# Beschreibung
# PROGRAMMKASSETTE R 0137
# WISSENSCHAFT UND TECHNIK
# KLEINCOMPUTER robotron Z 9001 / Z 9002

Die Seite A der PROGRAMMKASSETTE R0137 enthält 4 BASIC-Programme zur mathematischen Statistik und ein Programm mit allgemein verwendbaren BASIC-Unterprogrammen: zur Erzeugung von unterschiedlich verteilten Zufallszahlen.

**Kassetteninhalt (Seite A)**

| Programm-name | Kurzbezeichnung | Länge, ca in byte | Zählerstand[1)] |
|---|---|---|---|
| R+VARANA | Varianzanalyse | 12100 | … |
| R+KTEST | Kolmogorov-Anpassungstest | 8800 | ... |
| R+CLUST | Clusteranalyse | 12800 | ... |
| R+ZUFALL | Erzeugung von Zufallszahlen für verschiedene Verteilungen | 8800 | ... |
| R+zufall | BASIC-Unterprogramme zur Erzeugung von Zufallszahlen verschiedener Verteilungen | 1950 | … |
| R+VARANA | Varianzanalyse | 12100 | ... |
| R+KTEST | Kolmogorov-Anpassungstest | 8800 | … |
| R+CLUST | Clusteranalyse | 12800 | ... |
| R+ZUFALL | Erzeugung von Zufallszahlen für verschiedene Verteilungen | 8800 | ... |
| R+zufall | BASIC-Unterprogramme zur Erzeugung von Zufallszahlen verschiedener Verteilungen | 1950 | ... |

---
[1)] Bitte den jeweiligen Zählerstand selbst ermitteln und eintragen.
Der Programmanfang ist am Vorton (etwa 5 Sekunden) der Programme zu erkennen.

**VEB ROBOTRON-MESSELEKTRONIK >OTTO SCHÖN< DRESDEN**
**DDR-8012 Dresden, Lingnerallee 3, Postschließfach 211**
11/85a
lt G 353-63/D/85
III/9/319 (M)



R+VARNA

Kurzbezeichnung: Varianzanalyse

Voraussetzungen: Z 9001: BASIC-Modul oder RAM Erweiterungsmodul erforderlich
Z 9002: keine

Inhaltsbeschreibung

Die Varianzanalyse ist ein statistisches Analyseverfahren zur quantitativen Untersuchung von Einflüssen (Effekte) eines oder mehrerer Faktoren auf Versuchsergebnisse. Das vorliegende Programm ist für Modelle mit "festen Effekten" (vgl. z.B. Nollau,V., Statistische Analysen Leipzig 1975) ausgelegt. Dabei kann für eine Einfluggröße die EINFACHE KLASSIFIKATION mit gleicher oder ungleicher Anzahl von Wiederholungen und im Falle zweier Einflußgrößen die KREUZKLASSIFIKATION verwendet werden. Bei der Kreuzklassifikation liegen 3 Varianten vor:
a) ohne Wechselwirkung (d.h. ohne Wiederholungen)
b) mit Wechselwirkungen und gleicher Anzahl von Wiederholungen (balanziertes Modell)
c) mit Wechselwirkungen und ungleicher Anzahl von Wiederholungen (quasi balanziertes Modell)

Im Programm schließt sich der Ausgabe der Streuungszerlegungstabelle die Hypotheseprüfung auf Wirkung der Einflußgrößen in den Stufen an. Je nach gewählter Klassifikation werden die entsprechenden Testgrößen berechnet.

Hinweise zur Programmabarbeitung

- Nach Auswahl der Klassifikation werden die Anzahl der Stufen und Wiederholungen sowie anschließend (wahlweise über Tastatur oder von Kassette) die Meßwerte eingegeben.
  Bei Dateneingabe von Kassette sind in jeder Stufe die Meßwerte (Wiederholungen) in Form eines Feldes Y der Länge MX bereitzustellen, wobei MX die maximale Anzahl von Wiederholungen pro Stufe angibt. Man beachte daß die Nichtübereinstimmung von „maximaler Anzahl von Wiederholungen" und Feldlänge MX von Y zu Eingabefehlern führt.

- Fehlermeldungen erscheinen bei ungeeigneten Datenstrukturen:
  - lediglich eine Wiederholung pro Stufe bei EINFACHER KLASSIFIKATION
  - Quasi-Balanzierungsbedingung verletzt
  - Reststreuung MQR = 0

- Auf Grund der sequentiellen Verarbeitung der Meßwertdaten ist eine Wiederholung der Rechnung nur nach erneutem RUN und erneuter Dateneingabe möglich.

R+CLUST

Kurzbezeichnung: Clusteranalyse

Voraussetzungen: Z 9001: BASIC-Modul und ein RAM-Erweiterungsmodul erforderlich
Z 9002: ein RAM-Erweiterungsmodul erforderlich

Inhaltsbeschreibung

Das Programm ermöglicht die Durchführung verschiedener Clusteranalyse-Algorithmen (agglomerative und divisive Verfahren, Leader-Algorithmus)(vgl. z.B. Späth, H. Clusteranalyse-Algorithmen zur Objektklassifizierung, Oldenbourg 1977). Bei agglomerativen Verfahren wird von M gegebenen Objekten ausgegangen, die M einelementigen Clustern entsprechen. Schrittweise werden jeweils die beiden Cluster mit minimalem Abstand vereinigt und die Abstände zu allen anderen Clustern neu berechnet. Dabei sind sieben verschiedene Abstandsfunktionen wählbar. Werden Cluster p und Cluster q zum Cluster k vereinigt, dann können die neuen Abstände $d_{ki}$ vom Cluster k zu allen anderen Clustern i wie folgt berechnet werden:

1. $d_{ki} := \min(d_{pi}, d_{qi})$
2. $d_{ki} := \max(d_{pi}, d_{qi})$
3. $d_{ki} := ½ (d_{pi} + d_{qi})$
4. $d_{ki} := ½ (d_{pi} + d_{qi}) – ¼ d_{pq}$
5. $d_{ki} := \frac{1}{m_p + m_q}(m_p d_{pi} + m_q d_{qi})$
6. $d_{ki} := \frac{m_p}{m_p + m_q} d_{pi} + \frac{m_q}{m_p + m_q} d_{qi} + \frac{m_p m_q}{m_p + m_q} d_{pq}$
7. $d_{ki} := \frac{1}{m_p + m_q + 1}\left[(m_p + m_i)d_{pi} + (m_q + m_i)d_{qi} - m_i d_{pq}\right]$

wobei $m_j$ die Anzahl der Objekte im Cluster j ist.
Diese sieben Abstandsfunktionen entsprechen den sieben Verfahren des entsprechenden Programmteils. Jedes Verfahren endet, wenn alle M Objekte ein Cluster bilden.

Das divisive Verfahren geht von genau einem Cluster aus, in dem alle M Objekte enthalten sind. Dieses Cluster wird zunächst willkürlich in zwei Cluster aufgespalten. Mit Hilfe des Kmeans-Algorithmus werden nun die Clustermitglieder zwischen den beiden Clustern so ausgetauscht, daß die innerhalb der beiden Cluster gebildeten Summen der quadrierten Abstände der Clustermitglieder vom jeweiligen Clusterschwerpunkt minimiert werden. Im weiteren Verlauf des Verfahrens wird auf jeder Stufe jeweils ein Cluster mit der größten Abweichungsquadratsumme ausgewählt, auf das das beschriebene Verfahren neu angewendet wird. Der Algorithmus wird beendet, wenn die vorgegebene Clusterzahl erreicht ist oder alle Quadratsummen innerhalb der Cluster identisch Null sind.



Der Leader-Algorithmus ist ein rein heuristisches Verfahren. Es werden vom Nutzer N Objekte als Leader (Führungselemente) ausgezeichnet und damit N Cluster gebildet, indem die Objekte, die keine Leader sind, dem Leader zugeordnet werden, von dem sie den kleinsten Euklidischen Abstand besitzen. Danach wird auf die N entstandenen Cluster der Kmeans-Algorithmus angewendet, d.h., es werden die Abstandsquadratsummen innerhalb der Cluster durch sukzessiven Austausch von Clustermitgliedern minimiert.

Es ist generell zu beachtend daß alle vorliegenden Cluster-Analyse-Algorithmen heuristischer Natur sind. Das impliziert z .B. die Möglichkeit, daß die Reihenfolge der Objekte Einfluß auf die Struktur der Lösung hat. Außerdem erfordert es unbedingt, daß man erst einen der jeweiligen Problemstellung angepaßten Algorithmus auswählt und danach dessen Ergebnisse interpretiert. Es ist im jedem Fall unzweckmäßig, willkürlich ein Verfahren auszuwählen und dessen Ergebnisse interpretieren zu wollen.

Hinweise zur Programmabarbeitung

- Das Programm kenn maximal 45 Objekte mit höchstens 12 Variablen verarbeiten.
- Die Eingabe der Objekt- oder Abstandsmatrix kann wahlweise über Tastatur oder von Kassette erfolgen. Bei Dateneingabe von Kassette ist entweder eine Objektmatrix der Dimension (M, Q) (M: Anzahl der Objekte; Q: Anzahl der Merkmale) oder eine Abstandsmatrix der Dimension (M, M) bereitzustellen. Man beachte, daß die Nichtübereinstimmung der Werte M und Q und der Dimension des einzugebenden Feldes zu Eingabefehlern führt.

- Das divisive Verfahren und der Leader-Algorithmus können nur auf Grundlage einer Objektmatrix arbeiten. Bei der Berechnung der Abstandsmatrix aus der Objektmatrix im Agglomerativen Verfahren wird der Euklidische Abstand verwendet, deshalb können in diesem Fall nur Objekte mit metrischen Daten sinnvoll verarbeitet werden. Wird dagegen eine Abstandsmatrix beim Agglomerativen Verfahren eingegeben, so sind auch nominale, ordinale und gemischte Daten für die Objekte sinnvoll verwendbar.

- Ausgabe für agrlomerative Verfahren:
  Für jedes Verfahren werden die Werte I, A(I), B(I) und H(I) in Form einer Tabelle ausgegeben.
  Bedeutung dieser Werte:
  - I ist die Stufe des Verfahrens
  - A(I), B(I) sind Clusternummern
  - H(I) ist der Abstand zwischen dem Cluster mit der Nummer A(i) und dem Cluster mit der Nummer B(I)

  In der I-ten Stufe des Verfahrens werden die Cluster mit den Nummern A(I) und B(I) zu dem Cluster mit der Nummer A(I) vereinigt, dabei beträgt die Entfernung zwischen den beiden Clustern H(I). Aus den Vektoren A, B und H läßt sich ein Dendrogramm erstellen. Falls die Anzahl der Objekte M kleiner als 22 ist, so wird nach jedem Verfahren das entsprechende Dendrogramm auf dem Bildschirm dargestellt.

- Ausgabe für divisives Verfahren:
  Nach jeder Aufspaltung werden die Clusternummern der Objekte und die Gesamtabweichungsquadratsumme ausgegeben. Dabei ist der j-te ausgegebene Wert die Clusternummer des j-ten Objekts. Für die Abschlußkonfiguration wird noch folgendes ausgegeben:
  a) Clusternummern in Reihenfolge ihrer Aufspaltung
     (Wird ein Cluster gespalten, so hat eines der neuen Cluster die Nummer des gerade gespaltenen Clusters und das andere, neue Cluster hat die kleinste noch nicht an ein Cluster vergebene Nummer. Kennt man also die Clusternummern in der Regenfolge ihrer Aufspaltung, kann man sich leicht das zugehörige Dendrogramm erzeugen.)
  b) Abweichungsquadratsumme innerhalb der Cluster und Gesamtabweichungsquadratsumme
  c) Wahlweise kann für die Abschlußkonfiguration die Zuordnung der Objekte zu den Clustern ausgegeben werden.

- Ausgabe für Leader-Algorithmus
  Nach dem Anlagern aller Objekte an die Leader werden die Clusternummern der Objekte ausgegeben. Dabei ist der j-te ausgegebene Wert die Clusternummer des j-ten Objekts. Wahlweise kann die Zuordnung der Objekte zu. den Clustern ausgegeben werden.
  Nach der Anwendung des Kmeans-Algorithmus werden wiederum die Clusternummern der Objekte ausgegeben sowie die Abweichungsquadratsummen innerhalb der Cluster 1 bis N und die Gesamtquadratsumme. Wahlweise kann die Zuordnung der Objekte zu den Clustern ausgegeben werden.

R+KTEST

Kurzbezeichnung: Kolmogorov-Anpassungstest

Voraussetzungen: Z 9001: BASIC-Modul und ein RAM-Erweiterungsmodul erforderlich
Z 9002: keine

Inhaltsbeschreibung

Es wird die Hypothese $H_0$ geprüft, ob die (unbekannte) Verteilungsfunktion F einer Grundgesamtheit, der eine Stichprobe vom Umfang N entnommen wurde, gleich einer vorgegebenen hypothetischen Verteilungsfunktion $F_0$ ist.
Für den Rest dieser Hypothese

$$H_0: F = F_0$$

wird die Testgröße $TG = \max_{t \in R} |F_0(t) - FE(t)|$ gebildet, wobei FE(t) die (aus der Stichprobe gewonnene) empirische Verteilungsfunktion (an der Stelle t) ist.
Die Hypothese wird abgelehnt, falls

$$\sqrt{N} \cdot TG > K_{1-a}$$



gilt, wobei $K_{1-\alpha}$ das Quantil der Kolmogorov-Verteilung zum Signifikanzniveau a ist. Anderenfalls ist gegen die Hypothese nichts einzuwenden (vgl. z.B. Müller, P.H., Lexikon der Stochastik, Berlin 1975).

Hinweise zur Programmabarbeitung

- Nach dem Stichprobenumfang N (N > 1) sind die Stichprobenwerte (nicht notwendig der Größe nach geordnet) einzugeben (eine Korrekturmöglichkeit erlaubt die Redigierung fehlerhafter Dateneingaben) bzw. von Kassette einzulesen.

- Bei Dateneingabe von Kassette muß das einzugebende Feld die Länge N (N: Stichprobenumfang) besitzen. Man beachte daß die Nichtübereinstimmung von Stichprobenumfang und Feldlänge zu Eingabefehlern führt.

- In einem Menübild werden verschiedene hypothetische Verteilungsfunktionen (Normalverteilung, gleichmäßige Verteilung, Weibullverteilung, Exponentialverteilung) sowie die Möglichkeit der nutzereigenen Definition einer beliebigen stetigen Verteilungsfunktion $F_0(t)$, die natürlich den Bedingungen

  $$\lim_{t \to -\infty} F_0(t) = 0, \qquad \lim_{t \to \infty} F_0(t) = 1$$

  $$F_0(t_1) \le F_0(t_2) \text{ für } t_1 \le t_2$$

  genügen muß, angeboten.
  Als Standard ist die Normalverteilung mit dem Erwartungswert μ (Standard: 0) und der Streuung σ² (Standard: 1) gesetzt. Die gleichmäßige Verteilung ist standardmäßig auf das Intervall [0, 1] konzentriert. Bei den anderen Verteilungen ist die Eingabe der Parameterwerte notwendig.

- Die Durchführung des Tests wird abgelehnt, falls die Stichprobenwerte und die gewählte hypothetische Verteilung nicht verträglich sind, d.h., bei der gleichmäßigen Verteilung liegen Stichprobenwerte außerhalb des vorgegebenen Intervalls [A, B], oder trotz negativer Stichprobenwerte wurde eine Weibull- oder Exponentialverteilung als hypothetische Verteilung gewählt.

- Für kleinen Stichprobenumfang kann eine Verschärfung des Tests dadurch erfolgen, daß anstelle des (vom Programm verwendeten) Quantils der Kolmogorov-Verteilung das entsprechende Quantil der exakten Verteilung (s. z.B. Müller/Neumann/Storm: Tafeln der mathematischen Statistik, Fachbuchverlag Leipzig 1973) verwendet wird.

- Bei größerem Stichprobenumfang dauert die Sortierung der Stichprobe ggf, eine längere Zeit.

- Bei "sehr ungeeigneter" Wahl der speziellen Verteilungsparameter (z.B. bei Wahl eines Erwartungswertes der Normalverteilung, der kleiner als das Stichprobenminimum ist) treten ggf. numerische Fehler auf.

- Fehler aus einer vom Nutzer falsch definierten Verteilungsfunktion F (z.B. $\lim_{t\to-\infty} F(t) \neq 0$ oder $\lim_{t\to\infty} F(t) \neq 1$ werden i.a. nicht abgefangen.
- Sämtliche Hilfsgrößen in der vom Nutzer zu definierenden Verteilungsfunktion müssen mit dem Buchstaben T beginnen.
- Eine Veränderung der durch den Nutzer definierten Verteilungsfunktion unter Beibehaltung der Stichprobenwerte ist aus rechentechnischen Gründen (EDIT-Modus) nicht möglich.

Beispiel

Als hypothetische Verteilungsfunktion soll die Funktion

$$F_0(t) = \sin(t),\ t \in (0, \frac{p}{2})$$

benutzt werden.
Nach Auswahl der Kennziffer 5 im Menübild sind z.B. die folgenden Anweisungen einzugeben:

```
3020 IF T < 0THEN T1 = 0: GOTO 3050
3030 IF T > ATN(1)*2 THEN T1 = 1:GOTO 3050
3040 T1 = SIN(T)
3050 T = T1
```

Danach ist das Programm entsprechend den Bildschirmaufforderungen fortzusetzen.

R+ZUFALL

Kurzbezeichnung: Erzeugung von Zufallszahlen für verschiedene Verteilungen

Voraussetzungen: Z 9001: BASIC-Modul oder RAM-Erweiterungsmodul erforderlich
Z 9002: keine

Inhaltsbeschreibung

Auf der Grundlage des durch BASIC bereitgestellten Zufallszahlengenerators RND für gleichmäßig auf dem Intervall (0, 1) verteilte Zufallszahlen werden Zufallszahlen für beliebige gleichmäßig verteilte, normalverteilte, exponentialverteilte, weibullverteilte oder binomialverteilte Zufallsgrößen erzeugt.

Hinweise zur Programmabarbeitung

- Bei der Eingabe der Parameter sind die entsprechenden Voraussetzungen (z.B. Positivität) zu beachten. Der Parameter N der Binomialverteilung ist auf N = 500 beschränkt.
- Bei der Erzeugung weibullverteilter Zufallszahlen können OV-Errors auftreten. In einem solchen Fall ist die Rechnung zu wiederholen.



- Es können maximal 500 Zufallszahlen erzeugt werden. Bei der Datenausgabe auf Kassette ist die Länge des Zufallszahlenfeldes durch die Auswahl der erzeugten Zufallszahlen bestimmt.

Beispiel

Es sollen 50 exponentialverteilte Zufallszahlen (mit Parameter T) erzeugt und auf Kassette ausgegeben werden.
Nach Wahl der Exponentialverteilung und Eingabe des Parameters T erfolgt nach Erzeugung der Zufallszahlen die Aufforderung

```
CSAVE* "NAME"; XX
```

(wobei NAME eine beliebige Bezeichnung ist)
In einem anderen Programm können diese Zufallszahlen z.B. durch die Anweisungsfolge

```
100 DIM A(50)
110 CLOAD*"NAME"; A
```

wieder eingelesen werden.

R+zufall

Kurzbezeichnung: BASIC-Unterprogramme zur Erzeugung von Zufallszahlen verschiedener Verteilungen

Voraussetzungen: eigenes BASIC-Programm des Anwenders erforderlich

Inhaltsbeschreibung

"R+zufall" enthält vier BASIC-Unterprogramme, die zum Erzeugen von normalverteilten, exponentialverteilten, weibullverteilten und binomialverteilten Zufallszahlen genutzt werden können.

Hinweise zur Programmabarbeitung

- Die Datei "R+zufall", die mittels des Kommandos CLOAD "R+zufall" in den Speicher geladen wird, gliedert sich in
  Unterprogramm normalverteilte Zufallszahlen,
  Unterprogramm exponentialverteilte Zufallszahlen,
  Unterprogramm weibullverteilte Zufallszahlen,
  Unterprogramm binomialverteilte Zufallszahlen.
- Vorraussetzung für die nutzung der Unterprogramme ist die vorherige Belegung der entsprechenden Parameter im Hauptprogramm des Anwenders. Tabelle 1 gibt eine Übersicht über die Nutzungsbedingungen der Unterprogramme.
- Zu beachten ist dabei, daß die angegebenen Hilfsvariablen in den jeweiligen Unterprogrammen wertmäßig verändert werden.
  Die Variablen R1 und R2 dürfen im Anwenderprogramm nicht benutzt werden.
- Bei der Erzeugung weibullverteilter Zufallszahlen können OV-Errors auftreten. In einem solchen Fall ist die Rechnung zu wiederholen.

Tabelle 1

| Unterprogramm zur Erzeugung von Zufallszahlen folgenden Typs | Aufruf | Eingangsgrößen | Ergebnisgröße (Zufallszahl) | Hilfsvariable |
|---|---|---|---|---|
| normalverteilt: | GOSUB 40000 | P7 = 1: N(0,1)-Verteilung<br>2: N(M7,S7) – " | X7 | C1, C2, C3, C4, R1, R2, R3, R4, R5, R6, R7, R8 |
| Dichte: $f(x) = \frac{1}{\sqrt{2p}S7} \exp\left(\frac{(x-M7)^2}{2 \cdot S7}\right)$ | | M7: Mittelwert<br>S7: Streuung | | |
| exponentialverteilt: | GOSUB 40500 | L8: Parameter (L8 > 0) | X8 | |
| Dichte: $f(x) = \begin{cases} L8 \cdot \exp(-L8 \cdot x) & ; x > 0 \\ 0 & ; sonst \end{cases}$ | | | | |
| weibullverteilt: | GOSUB 40600 | B9: Formparameter (B9 > 0)<br>P9: Skalenparameter (P9 > 0) | X9 | |
| Dichte: $f(x) = \begin{cases} \frac{B9}{P9}(\frac{x}{P9})^{B9-1} \exp\left(-(\frac{x}{P9})^{B9}\right) & ; x > 0 \\ 0 & ; sonst \end{cases}$ | | | | |
| binomialverteilt: | GOSUB 40700 | N1, P1: Parameter (im Anwenderprogramm ist ein Feld PP(N1) zu vereinbaren) | X1 | R1, R2, R3, R4, R5 |
| Einzelwahrscheinlichkeit $P(X=k) = \binom{N1}{k} P1^k (1-P1)^{N1-k}$ | | | | |



Beispiel

In einem Anwenderprogramm sollen K binomialverteilte Zufallszahlen (mit den Parametern N1 = 30, P1 = 0.6) erzeugt werden.
Nach dem Kommando

```
CLOAD "R+zufall"
```

sind die Zufallszahlenunterprogramme (zusätzlich zum Anwenderprogramm) im Speicher verfügbar. Das Anwenderprogramm kann z.B. folgende Gestalt haben:

```
100 N1 = 30:P1 = 0.6
110 DIM PP(N1)
120 FOR I=1 TO K
130 GOSUB 40700
…
500 NEXT I
```

In den Zeilen von 140 bis 490 wird die jeweils erzeugte Zufallszahl X1 weiter verarbeitet.