

Mathematische Statistik

Einleitung

Innerhalb des Teilnehmer-Rechensystems TR 440 stehen dem Benutzer leistungsfähige Programmsysteme und umfangreiche Programm-Bibliotheken zur Verfügung. Eine dieser Bibliotheken ist die statistische. Sie enthält Lösungen für verschiedene statistische Probleme.

Die Programme sind fast alle in FORTRAN codiert und stehen allen FORTRAN und ALGOL Benutzerprogrammen uneingeschränkt zur Verfügung.

Die einzelnen Programme sind auf eine Zusammenarbeit untereinander abgestimmt. Häufig auftretende Berechnungen werden nur einmal durchgeführt und von den anderen Programmen benutzt (modularer Aufbau). Für größere Programmkomplexe können leicht Steuerprogramme erstellt werden.

Beschreibung der einzelnen Programm-Moduln

Organisationsprogramme
zum Einlesen und Aufbereiten des Datenmaterials.

Zufallszahlen

können zu den bekanntesten Verteilungen und zu beliebigen, vorgegebenen diskreten Verteilungen erzeugt werden.

Elementare Berechnungen

wie Mittelwert, Kovarianzen, höhere Momente und Häufigkeitsauszählungen. Die Häufigkeitsauszählungen können auf Wunsch grafisch als Histogramm oder Summenhäufigkeitskurve dargestellt werden.

Verteilungen und ihre Umkehrfunktionen

für Gauß-, Student, χ^2 -, Fisher-, Binomial- und Poisson-Verteilung.

Konfidenzintervalle

für die Wahrscheinlichkeit einer Binomialverteilung sowie den Mittelwert und die Streuung einer Normalverteilung.

(Als Konfidenzintervall wird ein ein- oder zweiseitiges Zufallsintervall bezeichnet, das mit der Wahrscheinlichkeit P den unbekannt Parameter einer Verteilung enthält.)

Parameter-tests

für die Parameter Mittelwert und Streuung der Normalverteilung.

(Ein Parameter-test testet eine vorzugebende Hypothese für einen Parameter einer bekannten Verteilung.)

Anpassungstests

für Normal-, Gleich-, Poisson-, Binomialverteilung und für vorgegebene Verhältniszahlen. Die Parameter der Verteilungen können unbekannt sein, sie werden dann mit Hilfe der Stichprobenwerte geschätzt.

(Ein Anpassungstest prüft, ob eine Stichprobe aus einer Grundgesamtheit mit einer bestimmten Verteilung stammen kann.)

Einfluß mehrerer Größen auf ein Ergebnis

Für Normalverteilung sind der T-Test (eine Einflußgröße) sowie die ein-, zwei- und drei-dimensionale Varianzanalyse (bis zu 3 Einflußgrößen) vorhanden. Für beliebige Verteilungen und eine Einflußgröße sind die bekannten verteilungsunabhängigen Tests von Mann-Whitney, Kruskal-Wallis, Wilcoxon, Friedman, McNemar, Cochran sowie der χ^2 - und Zeichentest programmiert. Die Baumstruktur dieser Fragestellung ist in dem Diagramm 1 dargestellt.

Abhängigkeit zwischen m-Größen

Korrelation

Für beliebige Verteilungen können der Kontingenzkoeffizient, der Spearmansche- oder der Kendallsche-Korrelationskoeffizient oder die Kendallsche Übereinstimmungszahl berechnet und auf Signifikanz geprüft werden.

Für Normalverteilungen können der Pearsonsche, multiple- oder partielle Korrelationskoeffizient berechnet und auf Signifikanz geprüft werden.

Außerdem kann für eine Stichprobe aus einer Normalverteilung die Korrelationsmatrix berechnet werden.

Regression

Im Unterschied zur Korrelation, bei der nicht gesagt werden kann, welche Größe von welcher abhängt, versucht die Regression einen formelmäßigen Zusammenhang herzustellen. Hierzu ist ein Programm zur multiplen linearen Regressionsanalyse erstellt.

Faktorenanalyse

Sie behandelt das Problem, die wesentlichen Einflüsse auf ein Ergebnis herauszufinden.

Zum Beispiel wurden an m Rechtecken n Eigenschaften (Länge, Breite, Diagonale, Umfang, Fläche, ...) gemessen. Die Faktorenanalyse liefert, daß von den n Eigenschaften k (hier 2) wesentlich sind. In diesem Beispiel würden die Länge und die Breite die größten Faktorenladungen erhalten.

Diskriminanzanalyse

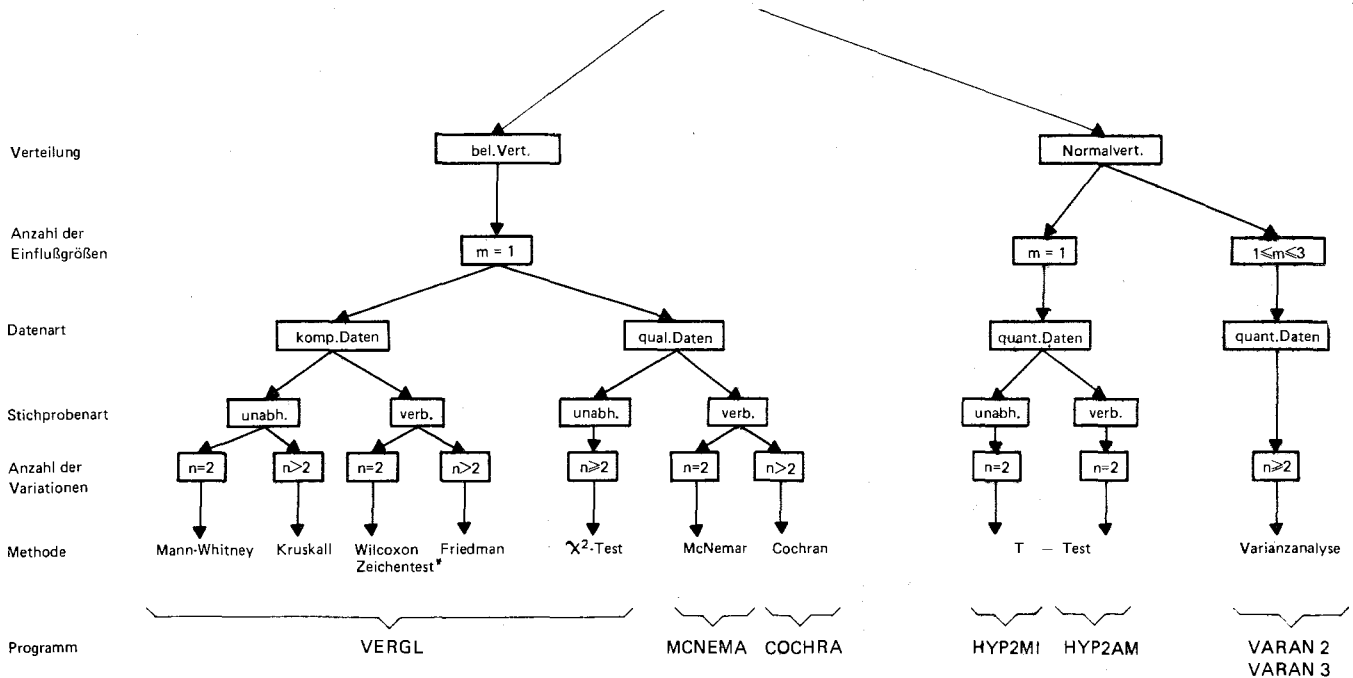
Sie löst die Aufgabe, die Trennebene von beliebig dimensionalen Stichproben zu bestimmen.

Es werden z. B. bei verschiedenen Rassen einige Merkmale gemessen, etwa Größe, Sitzgröße, Nasenbreite und Nasenlänge. Aus diesen Merkmalen bestimmt das Programm Diskriminanzfunktionen, die die verschiedenen Gruppen optimal trennen. Neue Daten könnten dann mit Hilfe dieser Diskriminanzfunktionen in eine der Gruppen eingeordnet werden.

Diagramm 1

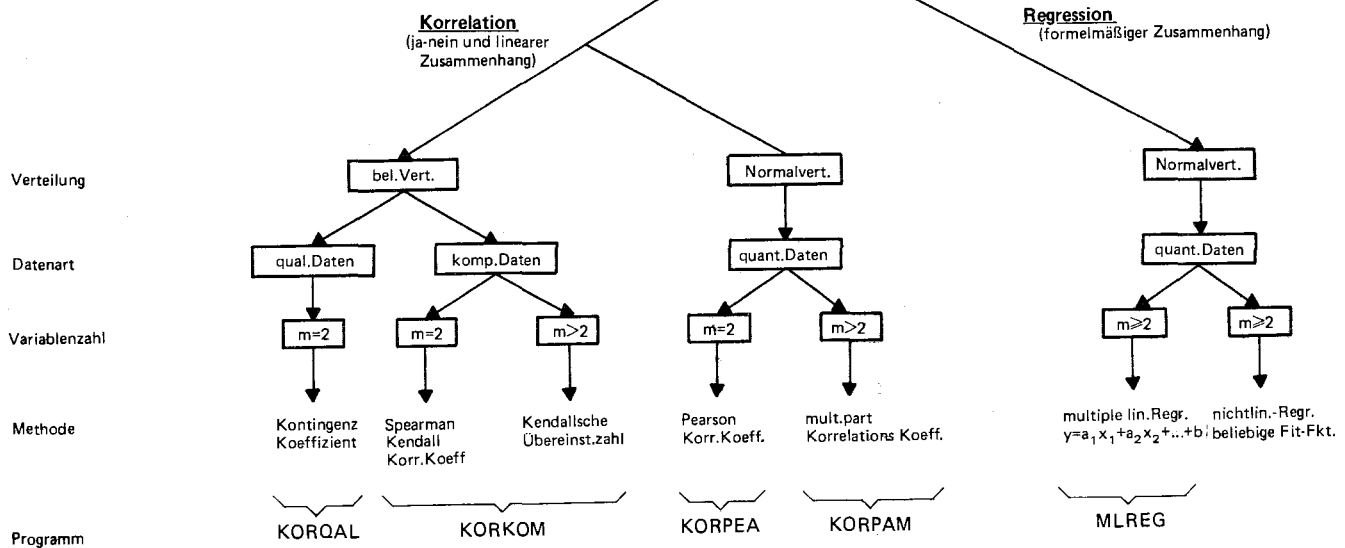
Einfluß von m-Größen in jeweils n-Variationen auf ein Ergebnis

Anwendung: $m = 1$: Haben n -verschiedene Methoden verschiedene Wirkungen? (z. B. 4 Lehrmethoden ... ($n = 4$))
 $m > 1$: Haben m -verschiedene Größen in n_j -Variationen oder eine Kombination von ihnen verschiedene Wirkungen? (z. B. 2 Wassersorten, 3 verschiedene Böden, 4 verschiedene Düngemittel.)
 Ergebnis: Ertrag von Weizen ($m = 3, n_1 = 2, n_2 = 3, n_3 = 4$)



* Zeichentest: nur die beiden Wertepaare müssen komparativ sein

Abhängigkeit zwischen m-Größen



Kontingenz-Koeffizient } gibt nur ja-nein Zusammenhang an
 Korrelations-Koeffizient }
 Übereinstimmungszahl } geben linearen Zusammenhang an

Fragebogenauswertung

Außer den bisher erwähnten Programmen, die sich auf spezielle mathematische Verfahren stützen, existieren in der Programmbibliothek auch Programme zur Auswertung von Fragebogen, die auf Lochkarten abgelocht sind. Bei diesen Programmen werden die Lochpositionen auf den Lochkarten ausgezählt. Die Daten des Fragebogens können dabei als Einfachlochung (numerische Verschlüsselung) oder als sogenannte Binärlochung (jedes Loch der Spalte kann ausgenutzt werden) auf Lochkarten übertragen werden. Es können Gesamtauswertungen, Tabellenauswertungen und Gesamtauswertungen unter Bedingungen vorgenommen werden.

Bei der Gesamtauswertung wird gezählt, wie häufig jede Antwort auftritt (Anzahl der Lochpositionen). Die Tabellenauswertung betrachtet die Daten als mehrdimensionale Tabelle und zählt etwa, wie häufig bestimmte Antworten mit anderen Antworten verknüpft vorkommen. Bei der Gesamtauswertung unter Bedingungen werden bestimmte Bereiche von Lochpositionen nur gezählt, wenn zusätzliche vorgebbare Bedingungen erfüllt sind. So kann man z. B. nach der Anzahl der Wähler einer Partei fragen, die männlich und über 40 Jahre alt sind. Für die Formulierung der Vorgaben bei der Tabellenauswertung und der Bedingungen bei der Gesamtauswertung unter Bedingungen benutzt man eine bestimmte Spezifizierungssprache.