Schließende Statistik

Neil Mañibo

Universität Bielefeld

19.07.2024

6.3: Das Testen von Hypothesen

Zwei-Stichproben-Tests

Doppelter t-Test

Wiederholung: Doppelter t-Test

- Seien X, Y zwei unabhängige ZVn, mit unbekannten EW $\mu_X = \mathbb{E}(X)$ und $\mu_Y = \mathbb{E}(Y)$
- Die Varianzen sind unbekannt, werden aber gleich vorausgesetzt.
$$\sigma^2 = V(X) = V(Y)$$
- Gegeben: eine Stichprobe von X (bzw. Y) von Umfang n_x (bzw. n_y)
- Mittelwerte: $\overline{x}, \overline{y}$
- empirische Varianzen: s_x^2, s_y^2

Wiederholung: Doppelter t-Test

- Hypothesenpaar: $H_0\colon \mu_X = \mu_Y$ gegen $H_A\colon \mu_X \neq \mu_Y$
- Prüfgröße: $\overline{D} := \overline{X} - \overline{Y}$
- Equivalentes Paar von H_0, H_A

$$H_0\colon \mu_D = 0 \quad \text{gegen} \quad H_A\colon \mu_D \neq 0$$

- Verteilung von $\overline{D}$ unter H_0: $\overline{D} \sim \mathcal{N}(0, \sigma^2_{\overline{D}})$, wobei die Varianz unbekannt ist

Wiederholung: Schätzwert für σ^2

$$\overline{D} = \overline{X} - \overline{Y} \implies V(\overline{D}) = V(\overline{X}) + V(\overline{Y})$$

$$V(\overline{X}) + V(\overline{Y}) = \frac{1}{n_x}V(X) + \frac{1}{n_y}V(Y) = \frac{n_x + n_y}{n_x n_y}\sigma^2 \qquad (6.20)$$

Nun müssen wir einen Schätzwert für σ^2 finden (aus der beiden Stichproben!) – **gewichtetes Mittel von s_x und s_y:**

$$\boxed{\widehat{\sigma}^2 = s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} \qquad (*)}$$

Wiederholung: Schätzwert für $\sigma_{\overline{D}}^2$

$$\widehat{\sigma}^2 = s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

$$\Longrightarrow \widehat{\sigma}_{\overline{D}}^2 = \frac{n_x + n_y}{n_x n_y} s^2$$

$\Longrightarrow$ die standardisierte Prüfgröße

$$\frac{\overline{D}}{S}\sqrt{\frac{n_x n_y}{n_x + n_y}}$$

ist unter H_0 Student-t-verteilt (t_k) mit $k = n_x + n_y - 2$ Freiheitsgraden.

Wiederholung: Doppelter t-Test, Ablehnungskriterium

$$H_0\colon \mu_D = 0 \quad \text{gegen} \quad H_A\colon \mu_D \neq 0$$

die Nullhypothese H_0 ist abzulehnen falls

$$\frac{|\overline{d}|}{s}\sqrt{\frac{n_x n_y}{n_x + n_y}} > t_{k,1-\frac{\alpha}{2}}$$

wobei

- $\overline{d} = \overline{x} - \overline{y}$: Differenz der emp. Mittelwerte
- s: Schätzung für σ aus $(*)$
- $t_{k,1-\frac{\alpha}{2}} = t_{n_x+n_y-2,\,1-\frac{\alpha}{2}}$

Beispiel

Beispiel

In einer Kölner Klinik wurden im Jahr 1985 $n_x = 269$ Mädchen und $n_y = 288$ Jungen geboren, mit Durschnittsgewicht $\overline{x} = 3050\,\mathrm{g}$ bzw. $\overline{y} = 3300\,\mathrm{g}$. Die zugehörigen empirischen Standardabweichungen waren $s_x = 460\,\mathrm{g}$ und $s_y = 470\,\mathrm{g}$.
Es soll die Nullhypothese H_0: "Das Geburtsgewicht von Jungen und Mädchen ist gleich" gegen
H_A: "Das Geburtsgewicht ist verschieden."
zum Signifikanzniveau $\alpha = 0.01$ getestet werden.

1. $\overline{d}$, k, und s^2 berechnen
2. das richtige Quantil $t_{k,1-\frac{\alpha}{2}}$ festlegen
3. Werte ins Ablehnungskriterium einseitzen

Beispiel

Beispiel

In einer Kölner Klinik wurden im Jahr 1985 $n_x = 269$ Mädchen und $n_y = 288$ Jungen geboren, mit Durschnittsgewicht $\overline{x} = 3050\,\text{g}$ bzw. $\overline{y} = 3300\,\text{g}$. Die zugehörigen empirischen Standardabweichungen waren $s_x = 460\,\text{g}$ und $s_y = 470\,\text{g}$.
Es soll die Nullhypothese H_0: "Das Geburtsgewicht von Jungen und Mädchen ist gleich" gegen
H_A: "Das Geburtsgewicht ist verschieden."
zum Signifikanzniveau $\alpha = 0.01$ getestet werden.

- $|\overline{d}| = |\overline{x} - \overline{y}| = 250, \quad k = n_x + n_y - 2 = 555$
- $s^2 = \frac{268 \cdot 460^2 + 287 \cdot 470^2}{555} = 216409.2$
- $\frac{|\overline{d}|}{s} \sqrt{\frac{n_x n_y}{n_x + n_y}} = 6.34 > 2.576 = z_{0.995} \simeq t_{555,0.995}$

$\Longrightarrow$ H_0 wird abgelehnt.

Doppelter, einseitiger t-Test

$$H_0 \colon \mu_D \leqslant 0 \quad \text{gegen} \quad H_A \colon \mu_D > 0$$

die Nullhypothese H_0 ist abzulehnen falls

$$\frac{\bar{d}}{s}\sqrt{\frac{n_x n_y}{n_x + n_y}} > t_{k,1-\alpha}$$

Doppelter, einseitiger t-Test

$$H_0\colon \mu_D \geqslant 0 \quad \text{gegen} \quad H_A\colon \mu_D < 0$$

die Nullhypothese H_0 ist abzulehnen falls

$$\frac{\overline{d}}{s}\sqrt{\frac{n_x n_y}{n_x + n_y}} < -t_{k,1-\alpha}$$

t-Differenztest

Gepaarte Stichproben: t-Differenztest

- An n unabhängigen Objekten werden jeweils die Mermale X und Y gemessen.
  Annahme: X, Y sind normalverteilt
- n Paare $(X_1, Y_1), \ldots, (X_n, Y_n)$ von ZVn
- Beispiele
  - An n Personen:
    X: Länge des linken Bein
    Y: Länge des linken Bein
  - Bei n verschiedengeschlechtlichen Zwillingsgeburten:
    X: Geburtsgewicht des Mädchens
    Y: Geburtsgewicht des Jungen
- X_i und Y_i sind nicht unabhängig
  X_i ist aber von X_j, und Y_j, $j \neq i$ unabhängig!

Gepaarte Stichproben: t-Differenztest

- Die Erwartungswerte $\mu_X = \mathbb{E}(X)$ und $\mu_Y = \mathbb{E}(Y)$ sind unbekannt (ebenso die Varianzen)
- Hypothesenpaar:

$$H_0\colon \mu_X = \mu_Y \qquad \text{gegen} \qquad H_A\colon \mu_X \neq \mu_Y$$

- Betrachte die neue ZV $D := X - Y$.
  Unter H_0 ist $D \sim \mathcal{N}(0, \sigma_D^2)$
- einfachen t-Test für D durchführen!

$$H_0\colon \mu_D = 0 \qquad \text{gegen} \qquad H_A\colon \mu_D \neq 0$$

Gepaarte Stichproben: t-Differenztest

$$H_0\colon \mu_D = 0 \qquad \text{gegen} \qquad H_A\colon \mu_D \neq 0$$

- Prüfgröße: $\overline{D} = \frac{1}{n}\sum_{i=1}^{n}(D_1 + D_2 + \cdots + D_n) = \overline{X} - \overline{Y}$
- Unter H_0: $\frac{D\sqrt{n}}{S_D} \sim t_{n-1}$,
  wobei S_D: Standardabweichung der ZV D
- Stichprobe für D: $d_i = x_i - y_i$
- emp. Mittelwert: $\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \overline{x} - \overline{y}$
- emp. Varianz: $s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2$

Gepaarte Stichproben: t-Differenztest

$$H_0\colon \mu_D = 0 \qquad \text{gegen} \qquad H_A\colon \mu_D \neq 0$$

Ablehnungskriterium: H_0 ablehnen falls

$$\frac{|\bar{d}|}{s_d}\sqrt{n} > t_{n-1,1-\frac{\alpha}{2}}$$

Beispiel

Beispiel

Der Ertrag X (in Zentnern) von 8 Kirschbäumen im Jahr 1990 wird mit dem Ertrag Y derselben 8 Bäume im Jahr 1991 verglichen. Es soll die Nullhypothese
H_0: "Der Ertrag war in beiden Jahren gleich" gegen die Alternativhypothese H_A: "Der Ertrag ist verschieden" zum Signifikanzniveau 0.05 getestet werden.

	1	2	3	4	5	6	7	8
x_i	3.6	3.1	3.4	3.2	3.5	3.1	3.2	3.5
y_i	3.5	3.5	3.4	3.6	4.0	3.5	3.3	3.1

Beispiel

Beispiel

Der Ertrag X (in Zentnern) von 8 Kirschbäumen im Jahr 1990 wird mit dem Ertrag Y derselben 8 Bäume im Jahr 1991 verglichen. Es soll die Nullhypothese
H_0: "Der Ertrag war in beiden Jahren gleich" gegen die Alternativhypothese H_A: "Der Ertrag ist verschieden" zum Signifikanzniveau 0.05 getestet werden.

1. Tabelle für die Werte x_i, y_i, $d_i = x_i - y_i$ erstellen
2. $\overline{x}$, $\overline{y}$, $\overline{d}$ berechnen
3. die empirische Varianz s_d^2 aus d_i und $\overline{d}$ berechen
4. $\overline{d}, s_d$ und $\sqrt{n}$ ins Ablehnungskriterium einsetzen

Schritt 1: d_i ausrechnen

	1	2	3	4	5	6	7	8
x_i	3.6	3.1	3.4	3.2	3.5	3.1	3.2	3.5
y_i	3.5	3.5	3.4	3.6	4.0	3.5	3.3	3.1
d_i	0.1	-0.4	0.0	-0.4	-0.5	-0.4	-0.1	0.4

Schritt 2: $\overline{x}, \overline{y}, \overline{d}$ ausrechnen

	1	2	3	4	5	6	7	8	
x_i	3.6	3.1	3.4	3.2	3.5	3.1	3.2	3.5	$\overline{x} = 3.33$
y_i	3.5	3.5	3.4	3.6	4.0	3.5	3.3	3.1	$\overline{y} = 3.48$
d_i	0.1	-0.4	0.0	-0.4	-0.5	-0.4	-0.1	0.4	$\overline{d} = -0.15$

Schritt 3: empirische Varianz s_d^2 ausrechnen

	1	2	3	4	5	6	7	8	
x_i	3.6	3.1	3.4	3.2	3.5	3.1	3.2	3.5	$\overline{x} = 3.33$
y_i	3.5	3.5	3.4	3.6	4.0	3.5	3.3	3.1	$\overline{y} = 3.48$
d_i	0.1	−0.4	0.0	−0.4	−0.5	−0.4	−0.1	0.4	$\overline{d} = -0.15$

$$s_d^2 = \frac{1}{7}\sum_{i=1}^{8}(d_i - \overline{d})^2 = 0.104$$

Schritt 4: Ablehnungskriterium überprüfen

$$\frac{|\overline{d}|}{s_d}\sqrt{n} > t_{n-1,1-\frac{\alpha}{2}}$$

- $\alpha = 0.05 \implies t_{7,0.975} = 2.365$
- $\frac{|\overline{d}|}{s_d}\sqrt{n} = 1.32 < 2.365 = t_{7,0.975}$
- $\implies H_0$ wird beibehalten
- $\implies$ kein signifikanter Unterschied zwischen Erträgen aus den beiden Jahre

Bemerkung: Analog führt man den entsprechenden einseitigen Test durch! (mit $\frac{\overline{d}}{s_d}\sqrt{n}$ (ohne den Betrag), $t_{n-1,1-\alpha}$)

Der Chi-Quadrat-Anpassungstest im endlichen Fall

χ^2-Anpassungstest

- Y: diskrete ZV
- Mögliche Realisierungen von Y: $y_1, \dots, y_r$ (r Werte)
- Nullhypothese:

$$H_0: \text{ Es gilt } \mathbb{P}(Y = y_i) = p_i \quad \text{für alle} \quad i = 1, \dots, r$$

- Alternativhypothese:

$$H_A: Y \text{ folgt eine andere Verteilung}$$

- χ^2-Anpassungstest: ein Test darauf, dass eine ZV einer bestimmten Verteilung *nicht folgt*

χ^2-Anpassungstest

- In einer Stichprobe vom Umfang n kommt y_i n_i-mal vor.
- n_i: **beobachtete Haüfigkeit** für y_i (z.B. bei einem Säulendiagramm)
- Wäre H_0 richtig gewesen, findet man beim zufälligen Herausgreifen eines Objekts aus der Grundgesamtheit den Wert y_i mit Wahrscheinlichkeit p_i
- N_i: ZV für n_i
- $\mathbb{E}(N_i) = e_i = np_i$ (unter H_0): **erwartete Häufigkeit** für y_i

$$\chi^2 := \sum_{i=1}^{r} \frac{(n_i - e_i)^2}{e_i}$$

χ^2-Anpassungstest

$$\chi^2 := \sum_{i=1}^{r} \frac{(n_i - e_i)^2}{e_i}$$

- χ^2: Maß für die Abweichung der Stichprobe von der hypothetischen Verteilung
- Für groß genug n folgt die zugehörige ZV

$$\mathrm{X}^2 := \sum_{i=1}^{r} \frac{(N_i - e_i)^2}{e_i}$$

die χ^2-Verteilung (mit $k = r - 1$ Freiheitsgraden)

χ^2-Anpassungstest: Ablehnungskriterium

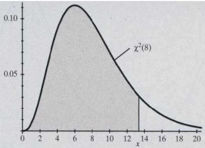
H_0 ist abzulehnen falls

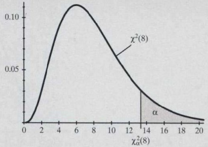
$$\boxed{\chi^2 > \chi^2_{k,1-\alpha}}$$

wobei $\chi^2_{k,q}$: q-Quantil der χ^2-Verteilung mit k Freiheitsgraden

- Test ist immer einseitig
- Die Ablehnung von H_0 zeigt einen statistischen Nachweis, dass Y eine andere Verteilung folgt!

Quantile der χ^2-Verteilung





$$P(X \leq x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw$$

	$P(X \leq x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21
11	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72
12	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22
13	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69
14	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14
15	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58
16	5.812	6.908	7.962	9.312	23.54	26.30	28.84	32.00
17	6.408	7.564	8.672	10.08	24.77	27.59	30.19	33.41
18	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.80
19	7.633	8.907	10.12	11.65	27.20	30.14	32.85	36.19

Faustregeln

- Der χ^2-Test darf nur benutzt werden
  - Im Fall $r \leqslant 8$, wenn alle $e_i \geqslant 5$ sind.
  - Im Fall $r > 8$, wenn alle $e_i \geqslant 1$
- Wenn die Daten (mind.) ordinalskaliert sind und n groß genug ist, lassen sich die Faustregeln erfüllen, indem man benachbarte Werte zu größeren Klassen zusammenfasst.
- Falls einen (oder mehrere Parameter) der hypotetischen Verteilung (unter H_0) erst aus der Stichprobe geschätzt werden muss, verringert sich die Zahl der Freiheitsgrade um a (die Zahl der geschätzten Parameter), also $k = r - a - 1$. (siehe das Beispiel aus S. 84 des Skripts)

Beispiel

Beispiel

Sei X eine diskrete Zufallsvariable mit Realisierungen in $S = \{0, 1, \ldots, 9\}$. Aus einer Stichprobe vom Umfang $n = 200$ bekommt man die folgende Häufigkeitstabelle.

x_i	0	1	2	3	4	5	6	7	8	9
n_i	15	18	29	19	10	25	12	20	21	31

Wir wollen einen Test durchführen (zum Signifikanzniveau $\alpha = 0.1$), um herauszufinden, ob X die Beobachtungen mit der Gleichverteilung auf $S = \{0, 1, \ldots, 9\}$ verträglich sind.

Beispiel

Beispiel

Sei X eine diskrete Zufallsvariable mit Realisierungen in $S = \{0, 1, \ldots, 9\}$. Aus einer Stichprobe vom Umfang $n = 200$ bekommt man die folgende Häufigkeitstabelle.

x_i	0	1	2	3	4	5	6	7	8	9
n_i	15	18	29	19	10	25	12	20	21	31

Wir wollen einen Test durchführen (zum Signifikanzniveau $\alpha = 0.1$), um herauszufinden, ob X die Beobachtungen mit der Gleichverteilung auf $S = \{0, 1, \ldots, 9\}$ verträglich sind.

- "H_0: X ist gleichverteilt auf S" gegen "H_A: X ist nicht gleichverteilt auf S"
- Unter H_0: $\mathbb{P}(X = x_i) = \frac{1}{10}$ für alle x_i

Beispiel

x_i	0	1	2	3	4	5	6	7	8	9
n_i	15	18	29	19	10	25	12	20	21	31
e_i	20	20	20	20	20	20	20	20	20	20
$(n_i - e_i)^2$	25	4	81	1	100	25	81	0	1	121

$$\chi^2 := \sum_{i=0}^{9} \frac{(n_i - e_i)^2}{e_i} = \frac{439}{20} = 21.95$$

Beispiel

x_i	0	1	2	3	4	5	6	7	8	9
n_i	15	18	29	19	10	25	12	20	21	31
e_i	20	20	20	20	20	20	20	20	20	20
$(n_i - e_i)^2$	25	4	81	1	100	25	81	0	1	121

$$\chi^2 := \sum_{i=0}^{9} \frac{(n_i - e_i)^2}{e_i} = \frac{439}{20} = 21.95$$

- $r = 10, a = 0 \implies k = r - a - 1 = 9$
- Quantil: $\chi^2_{k,1-\alpha} = \chi^2_{9,0.90} = 14.68$
- Da $21.95 > 14.68 \implies H_0$ wird **abgelehnt**
- Es gibt genug statistischen Nachweis, dass X **die Gleichverteilung *nicht* folgt**.

χ^2-Anpassungstest im allgemeinen Fall

- Diskrete ZVn mit unendlich viele Realiserungen sowie kontinuierliche ZVn
- Einteilung der Wertebereich der ZV in endlich viele Klassen!

Beispiel: Falls X Werte von $(-\infty, \infty)$ annimmt, kann man die folgende Zerlegung betrachten:

$$A_1 = (-\infty, x_1],\ A_2 = (x_1, x_2], \ldots, A_r = (x_{r-1}, \infty)$$

mit $x_1 < x_2 < \cdots < x_{r-1}$.

- Die Klasseneinteilung ist dabei so vorzunehmen, dass für $r \leqslant 8$ alle $e_i = np_i \geqslant 5$ sind oder falls $r > 8$ alle $e_i \geqslant 1$ sind.