

II. Zwei-Stichproben-Tests

1. **Gepaarte Stichproben: t-Differenzentest.** An n unabhängigen Objekten werden jeweils die Merkmale X und Y gemessen, die als normalverteilt vorausgesetzt werden. Wir erhalten also n Paare $(X_1, Y_1), \dots, (X_n, Y_n)$ mit Realisierungen $(x_1, y_1), \dots, (x_n, y_n)$. Hier ein paar Beispiele: An n Personen wird jeweils die Länge X des linken und die Länge Y des rechten Beins gemessen; bei n verschiedengeschlechtlichen Zwillingsgeburten wird das Geburtsgewicht X des Mädchens und das Geburtsgewicht Y des Jungen registriert; oder in $n = 12$ Kalendermonaten bezeichnen X und Y die Arbeitslosigkeit in zwei verschiedenen Jahren. X_i und Y_i sind dabei *nicht* unabhängig (wohl aber X_i von X_j und Y_j für $j \neq i$) — man spricht auch von *verbundenen* Stichproben). Die Erwartungswerte $\mu_X := \mathbb{E}(X)$ und $\mu_Y := \mathbb{E}(Y)$ seien unbekannt, ebenso die Varianzen. Man testet auf Gleichheit der Erwartungswerte, also $H_0 : \mu_X = \mu_Y$ gegen $H_A : \mu_X \neq \mu_Y$. Wir betrachten nun die neue ZV $D := X - Y$ und nehmen an, dass diese normalverteilt ist. Dann ist D unter H_0 nach $\mathcal{N}(0, \sigma_D^2)$ verteilt, wobei wir die Varianz σ_D^2 nicht kennen. Der Test lässt sich also auf einen einfachen t-Test zurückführen, und zwar $H_0 : \mu_D = 0$ versus $H_A : \mu_D \neq 0$. Die Prüfgröße ist dann der Stichprobenmittelwert von D , also $\bar{D} := \frac{1}{n}(D_1 + \dots + D_n) = \bar{X} - \bar{Y}$. Unter H_0 ist der standardisierte Stichprobenmittelwert

$$\frac{\bar{D}}{S_D} \sqrt{n}$$

nach t_{n-1} verteilt, wobei S_D^2 die empirische Varianz von D als Zufallsvariable bezeichnet.

Um uns Schätzwerte für $\bar{D}$ und σ_D^2 zu verschaffen, brauchen wir eine Stichprobe für D . Die ergibt sich aus der ursprünglichen Stichprobe über die Differenzen

$$d_i = x_i - y_i, \quad i = 1, \dots, n.$$

Der zugehörige Stichprobenmittelwert ist

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \bar{x} - \bar{y},$$

und die empirische Varianz lautet

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2.$$

H_0 wird abgelehnt, falls

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

Ganz analog werden entsprechende einseitige Tests durchgeführt.

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 “Der Ertrag war in beiden Jahren gleich” gegen die Alternativhypothese “Der Ertrag ist verschieden” zum Signifikanzniveau 0.05 getestet werden.

Baum-Nr. i	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	$\bar{x} = 3.33$
y_i	3.5	3.5	3.4	3.6	4.0	3.5	3.3	3.1	$\bar{y} = 3.48$
d_i	0.1	-0.4	0.0	-0.4	-0.5	-0.4	-0.1	0.4	$\bar{d} = -0.15$

Es ergibt sich $s_d^2 = 0.104$ und somit $\frac{|\bar{d}|}{s_d} \sqrt{n} = 1.32 < 2.365 = t_{7,0.975}$, also kein signifikanter Unterschied.

2. **Zwei unabhängige Stichproben: Doppelter t-Test.** X und Y seien nun *unabhängige*, normalverteilte ZV mit unbekannten Erwartungswerten $\mu_X := \mathbb{E}(X)$ und $\mu_Y := \mathbb{E}(Y)$. Die Varianzen sind ebenfalls unbekannt, werden aber als *gleich* vorausgesetzt: $\sigma^2 = \mathbb{V}(X) = \mathbb{V}(Y)$. (Im Fall ungleicher Varianzen verwendet man stattdessen den sogenannten Berens–Fisher–Test.) Gegeben sind Stichproben von X und Y vom Umfang n_x bzw. n_y . Ihre Mittelwerte sind $\bar{x}$ bzw. $\bar{y}$, und ihre empirischen Varianzen s_x^2 bzw. s_y^2 .

Wir testen $H_0 : \mu_X = \mu_Y$ gegen $H_A : \mu_X \neq \mu_Y$. Prüfgröße ist wieder die *Differenz der Stichprobenmittel*,

$$\overline{D} := \overline{X} - \overline{Y}.$$

Wegen der Unabhängigkeit von X und Y folgt aus dem Additionssatz der Normalverteilung, dass $\overline{D}$ unter H_0 eine normalverteilte ZV ist mit $\mathbb{E}(\overline{D}) = 0$ und

$$\sigma_{\overline{D}}^2 := \mathbb{V}(\overline{D}) = \mathbb{V}(\overline{X}) + \mathbb{V}(\overline{Y}) = \frac{1}{n_x} \mathbb{V}(X) + \frac{1}{n_y} \mathbb{V}(Y) = \frac{n_x + n_y}{n_x n_y} \sigma^2. \quad (6.20)$$

Schätzwert für $\overline{D}$ ist natürlich wieder $\bar{d} := \bar{x} - \bar{y}$. Um diesen standardisieren zu können, müssen wir uns nun einen Schätzwert für $\sigma_{\overline{D}}^2$ verschaffen. Aus (6.20) folgt, dass es ausreicht, einen Schätzwert für das unbekannte σ^2 zu gewinnen. Diesen erhalten wir aus den beiden Stichproben als *gewichtetes Mittel* von s_x^2 und s_y^2 mit Gewichtungsfaktoren $n_x - 1$ und $n_y - 1$:

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

Somit ergibt sich aus (6.20) der Schätzwert für die Varianz von $\overline{D}$ als

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

Es lässt sich nun zeigen, dass die standardisierte Prüfgröße

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

unter H_0 Student-t-verteilt mit $k = n_x + n_y - 2$ Freiheitsgraden. H_0 ist also zu verwerfen, falls

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

Analog konstruiert man entsprechende einseitige Tests.

Beispiel. In einer Kölner Klinik wurden im Jahr 1985 $n_x = 269$ Mädchen und $n_y = 288$ Jungen geboren, mit Durchschnittsgewicht $\bar{x} = 3050$ g bzw. $\bar{y} = 3300$ g. Die zugehörigen empirischen Standardabweichungen waren $s_x = 460$ g und $s_y = 470$ g. Es soll die Nullhypothese “Das Geburtsgewicht von Jungen und Mädchen ist *gleich*” zum Signifikanzniveau $\alpha = 0.01$ getestet werden gegen die Alternativhypothese “Das Geburtsgewicht ist *verschieden*”. Es ist $|\bar{d}| = |\bar{x} - \bar{y}| = |3050 - 3300| = 250$, $k = n_x + n_y - 2 = 555$, $s^2 = \frac{268 \cdot 460^2 + 287 \cdot 470^2}{555} = 216409.2$, $\frac{|\bar{x} - \bar{y}|}{s} \sqrt{\frac{n_x n_y}{n_x + n_y}} = 6.34 > 2.576 = z_{0.995} \simeq t_{555, 0.995}$. Die Nullhypothese wird also abgelehnt.

Bemerkung. Die Voraussetzung einer normalverteilten Grundgesamtheit kann man bei allen behandelten Tests fallenlassen, wenn die Stichprobe so groß ist, dass die zu vergleichenden Mittelwerte in guter Näherung normalverteilt sind, wie das beim GEBURTENBEISPIEL der Fall war.

6.3.2 Der Chi-Quadrat-Anpassungstest im endlichen Fall

Die bisher behandelten statistischen Tests betrafen Erwartungswerte von Verteilungen. Manche Fragen kann man anhand des Erwartungswerts allein aber nicht entscheiden. Zum Beispiel kann die erwartete Augenzahl eines gefälschten Würfels mit der eines idealen Würfels identisch sein. Um die beiden Würfel zu unterscheiden, reicht es also nicht, die Erwartungswerte zu vergleichen – man muss vielmehr die gesamte *Verteilung* untersuchen. Dies geschieht durch einen sogenannten **Anpassungstest**.

Wir betrachten nur den Fall diskreter Zufallsvariablen. Sei Y eine diskrete Zufallsvariable, die r verschiedene Werte $y_1, y_2, \dots, y_r$ annehmen kann. (Es reicht, wenn diese nominalskaliert sind!) Die Nullhypothese behauptet, dass die Verteilung von Y durch vorgegebene Wahrscheinlichkeiten $p_1, p_2, \dots, p_r$ beschrieben wird:

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

Die Alternativhypothese H_A besagt, dass Y eine andere als die von der Nullhypothese behauptete Verteilung hat. Beachten Sie noch einmal, dass wir mit einem