Statistik und Wahrscheinlichkeitsrechnung

– Wintersemester 2019/20 –

Kapitel 07: Testverfahren

Prof. Dr. Adrian Ulges

Angewandte Informatik (B.Sc.) /
Informatik - Technische Systeme (B.Sc.) /
Wirtschaftsinformatik (B.Sc.)

Fachbereich DCSM
Hochschule RheinMain

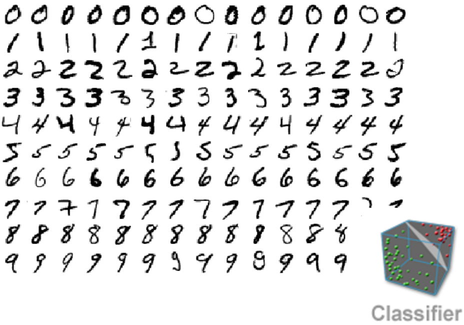
Outline

Beispiel: IT-Performance Bild: [1]

Unser Kunde behauptet, seit dem letzten Software-Update weise unser IT-Service Performance-Probleme auf, d.h. Anwender warten oft mehrere Sekunden auf eine Reaktion des Systems.

Wie ermitteln wir, ob das Update wirklich die Performance verschlechtert hat?

Beispiel: Machine Learning Bild: [2]



Wir variieren einen Parameter innerhalb eines komplexen Systems (z.B. eines OCR-Systems). Die Erkennungsrate steigt von 97.6% auf 97.9%.

Ist die Verbesserung zufallsbedingt oder "signifikant"?

Beispiel: Medikamente Bild: [4]

Ein Arzneimittel-Hersteller behauptet, dass sein neues Medikament nur “sehr selten” (bei < 1 von 10000 Behandelten) zur Nebenwirkung “Störungen der Blutbildung” führt. Ein Arzt und beobachtet bereits den zweiten Fall einer Blutbildungsstörung bei der 512. Behandlung.

Hat das Medikament in der Tat häufigere Nebenwirkungen?

Statistische Testverfahren[1]

- In allen obigen Beispielen geht es um die Prüfung einer Behauptung (oder *Hypothese*) anhand einer Stichprobe.
- Dies ist mit **statistischen Testverfahren** möglich!

Gegeben

- Eine *Behauptung* (oder **Hypothese**) $\mathcal{H}_0$ über die Verteilung der Daten:
  - "Die Wartezeit beträgt im Mittel $\mu \leq 0.75$ Sekunden."
  - "Die Paninibilder sind gleichverteilt."
- Eine **Stichprobe** $x_1, ..., x_n$ *(in der Regel i.i.d.)*.

Ansatz

- Wir wählen ein **Signifikanzniveau** α (und $\gamma := 1 - \alpha$).
- Beispiel: $\alpha = 1\%$, also $\gamma = 99\%$.
- **Zielsetzung**: **Prüfe, ob die Stichprobe mit Wahrscheinlichkeit γ mit der Hypothese $\mathcal{H}_0$ vereinbar ist**.

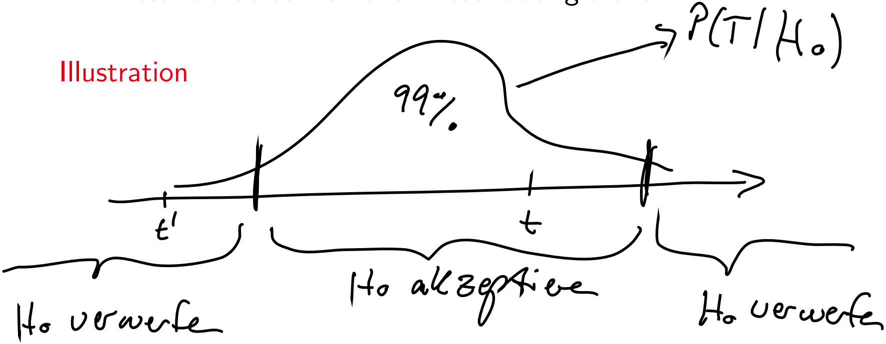
[1] Grundlage dieses Kapitels: Papula, Band 3, III, Kapitel 4-5

Testverfahren: Generelles Vorgehen

Vier Schritte

1. Hypothese aufstellen $H_0 = \Theta = 100$ (Bsp.)
2. Testvariable bestimmen T
3. Kritische Grenzen berechnen
4. Testwert berechnen und Entscheidung treffen

Illustration



Testverfahren: Generelles Vorgehen

Schritt 1: Hypothese aufstellen

- Die "Behauptung" $\mathcal{H}_0$ nennen wir die '**Null-Hypothese**'. Die Gegenbehauptung ist die **Alternativ-Hypothese** $\mathcal{H}_1$.
- **Fall 1** (*Zweiseitiger Test*): $\boxed{\mathcal{H}_0 : \theta = \theta_0, \mathcal{H}_1 : \theta \neq \theta_0}$
  - θ_0 ist der behauptete Wert
  - Im Beispiel: $\mathcal{H}_0 =$ "Die Schokotafeln wiegen im Mittel 100 g"
- **Fall 2** (*Einseitiger Test*): $\boxed{\mathcal{H}_0 : \theta \geq \theta_0,\ \mathcal{H}_1 : \theta < \theta_0}$
  - $\mathcal{H}_0 =$ "Die Schokotafeln wiegen im Mittel <u>mindestens</u> 100 g"
  - Geht auch umgekehrt ($\mathcal{H}_0 : \theta \leq \theta_0$)

Schritt 2: Testvariable bestimmen

- Wir bestimmen eine **Testvariable** $T := g(X_1, ..., X_n)$.
- Hierbei sind $X_1, ..., X_n$ die Zufallsvariablen, aus denen die Stichprobe $x_1, ..., x_n$ gewonnen wird.
- Die Berechnung der Variablen T **variiert je nach Test**!

Testverfahren: Generelles Vorgehen

Schritt 3: Kritische Grenzen berechnen

- Wir bestimmen zwei kritische Grenzen c_u, c_o so dass

$$P(c_u \leq T \leq c_o) = \gamma$$

- Oft lesen wir die Grenzen c_u, c_o aus **Quantiltabellen** ab
- Anmerkung: Bei **einseiten** Tests gibt es nur eine kritische Grenze c, d.h. wir fordern

$$P(T \geq c) = \gamma \qquad (\text{bzw. } P(T \leq c) = \gamma)$$

Schritt 4: Testwert berechnen und Testentscheidung

- Wir berechnen den Testwert $t := g(x_1, x_2, ..., x_n)$.
- Liegt t zwischen den Grenzen ("nicht-kritischer Bereich"), müssen wir die Hypothese $\mathcal{H}_0$ **akzeptieren**. Ansonsten **lehnen** wir $\mathcal{H}_0$ **ab**.

Testverfahren: Fehlerquellen

- **Ein Testverfahren gibt nie völlige Sicherheit!**
  Es ist möglich, dass wir die Hypothese $\mathcal{H}_0$ irrtümlich zurückweisen oder irrtümlich akzeptieren.
- Wir unterscheiden *zwei Arten* von Fehlern

Fehler erster Art

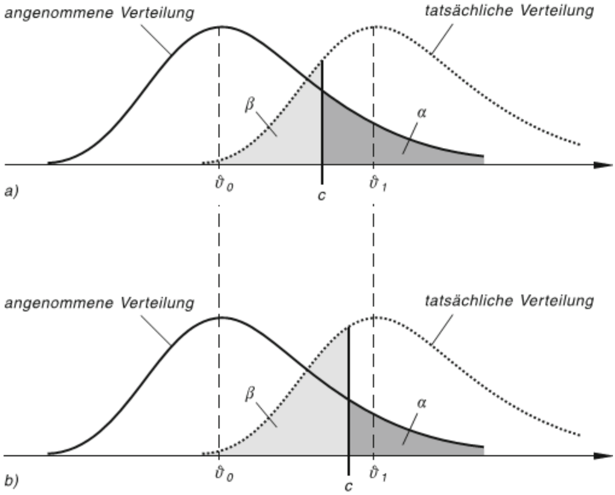
- Ablehnung der Hypothese $\mathcal{H}_0$, obwohl $\mathcal{H}_0$ gilt.
- Dies geschieht mit Wahrscheinlichkeit α.
- Wir nennen α auch das *Lieferantenrisiko*.

Fehler zweiter Art

- Die Stichprobe fällt zu günstig aus und $\mathcal{H}_0$ wird (obwohl falsch) beibehalten.
- Dies geschieht mit einer (unbekannten) Wahrscheinlichkeit β.
- Wir nennen β auch das *Konsumentenrisiko*.

Testverfahren: Fehlerquellen[2]

Eine Verringerung von α vergrößert β (und umgekehrt)



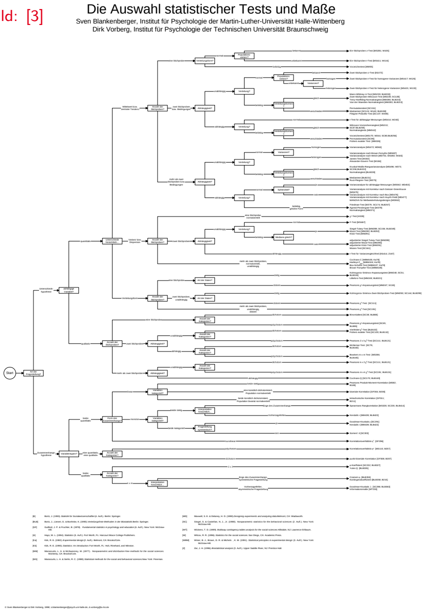
[2] Bildquelle: *Papula: Mathematik für Ingenieure und Naturwissenschaftler, Band 3., S.551-552*

Testverfahren: Fehlerquellen

Wie wählen wir α?

- Die Wahl des Signifikanzlevels α ist **domänenabhängig**.
- Häufige Wahl: $1\% \leq \alpha \leq 5\%$.
- Führt eine fälschliche **Ablehnung** der Hypothese $\mathcal{H}_0$ zu schwerwiegenden Folgen, wählt man α kleiner (z.B. 0.1%) und nimmt Fehler erster Art bewusst in Kauf.
  Beispiel: Unschuldsvermutung in Kriminalfällen.
- Führt eine fälschliche **Akzeptanz** der Hypothese $\mathcal{H}_0$ zu schwerwiegenden Folgen, wählt man α groß (z.B. 5%).
  Beispiel: Flugzeug-Bauteil.
- Eine **gleichzeitige Verringerung** von Fehlern 1. und 2. Art... ist möglich durch eine *Erhöhung des Stichprobenumfangs n*. Dies verbessert die *Trennschärfe* des Tests.

Vielfalt an Testverfahren Bild: [3]



- Es gibt eine Vielzahl von Testverfahren *(je nach Verteilungstyp, Parametern und weiteren Randbedingungen)*
- Hier: die zwei **häufigsten** Varianten
  - Prüfe den Erwartungswert μ *(Parametertest)*
  - Prüfe die Verteilung *(χ^2-Test)*

Outline

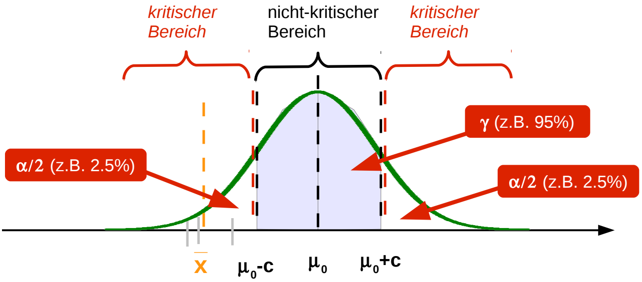
1. Motivation und Grundlagen
2. Normalverteilung, μ (bei bekanntem σ^2)
3. Normalverteilung, μ (bei unbekanntem σ^2)
4. Differenzentests
5. Verteilungstests: Der χ^2-Test
6. Der χ^2-Unabhängigkeitstest

Normalverteilung: Parametertest für μ (bekanntes σ^2) ✱

- X sei **normalverteilt** mit bekannter Standardabweichung σ und **unbekanntem Erwartungswert μ**.
- **Stichprobe**: $x_1, \ldots, x_n$ (i.i.d., gleiche Verteilung wie X).
- **Hypothese** $\mathcal{H}_0$: "Der Erwartungswert der Verteilung ist μ_0".

Herleitung

- Wir definieren den nicht-kritischen Bereich als $[\mu_0 - c, \mu_0 + c]$.



Normalverteilung: Parametertest für μ (bekanntes σ^2)

Satz (Test für μ bei bekanntem σ^2)

Gegeben sei eine Stichprobe normalverteilter i.i.d. Zufallsvariablen $X_1, \ldots, X_n$ mit Varianz σ^2 und unbekannten Erwartungswerts μ. Dann ist unter Annahme der Hypothese $\mathcal{H}_0$ ($\mu{=}\mu_0$) die Testvariable

$$U := \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

standardnormalverteilt. *Gegeben ein Signifikanzniveau $\alpha \in]0,1[$, akzeptieren wir die Hypothese $\mathcal{H}_0$, falls*

$$\mathbf{-x_{1-\alpha/2} \leq U \leq x_{1-\alpha/2}}.$$

Anmerkungen

- $x_{1-\alpha/2}$ bezeichnet hier das $(1-\alpha/2)$-Quantil der Standard-normalverteilung (z.B. falls $\alpha = 10\%$: das 95%-Quantil).
- Im Fall eines **einseitigen** Tests akzeptieren wir $\mathcal{H}_0$: "$\mu \leq \mu_0$" (bzw. "$\mu \geq \mu_0$"), falls $U \leq x_{1-\alpha}$ (bzw. $U \geq x_\alpha$).

Beweis

Um die kritischen Grenzen zu ermitteln, ermitteln wir eine Konstante c, so dass:

$$P(\mu_0 - c \leq \overline{x} \leq \mu_0 + c) = 1 - \alpha$$

Wir formen die Bedingung um:

Beweis

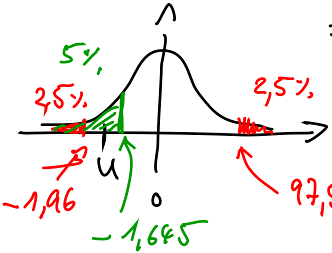
Beispiel: Schokolade Bild: [6]

- Wir werden mit Schokotafeln beliefert.
- Hypothese des Herstellers: Die Tafeln wiegen 100 g ($\sigma = 2$ g sei bekannt).
- Wir wiegen 10 Tafeln. Der Mittelwert beträgt $\overline{x} = 98.9$ g.
- Wir formulieren die Hypothese $\mathcal{H}_0$: "$\mu = 100$g".
- Wir wählen $\alpha = 5\%$.

$\mu_0 = 100$
$\bar{x} = 98{,}9$
$n = 10$

Testvariable = $U = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{98{,}9 - 100}{2/\sqrt{10}}$

$= -1{,}739$



97,5%-Quantil = 1,96

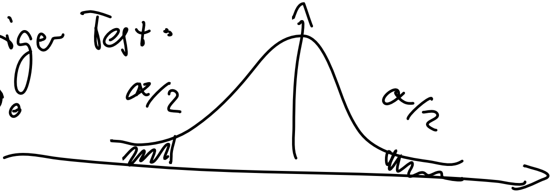
$\Rightarrow$ H_0 ~~akzeptiere~~!

$\Rightarrow$ H_0 verwerfen!

Beispiel: Schokolade

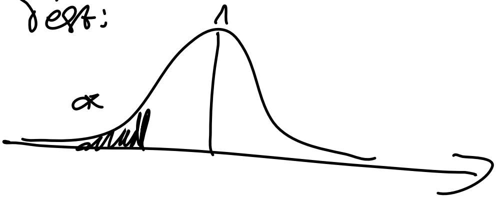
Beidseitiger Test:

$H_0: \mu = \mu_0$



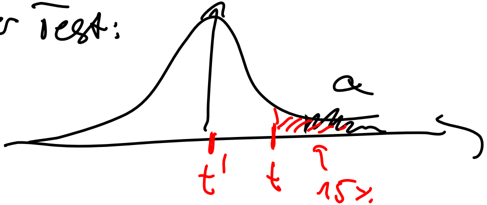
Linksseitiger Test:

$H_0: \mu \geq \mu_0$



Rechtsseitiger Test:

$H_0: \mu \leq \mu_0$



Beispiel: Schokolade (schon wieder)

- Der Hersteller behauptet:
  Die Tafeln wiegen <u>mindestens</u> 100g.
- Führen Sie einen **einseitigen** Test durch.
- Hier noch einmal die Eckdaten:
  $\overline{x} = 98.9\text{g}$, $\sigma = 2\text{g}$, $n = 10$, $\alpha = 5\%$.

α	x_α
0.001	-3.090
0.005	-2.576
0.01	-2.326
0.025	-1.960
0.05	-1.645
0.1	-1.282
0.9	1.282
0.95	1.645
0.975	1.960
0.99	2.326
0.995	2.576
0.999	3.090

Der p-Wert

Definition (p-Wert)

Wir führen einen Test mit einer Testvariable T durch und erhalten einen Testwert t. Dann bezeichnet der **p-Wert** *des Tests die* **Wahrscheinlichkeit**, *dass – gegeben die Hypothese $\mathcal{H}_0$ – die Testvariable T einen noch extremeren Wert als t annimmt. Wir berechnen den p-Wert folgendermaßen:*

	p-Wert
linksseitiger Test	$P(T \leq t \mid \mathcal{H}_0)$
rechtsseitiger Test	$P(T \geq t \mid \mathcal{H}_0)$
beidseitiger Test	$2 \cdot min\Big(P(T \leq t \mid \mathcal{H}_0), P(T \geq t \mid \mathcal{H}_0)\Big)$

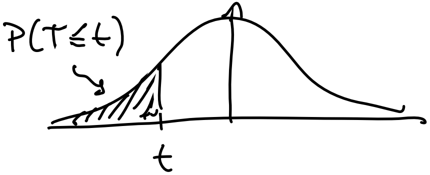
Anmerkungen

- Der p-Wert liegt immer zwischen 0 und 1.
- Je kleiner der p-Wert, desto *ungewöhnlicher/verdächtiger* ist die beobachtete Stichprobe.

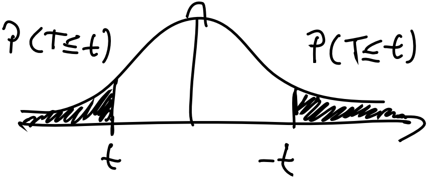
Der p-Wert

Illustration

linksseitiger Test



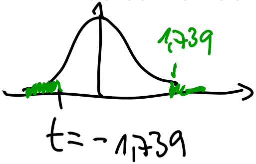
beidseitiger Test



p-Wert $= 2 \cdot P(T \leq t)$

Beispiel

Wir berechnen den p-Wert im obigen Beispiel (Schokolade):



Linksseitig: p-Wert $= P(T \leq -1{,}739)$

$= 1 - P(T \leq 1{,}739)$

$= 1 - 0{,}9591 \approx 4\%$

Beidseitig: $2 \times P(T \leq -1{,}739)$

$\approx 8\%$

Outline

Normalverteilung: Tests für μ (<u>un</u>bekanntes σ^2)

- Die Ausgangssituation ist dieselbe wie eben, nur sei die Standardabweichung σ diesmal unbekannt.
- Wie schon bei der Intervallschätzung (Kapitel 6) ersetzen wir σ durch die **korrigierte Stichprobenstandardabweichung $\mathbf{s}^*$**.

Herleitung

- Wir fordern wie eben *(hier für den zweiseitigen Test)*:

Die t-Verteilung von Student

Definition (Die t-Verteilung)

Die Verteilung einer stetigen Zufallsvariable T mit der Dichtefunktion

$$f(t;n) := B_n \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

ist die **(Student'sche) t-Verteilung** *mit n Freiheitsgraden.*

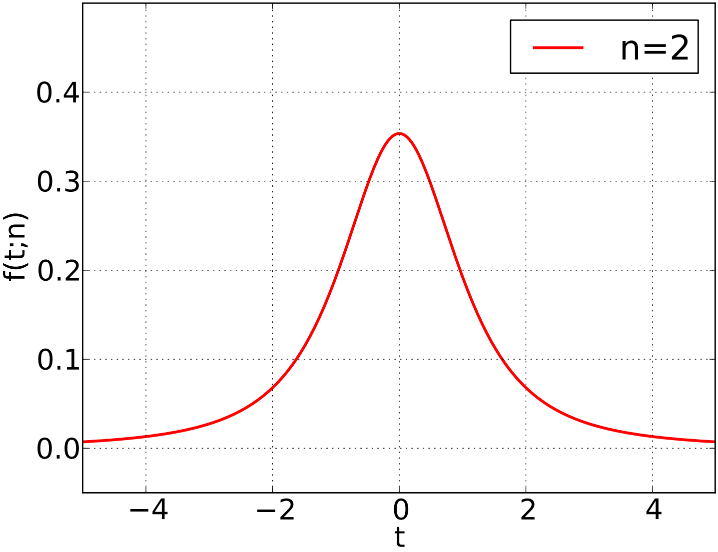
Anmerkungen

- Der Normierungsfaktor B_n ist so gewählt, dass die Dichte zu 1 integriert. Wir werden hierauf nicht näher eingehen.

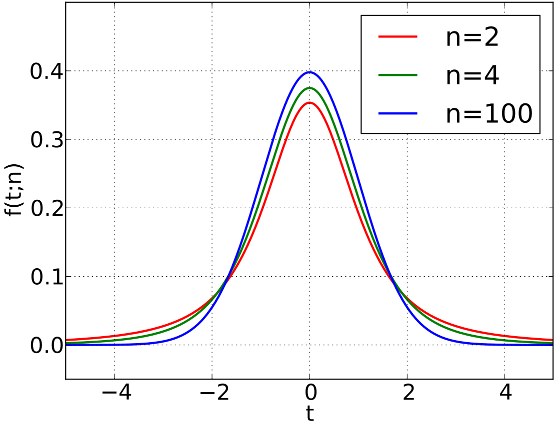
Parameter

- Einziger Parameter: $n (> 0)$, die Anzahl der Freiheitsgrade.

Die Student'sche t-Verteilung: Illustration



Die Student'sche t-Verteilung: Eigenschaften



- $f(t)$ ist symmetrisch zu $t = 0$, und $lim_{|t|\to\infty} f(t) = 0$
- Je höher der Freiheitsgrad n, desto besser wird die t-Verteilung durch die Standardnormalverteilung approximiert.
- **Faustregel**: Für $\mathbf{n} > \mathbf{30}$ können wir statt der t-Verteilung die Standardnormalverteilung verwenden.

Die Student'sche t-Verteilung: Quantile

Wir lesen die Werte der t-Verteilung üblicher Weise aus **Quantiltabellen**[3] ab:

	p				
f	0,90	0,95	0,975	0,99	0,995
1	3,078	6,314	12,707	31,820	63,654
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
22	1,321	1,717	2,074	2,508	2,819
24	1,318	1,711	2,064	2,492	2,797
26	1,315	1,706	2,056	2,479	2,779
28	1,313	1,701	2,048	2,467	2,763
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
100	1,290	1,660	1,984	2,364	2,626
200	1,286	1,653	1,972	2,345	2,601
500	1,283	1,648	1,965	2,334	2,586
⋮	⋮	⋮	⋮	⋮	⋮
∞	1,282	1,645	1,960	2,326	2,576

[3] Quelle: Papula, Mathematik für Ingenieure und Naturwissenschaftler, Band 3.

Normalverteilung: Test für μ (unbekanntes σ^2)

Satz (Tests für μ bei unbekanntem σ^2)
Gegeben sei eine Stichprobe normalverteilter i.i.d. Zufallsvariablen $X_1, \ldots, X_n$ unbekannter Varianz σ^2 und unbekannten Erwartungswerts μ. Dann ist die Testvariable

$$T := \frac{\overline{X} - \mu_0}{s^* / \sqrt{n}}$$

t-verteilt mit n-1 Freiheitsgraden. *Gegeben ein Signifikanzniveau $\alpha \in]0,1[$, akzeptieren wir die Hypothese $\mathcal{H}_0$: "$\mu = \mu_0$", falls*

$$-t^{n-1}_{1-\alpha/2} \leq T \leq t^{n-1}_{1-\alpha/2}.$$

Anmerkungen

- $t^{n-1}_{1-\alpha/2}$ bezeichnet nun das $(1-\alpha/2)$-Quantil der **t-Verteilung** mit $n-1$ Freiheitsgraden.
- **Einseitige Tests**: Analog zum Fall bekannter Varianz.

Beispiel: Zylinderscheiben

- Wir entnehmen einer Produktion $n = 16$ Zylinderscheiben normalverteilten Durchmessers
- Stichprobenkennwerte: $\overline{x} = 20.6\text{mm}$, $s^* = 0.5\text{mm}$
- Behauptung $\mathcal{H}_0$: "$\mu = 20.2\text{mm}$"
- Signifikanzniveau $\alpha = 5\%$

Beispiel: Zylinderscheiben[4]

	p				
f	0,90	0,95	0,975	0,99	0,995
1	3,078	6,314	12,707	31,820	63,654
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
22	1,321	1,717	2,074	2,508	2,819
24	1,318	1,711	2,064	2,492	2,797
26	1,315	1,706	2,056	2,479	2,779
28	1,313	1,701	2,048	2,467	2,763
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
100	1,290	1,660	1,984	2,364	2,626
200	1,286	1,653	1,972	2,345	2,601
500	1,283	1,648	1,965	2,334	2,586
⋮	⋮	⋮	⋮	⋮	⋮
∞	1,282	1,645	1,960	2,326	2,576

[4] Bildquelle: *Papula: Mathematik für Ing. und Naturwissensch., Band 3.*

Outline

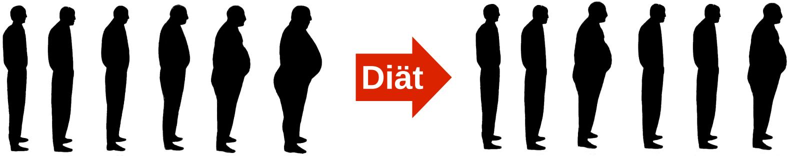
1. Motivation und Grundlagen
2. Normalverteilung, μ (bei bekanntem σ^2)
3. Normalverteilung, μ (bei unbekanntem σ^2)
4. Differenzentests
5. Verteilungstests: Der χ^2-Test
6. Der χ^2-Unabhängigkeitstest

Motivation Bild: [1]

Oft wollen wir mit Testverfahren untersuchen, ob ein **Unterschied zwischen zwei Verteilungen** besteht:

- Ist die “All-you-can-eat”-Diät wirkungsvoll?
  $\rightarrow$ *Wir untersuchen das Körpergewicht vor/nach der Diät.*
- Sind Autos der Marke A sicherer als die der Marke B?
  $\rightarrow$ *Wir vergleichen die Unfallstatistiken.*
- Hat das System-Update die Performance verschlechtert?
  $\rightarrow$ *Wir untersuchen die Wartezeiten vor/nach dem Update.*
- Sind die Messgeräte A und B gleichwertig?
  $\rightarrow$ *Wir vergleichen die Messwerte auf denselben Objekten.*

Ausgangssituation im Folgenden



- Zwei normalverteilte Zufallsvariablen X, Y
  (z.B. Körpergewicht vor/nach der Diät).
- Die Erwartungswerte μ_1, μ_2 sind unbekannt.
- Wir entnehmen zwei Stichproben:

$$x_1, \ldots, x_n \qquad y_1, \ldots, y_n$$

  (z.B. das Gewicht von n Probanden vor/nach der Diät).
- Gilt die Hypothese $\mathcal{H}_0$: "$\mu_1 = \mu_2$"?
  (im Beispiel hieße das: Die Diät ist wirkungslos).

Abhängige vs. unabhängige Stichproben

Definition (Abhängigkeit von Stichproben)

Zwei Stichproben $x_1, \ldots, x_n$ *und* $y_1, \ldots, y_n$ *heißen* **abhängig**, *wenn sich jeder Wert der einen Stichprobe genau einem Wert der anderen Stichprobe zuordnen lässt.*
Praktisch bedeutet dies: "Beide Stichproben wurde auf denselben Objekten erhoben".

Abhängige Stichproben: Beispiele

- Körpergewicht derselben Probanden vor/nach der Diät.
- Gewicht derselben Personen (gemessen mit anderen Waagen).

Unabhängige Stichproben: Beispiele

- Blutdruck verschiedener Patienten mit/ohne Übergewicht.
- Wartezeiten von Usern vor/nach dem System-Update.

Fall 1: Abhängige Stichproben (Varianz unbekannt)

Satz (Differenzentest (abhängige Stichproben, Varianz unbekannt))
Gegeben seien zwei abhängige Stichproben normalverteilter i.i.d. Zufallsvariablen, $X_1, \dots, X_n$ *und* $Y_1, \dots, Y_n$, *unbekannter Varianzen* σ_1^2, σ_2^2 *und unbekannter Erwartungswerte* μ_1, μ_2.
Wir definieren $Z_1, \dots, Z_n$ *mit* $Z_i := X_i - Y_i$, *und es sei* $\mathbf{s}^*$ *die* **korrigierte Standardabweichung** *der zugehörigen* **Stichprobe** $z_1, \dots, z_n$. *Gilt die Hypothese* $\mathcal{H}_0$: "$\mu_1 = \mu_2$", *so ist die Testvariable*

$$T := \frac{\overline{Z}}{s^*/\sqrt{n}}$$

t-verteilt *mit n-1 Freiheitsgraden. Gegeben ein Signifikanzniveau* $\alpha \in]0,1[$, *akzeptieren wir* $\mathcal{H}_0$, *falls*

$$-\mathbf{t}^{\mathbf{n-1}}_{\mathbf{1-\alpha/2}} \leq \mathbf{T} \leq \mathbf{t}^{\mathbf{n-1}}_{\mathbf{1-\alpha/2}}$$

Anmerkungen

- Analogie zu oben: Unbekannte Varianz $\rightarrow$ t-Verteilung.

Fall 1: Abhängige Stichproben

Beweis

Fall 1: Abhängige Stichproben (Beispiel)

- Wir vermessen 6 elektrische Widerstände mit zwei unterschiedlichen Messgeräten $\rightarrow X$ und Y.
- Hypothese $\mathcal{H}_0 : \mu_1 - \mu_2 = 0$
  (d.h., die Messgeräte liefern im Mittel dasselbe Ergebnis).
- **Varianzen** σ_1^2, σ_2^2 unbekannt. **Signifikanzniveau**: $\alpha = 0.01$

i	1	2	3	4	5	6
x_i	100.5	102.0	104.3	101.5	98.4	102.9
y_i	98.2	99.1	102.4	101.1	96.2	101.8
z_i	2.3	2.9	1.9	0.4	2.2	1.1

- Wir berechnen aus $z_1, ..., z_n$: $\bar{z} = 1.8$, $s^* = 0.903$
- Testwert t:

$$t = \frac{\bar{z}}{s^*/\sqrt{n}} = \frac{1.8}{0.903/\sqrt{6}} = 4.833$$

- Nicht-kritischer Bereich *(lese $t_{0.995}$ aus Quantiltabelle der t-Verteilung mit 5 Freiheitsgraden ab):* $[-4.032, 4.032]$
- Wir lehnen die Hypothese $\mathcal{H}_0$ somit ab!

Fall 2: Unabhängige Stichproben***

Hier betrachten wir nur den Fall unbekannter aber gleicher Varianzen $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Dieser Test ist auch bekannt als **Zweistichproben-t-Test**:

Satz (Zweistichproben-t-Test)

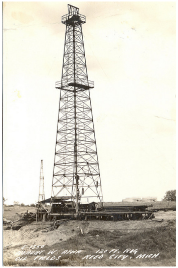
Gegeben seien zwei unabhängige Stichproben normalverteilter i.i.d. Zufallsvariablen, $X_1, \dots, X_{n_1}$ *und* $Y_1, \dots, Y_{n_2}$ *unbekannter, aber gleicher Varianz* σ^2 *und unbekannter Erwartungswerte* μ_1, μ_2*. Es seien* $\mathbf{s_1^*}, \mathbf{s_2^*}$ *die* **korrigierten Standardabweichungen** *der beiden Stichproben. Dann ist die Testvariable*

$$T := \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\overline{X} - \overline{Y}}{\sqrt{(n_1 - 1) \cdot s_1^{*2} + (n_2 - 1) \cdot s_2^{*2}}}$$

t-verteilt *mit* $n_1 + n_2 - 2$ *Freiheitsgraden. Gegeben ein Signifikanzniveau* $\alpha \in]0, 1[$*, akzeptieren wir die Hypothese* $\mathcal{H}_0$*: "*$\mu_1 = \mu_2$*", falls*

$$-\mathbf{t_{1-\alpha/2}^{n_1+n_2-2}} \leq \mathbf{T} \leq \mathbf{t_{1-\alpha/2}^{n_1+n_2-2}}$$

Unabhängige Stichprobe: Beispiel*** Bild: [5]

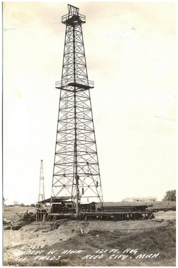


Wir wollen die Leistung zweier **Ölpumpen** miteinander vergleichen. X/Y bezeichne das pro Minute geförderte Ölvolumen von Pumpe A/B. Wir erheben Stichproben mit folgenden Kennwerten *(die Varianzen σ_1^2, σ_2^2 seien gleich aber unbekannt)*:

Pumpe A	$n_1 = 10$	$\bar{x} = 50$	$s_1^* = 4.21$
Pumpe B	$n_2 = 12$	$\bar{y} = 45$	$s_2^* = 4.81$

- Wir testen die Hypothese $\mathcal{H}_0 : \mu_1 \leq \mu_2$ *("Pumpe B leistet nicht weniger als Pumpe A")* und legen $\alpha = 0.05$ fest.

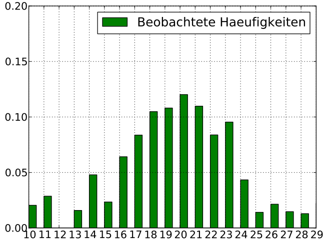
Unabhängige Stichproben: Beispiel***

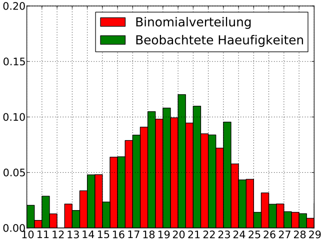


Outline

Von Parametertests zu Verteilungstests

- **Bisher**: Sogenannte **Parametertests**, d.h. wir prüfen eine Aussage bezüglich eines *Parameters* (z.B. μ) einer Verteilung.
- **Jetzt**: **Verteilungstests**, d.h. wir prüfen ob die Verteilung selbst der Hypothese entspricht.
- **Beispiel**: $\mathcal{H}_0 =$ "Die Ergebnisse eines gegebenen Würfels sind gleichverteilt, d.h. der Würfel ist fair".
- **Schlüsselfrage**: *Wie gut passen die beobachteten* **Häufigkeiten** $\mathbf{h_1}, \ldots, \mathbf{h_m}$ *auf die* **Wahrscheinlichkeiten** *der Verteilung,* $\mathbf{p_1}, \ldots, \mathbf{p_m}$*?*





Der χ^2-Verteilungstest

Satz (χ^2-Verteilungstest)

$X_1, \ldots, X_n$ seien **beliebig verteilte** **diskrete** **(i.i.d) Variablen** *mit m Ausprägungen. Gegeben eine Stichprobe $x_1, \ldots, x_n$, zählen wir die Vorkommen der Ausprägungen und erhalten absoluten Häufigkeiten $H_1, \ldots, H_m$. Die* **Hypothese** *$\mathcal{H}_0$ lautet:*

> *"$X_1, \ldots, X_n$ folgen einer gegebenen Verteilung $p_1, \ldots, p_m$, d.h. für alle $k{=}1, \ldots, m$ gilt: $P(X{=}k) = p_k$".*

Dann ist die Testgröße:

$$T_{\chi^2} := \sum_{k=1}^{m} \frac{(H_k - n{\cdot}p_k)^2}{n{\cdot}p_k}$$

$\boldsymbol{\chi}^2$**-verteilt** *mit $m-1$ Freiheitsgraden. Wir akzeptieren die Hypothese $\mathcal{H}_0$ mit Signifikanzniveau α, falls*

$$\mathbf{T}_{\boldsymbol{\chi}^2} \leq \mathbf{c}_{\mathbf{1}-\boldsymbol{\alpha}}^{\mathbf{m}-\mathbf{1}}$$

Anmerkungen

- $c_{1-\alpha}^{m-1}$ ist das $(1-\alpha)$-Quantil der χ^2-Verteilung.

Der χ^2-Verteilungstest: Idee

Die χ^2-Verteilung

Definition (Die χ^2-Verteilung)

Es sei $n \in \mathbb{N}^+$. Dann nennen wir eine stetige Zufallsvariable X mit der Dichtefunktion

$$f(x; n) = \begin{cases} A_n \cdot x^{\frac{n-2}{2}} \cdot e^{-\frac{x}{2}} & \textit{falls } x \geq 0 \\ 0 & \textit{sonst} \end{cases}$$

***χ^2-verteilt (sprich: “Chi-Quadrat”) mit n Freiheitsgraden**.*

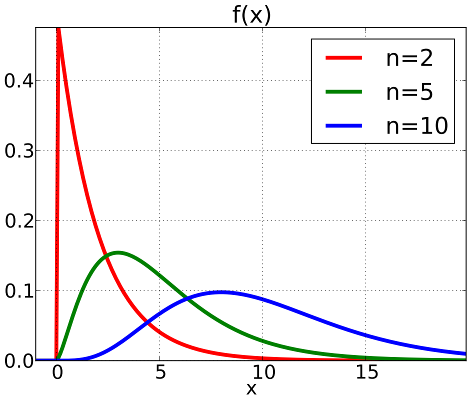
Anmerkungen

- Der Normalisierungsfaktor A_n wird mit Hilfe der **Gamma-Funktion** berechnet *(hier nicht im Detail)*.

Erwartungswert und Varianz

- Erwartungswert: $E(X) = n$
- Varianz: $Var(X) = 2n$

Die χ^2-Verteilung: Grafische Darstellung



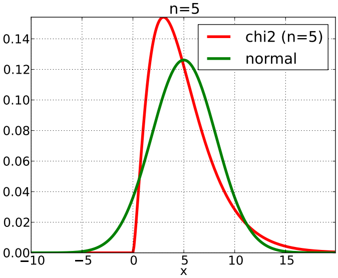
Anmerkungen

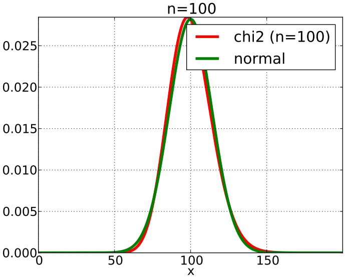
- Die χ^2-Verteilung ist asymmetrisch.
- Für $n > 2$ besitzt die Verteilung ein lokales Maximum bei $n - 2$. Für $n \leq 2$ ist die Verteilung monoton fallend.

Die χ^2-Verteilung: Eigenschaften

Anmerkungen (cont'd)

- Wir können die χ^2-Verteilung durch eine Normalverteilung $\varphi(x; \mu = n, \sigma^2 = 2n)$ annähern.
- **Bedingung**: n ist groß
- **Faustregel**: $n > 100$





Beispiel: Fairness-Prüfung eines Würfels

$$T_{\chi^2} := \sum_{k=1}^{m} \frac{(H_k - n \cdot p_k)^2}{n \cdot p_k}$$

$p_k = \frac{1}{6}$

- Wir haben $n =$ 120 mal gewürfelt und folgende Häufigkeiten beobachtet:

k	1	2	3	4	5	6
H_k	15	19	22	21	17	26

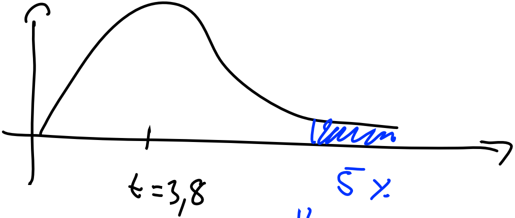
- Wir wählen das Signifikanzniveau $\alpha = 5\%$
- Wir berechnen den Testwert t_{χ^2}...

$t_{\chi^2} = \frac{5^2}{20} + \frac{1^2}{20} + \frac{2^2}{20} + \dots$

$\approx 3{,}8$

K	1	2	3	4	5	6
p_k	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
$n \cdot p_k$	20	20	20	20	20	20
$(H_k - n \cdot p_k)^2$	5^2	1^2	2^2	1^2	3^2	6^2

Beispiel: Fairness-Prüfung eines Würfels



$\Downarrow$

$c^5_{95\%}$ 5 Freiheitsgrade (weil 6 Realisierungen)

$= 11{,}07$ (Tabelle)

$\Longrightarrow$ Hypothese akzeptieren, weil $t \le c^5_{95\%}$ (Würfel ist ok).

Beispiel: Fairness-Prüfung eines Würfels[5]

	p									
f	0,005	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99	0,995
1	0,000	0,000	0,001	0,004	0,016	2,71	3,84	5,02	6,63	7,88
2	0,01	0,020	0,051	0,103	0,211	4,61	5,99	7,38	9,21	10,60
3	0,07	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,35	12,84
4	0,21	0,297	0,484	0,711	1,064	7,78	9,49	11,14	13,28	14,86
5	0,41	0,554	0,831	1,15	1,16	9,24	11,07	12,83	15,09	16,75
6	0,68	0,872	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	12,02	14,06	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,96
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,57	5,58	17,28	19,67	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
22	8,6	9,5	11,0	12,3	14,0	30,8	33,9	36,8	40,3	42,8
24	9,9	10,9	12,4	13,8	15,7	33,2	36,4	39,4	43,0	45,6
26	11,2	12,2	13,8	15,4	17,3	35,6	38,9	41,9	45,6	48,3
28	12,5	13,6	15,3	16,9	18,9	37,9	41,3	44,5	48,3	51,0
30	13,8	15,0	16,8	18,5	20,6	40,3	43,8	47,0	50,9	53,7
40	20,7	22,2	24,4	26,5	29,1	51,8	55,8	59,3	63,7	66,8
50	28,0	29,7	32,4	34,8	37,7	63,2	67,5	71,4	76,2	79,5
60	35,5	37,5	40,5	43,2	46,5	74,4	79,1	83,3	88,4	92,0
70	43,3	45,4	48,8	51,7	55,3	85,5	90,5	95,0	100,4	104,2
80	51,2	53,5	57,2	60,4	64,3	96,6	101,9	106,6	112,3	116,3
90	59,2	61,8	65,6	69,1	73,3	107,6	113,1	118,1	124,1	128,3
100	67,3	70,1	74,2	77,9	82,4	118,5	124,3	129,6	135,8	140,2

[5] Bildquelle: *Papula: Math. für Ing. und Nwsch., Band 3.*

Outline

Beispiel: Sind diese Variablen unabhängig?

- 100 Probanden wurden nach Ihrer Lieblings-Eissorte gefragt.
- Sind die Variablen *G (Geschlecht)* und *E (Lieblings-Eis)* unabhängig?

	E=Schoko	E=Vanille	E=Erdbeer	Häufigkeit
G=weiblich	11	16	3	30
G=männlich	39	19	12	70
Häufigkeit	50	35	15	

Idee des χ^2-Unabhängigkeitstests

Idee des χ^2-Unabhängigkeitstests

Der χ^2-Unabhängigkeitstest

Satz (χ^2-Unabhängigkeitstest)

$(X_1, Y_1), ..., (X_n, Y_n)$ *seien* **Paare diskreter (i.i.d) Variablen**. *Die X-Werte besitzen p Ausprägungen, die Y-Werte q Ausprägungen. Aus einer Stichprobe* $(x_1, y_1), ..., (x_n, y_n)$ *leiten wir die absoluten Häufigkeiten* H_{ij} *(mit* $i{=}1, .., p$ *und* $j{=}1, .., q$*) ab. Ebenso berechnen wir die Randhäufigkeiten der Realisierungen von* X, $H_1, .., H_p$, *und der Realisierungen von* Y, $H'_1, .., H'_q$. *Wir definieren:*

$$H^*_{ij} := \frac{H_i \cdot H'_j}{n}$$

Dann ist die Testgröße

$$T_{\chi^2} := \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(H_{ij} - H^*_{ij})^2}{H^*_{ij}}$$

$\boldsymbol{\chi}^2$**-verteilt** *mit* $(p-1)(q-1)$ *Freiheitsgraden. Wir akzeptieren die Hypothese* $\mathcal{H}_0$ *(= "X und Y sind unabhängig") bei Signifikanzniveau* α, *falls*

$$\mathbf{T}_{\boldsymbol{\chi}^2} \leq \mathbf{c}_{\mathbf{1-\boldsymbol{\alpha}}}^{(\mathbf{p-1})(\mathbf{q-1})}$$

Anmerkungen

- $c_{1-\boldsymbol{\alpha}}^{(p-1)(q-1)}$: $(1-\alpha)$-Quantil der χ^2-Vtlg. mit $(p-1)(q-1)$ Freiheitsgraden

χ^2-Unabhängigkeitstest: Beispiel

	E=Schoko	E=Vanille	E=Erdbeer	Häufigkeit
G=weiblich	11	16	3	30
G=männlich	39	19	12	70
Häufigkeit	50	35	15	

Tabelle mit Idealwerten H_{ij}^*

	E=Schoko	E=Vanille	E=Erdbeer	Häufigkeit
G=weiblich				30
G=männlich				70
Häufigkeit	50	35	15	

χ^2-Unabhängigkeitstest: Beispiel

χ^2-Unabhängigkeitstest: Beispiel

The End

References

[1] Dylan Meconis: Parent Line.
https://flic.kr/p/5R2khv (changed to black and white, CC license, retrieved: Dec 2016).

[2] The MNIST Database of Handwritten Digits.
http://yann.lecun.com/exdb/mnist/ (retrieved: Oct 2016).

[3] Sven Blankenberger and Dirk Vorberg.
Die Auswahl statistischer Tests und Maße.
http://www.students.uni-marburg.de/~Cal/Zeug/Entscheidungsbaum.pdf (retrieved: Oct 2016).

[4] Christian Schnettelker.
Pills / Pillen II.
https://flic.kr/p/t6rAvc (released under CC 2.0 license, retrieved: Jan 2017).

[5] Don...The UpNorth Memories Guy... Harrison.
Reed City MI RPPC Atha Oil Fields 120ft Rig LL Cook E-1350 Postmarked 1943.
https://flic.kr/p/5StUEZ (released under CC 2.0 license, retrieved: Jan 2017).

[6] Robert Müller (robert.molinarius).
mouth-watering !
https://flic.kr/p/9g7ZPn (released under CC 2.0 license, cropped, horizontally flipped, background whitened, retrieved: Jan 2017).