

Zusammenfassung: Forschungsmethoden II

SS04 - Prof. Dr. Zysno

Steffens, Christian - christian@steffens.cn

Revision: June 28, 2006

Contents

| | | |
|----------|--|-----------|
| 1 | Typisierung und Darstellung von Daten | 5 |
| 1.1 | Skalenniveau | 5 |
| 1.2 | Nominal | 5 |
| 1.3 | Ordinal | 5 |
| 1.4 | Intervall | 6 |
| 1.5 | Ratio/Verhältnisskala | 6 |
| 2 | Mittelwerte / Zentrale Tendenz | 7 |
| 2.1 | Arithmetisches Mittel | 7 |
| 2.2 | Gewogenes Arithmetisches Mittel | 7 |
| 2.3 | Median | 7 |
| 2.4 | Modus | 7 |
| 3 | Streuungsmaße | 8 |
| 3.1 | Spannweite | 8 |
| 3.2 | Average Deviation | 8 |
| 3.3 | Varianz | 8 |
| 4 | Korrelation | 9 |
| 4.1 | Kovarianz | 9 |
| 4.2 | Korrelation | 9 |
| 5 | Regression | 10 |
| 6 | Elementare Wahrscheinlichkeit: Kombinatorik | 11 |
| 6.1 | Permutation ohne Wiederholung | 11 |
| 6.2 | Permutation mit Wiederholung | 11 |
| 6.3 | Variation ohne Wiederholung | 11 |
| 6.4 | Variation mit Wiederholung | 12 |
| 6.5 | Kombination ohne Wiederholung | 12 |
| 6.6 | Kombination mit Wiederholung | 12 |
| 6.7 | Zusammenfassung: | 12 |
| 7 | Elementare Wahrscheinlichkeiten: Grundlagen | 13 |
| 7.1 | Bezeichner | 13 |
| 7.2 | Axiomatische Definition | 13 |
| 7.2.1 | spezieller Additionssatz | 13 |
| 7.2.2 | spezieller Multiplikationssatz | 13 |
| 7.3 | Komplementär Ereignisse | 13 |

| | | |
|-----------|--|-----------|
| 8 | Diskrete Verteilungen | 14 |
| 8.1 | Binomialverteilung | 14 |
| 8.2 | Wahrscheinlichkeits- / Dichtefunktion | 14 |
| 8.3 | Mittelwert / Erwartungswert bei Binomialverteilung | 15 |
| 8.4 | Streuung bei Binomialverteilung | 15 |
| 9 | Stetige Verteilungen | 16 |
| 9.1 | Einsatz von theoretischen Verteilungen | 16 |
| 9.2 | Chi-Quadrat-Test | 16 |
| 9.3 | Normalverteilung | 17 |
| 9.4 | Die z-Transformation | 17 |
| 9.4.1 | Beispiel 1: | 17 |
| 9.4.2 | Beispiel 2: | 17 |
| 9.4.3 | F-Werte für ausgewählte z_u und z_o | 18 |
| 10 | Statische Inferenz | 19 |
| 10.1 | Stichprobenfehler / Standardfehler | 19 |
| 10.1.1 | Standardfehler des arithmetischen Mittels | 19 |
| 10.2 | Konfidenzintervall | 19 |
| 10.3 | Hypothesen | 19 |
| 10.4 | Beispiel | 20 |
| 10.5 | Fehler | 20 |

Vorwort

Ich stelle hier ein begleitendes Dokument zur Vorlesung "*Forschungsmethoden II - Einführung in die Statistik*" bereit. Es handelt sich hierbei nicht um ein Skript, sondern um eine ergänzende Zusammenfassung des bestehenden Unterrichts, erweitert um einige zusätzliche Erklärungen.

Den Großteil der zusätzlichen Erklärungen und Beispielen habe ich der Wikipedia Seite entnommen (<http://www.matheboard.de/lexikon/>). Diese Seite empfehle ich übrigens auch, wenn ich noch weitere Informationen sucht, die Erläuterungen dort sind wirklich sehr gut. Natürlich gilt auch bei diesem Dokument die übliche Geschichte: Ich erhebe keinen Anspruch auf Vollständigkeit/Korrektheit und ich hafte auch nicht für irgendwelche entstandenen "Schäden".

Falls jemand einen Fehler im Dokument entdeckt, sollte er dies doch bitte melden - vorzugsweise per EMail an christian@steffens.cn.

1 Typisierung und Darstellung von Daten

1.1 Skalenniveau

Das Skalenniveau bestimmt:

- die (mathematischen) Operationen, die mit einer entsprechend skalierten Variablen zulässig sind. Dabei können Operationen, die bei Variablen eines bestimmten Skalenniveaus zulässig sind, grundsätzlich auch auf Variablen aller höheren Skalenniveaus durchgeführt werden.
- welche Transformationen mit entsprechend skalierten Variablen durchgeführt werden können, ohne Information zu verlieren bzw. zu verändern.
- welche Information das entsprechende Merkmal liefert, welche Interpretationen Ausprägungen des entsprechenden Merkmals zulassen.

| SKALENNIVEAU | ZUVERLÄSSIGE KENNWERTE |
|--------------|------------------------------|
| Nominal | Gleichheit, Ungleichheit |
| Ordinal | Relationale Operationen |
| Intervall | Gleichheit von Differenzen |
| Ratio | Gleichheit von Verhältnissen |
| Absolut | Häufigkeit |

1.2 Nominal

Bei nominalskalierten Merkmalen wird der Untersuchungseinheit für das entsprechende Merkmal (genau) *ein Name* bzw. (genau) *eine Kategorie zugeordnet*. Für die Namen oder Kategorien werden dabei oft auch Zahlen verwendet.

| | MERKMAL | KATEGORIEN |
|------------------|--------------|----------------------------------|
| Beispiel: | Jahreszeit | Frühjahr, Sommer, Herbst, Winter |
| | Kreditwürdig | ja, nein |
| | Farbe | rot, gelb, grün |

1.3 Ordinal

Bei ordinalskalierten Merkmalen wird der Untersuchungseinheit für das entsprechende Merkmal (genau) ein Name bzw. (genau) eine Kategorie zugeordnet. Hier lassen sich die einzelnen Kategorien/Namen *in eine Reihenfolge bringen*. Für die Namen oder Kategorien werden dabei oft auch Zahlen verwendet. *Allerdings* lassen sich die *Abstände* zwischen den einzelnen Kategorien *nicht sinnvoll interpretieren*.

| | MERKMAL | KATEGORIEN |
|------------------|---------------|--|
| Beispiel: | Schulnoten | 1 besser als 2 besser als 3 besser als 4 besser als 5 besser als 6 |
| | Einkommen | hoch > mittel > niedrig |
| | Zufriedenheit | sehr zufrieden, zufrieden, unzufrieden, sehr unzufrieden |

1.4 Intervall

Bei intervallskalierten Merkmalen lassen sich zusätzlich zu den Eigenschaften der Ordinalskala die *Abstände* zwischen den verschiedenen Merkmalsausprägungen *exakt bestimmen*. Allerdings *existiert kein natürlicher Nullpunkt* für die Skala. Willkürlich definierte Nullpunkte, wie z.B. bei der Celsius-Temperaturskala zählen hier nicht als natürlicher Nullpunkt, während der Nullpunkt der Kelvin-Temperaturskala, der dem absoluten Nullpunkt entspricht, ein natürlicher Nullpunkt ist.

Beispiel:

- Jahreszahlen
- Temperaturen in Kelvin

1.5 Ratio/Verhältnisskala

Verhältnisskalierte Merkmale besitzen neben der Eigenschaften von Intervallskala Merkmalen einen *natürlichen Nullpunkt*.

Beispiel:

- Entfernung - abs. Nullpunkt: keine Entfernung
- Zeit für einen Vorgang - abs. Nullpunkt: keine Zeit

2 Mittelwerte / Zentrale Tendenz

2.1 Arithmetisches Mittel

Das arithmetische Mittel (auch Durchschnitt) ist der am häufigsten benutzte Mittelwert und wird deshalb auch als Standardmittelwert bezeichnet.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Das arithmetische Mittel einer Stichprobe ist nach vielen Kriterien eine geeignete Schätzung für den Erwartungswert der Verteilung, aus der die Stichprobe stammt.

2.2 Gewogenes Arithmetisches Mittel

x_k Name der Klasse

n_k Klassenhäufigkeit

m Anzahl Klassen

Das gewichtete Mittel wird verwendet, wenn man Mittelwerte aus Stichproben der gleichen Grundgesamtheit mit verschiedenen Stichprobenumfängen miteinander kombinieren will:

$$\bar{x} = \frac{\sum_{k=1}^m n_k \cdot x_k}{\sum_{k=1}^m n_k}$$

2.3 Median

Sortiert man eine Reihe von Messwerten der Größe nach, so ist der Wert, der in der Mitte dieser Reihe liegt, der Median. Hat man eine gerade Anzahl von Werten, ergibt sich der Median als arithmetisches Mittel der beiden mittleren Werte. Die eine Hälfte der Werte ist größer, die andere Hälfte kleiner als der Median. Der Median ist also das 50%-Perzentil.

$$x_{median} = \begin{cases} x_{(n+1)/2} & , \text{ bei ungeraden } n \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & , \text{ bei geraden } n \end{cases}$$

Im Gegensatz zum arithmetischen Mittelwert, auch Durchschnitt genannt, verändert sich der Median durch einzelne Extremwerte kaum. So ist der Median der Zahlenreihen 1,2,3,4,5 und 1,2,3,4,100 jeweils 3; 3 ist in der Mitte, mit je zwei Nachbarn. Der Mittelwert ist im ersten Fall ebenfalls 3, bei der zweiten Reihe verschiebt der "Ausreißer" 100 den Mittelwert auf 22.

2.4 Modus

Der Modus oder Modalwert ist der häufigste Wert einer Häufigkeitsverteilung. Da eine Verteilung mehrgipflig sein kann, können einer Verteilung auch mehrere Modi zugeordnet sein.

$$Mod = x_{max(n_k)}$$

3 Streuungsmaße

3.1 Spannweite

Unterschied zwischen maximalen und minimalen Wert.

$$SW = x_{max} - x_{min}$$

3.2 Average Deviation

Durchschnittliche absolute Abweichung.

$$AD = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

3.3 Varianz

Varianz Die Varianz ist in der Statistik ein Streuungsmaß, d.h. *ein Maß für die Abweichung einer Zufallsvariable X von ihrem Erwartungswert $E[X]$* . Ihr Nachteil ist, dass sie eine andere Einheit als die Daten besitzt. Man verwendet daher oft auch die Standardabweichung, die als Quadratwurzel aus der Varianz definiert ist.

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

Kurz: Summe der Abweichungsquadrate vom arithmetischen Mittel.

Varianz ohne arithmetisches Mittel

$$s^2 = \frac{n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n \cdot (n - 1)}$$

Varianz klassierter Daten

$$s^2 = \frac{\sum_{j=1}^k (x_j - \bar{x})^2 \cdot n_j}{(\sum_{j=1}^k n_j) - 1}$$

4 Korrelation

4.1 Kovarianz

Die Kovarianz ist eine Maßzahl der Statistik, die ausdrückt, wie stark die Korrelation zweier Größen ist, das heißt wie stark diese Größen voneinander abhängen. Der Korrelationskoeffizient ist eine Normierung der Kovarianz und schwankt daher zwischen -1 und +1 schwanken.

- Ein Korrelationskoeffizient von 0 drückt dabei aus, dass Änderungen der einen Größe keinen Einfluss auf Änderungen der anderen Größe haben.
- Ein Korrelationskoeffizient von +1 bedeutet, dass immer, wenn eine Größe zu- oder abnimmt, die andere Größe ebenfalls zu- oder abnimmt.
- Ein Korrelationskoeffizient von -1 bedeutet dagegen, dass die Zunahme einer Größe eine Abnahme der anderen Größe hervorruft und umgekehrt.

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n - 1} = \text{cov}(x, y)$$

4.2 Korrelation

$$\text{Korrelation} = \frac{\text{Kovarianz}}{\text{Gesamtvarianz}}$$

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

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

$$S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{(n - 1)}}$$

$$\begin{aligned} r &= \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \cdot \sqrt{n \cdot \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \end{aligned}$$

Beispiele:

- *positive Korrelation* (je mehr, desto mehr): Je mehr Futter, desto dickere Kühe.
- *negative Korrelation* (je mehr, desto weniger): Je mehr Verkauf von Regenschirmen, desto weniger Verkauf von Sonnencreme.

5 Regression

Regressionsanalysen sind Techniken, mit denen für eine Gleichung $y = f(x)$ die Parameter so angeglichen werden, dass minimale Abweichungen zwischen experimentellen und kalkulierten Werten entstehen.

Y-Wert mittels X abschätzen:

$$a_{\hat{y}x} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_{\hat{y}x} = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i \cdot y_i}{n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \bar{y} - a_{\hat{y}x} \bar{x}$$

$$\hat{y} = a_{\hat{y}x} \cdot x + b_{\hat{y}x}$$

X-Wert mittels Y abschätzen:

$$a_{\hat{x}y} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2} = \frac{S_{xy}}{S_y^2}$$

$$b_{\hat{x}y} = \frac{\sum_{i=1}^n y_i^2 \sum_{i=1}^n x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i \cdot y_i}{n \cdot \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2} = \bar{x} - a_{\hat{x}y} \bar{y}$$

$$\hat{x} = a_{\hat{x}y} \cdot y + b_{\hat{x}y}$$

Ist der Korrelationskoeffizient gleich 1 sind die beiden gerade gleich, ist er gleich 0 stehen sie senkrecht zueinander.

Bemerkung Andere Möglichkeit zur Berechnung des Korrelationskoeffizienten:

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_x^2 \cdot S_y^2}}$$

$$r_{xy} = \frac{\text{Kovarianz}}{\text{Gesamtvarianz}}$$

$$a_{\hat{y}x} = \frac{\text{Kovarianz}}{\text{Basisvarianz}} = \frac{S_{xy}}{S_x^2}$$

$$a_{\hat{x}y} = \frac{\text{Kovarianz}}{\text{Basisvarianz}} = \frac{S_{xy}}{S_y^2}$$

$$\Rightarrow r_{xy} = \sqrt{a_{\hat{y}x} \cdot a_{\hat{x}y}}$$

6 Elementare Wahrscheinlichkeit: Kombinatorik

6.1 Permutation ohne Wiederholung

In der Kombinatorik versteht man unter einer n -stelligen Permutation, bzw. einer n -stelligen Permutation ohne Wiederholung die Anordnung einer Menge mit n Elementen. Beispielsweise sind (cba) und (bca) zwei unterschiedliche Permutationen der Menge $\{a,b,c\}$.

$$P_n = n! = \prod_{i=0}^{n-1} (n - i)$$

6.2 Permutation mit Wiederholung

Für die Zahl der möglichen Anordnungen von Objekten aus mehreren Klassen, die untereinander jeweils innerhalb einer Klasse nicht unterscheidbar sind, ist es hilfreich, zunächst die mögliche Zahl der Anordnungen der Objekte zu betrachten und dann zu überlegen, wieviele dieser Anordnungen nicht unterscheidbar sind. Die Zahl der möglichen Anordnungen bei unterscheidbaren Objekte wird durch die Zahl der nicht unterscheidbaren Anordnungen geteilt.

$$P_n = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!}$$

Beispiel: Zug

- 1 Postwagen
- 4 Reisewagen
- 2 Sonderwagen
- 1 Speisewagen

Gesucht Anzahl Möglichkeiten die verschiedenen Wagons zu variieren: $\frac{8!}{1! \cdot 4! \cdot 2! \cdot 1!} = \underline{840}$

6.3 Variation ohne Wiederholung

$$V_n = n \cdot (n - 1) \cdot \dots \cdot (n - (k - 1)) = \prod_{i=0}^{k-1} (n - i)$$

Erweitern mit $(n - k) \cdot (n - k - 1) \cdot \dots \cdot 2 \cdot 1$:

$$V_n = \frac{n \cdot (n - 1) \cdot \dots \cdot (n - k + 1) \cdot (n - k) \cdot (n - k - 1) \cdot \dots \cdot 2 \cdot 1}{(n - k) \cdot (n - k - 1) \cdot \dots \cdot 2 \cdot 1}$$

$$V_n = \frac{n!}{(n - k)!}$$

6.4 Variation mit Wiederholung

Wenn aus n Objekten k Objekte mit Zurücklegen und mit Beachtung der Reihenfolge ausgewählt werden sollen, dann kann jedes der n Objekte auf jedem der k Plätze der Auswahl erscheinen, es gibt demzufolge n_k mögliche Auswahlen.

$$V_n = n^k$$

Beispiel: Pincode

4 stellig aus 10 Ziffern wählbar mit Wiederholung: $10^4 = 10000$

6.5 Kombination ohne Wiederholung

Im Gegensatz zu den Variationen werden bei den Kombinationen die Anordnungen außer acht gelassen, d.h. "abc" ist gleichwertig mit "bca". Es muss also weniger Kombinationen als Variationen geben.

$$C_n = \frac{V_n}{k!} = \frac{\prod_{i=0}^{k-1} (n-i)}{\prod_{i=0}^{k-1} (k-i)} = \frac{\frac{n!}{(n-k)!}}{k!} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}$$

6.6 Kombination mit Wiederholung

Allgemein: Wieviele Zusammenstellungen von k beliebigen (gleichen oder ungleichen) Elementen in n Zellen sind möglich.

$$C_n = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k! \cdot (n-1)!}$$

Beispiel: Anzahl verschiedener Augenkombinationen beim Werfen von 3 Würfeln

$$\binom{6+3-1}{3} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 54$$

6.7 Zusammenfassung:

| | OHNE WIEDERHOLUNG | MIT WIEDERHOLUNG |
|-------------|---------------------------|---------------------------------------|
| PERMUTATION | $P_n = n!$ | $P_n = \frac{n!}{\prod_{i=1}^m k_i!}$ |
| VARIATION | $V_n = \frac{n!}{(n-k)!}$ | $V_n = n^k$ |
| KOMBINATION | $C_n = \binom{n}{k}$ | $C_n = \binom{n+k-1}{k}$ |

7 Elementare Wahrscheinlichkeiten: Grundlagen

7.1 Bezeichner

$f(Z)$ Eintreten von Ereignis Z

$f(\neg Z)$ Nicht-Eintreten von Ereignis Z

$rh(Z)$ relative Häufigkeit des Eintretens von Ereignis Z , liegt zwischen $[0 \dots 1]$.

$f_{cum}(Z)$ kumulierte Häufigkeiten des Eintretens von Ereignis Z .

$rh_{cum}(Z)$ kumulierte relative Häufigkeit.

7.2 Axiomatische Definition

Axiom 1:

$$P(X) \geq 0$$

Axiom 2:

$$P(\Omega) = 1$$

Axiom 3:

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

7.2.1 spezieller Additionssatz

$$P(X \cup Y \cup Z) = P(X) + P(Y) + P(Z) - P(X \cap Y) - P(X \cap Z) - P(Y \cap Z) + P(X \cap Y \cap Z)$$

7.2.2 spezieller Multiplikationssatz

Für zwei unabhängige Ereignisse X und Y gilt:

$$P(X \cap Y) = P(X) \cdot P(Y)$$

7.3 Komplementär Ereignisse

Beispiel: Ein Arbeiter bedient 3 Webstühle. Die Wahrscheinlichkeit für einständigen störungsfreien Lauf:

$$P_1 = 0,9$$

$$P_2 = 0,8$$

$$P_3 = 0,7$$

Wie groß ist die Wahrscheinlichkeit, dass wenigstens einer der drei Webstühle versagt?

$$P = 1 - P_1 \cdot P_2 \cdot P_3$$

Wie groß ist die Wahrscheinlichkeit, dass wenigstens einer der drei Webstühle läuft?

$$P = 1 - Q_1 \cdot Q_2 \cdot Q_3 = 1 - (1 - P_1) \cdot (1 - P_2) \cdot (1 - P_3)$$

8 Diskrete Verteilungen

8.1 Binomialverteilung

Herleitung:

Die Binomialverteilung ist eine der wichtigsten diskreten Wahrscheinlichkeitsverteilungen. Sie beschreibt Ergebnisse von Bernoulli-Prozessen.

Ein Bernoulli-Prozess (auch: Bernoulli-Kette) besteht aus einer Abfolge mehrerer, unter gleichbleibenden Bedingungen durchgeführter Bernoulli-Versuche.

Ein Bernoulli-Versuch (auch: Bernoulli-Experiment) ist ein Zufallsversuch mit genau zwei möglichen Ergebnissen. Die Wahrscheinlichkeit dieser Ergebnisse wird durch die Bernoulli-Verteilung beschrieben.

Beispiel: Wahrscheinlichkeit für keine Sechs bei 4 Würfeln.

$$P_4(\neg \text{Sechs}) = \left(\frac{5}{6}\right)^4 = \frac{625}{1296} = 0,482$$

$$P_n(x) = p^n$$

$$Q_n(x) = q^n$$

Beispiel: Wie gross ist die Wahrscheinlichkeit, dass bei 5 Würfeln die beiden ersten eine Sechs und die drei letzten keine Sechs sind.

$$P_n(2|3) = \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3 = 0,016$$

$$P_n(x) = p^x \cdot q^{n-x}$$

8.2 Wahrscheinlichkeits- / Dichtefunktion

Beispiel: Wie groß ist die Wahrscheinlichkeit, dass Familien mit fünf Kindern 0..5 Mädchen haben:

$$f(x|5; \frac{1}{2}) = \binom{5}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} = \binom{5}{x} \left(\frac{1}{2}\right)^5$$

$$f(x|n; p) = \binom{n}{k} p^x q^{n-x}$$

Falls

$$p = \frac{1}{2}$$

$$f(x|n; \frac{1}{2}) = \binom{n}{k} \frac{1}{2^n}$$

Beispiel:

Wahrscheinlichkeit bei 5 Würfeln mindestens zwei Sechsen zu erhalten.
Lösung: Addition der Wahrscheinlichkeiten für 2, 3, 4 und 5 Sechsen.

$$F(x \geq 2|n = 5; p = \frac{1}{6}) \approx 20\%$$

allg. Verteilungsfunktion:

$$F(x|n; p) = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}$$

8.3 Mittelwert / Erwartungswert bei Binomialverteilung

Beispiel: Welche Häufigkeit ist bei 100-maligen Münzwurf für "Zahl" zu erwarten?

$$\mu = 100 \cdot \frac{1}{2} = 50$$

$$\mu = n \cdot p$$

8.4 Streuung bei Binomialverteilung

Beispiel: Welche Streuung ist bei 100-maligen Münzwurf zu erwarten?

$$\sigma = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{25} = 5$$

$$\sigma^2 = n \cdot p \cdot q$$

$$\sigma = \sqrt{npq} = \sqrt{\mu q}$$

9 Stetige Verteilungen

9.1 Einsatz von theoretischen Verteilungen

Vergleich Empirische und theoretische Verteilung werden verglichen.

Vertrauensbereich Bestimmen des Vertrauensbereich für einen Kennwert.

Hypothesentest Signifikanz von empirischen Ergebnissen.

Hypothesentest Signifikanz von Unterschieden zwischen den Kennwerten zweier Stichproben.

9.2 Chi-Quadrat-Test

$$\chi^2 = \sum_{k=1}^m \frac{(f_{beob,k} - f_{erw,k})^2}{f_{erw,k}}$$

Beispiel: Überprüfung eines Würfels auf Gleichverteilung.

H_0 Nullhypothese - Der Würfel ist nicht gezinkt.

H_1 Alternativhypothese - Der Würfel ist gezinkt.

| k | f_{beob} | f_{erw} | $f_{beob} - f_{erw}$ | $(f_{beob} - f_{erw})^2$ | $\frac{(f_{beob} - f_{erw})^2}{f_{erw}}$ |
|-----|------------|-----------|----------------------|--------------------------|--|
| 1 | 11 | 10 | 1 | 1 | 0,1 |
| 2 | 7 | 10 | -3 | 9 | 0,9 |
| 3 | 9 | 10 | -1 | 1 | 0,1 |
| 4 | 12 | 10 | 2 | 4 | 0,4 |
| 5 | 11 | 10 | 1 | 1 | 0,1 |
| 6 | 10 | 10 | 0 | 0 | 0,0 |
| | | | | | $\chi_{err}^2 = 1,6$ |

Vorgabe: $\alpha = 0,05$; $df = 5$

Tabelle: $\chi_{krit}^2 = 11,...$

Entscheidung: H_0 wird beibehalten mit einer Irrtumswahrscheinlichkeit von 5 %.

$$\chi_{err}^2 < \chi_{krit}^2$$

Erhöhter Stichprobenumfang: hat Alternative Hypothese zur Folge. Das Chi-Quadrat ist nun gleich 16 und somit größer als der entsprechende Werte aus der Tabelle, es wird daher H_1 angenommen.

9.3 Normalverteilung

$$\begin{aligned}
 f(x|\mu;\sigma) &= \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \\
 &= \frac{1}{\sqrt{2\pi}\sigma \cdot e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}} \\
 f(10|10,2) &= \frac{1}{2,5 \cdot 2e^{\frac{1}{2}\left(\frac{10-10}{2}\right)^2}} \\
 &= \frac{1}{5}
 \end{aligned}$$

9.4 Die z-Transformation

$$\begin{aligned}
 z_i &= \frac{x_i - \bar{x}}{s} \\
 x_i &= \bar{x} + s \cdot z_i
 \end{aligned}$$

9.4.1 Beispiel 1:

Bestimmung eines z-Wertes und der zugehörigen Wahrscheinlichkeit. Wie groß ist die Wahrscheinlichkeit, dass ein Mann aufrecht durch einer 2 Meter hohe Tür gehen kann? Der Mittelwert der Körpergröße sei 180cm, die Standardabweichung $s=10$ cm.

$$\begin{aligned}
 z_i &= \frac{200 - 180}{10} = 2 \\
 P(z \leq 2) &= 0,9772 \quad \text{siehe Tabelle}
 \end{aligned}$$

9.4.2 Beispiel 2:

Bestimmung eines Intervalls und der zugehörigen Wahrscheinlichkeit. Eine Kartoffelsortiermaschine umfasst 5 Gitter, die von oben nach unten Kartoffeln vom Querschnitt 6,5,4,3,2cm durchlassen. Der Durchmesser der "Helene" ist normalverteilt mit einem Mittelwert von 4cm und einer Standardabweichung von 1cm.

Wieviele Kartoffeln fallen vorraussichtlich in die Abteilung mit der Größe 5-6cm?

$$\begin{aligned}
 z_6 &= \frac{6 - 4}{1} = 2 \\
 z_5 &= \frac{5 - 4}{1} = 1 \\
 P^{(z_6)} &= 0,9772 \\
 P^{(z_5)} &= 0,8413 \\
 P^{(z_6 - z_5)} &= 0,136 \\
 &= 13,6\% \quad \text{fallen in die Abteilung 5-6cm}
 \end{aligned}$$

Bemerkung Um auf Normalverteilung hinzu prüfen, empfiehlt es sich den χ^2 -Test durchzuführen.

9.4.3 F-Werte für ausgewählte z_u und z_o

| z_u | z_o | F |
|-----------|-----------|---------------|
| $-\infty$ | ∞ | 1 |
| $-\infty$ | $\mu = 0$ | $\frac{1}{2}$ |
| $-\sigma$ | $+\sigma$ | 0,6826 |

10 Statische Inferenz

10.1 Stichprobenfehler / Standardfehler

10.1.1 Standardfehler des arithmetischen Mittels

Notation $\bar{x} \cong$ Mittelwert der Mittelwertverteilung
 $\sigma_{\bar{x}} \cong$ Standardabweichung der Mittelwerte

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \quad n \rightarrow \infty \rightarrow 0$$

10.2 Konfidenzintervall

$$\mu = \bar{x} \pm z_{\alpha} \cdot \sigma_{\bar{x}}$$

Beispiel In einer Zufallsstichprobe von 625 Personen, wurde ein durchschnittlicher IQ von 80 Punkte erreicht. Die Standardabweichung betrug $s = 10$. Geben Sie das 90 %-Intervall für den Populationsparameter μ an.

$$\begin{aligned} \mu &= 80 \pm 2,3 \cdot \sigma_{\bar{x}} \\ &= 80 \pm 2,33 \cdot 0,4 \\ &= 80 \pm 0,932 \\ \sigma_{\bar{x}} &= \frac{10}{\sqrt{625}} \end{aligned}$$

10.3 Hypothesen

Alternativhypothese H_1 : Blinde können im Frequenzband zwischen 200-2000 Herz mehr Töne unterscheiden als Normalsichtige.

Nullhypothese logisch $H_0 : \neg H_1$ H_1 : Es nicht wahr, dass Blinde im Frequenzband zwischen 200-2000 Herz mehr Töne unterscheiden als Normalsichtige.

Nullhypothese sprachlich H_0 : Blinde können im Frequenzband zwischen 200-2000 Herz nicht mehr Töne unterscheiden als Normalsichtige.

statistische Kurzform

$$H_0 : \mu_{\text{norm}} \geq \mu_{\text{blind}}$$

$$H_1 : \mu_{\text{norm}} < \mu_{\text{blind}}$$

10.4 Beispiel

Untersuchungsergebnisse:

$$n = 100$$

$$\mu_{\text{Norm}} = 720$$

$$\sigma_{\text{Norm}} = 120$$

$$\bar{x}_{\text{Blind}} = 750$$

$$\begin{aligned} z &= \frac{\bar{x} - \mu_o}{\sigma_{\bar{x}}} \\ &= \frac{750 - 720}{12} = \frac{5}{2} \end{aligned}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{120}{10}$$

$$z_{\text{errechnet}} = 2,5$$

$$F(z = 2,5; \text{einseitig}) = 99,38\%$$

10.5 Fehler

α Fehler erster Art: Die H1 wird angenommen, obwohl eigentlich die H0 gilt.

β Fehler zweiter Art : Die H0 wird beibehalten, obwohl eigentlich die H1 gilt.