Stochastik für die Informatik, Vorlesung 21

Inhalt

- Hypothesentests
- p-Wert
- t-Test
- χ^2-Test

Lernziele

- Grundprinzipien von Hypothesentests kennen
- p-Werte berechnen können
- Die Anwendungsbereiche des $t-$Test kennen, den Test durchführen und interpretieren können
- Die Anwendungsbereiche des χ^2-Test kennen, den Test durchführen und interpretieren können

Vorkenntnisse Stoff der vorherigen Vorlesungen, insbesondere Verteilungen, Quantile, t-Verteilung.

Testen von Hypothesen

Vorgehen: Situation: $x_1, ..., x_n$ Messwerte, Grundannahmen festhalten.

1. Formuliere die Nullhypothese H_0. (Bsp).
2. Konstruiere den Annahmebereich (bzw. Ablehnungsbereich).
3. Führe das Experiment durch.
4. Überprüfe, ob die Daten im Annahmebereich liegen.

Ergebnis von 4.:

- Falls nein: Nullhypothese verwerfen
- Falls ja: Nullhypothese nicht verwerfen, bzw. annehmen

Falls Nullhypothese verworfen: Alternativhypothese $H_A := \neg H_0$ annehmen.

Testen von Hypothesen

Vorgehen: Situation: $x_1, \dots, x_n$ Messwerte, Grundannahmen festhalten.

1. Formuliere die Nullhypothese H_0. (Bsp).
2. Konstruiere den Annahmebereich (bzw. Ablehnungsbereich).
3. Führe das Experiment durch.
4. Überprüfe, ob die Daten im Annahmebereich liegen.

Ergebnis von 4.:

- Falls nein: Nullhypothese verwerfen
- Falls ja: Nullhypothese nicht verwerfen, bzw. annehmen

Falls Nullhypothese verworfen: Alternativhypothese $H_A := \neg H_0$ annehmen.

Testen von Hypothesen

(Def.10.1) Seien $X_1, ..., X_n$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$. Seien $x_1, ..., x_n \in \mathbb{R}$ Messwerte. Eine Nullhypothese über $X_1, ..., X_n$ ist eine Aussage, deren Wahrheitsgehalt nur von der gemeinsamen Verteilung von $X_1, ..., X_n$ abhängt. Die Alternativhypothese H_A ist die Negation von H_0.

Ein Annahmebereich zum Fehlerniveau $\alpha \in [0, 1]$ für die Nullhypothese H_0 ist eine Teilmenge A von $\mathbb{R}^n$, so dass

$$\mathbb{P}((X_1, ...X_n) \in A \,|\, H_0 \text{ ist wahr }) \geq 1 - \alpha$$

ist.

Fehler 1. und 2. Art

(Def. 10.2) Sei H_0 eine Nullhypothese, und A ein Annahmebereich. Dann sind zwei Typen von "Fehlern" möglich:

- Der Fehler 1. Art: Die Nullhypothese wird verworfen, obwohl sie eigentlich richtig wäre
- Der Fehler 2. Art: Die Nullhypothese wird angenommen, obwohl sie eigentlich falsch ist

p-Wert

(Def.) Eine Teststatistik ist eine Zufallsvariable Y, welche als Funktion

$$Y = f(X_1, ..., X_n)$$

der zum Experiment gehörigen Zufallsvariablen geschrieben werden kann. Der Testwert $y = f(x_1, ..., x_n)$ ist der Wert der Teststatistik für die gemessenen Daten.

(Def. 10.3) Sei Y eine Teststatistik, und H_0 eine Nullhypothese zu den entsprechenden Daten. Der rechtsseitige p-Wert für $y \in Y(\Omega)$ ist definiert als

$$p(y) = \mathbb{P}(Y \geq y \mid H_0),$$

der linksseitige p-Wert als

$$p(y) = \mathbb{P}(Y \leq y \mid H_0),$$

und der beidseitige p-Wert als

$$p(y) = 2 \min\{\mathbb{P}(Y \geq y \mid H_0), \mathbb{P}(Y \leq y \mid H_0)\}.$$

p-Wert

- Der $p-$ Wert gibt die Wahrscheinlichkeit für Werte "gleich oder extremer als die Teststatistik an der Stelle der Messwerte $x_1, ..., x_n$"
- $p(y)$ groß bedeutet: Die Daten $(x_1, ..., x_n)$ sind "typische" (und nicht extreme) Realisierungen der Zufallsvariablen $X_1, ..., X_n$. Große $p-$Werte stützen also die Nullhypothese.

(Satz) Sei $p(y)$ ein $p-$Wert zu einer Teststatistik $Y = f(X_1, ..., X_n)$. Dann ist

$$A_\alpha := \{x : p(x) \geq \alpha\}$$

ein Annahmebereich zum Fehlerniveau α.

- H_0 wird also angenommen wenn $p \geq \alpha$ ist, und abgelehnt wenn $p < \alpha$ ist.
- (Beispiel 10.2: p-Wert im Materialbelastungstest)

$t-$Test

Der t-Test ist ein konkreter Test, welcher dem allgemeinen Vorgehen folgt und auf der t-Verteilung basiert.

Verwendung:

- Ein Merkmal: Test ob der Erwartungswert einer Zufallsvariablen gleich (kleiner gleich, größer gleich) einer gegebenen Zahl μ_0 ist
- Zwei Merkmale: Test ob zwei unabhängige Zufallsvariablen denselben Erwartungswert haben

Ein Merkmal: Gegeben: Daten $x_1, ..., x_n$

Annahme Daten sind Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen $X_1, ..., X_n$. Diese sind (annähernd) normalverteilt (z.B. weil der zentrale Grenzwertsatz gilt)

t-Test für ein Merkmal

Zwei Situationen:

- Einseitiger t-Test: Test ob der Erwartungswert μ der X_i größer (kleiner) als eine gegebenen Zahl μ_0 ist, d.h.

$$H_0 : \quad \mu \geq \mu_0$$

  bzw.

$$H_0 : \quad \mu \leq \mu_0.$$

- Zweiseitiger t-Test: Test ob der Erwartungswert μ der X_i gleich einer gegebenen Zahl μ_0 ist, d.h. Nullhypothese

$$H_0 : \quad \mu = \mu_0$$

Ein Merkmal, einseitiger t-Test: Vorgehen

1. Stelle die Nullhypothese auf: H_0: $\mu \leq \mu_0$
2. Wähle Fehlerniveau $\alpha \in (0, 1)$
3. Erhebe die Daten $x_1, ..., x_n$.
4. Bestimme den sogenannten *Freiheitsgrad* f: Im t-Test ist $f = n - 1$
5. Bestimme $t_{1-\alpha,f}$, das $1 - \alpha-$Quantil der $t-$Verteilung zum Parameter f
6. Berechne $\bar{\mu}_n$ und $\bar{\sigma}_n = \sqrt{\bar{\sigma}_n^2}$ aus den Daten
7. Berechne den Testwert
$$t = \sqrt{n} \cdot \frac{\bar{\mu}_n - \mu_0}{\bar{\sigma}_n}$$
8. Vergleiche den Testwert t mit dem Quantil $t_{1-\alpha,f}$.
   Entscheidungsregel:
   - Ist $t \geq t_{1-\alpha,f}$, so wird H_0 abgelehnt
   - Andernfalls wird H_0 nicht abgelehnt (angenommen).

Ein Merkmal, einseitiger $t-$Test

Begründung des Vorgehens: Nach Satz 9.4 ist unter den getroffenen Annahmen, und falls H_0 gilt, die Zufallsvariable

$$T := \sqrt{n}\frac{\bar{\mu}_n - \mu_0}{\bar{\sigma}_n}$$

(annähernd) t-verteilt mit Parameter $n-1$.

Falls der Testwert t größer als $t_{1-\alpha,n-1}$ ist, bedeutet das, dass im Falle dass H_0 gilt, die Daten $x_1, ..., x_n$ sehr unwahrscheinlich sind. Deshalb lehnt man in diesem Fall H_0 ab.

- Nullhypothese $\mu \geq \mu_0$: Umgekehrte Entscheidungsregel $t \leq t_{\alpha,f} = -t_{1-\alpha,f}$ für Ablehnung
- Alternative Entscheidungsregeln: $p-$ Wert, Konfidenzintervall
- (Beispiel 10.4)

Ein Merkmal, einseitiger $t-$Test: Beispiel 10.3

Beispiel: Wirkung eines fiebersenkenden Medikamentes

Nr. Patient	Temp. Einnahme	Temp. 3 h später	Differenz
i	u_i	v_i	$x_i = u_i - v_i$
1	39.1	38.7	0.4
2	38.3	38.1	0.2
3	37.6	37.9	-0.3
4	38.0	37.5	0.5
5	40.1	39.2	0.9
6	39.5	39.1	0.4
7	38.7	38.7	0.0
8	37.9	37.5	0.4
9	39.2	38.2	1.0
10	38.0	37.4	0.6

Ein Merkmal: Zweiseitiger $t-$Test

1. Stelle die Nullhypothese auf: H_0: $\mu = \mu_0$
2. Wähle Fehlerniveau $\alpha \in (0, 1)$
3. Erhebe die Daten $x_1, ..., x_n$.
4. Bestimme den sogenannten *Freiheitsgrad* f: Im t-Test ist $f = n - 1$
5. Bestimme $t_{1-\alpha/2,f}$, das $1 - \alpha/2-$Quantil der $t-$Verteilung zum Parameter f
6. Berechne $\bar{\mu}_n$ und $\bar{\sigma}_n$ aus den Daten
7. Berechne den Testwert
$$t = \sqrt{n} \cdot \frac{\bar{\mu}_n - \mu_0}{\bar{\sigma}_n}$$
8. Vergleiche den Testwert t mit dem Quantil $t_{1-\alpha/2,f}$. Entscheidungsregel:
    - Ist $|t| \geq t_{1-\alpha/2,f}$, so wird H_0 abgelehnt
    - Andernfalls wird H_0 angenommen

$t-$Test für zwei Merkmale

Test ob zwei unabhängige Zufallsvariablen denselben Erwartungswert haben

Gegeben: Daten $x_1, ..., x_n$ und $y_1, ..., y_m$

Annahme:

- Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen $X_1, ..., X_n$, bzw. von unabhängigen, identisch verteilten Zufallsvariablen $Y_1, ..., Y_m$.
- X_1 und Y_1 sind (annähernd) normalverteilt (z.B. weil der zentrale Grenzwertsatz gilt)
- $\mu_x = \mathbb{E}[X_1], \mu_y = \mathbb{E}[Y_1]$, $\sigma_x^2 = \sigma_y^2$

Nullypothese H_0: $\mu_x = \mu_y$

(Bem.: Einseitiger/zweiseitiger Test)

$t-$Test für zwei Merkmale, zweiseitig

1. Stelle die Nullhypothese auf: H_0: $\mu_x = \mu_y$
2. Wähle Fehlerniveau $\alpha \in (0, 1)$
3. Erhebe die Daten $x_1, \dots, x_n, y_1, \dots, y_m$.
4. Bestimme den sogenannten *Freiheitsgrad* f: Im t-Test mit zwei Größen ist $f = n + m - 2$
5. Bestimme $t_{1-\alpha/2,f}$, das $1 - \alpha/2-$Quantil der $t-$Verteilung zum Parameter f
6. Berechne die empirischen Mittel und empirischen Varianzen $\bar{\mu}_x, \bar{\mu}_y$ und $\bar{\sigma}_x^2, \bar{\sigma}_y^2$ aus den Daten, sowie
$$S^2 := \frac{(n-1)\bar{\sigma}_x^2 + (m-1)\bar{\sigma}_y^2}{n+m-2}$$
7. Berechne den Testwert
$$t = \sqrt{\frac{nm}{n+m}} \frac{\bar{\mu}_x - \bar{\mu}_y}{S}$$
8. Entscheidungsregel:
   - Ist $|t| \geq t_{1-\alpha/2,f}$, so wird H_0 abgelehnt
   - Andernfalls wird H_0 angenommen

Beispiel 10.4: $t-$Test für zwei Merkmale, zweiseitig

- H_0: $\mu_x = \mu_y$
- Fehlerniveau $\alpha = 0.05$
- Beispiel mit R: Untersuchung zur Wirkung von Marihuanakonsum auf das Kurzzeitgedächtnis. Daten: Anzahl erfolgreich ausgeführter Aufgaben
  - x=nonsmokers = c(18,22,21,17,20,17,23,20,22,21)
  - y= smokers = c(16,20,14,21,20,18,13,15,17,21)
- R-Befehl: t.test (Bem: optionale Eingaben)

Beispiel: $t-$Test für zwei Merkmale, zweiseitig

```
> t.test(nonsmokers,smokers,var.equal=T)

    Two Sample t-test

data:  nonsmokers and smokers
t = 2.2573, df = 18, p-value = 0.03665
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1801366 5.0198634
sample estimates:
mean of x mean of y 
     20.1      17.5 
```

$t_{0.975,18} = 2,1009$

Die Nullhypothese wird also abgelehnt, da $|t| > t_{0.975,18}$,

Alternative Entscheidungsregeln: Ablehnung, da
$p = 0.0366 < 0.05 = \alpha$, bzw. da das Konfidenzintervall die 0 nicht umfasst.

χ^2-Test: Beispiel

Beispiel 10.6

(Daten: Gewinnspiel einer Getränkemarke zur WM 2014, eigene Erhebung)

Gruppe	A	B	C	D	E	F	G	H
Anzahl	2	4	2	3	3	7	5	4

Gesamtzahl: 30

Fragestellung: Sind alle Gruppen gleich häufig vertreten?

χ^2-Test

Der χ^2-Test kann in folgenden Situationen angewandt werden:

- Test ob die Daten einer bestimmten Verteilung folgen
- Test auf Unabhängigkeit

Grundprinzip (Test auf feste Verteilung):

- Gegeben: Daten $x_1, ..., x_n$, gruppiert in k Klassen mit Häufigkeiten $N_1, ..., N_k$.
- Grundannahme: Realisierungen unabhängiger Zufallsvariablen
- Nullhypothese: Die Daten folgen einer bestimmten Verteilung (Normalverteilung, uniforme Verteilung....)
- Falls die Klassengrößen alle hinreichend groß sind, so folgt die mittlere quadratische Abweichung der gemessenen Häufigkeiten von den theoretischen einer χ^2-Verteilung.

χ^2-Test

Der χ^2-Test kann in folgenden Situationen angewandt werden:

- Test ob die Daten einer bestimmten Verteilung folgen
- Test auf Unabhängigkeit

Grundprinzip (Test auf feste Verteilung):

- Gegeben: Daten $x_1, ..., x_n$, gruppiert in k Klassen mit Häufigkeiten $N_1, ..., N_k$.
- Grundannahme: Realisierungen unabhängiger Zufallsvariablen
- Nullhypothese: Die Daten folgen einer bestimmten Verteilung (Normalverteilung, uniforme Verteilung....)
- Falls die Klassengrößen alle hinreichend groß sind, so folgt die mittlere quadratische Abweichung der gemessenen Häufigkeiten von den theoretischen einer χ^2-Verteilung.

χ^2-Test auf Verteilung

1. Gruppieren der Daten in k Klassen $A_1, ..., A_k$, wobei die Klassen A_i Teilmengen von $\mathbb{R}$ sind, so dass $A_i \cap A_j = \emptyset$ gilt, und so dass jeder Messwert x_j in einer Klasse A_i enthalten ist. Bestimmung der Häufigkeiten
$$N_i := |\{j : x_j \in A_i\}|, i = 1, ..., k.$$
2. Aufstellen der Nullhypothese: H_0 : Die Daten folgen einer bestimmten Verteilung (z.B. Normalverteilung,...). Ev. Schätzung der Parameter der Verteilung.
3. Wahl des Fehlerniveaus $\alpha \in (0, 1)$
4. Berechnung des Freiheitsgrades: $f = k - m - 1$, wobei m die Anzahl geschätzert Parameter in der angenommenen Verteilung ist.
5. Bestimmung des $1 - \alpha-$Quantils $\chi^2_{1-\alpha,f}$ der χ^2-Verteilung mit Parameter $k - m - 1$ als Vergleichswert.

χ^2-Test auf Verteilung

6. Berechung der theoretischen Häufigkeiten unter der Nullhypothese:

$$F_i = n \cdot \mathbb{P}(X \in A_i \mid H_0), i = 1, ..., k$$

7. Berechnung des Testwerts

$$\chi^2 := \sum_{i=1}^{k} \frac{(N_i - F_i)^2}{F_i}.$$

8. Vergleich des Testwerts χ^2 mit dem Quantil $\chi^2_{1-\alpha,f}$.
   Entscheidungsregel:
   - Ist $\chi^2 > \chi^2_{1-\alpha,f}$ so wird H_0 verworfen.
   - Andernfalls wird H_0 angenommen.

Beispiel 10.6:

Gruppe	A	B	C	D	E	F	G	H
Häufigkeiten	2	4	2	3	3	7	5	4

Nullhypothese: Die Daten sind diskret gleichverteilt

Fehlerniveau $\alpha = 0.05$

Freiheitsgrad 8-1=7 (8 Klassen, keine geschätzen Parameter)

Theoretische Häufigkeiten: $F_i = 30/8 = 3.75, i = 1, ..., 8$
Testwert

$$\chi^2 = \frac{1}{3.75}\left((2 - 3.75)^2 + (4 - 3.75)^2 + ... + (4 - 3.75)^2\right) = 5.2$$

Tabelle: Quantile der χ^2-Verteilung

Tabelle der Quantile der Chiquadrat- und Gamma-Verteilung

α-Quantile $\chi^2_{n;\alpha}$ der Chiquadrat-Verteilungen $\boldsymbol{\chi}^2_n = \Gamma_{1/2,n/2}$ mit n Freiheitsgraden. $\chi^2_{n;\alpha}$ ist der Wert $c > 0$ mit $\boldsymbol{\chi}^2_n([0,c]) = \alpha$. Durch Skalierung erhält man die Quantile der Gamma-Verteilung $\Gamma_{\lambda,r}$ mit $\lambda > 0$ und $2r \in \mathbb{N}$. Notation: $^{-5}3.9 \equiv 3.9 \cdot 10^{-5}$.

$\alpha =$	0.005	0.01	0.02	0.05	0.1	0.9	0.95	0.98	0.99	0.995
$n = 1$	$^{-5}3.9$	$^{-4}1.6$	$^{-4}6.3$	$^{-3}3.9$	0.0158	2.706	3.842	5.412	6.635	7.879
2	0.0100	0.0201	0.0404	0.1026	0.2107	4.605	5.991	7.824	9.210	10.60
3	0.0717	0.1148	0.1848	0.3518	0.5844	6.251	7.815	9.837	11.34	12.84
4	0.2070	0.2971	0.4294	0.7107	1.064	7.779	9.488	11.67	13.28	14.86
5	0.4117	0.5543	0.7519	1.145	1.610	9.236	11.07	13.39	15.09	16.75
6	0.6757	0.8721	1.134	1.635	2.204	10.65	12.59	15.03	16.81	18.55
7	0.9893	1.239	1.564	2.167	2.833	12.02	14.07	16.62	18.48	20.28
8	1.344	1.646	2.032	2.733	3.490	13.36	15.51	18.17	20.09	21.95
9	1.735	2.088	2.532	3.325	4.168	14.68	16.92	19.68	21.67	23.59
10	2.156	2.558	3.059	3.940	4.865	15.99	18.31	21.16	23.21	25.19
11	2.603	3.053	3.609	4.575	5.578	17.28	19.68	22.62	24.72	26.76
12	3.074	3.571	4.178	5.226	6.304	18.55	21.03	24.05	26.22	28.30
13	3.565	4.107	4.765	5.892	7.042	19.81	22.36	25.47	27.69	29.82
14	4.075	4.660	5.368	6.571	7.790	21.06	23.68	26.87	29.14	31.32
15	4.601	5.229	5.985	7.261	8.547	22.31	25.00	28.26	30.58	32.80
16	5.142	5.812	6.614	7.962	9.312	23.54	26.30	29.63	32.00	34.27
17	5.697	6.408	7.255	8.672	10.09	24.77	27.59	31.00	33.41	35.72

Beispiel 10.6:

Gruppe	A	B	C	D	E	F	G	H
Häufigkeiten	2	4	2	3	3	7	5	4

Nullhypothese: Die Daten sind diskret gleichverteilt

Fehlerniveau $\alpha = 0.05$

Freiheitsgrad $f = 8 - 1 = 7$ (8 Klassen, keine geschätzen Parameter)

Theoretische Häufigkeiten: $F_i = 30/8 = 3.75, i = 1, ..., 8$

Testwert

$$\chi^2 = \frac{1}{3.75}\Big((2 - 3.75)^2 + (4 - 3.75)^2 + ... + (4 - 3.75)^2\Big) = 5.2$$

Das 0.95-Quantil der χ^2-Verteilung mit Parameter $n = 7$ ist

$\chi^2_{0.95,7} = 14.07$

Die Nullhypothese wird also nicht abgelehnt.