Regressionsanalyse

POOL F2-129
POOL F2-130

Funktion

Für eine Anzahl von Wertesätzen, die jeweils aus n unabhängigen Werten bestehen, werden die Regressions-Koeffizienten berechnet und die entsprechenden statistischen Kenngrößen ermittelt. Das Programm baut sich aus einer Anzahl von Unterprogrammen auf, die einerseits der normalen LGP-30-Bibliothek entnommen sind und andererseits speziell angefertigt wurden.

Die Eingangsdaten können in Festkommadarstellung vorliegen (F2-129) oder als Gleitkommazahlen gegeben sein (F2-130). In beiden Fällen wird aber im Gleitkomma gerechnet, lediglich die Ausgabe der Werte ist dann wieder in den entsprechenden Formaten. Nur Zwischenergebnisse erscheinen bei beiden Programmen in Gleitkommadarstellung.

Im einzelnen leisten die speziellen statistischen Unterpläne folgendes:

1. Unterplan UP 1 bildet die obere Dreiecksmatrix aus N Sätzen von je n Werten X_1, X_2 ... X_n nach folgendem Schema

$$
\begin{array}{lllll}
N\ \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \cdots\cdots & \Sigma X_n \\
\ \ \Sigma X_1^2 & \Sigma X_1X_2 & \Sigma X_1X_3 & \cdots\cdots & \Sigma X_1X_n \\
 & \Sigma X_2^2 & \Sigma X_2X_3 & \cdots\cdots & \Sigma X_2X_n \\
 & & \Sigma X_3^2 & \cdots\cdots & \Sigma X_3X_n \\
 & & & & \vdots \\
 & & & & \Sigma X_n^2
\end{array}
$$

Eingabe: N Sätze von je n Werten X_1, X_2 ... X_n

Ausgabe: Druck der ∇-Matrix in Fest- oder Gleitkomma

2. <u>Unterplan UP 2</u> skaliert die ∇-Matrix aus UP 1, falls die Werte X_i mit Maßstabsfaktoren C_i behaftet waren, wenn also statt X_i der Wert $\frac{X_i}{C_i}$ eingelesen wurde. Nach der Skalierung steht die Matrix dann genau wie bei UP 1 im Speicher. Beispielsweise entstehen aus den Elementen

$$\sum \frac{X_1}{C_1}, \quad \sum \frac{X_1 X_2}{C_1 C_2}, \quad \sum \frac{X_2^2}{C_2^2} \quad \text{usw.,}$$

die mit UP 1 gebildet wurden, als Ergebnis von UP 2 die Elemente

$$\sum X_1, \quad \sum X_1 X_2, \quad \sum X_2^2 \quad \text{usw.}$$

<u>Eingabe:</u> Elemente der Matrix aus UP 1; Maßstabsfaktoren C_1, $C_2 \ldots C_n$

<u>Ausgabe:</u> ∇-Matrix in gleicher Form wie bei UP 1

3. <u>Unterplan UP 3</u> berechnet aus den Elementen der Matrix aus UP 2 bzw. UP 1 folgende Größen:

a) Mittelwerte $\bar{X}_i = \frac{1}{N} \sum_{k=1}^{N} X_{i,k}$

b) Mittlere Streuung $\sigma_i = \sqrt{\frac{\sum X_{i,k}^2}{N} - \bar{X}_i^2}$

c) Korrelations-Koeffizienten $R_{i,j} = \frac{\sum X_{i,k} X_{j,k}/N - \bar{X}_i \bar{X}_j}{\sigma_i \cdot \sigma_j}$

Die Indices i, j entsprechen denen der Kreuzprodukte in der ∇-Matrix, an deren Stelle die Elemente $R_{i,j}$ treten. Nur die erste Zeile bleibt unverändert.

4. <u>Unterplan UP 4a</u> ergänzt die ∇-Matrix aus UP3 zu einer symmetrischen $\square$-Matrix.

$$R_{j,i} = R_{i,j} \, .$$

Die Adressen der ersten Elemente der ∇-Matrix und der $\square$-Matrix sind gleich.

5. <u>Unterplan UP 5</u> bildet aus der Matrix $\mathcal{R}$ von UP 4a die inverse Matrix $\mathcal{R}^{-1}$. Falls gewünscht (Sprungtaste nicht gedrückt), kann die inverse Matrix ausgedruckt werden.

6. Unterplan UP 6a berechnet aus der inversen Matrix $\mathcal{R}^{-1}$ der Korrelations-Koeffizienten (UP 5) die Regressions-Koeffizienten b_i der beschreibenden Gleichung und die zugehörigen statistischen Kenngrößen. Hierbei wird vorausgesetzt, daß der Wert für die abhängige Variable an letzter Stelle eines Wertesatzes steht, also identisch ist mit X_n; $X_n \equiv Y$

Im einzelnen werden berechnet:

a) Die β-Gewichte durch Division der Spaltenelemente durch das Hauptdiagonalelement der betr. Spalte

b) Die Regressions-Koeffizienten b_i aus den Beziehungen

$$b_o = \overline{Y} - \overline{X}_1 \frac{\sigma y}{\sigma x_1} \beta_1 - \overline{X}_2 \frac{\sigma y}{\sigma x_2} \beta_2 - \ldots\ldots$$

$$b_1 = \beta_1 \frac{\sigma y}{\sigma x_1}$$

$$b_2 = \beta_2 \frac{\sigma y}{\sigma x_2}$$

usw., wobei die $\beta_1, \beta_2 \ldots \beta_{n-1}$ die β-Gewichte der letzten Spalte (Y-Spalte) sind.

c) Die partiellen Korrelations-Koeffizienten r durch Division des entsprechenden Elements der inversen Matrix $\mathcal{R}^{-1}$ über der Hauptdiagonalen durch die Quadratwurzel aus dem Produkt der beiden Hauptdiagonalglieder in der zugehörigen Spalte und Zeile.

$$r_{ik} = \frac{a_{ik}}{\sqrt{a_{ii} \cdot a_{kk}}}$$

d) Die Mehrfach-Korrelations-Koeffizienten R_i aus der Beziehung

$$R_i = \sqrt{1 - \frac{1}{a_{ii}}} \quad , \qquad \text{insbesondere}$$

$$R = \sqrt{1 - \frac{1}{a_y}}$$

e) Der mittlere Fehler des erwarteten Wertes

$$S = \sigma_y \sqrt{1 - R^2}$$

f) Der allgemeine Mehrfach-Korrelationskoeffizient

$$\hat{R} = \sqrt{1 - (1-R^2) \frac{N-1}{N-n}}$$

g) Der allgemeine mittlere Fehler des erwarteten Wertes

$$\hat{S} = \sigma_y \sqrt{(1-\hat{R}^2) \frac{N}{N-1}}$$

h) Der mittlere Fehler des Regressions-Koeffizienten

$$S_{bi} = \hat{S} \sqrt{\frac{1 - r_{i,n}^2}{N \cdot \sigma_i^2 \; (1 - R_i^2)}}$$

<u>Ausgabe:</u> Ausgeschrieben werden die Werte β_i, b_i, R_i, S_{bi}, R, S, $\hat{R}$, $\hat{S}$

7. <u>Unterplan UP 7</u> benutzt die errechneten Regressions-Koeffizienten b_i, um für jeden der N gegebenen Wertesätze X_i jeweils den Wert der abhängigen Variablen ${}_kY$ zu berechnen, und mit dem gegebenen Wert Y zu vergleichen.

$${}_kY = b_o + \sum_{i=1}^{n-1} b_i X_{i,k} \qquad k = 1, 2 \ldots N$$

$$\Delta = Y - {}_kY$$

<u>Eingabe:</u> N Wertesätze X_1, X_2 ... X_{n-1}, Y in Festkommadarstellung (F2-129) oder in Gleitkommaformat (F2-130).

<u>Ausgabe:</u> Für jeden der N Wertesätze werden die Größen Y, ${}_kY$ und Δ geschrieben.

Übrige Unterprogramme, Speicherbelegung

Außer den oben beschriebenen statistischen Unterplänen werden folgende Standard-Unterprogramme benutzt:

1)	Programmeingabe 10.4	0000-0263
2)	Dateneingabe 11.2	0300-0563
3)	Datenausgabe 12.1a	0600-0850
4)	Gleitkommaumwandlung 25.0R	0900-1163
5)	Gleitkommasystem 24.0	1200-2163
6)	Gleitkomma-Eingabe-Ausgabe 11.6/12.6	2200-2763
7)	Matrix-Inversion GK 29.0	2800-2963

Die statistischen Unterpläne belegen die Speicherplätze 3000-4060. Die Plätze 4100 bis 6131 stehen für Datenspeicherung zur Verfügung. Das genügt für Wertesätze bis n = 32, wobei die Zahl N der Wertesätze beliebig ist und nur durch die Größenordnung der aufgelaufenen Summen begrenzt ist.