DK 519.2 : 311.1/.2 : 001.4 DEUTSCHE NORM März 1985

Statistische Auswertungen

Zählbare (diskrete) Merkmale

DIN 53 804 Teil 2

Statistical evaluations; countable (discrete) characteristics

Estimations statistiques; caractères discrete dénombrable

Inhalt

1 Anwendungsbereich und Zweck

Die Eigenschaften von Produkten oder Tätigkeiten werden durch Merkmale erfaßt. Den Merkmalsausprägungen werden Werte einer jeweils geeigneten Skale zugeordnet. Die Skalenwerte sind

- bei meßbaren (kontinuierlichen) Merkmalen beliebige reelle Zahlen (als Zahlenwerte von physikalischen Größen)
- bei zählbaren (diskreten) Merkmalen ganze Zahlen (Zählwerte)
- bei Ordinalmerkmalen Eigenschaftskategorien, die einer Rangordnung folgen (z. B. glatt, etwas verknittert, stark verknittert)
- bei Attributmerkmalen Attribute (z. B. vorhanden/nicht vorhanden oder rot/gelb/blau).

Meßbare und zählbare Merkmale werden als quantitative, Ordinalmerkmale und Attributmerkmale als qualitative (beurteilbare) Merkmale bezeichnet. Diese Merkmalsarten entsprechen den Grundbegriffen der Meßtechnik: Messen, Zählen, Sortieren und Klassieren (siehe DIN 1319 Teil 1).

Da es im allgemeinen nicht sinnvoll ist, Merkmalswerte an allen Einheiten einer Grundgesamtheit zu bestimmen, werden Stichproben gezogen und Kennwerte der Stichproben ermittelt.

Parameter der Wahrscheinlichkeitsverteilung, die das Verhalten des Merkmals in der Grundgesamtheit beschreiben, werden mit Hilfe von Kennwerten der Stichprobe geschätzt. Die Schätzung ist mit einer angebbaren Unsicherheit behaftet. Hypothesen über die durch eine Stichprobe untersuchte Grundgesamtheit können mittels statistischer Tests geprüft werden.

Diese Norm beschreibt statistische Verfahren, mit denen Merkmalswerte aufbereitet und Parameter der zugrundeliegenden Wahrscheinlichkeitsverteilung geschätzt oder getestet werden können.

Die statistischen Verfahren richten sich nach der benutzten Skalenart. Die Normreihe erscheint deshalb in 4 Teilen. In DIN 53 804 Teil 1 werden meßbare Merkmale, in Teil 3 Ordinalmerkmale und in Teil 4 Attributmerkmale behandelt.

Die vorliegende Norm befaßt sich mit zählbaren Merkmalen. Sie beschreibt statistische Verfahren, mit denen Zählwerte (Anzahl von Vorkommnissen, z. B. Unfälle, Fadenbrüche) aufbereitet und die Parameter der Wahrscheinlichkeitsverteilung, hier der Parameter der Poisson-Verteilung, geschätzt und getestet werden können.

2 Begriffe

Die in der vorliegenden Norm benutzten statistischen Begriffe sind den Normen DIN 13 303 Teil 1 und Teil 2 und DIN 55 350 Teil 12, Teil 14 (z. Z. Entwurf), Teil 21, Teil 22, Teil 23 und Teil 24 zu entnehmen. Außerdem gelten die nachstehend aufgeführten Begriffe:

Zählabschnitt (Beobachtungsabschnitt) ist die Beobachtungseinheit, in der das Auftreten bestimmter Vorkommnisse gezählt wird. Der Zählabschnitt (oder die Zählabschnitte, wenn mehrere ausgezählt werden) bildet die aus der Grundgesamtheit gezogene Stichprobe.

Zählbares Merkmal ist die Anzahl der Vorkommnisse auf einem Zählabschnitt.

Zählwert

Der Zählwert x_i ist ein Einzelwert des zählbaren Merkmals.

Fortsetzung Seite 2 bis 11

Normenausschuß Materialprüfung (NMP) im DIN Deutsches Institut für Normung e.V.
Ausschuß Qualitätssicherung und angewandte Statistik (AQS) im DIN

Anmerkung: Dem Zählwert x_i entspricht bei meßbaren Merkmalen (siehe DIN 53 804 Teil 1) der Einzelwert x_i.

Der Größe nach geordnete Zählwerte werden mit $x_{(i)}$ bezeichnet.

Bei den in der vorliegenden Norm behandelten zählbaren (diskreten) Merkmalen werden die Merkmalswerte auf einer diskreten Skale dargestellt [6].

Beispiele für Zählabschnitte und zählbare Merkmale

Grundgesamtheit	Zählabschnitt	Zählbares Merkmal
Zeitspanne von 1960 bis 1980	1 Jahr	Anzahl der Unfälle in einem Jahr
Öffnungszeiten des Postschalters	1 Stunde	Anzahl der in einer Stunde am Postschalter ankommenden Kunden
Die ersten 2 Stunden der Anschaltzeit	1 Minute	Anzahl der in einer Minute von einer erhitzten Kathode emittierten Elektronen
Jahr 1980	24 Stunden	Anzahl der in 24 Stunden eine Mautstelle passierenden Fahrzeuge
Kabelproduktion im Monat Mai 1981	1000 m Kabel	Anzahl Isolationsfehler an 1000 m Kabel
Garnlieferung	100 000 m Garn	Anzahl Fadenbrüche auf 100 000 m Garn
Stücke einer Gewebekette	1 Gewebestück	Anzahl Fehler im Gewebestück
Blutprobe	Zählfeld vorgegebener Fläche	Anzahl der Erythrozyten im Zählfeld
Zellkultur	1 cm^3 Suspension	Anzahl Hefezellen in einem Kubikzentimeter Suspension
Fachbuch	4 Seiten	Anzahl Druckfehler auf 4 Seiten
Tagesproduktion von Tischplatten	3 Tischplatten	Anzahl Oberflächenfehler auf drei Tischplatten
Backofenfüllung von Rosinenbroten	1 Rosinenbrot	Anzahl Rosinen im Rosinenbrot

Anmerkung: Unsicherheiten bei der Abgrenzung des Zählabschnitts können das Zählergebnis beeinflussen; ihre quantitative Erfassung ist aber nicht Gegenstand dieser Norm.

3 Poisson-Verteilung

Für das Berechnen von Vertrauensbereichen und Testen von Hypothesen wird in dieser Norm vorausgesetzt, daß das zählbare Merkmal einer Poisson-Verteilung[1]) folgt. Die Wahrscheinlichkeitsfunktion der Poisson-Verteilung legt fest, mit welcher Wahrscheinlichkeit das Vorkommnis auf einem Zählabschnitt nullmal, einmal, zweimal, ..., m-mal, ... auftritt.

Einziger Parameter der Poisson-Verteilung ist der Erwartungswert μ; er gibt an, wie oft das Vorkommnis im Mittel auf dem betrachteten Zählabschnitt auftritt. Der Erwartungswert μ ist proportional zur Größe des Zählabschnitts, d. h. bei a-facher Vergrößerung des Zählabschnitts beträgt der Erwartungswert $a \cdot \mu$.

Varianz und Erwartungswert sind bei der Poisson-Verteilung stets gleich groß:

$$\sigma^2 = \mu \qquad (1)$$

Diese Beziehung kann zum Testen auf Poisson-Verteilung herangezogen werden. Statistische Tests auf Vorliegen einer Poisson-Verteilung siehe [1].

4 Kennwerte einer aus n Zählwerten bestehenden Stichprobe

Mittelwert:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (2)$$

Weitere Kennwerte siehe DIN 53 804 Teil 1.

5 Graphische Darstellung der Zählwerte

Es kann sinnvoll sein, die n Zählwerte nach gleichen Zahlenwerten zu sortieren. Dabei wird mit n_l ($l = 0, 1, 2, \ldots, k$) die Anzahl gleicher Zählwerte mit dem Zahlenwert l bezeichnet und es gilt:

$$\sum_{l=0}^{k} n_l = n \qquad (3)$$

Tabelle 1 zeigt ein Beispiel hierfür (siehe auch Beispiel A.4).

Tabelle 1. **Sortierte Zählwerte**

Zahlenwert des zählbaren Merkmals l	Anzahl gleicher Zählwerte n_l	Summe der Anzahlen $G_l = \sum_{i=0}^{l} n_i$
0	13	13
1	7	20
2	5	25
3	3	28
4	4	32
5	2	34
6	0	34
7	1	35
8	0	35
	$\sum_{l=0}^{k} n_l = 35$	

[1]) Siehe Seite 3

Werden über den Zahlenwerten l Stäbe mit den Höhen n_l aufgetragen, entsteht ein Stabdiagramm für die n Zählwerte (siehe Bild 1).

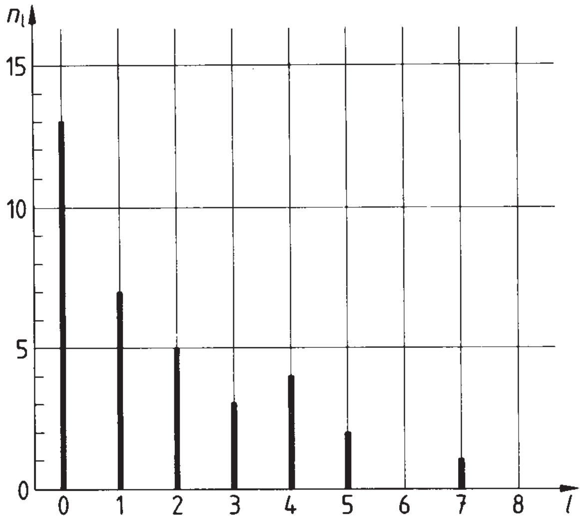

Bild 1. Stabdiagramm für die sortierten Zählwerte der Tabelle 1

Das Stabdiagramm liefert anschauliche Hinweise auf Eigenschaften der Verteilung (Symmetrie, Ausreißer, usw.).

Neben dem Stabdiagramm empfiehlt sich die Darstellung als Summentreppe, deren Sprünge bei den Zahlenwerten l liegen und deren Sprunghöhen gleich n_l sind (siehe Bild 2).

An der Summentreppe kann abgelesen werden, wieviele Zählwerte kleiner als ein vorgegebener Wert oder ihm gleich sind.

Anstelle der Anzahlen n_l können auch die relativen Häufgkeiten n_l/n für Stabdiagramm und Summentreppe verwendet werden. Dabei ist die Anzahl n der Zählwerte anzugeben.

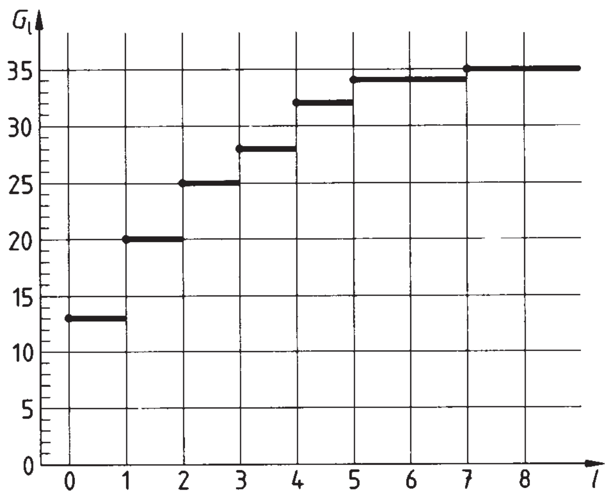

Bild 2. Summentreppe für die Zählwerte der Tabelle 1

1) Die Poisson-Verteilung wird mitunter als Verteilung seltener Ereignisse bezeichnet. Wie dabei das Wort „selten" zu verstehen ist, kann z. B. in [3] nachgelesen werden.

6 Schätzwert und Vertrauensbereich für den Erwartungswert bei Poisson-Verteilung

6.1 Bei einmaligem Auszählen

Der Schätzwert $\hat{\mu}$ (sprich: mü Dach) für den Erwartungswert μ der Anzahl der Vorkommnisse ist

$$\hat{\mu} = x \tag{4}$$

Die Vertrauensgrenzen μ_{un} und μ_{ob} des Vertrauensbereichs für μ bei zweiseitiger Abgrenzung,

$$\mu_{un} \leq \mu \leq \mu_{ob} \tag{5}$$

oder bei einseitiger Abgrenzung,

$$\mu_{un} \leq \mu \text{ bzw. } \mu_{ob} \geq \mu, \tag{6}$$

können in Abhängigkeit vom Zählwert x für das Vertrauensniveau $1 - \alpha = 0{,}95$ der Tabelle 2 entnommen werden (siehe Beispiel A.1).

Exakte Formeln zur Berechnung von μ_{un} und μ_{ob} sind bei zweiseitiger Abgrenzung

$$\mu_{un} = \frac{1}{2}\chi^2_{2x;\alpha/2}$$

und (7)

$$\mu_{ob} = \frac{1}{2}\chi^2_{2(x+1);1-\alpha/2}$$

bei einseitiger Abgrenzung

$$\mu_{un} = \frac{1}{2}\chi^2_{2x;\alpha}$$

bzw. (8)

$$\mu_{ob} = \frac{1}{2}\chi^2_{2(x+1);1-\alpha}$$

Die Werte $\chi^2_{f;\alpha}$ können dem Schrifttum, z. B. [1], [2], entnommen werden.

Für größere Werte von x (etwa $x \geq 40$) können die Vertrauensgrenzen näherungsweise

bei zweiseitiger Abgrenzung nach

$$\mu_{un} = \left(\sqrt{x} - \frac{1}{2}u_{1-\alpha/2}\right)^2$$

und (9)

$$\mu_{ob} = \left(\sqrt{x+1} + \frac{1}{2}u_{1-\alpha/2}\right)^2$$

bei einseitiger Abgrenzung nach

$$\mu_{un} = \left(\sqrt{x} - \frac{1}{2}u_{1-\alpha}\right)^2$$

bzw. (10)

$$\mu_{ob} = \left(\sqrt{x+1} + \frac{1}{2}u_{1-\alpha}\right)^2$$

berechnet werden [4].

Dabei sind $u_{1-\alpha}$ und $u_{1-\alpha/2}$ Tabellenwerte (Quantile) der standardisierten Normalverteilung, die in Tabelle 3 auszugsweise aufgeführt sind.

Eine bessere Näherung liefern die Formeln (11) bis (14), die für $x \geq -4 \cdot \ln \alpha$ (d. h. im Fall $1 - \alpha = 0{,}95$ für $x \geq -4 \cdot \ln 0{,}05 = 11{,}98$) Vertrauensgrenzen mit einem relativen Fehler von höchstens 0,1 % ergeben.

Bei zweiseitiger Abgrenzung werden mit dem Hilfswert

$$d_{\alpha/2} = \frac{u^2_{1-\alpha/2} - 15}{54} \tag{11}$$