianz.

Es hat folgenden Parameter :

(1) <merkmalsname>

Es werden zehn Eingabefelder angeboten, in die Namen der auszuwertenden Merkmale eingetragen werden müssen. Es muß mindestens ein Merkmalsname angegeben werden, es können aber auch Eingabefelder leerbleiben.

Anmerkung: Standardabweichung und Varianz werden mit Vorfaktor $1/(n-1)$ berechnet.

2.3. Eindimensionale Häufigkeitsauszählung

Das Kommando "eindimensionale Häufigkeitsauszählung" dient dazu, eine Häufigkeitsauszählung für ein Merkmal zu erstellen. Die Auszählung kann für ein stetiges oder ein diskretes Merkmal durchgeführt werden. Bei stetigem Merkmal müssen Intervalle angegeben werden; für jedes Intervall wird dann die Zahl der Werte ausgezählt, die in diesem Intervall liegen. Bei diskretem Merkmal wird zunächst die absolute Häufigkeit von den verschiedenen Werten der Stichprobe ausgezählt.

Nach der Anwahl dieses Kommandos muß man zuerst eingeben, ob man die Häufigkeitsauszählung für ein diskretes oder ein stetiges Merkmal durchführen will. Mit Anwahl eines stetigen Merkmals gibt man gleichzeitig an, ob die oben genannten Intervalle, deren Grenzen noch einzugeben sind, rechts oder links offen sind.

Man kann hinsichtlich der Darstellung der Häufigkeiten zwischen einer tabellarischen und einer graphischen (waagerechtes Stabdiagramm) Darstellung wählen. Die Ausgabe des Ergebnisses erfolgt in eine Datei. Zusätzlich kann man die Häufigkeiten graphisch auf dem Bildschirm darstellen.

In den nächsten zwei Unterabschnitten werden Hinweise zum Ausfüllen der Bildschirm - Formulare für die Häufigkeitsauszählung gegeben.

2.3.1. Häufigkeitsauszählung bei stetigem Merkmal

Das Kommando "Häufigkeitsauszählung bei stetigem Merkmal" hat folgende 2 Parameter :

(1) <Merkmalsname>

Hier wird der Name des Merkmals eingetragen, für das eine Auszählung erstellt werden soll.

(2) <Intervallgrenzen>

Die Eingabe der Intervalle kann auf zwei verschiedene Arten geschehen :

1. Eingabe von äquidistanten Intervallgrenzen

Es müssen die drei Eingabefelder "erste Intervalluntergrenze", "Länge der Intervalle" und "Anzahl der Intervalle" ausgefüllt werden. Aus diesen Werten erzeugt das Programm dann die Folge der Intervallobergrenzen:

$$g, g + 1 \cdot l, g + 2 \cdot l, \dots, g + n \cdot l$$

Dabei bedeutet:

- g – Erste endliche Intervallgrenze
- l – Länge der endlichen Intervalle
- n – Anzahl der Intervalle ($n \geq 0$)

2. Eingabe von einzelnen Intervallobergrenzen

Hier werden die Intervallgrenzen in aufsteigender Folge sortiert eingetragen, und zwar ist für jedes Intervall die Obergrenze anzugeben, die Untergrenze eines Intervalls ist die Obergrenze des vorausgehenden Intervalls. Es ist mindestens eins der zwanzig Eingabefelder für die Intervallobergrenzen auszufüllen. Es können auch Felder leergelassen werden.

Hinweise zur Ausgabe

Die Ausgabe in eine Datei umfaßt :

1. Den Merkmalsnamen
2. Den Stichprobenumfang
3. Je nach gewünschter Darstellungsart :
 - (A) Die tabellarische Darstellung :
 1. Die Intervallgrenzen
 2. Die absoluten Häufigkeiten
 3. Die prozentualen Häufigkeiten bezogen auf den Stichprobenumfang
 - (B) Die graphische Darstellung :
 1. Die Intervallgrenzen
 2. Stäbe, deren Länge die prozentualen Häufigkeiten in den Intervallen darstellen

2.3.2. Häufigkeitsauszählung bei diskretem Merkmal

Das Kommando "Häufigkeitsauszählung bei diskretem Merkmal" hat den folgenden Parameter:

(1) <Merkmalsname>

Hier wird der Name des auszuzählenden Merkmals eingetragen.
Wenn er länger als sieben Zeichen ist, wird er abgeschnitten.

Hinweise zur Ausgabe

Die Ausgabe in eine Datei umfaßt:

1. Den Merkmalsnamen
2. Den Stichprobenumfang
3. Je nach gewünschter Darstellungsart :

(A) Die tabellarische Darstellung :

1. Die verschiedenen, der Größe nach geordneten Zahlenwerte
2. Die absoluten Häufigkeiten
3. Die prozentualen Häufigkeiten

(B) Die graphische Darstellung :

1. Die verschiedenen, der Größe nach geordneten Zahlenwerte
2. Stäbe, deren Länge die prozentualen Häufigkeiten der verschiedenen Merkmalswerte darstellen

2.4. Korrelationsmatrix

Das Kommando "Korrelationsmatrix" dient zur Berechnung der Pearsonschen Stichproben – Korrelationskoeffizienten zwischen bestimmten anzugebenden Merkmalen.

Das Kommando hat folgende Parameter:

(1) <merkmalsname>

Von den zehn Eingabefeldern sind mindestens zwei mit einem Merkmalsnamen auszufüllen. Auch hier können wieder Eingabefelder leergelassen werden. Es ist darauf zu achten, daß kein Merkmalsname doppelt eingegeben wird.

Hinweis zur Ausgabe

Die Korrelationskoeffizienten werden in Matrixform ausgegeben und nicht über die 80ste Spalte hinaus geschrieben. So ist sichergestellt, daß man die Korrelationsmatrix auch auf Druckern ab einer Zeilenbreite von 80 Zeichen ausdrucken kann.

2.5. Lineare Regression

Das Kommando "Lineare Regression" bewirkt das Zeigen eines Bildschirmformulars, auf dem das Feld 'abhängiges Merkmal' und mindestens eines der neun Felder ausgefüllt werden müssen, die für die Namen der unabhängigen Merkmale vorgesehen sind. Dabei darf kein Merkmalsname doppelt auftauchen.

Die Eingabe veranlaßt die Berechnung aller Regressionskoeffizienten und des Achsenabschnitts nach der Methode der kleinsten Quadrate. Außerdem erfolgen zweiseitige T-Tests und ein F-Test auf die Regressionskoeffizienten, die in den entsprechenden Nullhypothesen entweder einzeln oder insgesamt zu Null angenommen werden.

Die Ausgabe umfaßt :

- eine Tabelle der Mittelwerte aller in die Berechnung einbezogenen Merkmale
- die vollständige Regressionsgleichung
- das Bestimmtheitsmaß und den multiplen Korrelationskoeffizienten
- die Tabelle aller abhängigen Merkmale mit Regressionskoeffizienten, entsprechenden T-Prüfgrößen, zweiseitigen Überschreitungswahrscheinlichkeiten (\bar{u}_w) und einer Signifikanzentscheidung. Diese ist "ja", falls $\bar{U}_W < \text{Irrtumswahrscheinlichkeit} = 0.05$, sonst "nein".
- F-Prüfgröße mit Anzahl der Freiheitsgrade.

Anmerkung : Die Überschreitungswahrscheinlichkeiten bei T-Tests werden nur bei mehr als 15 Freiheitsgraden berechnet, andernfalls wird eine Signifikanz-Entscheidung anhand einer einprogrammierten Tabelle von alpha-Quantilen ermittelt. Beim F-Test ist nach bisherigem Stand eine Signifikanz-Entscheidung vom Anwender unter Heranziehung der F-Prüfgröße und einer Vertafelung von Quantilen zu F-Verteilungen selbst zu treffen.

2.6. Verarbeitung der Ausgaben

Nach Anwahl des Kommandos "Verarbeitung der Ausgaben" kann man die EUMEL-Textdatei "ausgabe" - in dieser Datei befinden sich die Ergebnisse der durchgeführten DASYS Funktionen - editieren, löschen und auf einem EUMEL-Drucker ausdrucken.

2.7. Unterbrechen

Das Kommando "Unterbrechen" dient zum Unterbrechen von DASYS. Nach erneutem Umladen bzw. nach dem Fortsetzen einer Task befindet man sich wieder im DASYS-Monitor.

3. Elan – Kommando – Monitor

Erscheint auf dem Bildschirm die Aufforderung "gib kommando", so kann ein beliebiges Elan-Kommando (wie z. B. "fetch") an den Elan-Compiler übergeben werden. Dieser übersetzt das Kommando und führt es aus. Damit ist es möglich, auszuwertende Daten zu archivieren oder von Archiven zu laden, die DASYS-Ausgaben zu archivieren und den EUMEL-Drucker auf bestimmte Parameter einzustellen.

4. Verhalten bei Fehlern

Tritt bei der Ausführung einer DASYS-Funktion ein Fehler auf (z. B. wenn eine DASYS-Funktion keine Daten vorfindet), so fängt der DASYS-Monitor diese ab und zeigt die entsprechende Fehlermeldung in der untersten Zeile. Nach Betätigen einer beliebigen Taste erscheint wieder das letzte Menü.

Anhang zum Handbuch : Beispielergebnisse

Auf den folgenden Seiten sind einige Ergebnisse der DASYS-Funktionen zusammengestellt, die sich aus der Auswertung der Daten in der Eingabedatei "testdaten" aus Kapitel 2.1. ergeben.

Auswertung des Merkmals grz

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	72.0217
Summe :	1656.5000
Standardabweichung :	36.8527
Varianz :	1358.1181

Minimum :	22.7000
Median :	62.6000
Maximum :	144.2000
Spannweite :	121.5000

Auswertung des Merkmals abs

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	1.0478
Summe :	24.1000
Standardabweichung :	.8665
Varianz :	.7508

Minimum :	.4000
Median :	.7000
Maximum :	4.2000
Spannweite :	3.8000

Auswertung des Merkmals laen

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	146.2174
Summe :	3363.0000
Standardabweichung :	54.1924
Varianz :	2936.8142

Minimum :	62.0000
Median :	150.0000
Maximum :	260.0000
Spannweite :	198.0000

Auswertung des Merkmals hub

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	234.3478
Summe :	5390.0000
Standardabweichung :	57.6097
Varianz :	3318.8735

Minimum :	145.0000
Median :	230.0000
Maximum :	350.0000
Spannweite :	205.0000

Auswertung des Merkmals dmes

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	103.9130
Summe :	2390.0000
Standardabweichung :	24.0717
Varianz :	579.4466

Minimum :	60.0000
Median :	110.0000
Maximum :	140.0000
Spannweite :	80.0000

Auswertung des Merkmals modu

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	3.6717
Summe :	84.4500
Standardabweichung :	1.6658
Varianz :	2.7750

Minimum :	1.0000
Median :	3.8500
Maximum :	6.9000
Spannweite :	5.9000

Auswertung des Merkmals teza

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Arithmetisches Mittel :	12.3913
Summe :	285.0000
Standardabweichung :	1.6442
Varianz :	2.7036

Minimum :	10.0000
Median :	12.0000
Maximum :	15.0000
Spannweite :	5.0000

Auszählung bezüglich des Merkmals grz

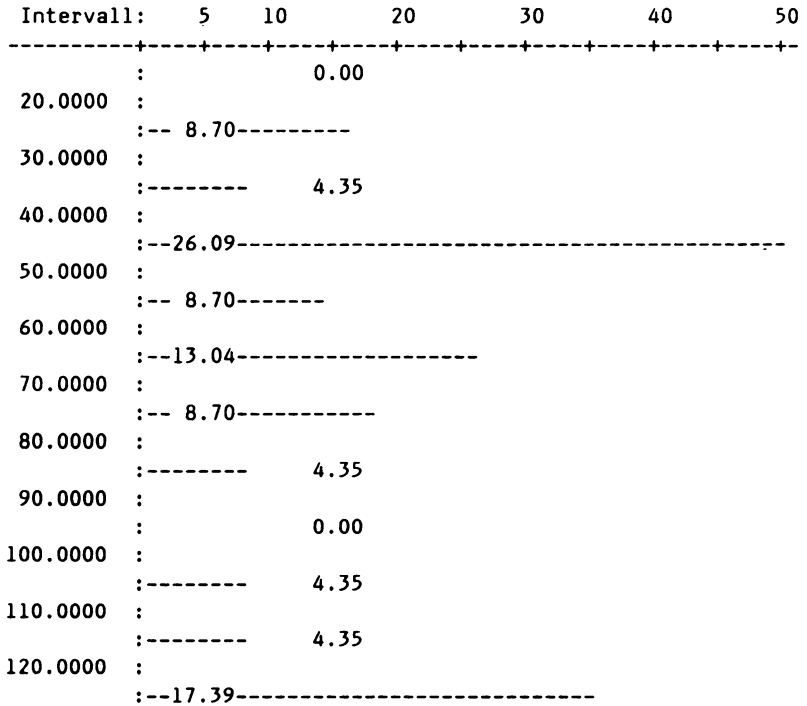
Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Intervall	: abs. Häufigk.	: proz Häufigkeit
20.0000 < x <= 30.0000	: 2	: 8.6957 %
30.0000 < x <= 40.0000	: 1	: 4.3478 %
40.0000 < x <= 50.0000	: 6	: 26.0870 %
50.0000 < x <= 60.0000	: 2	: 8.6957 %
60.0000 < x <= 70.0000	: 3	: 13.0435 %
70.0000 < x <= 80.0000	: 2	: 8.6957 %
80.0000 < x <= 90.0000	: 1	: 4.3478 %
90.0000 < x <= 100.0000	: 0	: 0.0000 %
100.0000 < x <= 110.0000	: 1	: 4.3478 %
110.0000 < x <= 120.0000	: 1	: 4.3478 %
120.0000 < x <=	: 4	: 17.3913 %

Graphische Häufigkeitsdarstellung für das Merkmal grz

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

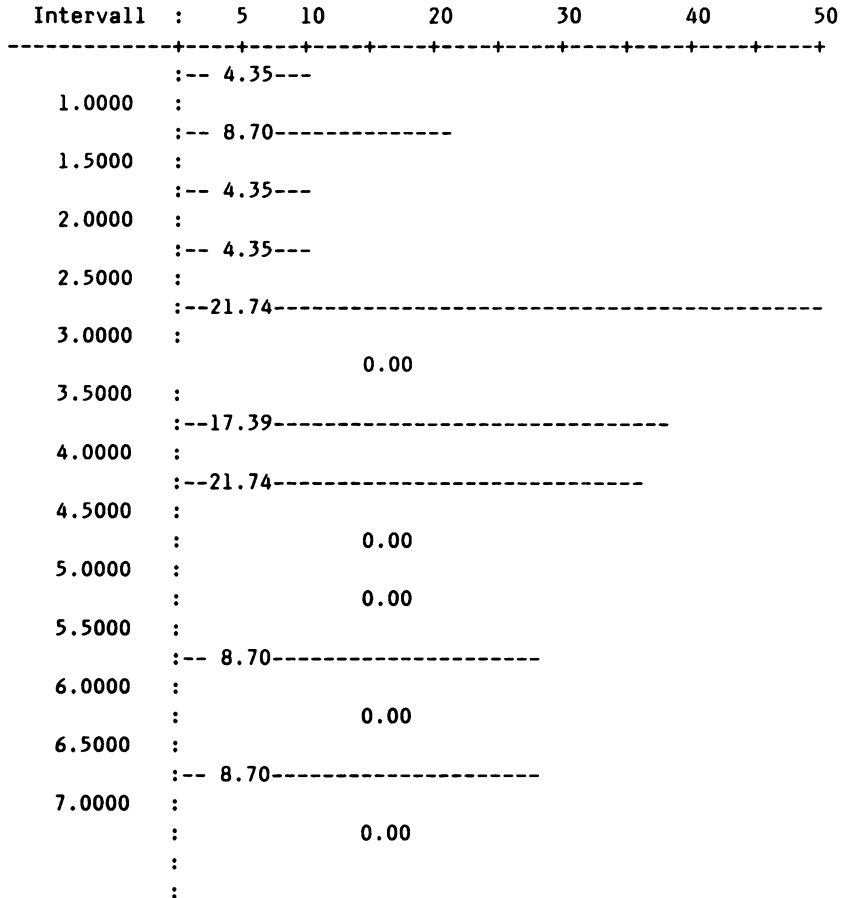
Die Intervalle sind jeweils unten offen, oben geschlossen



Graphische Häufigkeitsdarstellung für das Merkmal modu

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Die Intervalle sind jeweils unten offen, oben geschlossen



Auszählung bezüglich des Merkmals laen

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Intervall	: abs. Häufigk.	: proz Häufigkeit
60.0000 < x <= 90.0000	: 5	: 21.7391 %
90.0000 < x <= 120.0000	: 1	: 4.3478 %
120.0000 < x <= 150.0000	: 11	: 47.8261 %
150.0000 < x <= 180.0000	: 0	: 0.0000 %
180.0000 < x <= 210.0000	: 3	: 13.0435 %
210.0000 < x <= 240.0000	: 2	: 8.6957 %
240.0000 < x <= 270.0000	: 1	: 4.3478 %
:	:	:

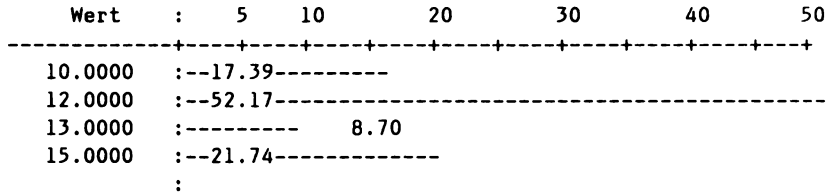
Auszählung bezüglich des Merkmals teza

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten

Merkmalswerte d. Größe nach	: abs. Häufigk.	: proz. Häufigkeit
x = 10.0000	: 4	: 17.3913 %
x = 12.0000	: 12	: 52.1739 %
x = 13.0000	: 2	: 8.6957 %
x = 15.0000	: 5	: 21.7391 %

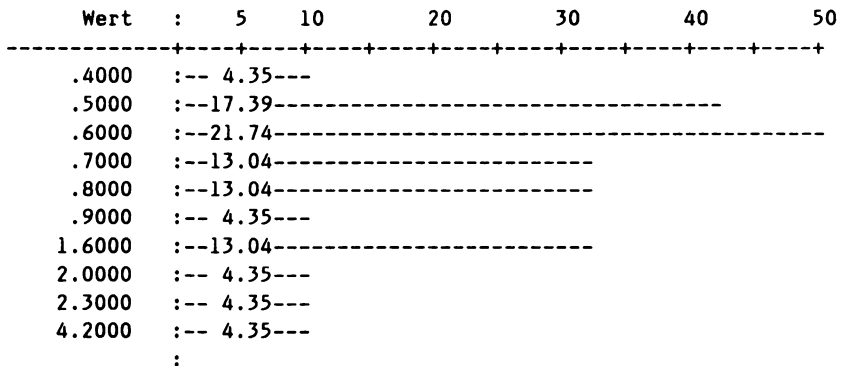
Graphische Häufigkeitsdarstellung für das Merkmal teza

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten



Graphische Häufigkeitsdarstellung für das Merkmal abs

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten



Korrelationsmatrix

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten.

	+ grz	+ abs	+ teza	+ hub	+ modu	+
grz	+ 1.000 +	.813 +	.439 +	.545 +	.089 +	
abs	+ .813 +	1.000 +	.060 +	.153 +	.005 +	
teza	+ .439 +	.060 +	1.000 +	.224 +	-.234 +	
hub	+ .545 +	.153 +	.224 +	1.000 +	.182 +	
modu	+ .089 +	.005 +	-.234 +	.182 +	1.000 +	

Korrelationsmatrix

Die Auswertung erfolgte mittels einer Stichprobe von 23 Objekten.

	+ grz	+ laen	+ dmes	+
grz	+ 1.000 +	.550 +	.301 +	
laen	+ .550 +	1.000 +	.420 +	
dmes	+ .301 +	.420 +	1.000 +	

* * * Lineare Regression * * *

Stichprobenumfang : 23

Zahl der unabhängigen Variablen : 6

Merkmalsname	Mittelwert	Standardabweichung
grz	72.0217	36.8527
abs	1.0478	.8665
laen	146.2174	54.1924
hub	234.3478	57.6097
dmes	103.9130	24.0717
modu	3.6717	1.6658
teza	12.3913	1.6442

Es ergibt sich folgende Gleichung :

$$\begin{aligned}
 \text{grz} &= \\
 &31.5261 \text{ abs} && + \\
 &- .1380 \text{ laen} && + \\
 &.3418 \text{ hub} && + \\
 &.0002 \text{ dmes} && + \\
 &2.5432 \text{ modu} && + \\
 &7.8241 \text{ teza} && + \\
 &- 127.2410
 \end{aligned}$$

Multipler Korrelationskoeffizient : .972

Bestimmtheitsmaß : .946

Zweiseitige T-Tests bezüglich der einzelnen Regressionskoeffizienten mit Irrtumswahrscheinlichkeit $\alpha = 0.05$ und mit 16 Freiheitsgraden, Nullhypothese :

"Der jeweilige Regressionskoeffizient der Grundgesamtheit ist Null"

Merkmal	Regressions- koeffizient	Prüfgr. t0	Wahrscheinlichk. P ($ t \geq t0 $)	Signifi- kanz
abs	31.5261	12.056	0.000	Ja
laen	- .1380	- .3809	.708	Nein
hub	.3418	1.0034	.331	Nein
dmes	.0002	.0009	.999	Nein
modu	2.5432	.7402	.470	Nein
teza	7.8241	5.4055	0.000	Ja

F-Prüfgröße für einen F-Test mit (6, 16) Freiheitsgraden, Nullhypothese :

"Die Regressionskoeffizienten sind gemeinsam gleich Null"

F - Prüfgröße 46.451

Für die folgende Regressionsanalyse wurden nur die ersten 15 Wertetupel der Eingabedatei als Stichprobe genommen, daher ergeben sich auch nur 13 Freiheitsgrade bei den T-Tests.

* * * Lineare Regression * * *

Stichprobenumfang : 15

Zahl der unabhängigen Variablen : 1

Merkmalsname	Mittelwert	Standardabweichung
grz	49.7933	15.8721
abs	.6133	.1246

Es ergibt sich folgende Gleichung :

$$\text{grz} = 79.1012 \text{ abs} + 1.2779$$

Multipler Korrelationskoeffizient : .621

Bestimmtheitsmaß : .386

Zweiseitige T-Tests bezüglich der einzelnen Regressionskoeffizienten mit Irrtumswahrscheinlichkeit $\alpha = 0.05$ und mit 13 Freiheitsgraden, Nullhypothese :

"Der jeweilige Regressionskoeffizient der Gesamtheit ist Null"

Merkmalsname	Regressionskoeffizient	Prüfgr. t0	oberes t-Quantil P (/t/ = /t0/)	Signifikanz
abs	79.1012	2.8562	2.160	Ja

F-Prüfgröße für einen F-Test mit (1,13) Freiheitsgraden, Nullhypothese : "Die Regressionskoeffizienten sind gemeinsam gleich Null"

F - Prüfgröße 8.158

Teil II

Beschreibung für den Programmierer :

1. Einführung

Das Datenauswertungssystem ist als ein Fünf-Schichten-Modell in ELAN realisiert. Jede Schicht entspricht einem ELAN-Paket. Änderungen und Erweiterungen sind auf jeder der Ebenen möglich, die Spracherweiterungen der darunterliegenden Ebenen stehen dann jeweils zur Verfügung. Der Zweck der einzelnen Schichten so wie ihre Beziehungen untereinander sind in der folgenden Graphik grob dargestellt.

D A S Y S - A U F B A U

I Daten- verwaltung	Einlesen von Eumeldateien, sequentieller Zugriff auf die Objekte, direkter Zugriff auf die Merkmale des jeweils aktuellen Objekts.
---------------------------	---

(II) verwendet ausschließlich (I) für
Datenzugriff, sichert gültige Adressen zu.

II elementare Statistik	Prozeduren zur Berechnung einzelner Funktionen auf Merkmalen aller Objekte parametrisiert, ohne I/O Operationen.
-------------------------------	--

Mehrere (III)-Prozeduren verwenden jeweils
eine (II)-Proz. zur Ermittlung entspr. Werte.

III Statistik Funktionen	Prozeduren, die jeweils eine DASYS- Funktion realisieren und entsprechend formatierten Output erzeugen. Eingaben parametrisiert.
--------------------------------	---

Jede (IV)-Prozedur basiert auf genau einer
(V)-Prozedur.

IV Dialog	Parameterlose Prozeduren, die jeweils eine DASYS-Funktion realisieren und alle notwendigen Eingaben im Dialog erfragen.
--------------	--

Die Monitor-Prozedur verwendet genau alle
(IV)-Prozeduren.

V Monitor	Parameterlose Prozedur, die in Menü- technik die Auswahl der DASYS- Funktionen erlaubt.
--------------	---

Im folgenden sind die Pakete im einzelnen beschrieben. Die Spezifikation ihrer Prozedurschnittstellen stellt zugleich die Minimalanforderungen der darauf aufzubauenden Pakete dar. Das Dasysssystem kann man mit dem Kommando 'monitor' verlassen und mit 'monitor dasys' neu starten.

2. Verwaltung

Das Paket "verwaltung" dient dazu, die Eingabedaten aus einer Eumel – Textdatei nach syntaktischer Prüfung in einer internen Datenstruktur zu speichern und effiziente Zugriffe auf diese zu ermöglichen. Da für statistische Auswertungen die Daten sequentiell abgearbeitet werden, stellt das Paket die Daten in einer Reihenfolge zur Verfügung, die durch ein Pointerarray definiert wird. Diese Methode ermöglicht insbesondere ein effizientes Umsortieren des Datenbestandes mit minimaler Zugriffszahl.

Die Prozeduren im einzelnen:

PROC get data (TEXT CONST filename) :

Liest die Daten aus der durch 'filename' spezifizierten Datei und überträgt sie in die interne Struktur. Enthält die Eingabedatei syntaktische Fehler, wird durch einen errorstop mit einer Klartext – Fehlermeldung abgebrochen, aus der auch die Position des Fehlers in der Eingabedatei ersichtlich ist.

Fehlen einzelne Werte relevanter Merkmale, wird eine entsprechende Meldung ausgegeben und das entsprechende Wertetupel bei der späteren Auswertung ignoriert.

INT PROC objects :

liefert die Anzahl der als gültig erkannten Wertetupel.

INT PROC values per obj :

liefert die Anzahl der gültigen Merkmale (vergl. DASYS Handbuch, Format der Eingabedatei)

PROC sort (INT CONST merkmal) :

sortiert alle Daten nach der Größe der Werte von 'merkmal'. Es wird der rekursive Quicksort – Algorithmus verwendet.

PROC reset obj :

setzt das aktuelle Objekt der aktuellen Sortierreihenfolge.

PROC next obj :

setzt das aktuelle Objekt auf das nächste Objekt der aktuellen Sortierreihenfolge.

PROC skip obj (INT CONST i) :

bewirkt $i + 1$ - maliges Aufrufen von 'next obj'.

REAL PROC value (INT CONST merkmal) :

liefert den Wert des Merkmals 'merkmal' des aktuellen Objekts.

BOOL PROC is last obj :

liefert 'TRUE', wenn das aktuelle Objekt das letzte ist, sonst 'FALSE'.

INT PROC value number (TEXT CONST name) :

liefert die Nummer zum Merkmalsnamen 'name', mit der das Merkmal programm-intern identifiziert wird.

TEXT PROC value name (INT CONST col) :

liefert den Klartext des Merkmals, das durch 'col' identifiziert wird.

3. Einfache Statistik

Mit den Prozeduren aus diesem Paket können elementare statistische Funktionen auf alle Werte eines Merkmals ausgeführt werden. Der Zugriff auf diese Werte geschieht durch die in dem Paket "verwaltung" deklarierten Prozeduren.

3.1. Grundfunktionen eines Merkmals

An Grundfunktionen stehen im einzelnen zur Verfügung:

Suchen des Minimums, des Maximums und des Medians eines Merkmals, Berechnung der Summe, des arithmetischen Mittels, der Varianz und der Standardabweichung eines Merkmals.

Jede dieser Funktionen prüft zuerst, ob die entsprechende Kenngröße schon einmal berechnet worden ist. Ist dies nicht der Fall, so werden mit Hilfe der Prozeduren aus dem Paket "verwaltung" sämtliche Werte des aktuellen Merkmals eingelesen. Dann werden alle statistischen Kenngrößen für dieses Merkmal berechnet und abgespeichert. Anschließend wird auf jeden Fall der berechnete Funktionswert geliefert.

Da immer alle statistischen Kenngrößen auf einmal berechnet werden, ist darauf zu achten, daß mindestens zwei Werte für ein Merkmal vorhanden sind, damit Varianz und Standardabweichung berechnet werden können (Obwohl sich für Merkmale mit nur einem Objekt durchaus sinnvolle Werte für Minimum, Maximum, Median, Summe und arithmetisches Mittel ergeben, führt die gleichzeitige Berechnung von Standardabweichung und Varianz bei weniger als zwei Werten zum Fehlerabbruch).

Um die neue Berechnung der statistischen Werte zu erzwingen (z.B. bei Eingabe neuer Daten) gibt es die Prozedur "loesche alle berechneten daten". Beim nächsten Aufruf einer der statistischen Funktionen werden wieder alle statistischen Kenngrößen für das Merkmal neu berechnet und abgespeichert.

Die oben beschriebenen Prozeduren dieses Pakets bekommen das Merkmal in Form seiner internen Nummer als Parameter übergeben.

3.2. Funktionen von zwei Merkmalen

Dies betrifft zunächst nur die Berechnung des Pearsonschen Korrelationskoeffizienten zwischen zwei Merkmalen.

Die Prozedur "korrelation" bekommt als Parameter zwei Merkmalsnummern übergeben, für die der Korrelationskoeffizient zu berechnen ist. Als erstes wird geprüft, ob der gewünschte Koeffizient schon einmal berechnet wurde und daher ohne Rechnung zur Verfügung steht. Sonst wird die Berechnung tatsächlich durchgeführt.

Eine erneute Berechnung wird ebenfalls durch die Prozedur "loesche alle berechneten daten" (vergl. oben) erzwungen.

3.3. Häufigkeitsauszählungen eines Merkmals

Die Prozedur "haeufigkeitsauszaehlung" sortiert zuerst die Werte, für die die Auszählung durchgeführt werden soll und veranlaßt ggf. auch die Berechnung aller elementaren Kenngrößen, um einen erneuten Sortiervorgang bei der späteren Bestimmung des Median überflüssig zu machen. Sind die Kenngrößen des Merkmals allerdings schon berechnet, so wird diese Spalte trotzdem neu sortiert (Sie wird nun mindestens zum zweiten Mal sortiert, da nicht sicher ist, ob durch andere Prozeduraufrufe die erste Sortierung nicht zerstört wurde).

In der Textkonstanten "intervallgrenzen" findet die Prozedur "haeufigkeitsauszaehlung" die jeweils nächste Obergrenze für das nächste Intervall. Als Untergrenze wird die jeweils letzte Obergrenze benutzt. Hierbei wird für das erste Intervall die Untergrenze minus unendlich angenommen, so daß man die Zahl der Werte kleiner bzw. kleiner gleich der ersten Obergrenze als erste Häufigkeit erhält. Außerdem wird an den Text, der die Obergrenzen der Intervalle ("intervallgrenzen") enthält, noch ein Intervall angehängt, dessen Intervallgrenzen die letzte Obergrenze und plus unendlich sind. Dadurch erhält man als letzte Häufigkeit die Zahl der Werte, die größer bzw. größer gleich der letzten übergebenen Obergrenze sind.

An der **BOOL CONST "kleiner"** kann die Prozedur erkennen, ob die jeweilige Obergrenze noch zum Intervall gehört (**kleiner = FALSE**, d.h. die Bedingung für die Obergrenze ist kleiner gleich und damit ist die Bedingung für die Untergrenze größer) oder ob die Obergrenze nicht mehr zum Intervall gehört (**kleiner = TRUE**, d.h. die Bedingung für die Obergrenze ist kleiner und damit die Bedingung für die Untergrenze größer gleich).

Daneben existiert eine weitere Prozedur "häufigkeitsauszählung", die nur die Spalte, die ausgezählt werden soll, übergeben bekommt. Sie dient der Auszählung diskreter Merkmale. Diese Spalte wird nach dem gleichen Verfahren sortiert, wie schon oben beschrieben wurde. In zwei Texten, die als **VAR** Parameter übergeben werden, gibt die Prozedur "häufigkeitsauszählung" die Ergebnisse zurück. Im Text "ausgezählte werte" stehen, in aufsteigender Reihenfolge sortiert, die verschiedenen Werte, die in der Spalte vorhanden sind. Im zweiten Text "häufigkeiten" steht nun an entsprechender Stelle, an der im ersten Text ein bestimmter Wert zu finden ist, dessen absolute Häufigkeit.

3.4.Schnittstelle des Pakets "elementare statistik"

Das Paket "elementare statistik" stellt folgende **ELAN** – Prozeduren zur Verfügung:

PROC loesche alle berechneten daten :

Erzwingt die Neuberechnung aller elementaren statistischen Kenngrößen und Korrelationskoeffizienten.

REAL PROC summe (INT CONST spalte) :

Liefert die Summe aller Objekte eines Merkmals.

REAL PROC minimum (INT CONST spalte) :

Liefert das Minimum einer Spalte.

REAL PROC maximum (INT CONST spalte) :

Liefert das Maximum einer Spalte.

REAL PROC median (INT CONST spalte) :

Liefert den Median einer Spalte.

REAL PROC mittelwert (INT CONST spalte) :

Liefert das arithmetische Mittel einer Spalte.

REAL PROC standardabweichung (INT CONST spalte) :

Liefert die Standardabweichung einer Spalte. Dabei wird die folgende Formel verwendet, in der das arithmetische Mittel nicht als bekannt vorausgesetzt wird (da alle elementaren statistischen Werte auf einmal berechnet werden, ist zum Zeitpunkt der Berechnung der Standardabweichung das arithmetische Mittel noch nicht bekannt) :

$$\text{Varianz}(X) = \left(\sum_{i=1}^n (X(i))^2 \right) - \left(\sum_{i=1}^n X(i) \right)^2 / n \bigg/ (n-1)$$

$$\text{Standardabweichung}(X) = \text{sqrt} \left(\text{Varianz}(X) \right)$$

Hierbei bedeutet:

n - Stichprobenumfang

$\sum_{i=1}^n$ - steht für die Summenfunktion, bei der der Index i die Werte von 1 bis n durchläuft

X(i) - Merkmal X des i-ten Objekts

sqrt - Quadratwurzel

** - Potenzfunktion

REAL PROC varianz (INT CONST spalte) :

Liefert die Varianz einer Spalte. Algorithmus siehe "standardabweichung".

REAL PROC korrelation (INT CONST spalte x, spalte y) :

Liefert den Pearsonschen Korrelationskoeffizienten zwischen spalte x und spalte y. Dazu wird zunächst die Kovarianz berechnet:

$$\text{Kovarianz (X,Y)} = \frac{\sum_{i=1}^n (X(i)*Y(i))}{n} - \frac{\sum_{i=1}^n X(i)}{n} \\ * \frac{\sum_{i=1}^n Y(i)}{n} / (n-1)$$

Aus der Kovarianz ergibt sich der Korrelationskoeffizient als:

$$\text{Korrelationskoeff (X,Y)} = \text{Kovarianz (X,Y)} / \\ (\text{Standardabweichung (X)} * \text{Standardabweichung (Y)})$$

Hierbei bedeutet:

n – Stichprobenumfang

$\sum_{i=1}^n$ – steht für die Summenfunktion, bei der der Index i die Werte von 1 bis n durchläuft

X(i) – Merkmal X des i – ten Objekts

Y(i) – Merkmal Y des i – ten Objekts

TEXT PROC haeufigkeitsauszaehlung (INT CONST spalte,
TEXT CONST intervallgrenzen,
BOOL CONST kleiner) :

Erstellt für die übergebenen "intervallgrenzen" eine Häufigkeitsauszählung einer Spalte. "intervallgrenzen" muß ein Text der Länge $8 * \text{Anzahl der eingegebenen Intervallobergrenze}$ sein. Jeweils 8 Zeichen des Textes müssen die interne REAL-Darstellung einer Intervallgrenze beinhalten (siehe Anhang, Prozedur 'replace'). Die BOOL CONST "kleiner" gibt an, ob die Obergrenze des Intervalls noch zum Intervall selbst dazugehört oder nicht.

Der Text, den die Prozedur liefert, hat die Länge $2 * \text{Anzahl der Intervalle}$. Jeweils 2 Zeichen des Textes beinhalten die interne INT-Darstellung der Häufigkeit (siehe Anhang, Operator 'ISUB').

PROC haeufigkeitsauszaehlung (INT CONST spalte,
TEXT VAR werte,
haeufigkeiten) :

Erstellt für die übergebene Spalte eine Häufigkeitsauszählung. Nach dem Prozeduraufruf stehen in "werte" die in der Spalte auftretenden, zahlenmäßig verschiedenen Werte (sortiert), in "haeufigkeiten" für die entsprechenden Werte ihre Häufigkeiten. Der Text "werte", den die Prozedur liefert, hat die Länge $8 * \text{Anzahl der verschiedenen Werte}$. Jeweils 8 Zeichen des Textes beinhalten die interne REAL-Darstellung der verschiedenen Werte (siehe Anhang, Operator 'RSUB').

Der Text "haeufigkeiten" hat die Länge $2 * \text{Anzahl der verschiedenen Werte}$ und jeweils 2 Zeichen des Textes beinhalten die interne INT-Darstellung der Häufigkeit (siehe Anhang, Operator 'ISUB').

4. Statistik Funktionen

Dieses Paket realisiert die einzelnen Dasys-Kommandos ohne Dialog, also in einer Form, in der sie von Anwenderprogrammen z.B. für Auswertungen im Stapelbetrieb verwendet werden können. Alle Eingaben sind beim Aufruf der Prozeduren als Parameter zu übergeben. Jede Prozedur erzeugt die vollständige Ausgabe des jeweiligen Kommandos auf dem Ausgabeboard. Das Ausgabeboard kann mit der Prozedur 'put board' auf dem Bildschirm oder in eine Datei ausgegeben werden. Die Berechnung der Ausgabewerte erfolgt, außer bei der Prozedur 'regression', durch die Funktionen des Paketes II und ist dort beschrieben.

PROC basiswerte (TEXT CONST merkmale) :

Gibt die Ergebnisse der 'Funktionen eines Merkmals' entsprechend den Formaten des Formulars 'basiswerte' aus.

PROC korrelation (TEXT CONST merkmale) :

Gibt die Matrix der Korrelationkoeffizienten aller Merkmale untereinander aus.

**PROC regression (INT CONST abhaengiges merkm,
TEXT CONST unabhaengige merkmale) :**

'merkmal' bzw. 'unabhängige merkmale' müssen Texte der Länge $2 * \text{Anzahl der Merkmale}$ sein. Jeweils 2 Zeichen der Texte müssen die interne INT - Darstellung der Spaltennummer des gewünschten Merkmals beinhalten (siehe Anhang, Prozedur 'replace'). Diese Prozedur führt eine multilineare Regressions-Analyse mit einem abhängigen und k unabhängigen Merkmalen durch. Dabei werden die folgenden numerischen Schritte durchgeführt:

Berechnung der Beta - Gewichte:

$$\text{beta} := \sum_{i=1}^k r_{ij} \cdot r_{iy}^{-1} \quad j=1, \dots, k$$

Dabei bedeutet:

r_{iy} – Korrelationskoeffizient zwischen abh. und i -tem unabh. Merkmal

– 1

r_{ij} – Das Element der invertierten Korrelationsmatrix der unabh. Merkmale in Zeile i und Spalte j

k – Anzahl der unabhängigen Merkmale

Berechnung der Regressionskoeffizienten:

$$b_j := \text{beta}_j \cdot s_y / s_j$$

Dabei stehen s_y und s_j für die Standardabweichung des unabh. bzw des j -ten abh. Merkmals.

Berechnung des Achsenabschnitts :

$$b_0 := m_y - \sum_{j=1}^k b_j \cdot m_j$$

Dabei stehen m_y und m_j für das arithm. Mittel des unabh. bzw des j -ten abh. Merkmals.

Berechnung des Bestimmtheitsmaßes:

$$R^2 := \sum_{i=1}^k \text{beta}_i \cdot r_{iy}$$

Der multiple Korrelationskoeffizient ergibt sich als die positive Wurzel aus dem Bestimmtheitsmaß.

Die Verfahren zur Durchführung der Tests sind der einschlägigen Fachliteratur zu entnehmen.

PROC haeufigkeit (INT CONST spalte, TEXT VAR intervallgrenzen,
 BOOL CONST links dabei, BOOL CONST diskret) :

Erstellt für die übergebenen Intervalle bzw. für die diskreten Werte eine Häufigkeitsauszählung der angegebenen Spalte.

PROC histogramm (INT CONST spalte, TEXT VAR intervallgrenzen,
 BOOL CONST links dabei,
 BOOL CONST diskret) :

Erstellt für die übergebenen Intervalle bzw. für die diskreten Werte die graphische Häufigkeitsdarstellung der angegebenen Spalte. 'intervallgrenzen' muß ein Text der Länge 8 * Anzahl der angegebenen Intervallobergrenze sein. Jeweils 8 Zeichen des Textes müssen die interne REAL - Darstellung einer Intervallgrenze beinhalten (siehe Anhang, Prozedur 'replace').

5. Dialog

Das Paket "dialog" liest bei jedem Aufruf einer DASYS Funktion mit Hilfe von MASK alle Eingaben ein, die für diese Funktion benötigt werden. Für jede DASYS Funktion existiert daher ein Formular, das beim Aufruf der entsprechenden Funktion gezeigt wird und vom Benutzer auszufüllen ist. Möchte man eine angewählte Funktion doch nicht ausführen, so kann man sie durch Drücken von <ESC> wieder verlassen. Ansonsten kann man auf jedes Eingabefeld positionieren und dort Eingaben eintragen oder schon vorhandene Eingaben verbessern. Glaubt man, alle nötigen Eingaben in das Formular eingetragen zu haben, so kann man die Formulareingabe durch Drücken von <TAB> aus dem letzten Eingabefeld heraus oder durch Drücken von <CR> aus der letzten Eingabezeile heraus verlassen (Eine genaue Beschreibung der Steuerzeichen zum Positionieren und zum Verbessern von Eingaben in einem Formular entnehme man einer MASK-Beschreibung). Nun werden die Eingaben des Benutzers auf ihre Richtigkeit überprüft. Sind Fehler in den Eingaben, so wird der Cursor auf das fehlerhafte Feld positioniert und in der untersten Bildschirmzeile wird ein entsprechender Kommentar gezeigt, der den Fehler näher erläutert. Mögliche Fehler sind z.B.:

- Eingabefeld nicht ausgefüllt
- In ein Zahlenfeld Buchstaben eingetragen
- Auswertung von nicht existenten Merkmalen

Die fehlerhaften Daten müssen nun berichtigt werden. Dabei können aber auch wieder alle anderen Eingaben geändert werden.

Sind alle Eingaben korrekt, so wird die entsprechende Prozedur aus dem Paket "statistik funktionen" aufgerufen, nachdem die Parameter, die dieser Prozedur übergeben werden müssen, aus der Formulareingabe des Benutzers zusammengestellt wurden.

Die Ausgabe der Ergebnisse geschieht auf die Pinwand, die vor dem Aufruf der Prozedur aus dem Paket "statistische funktionen" gelöscht wurde. Die komplette Pinwand mit den Ergebnissen der DASYS-Funktion wird nun an das Ende einer EUMEL-Datei ("ausgabe") angehängt. Diese Datei kann der Benutzer dann vom DASYS-Monitor aus drucken.

5.1. Schnittstelle des Paketes "dialog"

Das Paket "dialog" stellt folgende ELAN – Prozeduren zur Verfügung:

PROC basiswerte :

Berechnung der Kenngrößen (arithmetisches Mittel, Standardabweichung, Varianz, Minimum, Maximum, Median und Summe). Es werden zehn Eingabefelder angeboten in die jeweils ein auszuwertendes Merkmal einzutragen ist. Es muß mindestens eins der zehn Felder ausgefüllt werden, es können aber Felder leergelassen werden. Alle eingegebenen Merkmale müssen in den aktuellen Daten existieren.

PROC häufigkeit diskret :

Häufigkeitsauszählung bei diskretem Merkmal. Das Eingabeformular besteht aus einem Eingabefeld, in das der Name des auszählenden Merkmals eingetragen werden muß.

PROC häufigkeit stetig (BOOL CONST links dabei) :

Häufigkeitsauszählung bei stetigem Merkmal. In das Eingabeformular ist zunächst der Name des auszählenden Merkmals einzutragen. Nun müssen die Intervalle, für die die Auszählung vorgenommen werden soll, eingegeben werden. Dieses kann auf zwei verschiedene Arten geschehen:

1. Eingabe von äquidistanten Intervallgrenzen
oder
2. Eingabe der einzelnen Intervallgrenzen.

Es müssen also entweder die Eingabefelder für die Werte "erste obergrenze", "länge der intervalle" und "anzahl der intervalle" ausgefüllt werden oder man gibt mindestens eine Intervallobergrenze in das Feld der zwanzig Intervallgrenzen ein. Dabei ist zu beachten, daß die Intervallobergrenzen in aufsteigender Folge sortiert eingegeben werden müssen.

Für den Fall der Eingabe von äquidistanten Intervallgrenzen ergeben sich die Obergrenzen der Intervalle zu:

$$g, g + 1 \cdot l, g + 2 \cdot l, g + 3 \cdot l, \dots, g + n \cdot l$$

Dabei bedeutet:

- g – Erste endliche Intervallgrenze
- l – Länge der Intervalle ($l > 0$)
- n – Anzahl der Intervalle

PROC histogramm diskret :

Graphische Häufigkeitsdarstellung zu diskretem Merkmal. Das Eingabeformular besteht aus einem Eingabefeld, in das der Name des auszuwertenden Merkmals eingetragen werden muß.

PROC histogramm stetig (BOOL CONST links dabei) :

Graphische Häufigkeitsdarstellung von stetigen Werten. In das Eingabeformular ist zunächst der Name des auszuzählenden Merkmals einzutragen. Nun müssen die Intervalle, für die die Auszählung vorgenommen werden soll, eingegeben werden. Dieses kann auf zwei verschiedene Arten geschehen:

1. Eingabe von äquidistanten Intervallgrenzen
oder
2. Eingabe der einzelnen Intervallgrenzen.

Es müssen also entweder die Eingabefelder für die Werte "erste obergrenze", "laenge der intervalle" und "anzahl der intervalle" ausgefüllt werden oder man gibt mindestens eine Intervallobergrenze in das Feld der zwanzig Intervallgrenzen ein. Dabei ist zu beachten, daß die Intervallobergrenzen in aufsteigender Folge sortiert eingegeben werden müssen.

Für den Fall der Eingabe von äquidistanten Intervallgrenzen ergeben sich die Obergrenzen der Intervalle zu:

$$g, g + 1 \cdot l, g + 2 \cdot l, g + 3 \cdot l, \dots, g + n \cdot l$$

Dabei bedeutet:

- g - Erste endliche Intervallgrenze
- l - Länge der Intervalle ($l > 0$)
- n - Anzahl der Intervalle ($n \geq 0$)

PROC korrelation :

Das Eingabeformular für die Berechnung der Korrelationskoeffizienten besteht aus zehn Eingabefeldern, in die die Namen der Merkmale einzutragen sind, die ausgewertet werden sollen. Wie bei den Kenngrößen können Eingabefelder leer gelassen werden, es muß aber mindestens ein Merkmal eingegeben worden sein. Es ist auch darauf zu achten, daß kein Merkmal mehrfach eingegeben wird.

PROC regression :

Das Eingabeformular für die Regression besteht aus zehn Eingabefeldern: eines für den Namen des abhängigen Merkmals; neun weitere, in die unabhängige Merkmale einzutragen sind. Wie vorher beschrieben, können Eingabefelder leer gelassen werden, es muß aber mindestens ein abhängiges Merkmal eingegeben worden sein. Es ist auch darauf zu achten, daß kein Merkmal mehrfach eingegeben wird.

PROC dateiname einlesen :

Mit Hilfe eines Eingabeformulars wird der Name eines Textfiles erfragt, der die Daten enthält. Diese Datei muß schon existieren. Aus dem Monitor kann man entweder den Editor aufrufen, um diese Datei zu verändern oder man kann die Daten dieser Datei als aktuelle Werte einlesen (auf diese eingelesenen Daten beziehen sich alle Berechnungen).

TEXT PROC dateiname :

Liefert den mit 'dateiname einlesen' eingelesenen Dateinamen.

PROC kommando :

Liest einen Text ein, der vom Elan-Compiler übersetzt und dann ausgeführt wird. So kann man z.B. Daten und Ausgaben archivieren, den Eumel-Drucker auf bestimmte Werte einstellen

Anhang

- die Prozedur

PROC replace (TEXT VAR text, INT CONST index, REAL CONST code) :

Ersetzt den Teilttext von Position ($8^* \text{ index} - 7$) bis (8^* index) in 'text' durch die interne REAL - Darstellung von 'code'.

- die Prozedur

PROC replace (TEXT VAR text, INT CONST index, value) :

Ersetzt den Teilttext von Position ($2^* \text{ index} - 1$) bis Position (2^* index) in 'text' durch die interne INT - Darstellung von 'value'.

- der Operator

INT OP ISUB (TEXT CONST text, INT CONST index) :

Liefert den INT - Wert des Teilttextes von Position ($2^* \text{ index} - 1$) bis (2^* index) aus 'text'.

- der Operator

REAL OP RSUB (TEXT CONST text, INT CONST index) :

Liefert den REAL - Wert des Teilttextes von Position ($8^* \text{ index} - 7$) bis (8^* index) aus 'text'.

Beispiel:

```
PROC trage realwert ein (TEXT VAR destination, INT CONST index,  
REAL CONST value) :
```

```
WHILE length (destination) < index * 8 REP
```

```
destination CAT 8* ""0"
```

```
PER; replace (destination, index, value);
```

```
END PROC trage realwert ein;
```

```
REAL PROC hole realwert (TEXT CONST source, INT CONST index) :
```

```
source RSUB index
```

```
END PROC hole realwert;
```