Zwei-Stichproben-t-Test für gepaarte Stichproben

statistik-online.ch

Der **Zwei-Stichproben-t-Test für gepaarte Stichproben** prüft, ob sich zwei verbundene Messungen im Durchschnitt unterscheiden.

1 Grundidee

Wir vergleichen nicht zwei unabhängige Gruppen, sondern die Differenzen innerhalb zusammengehöriger Paare.

Typische Fragestellung

Es gibt pro Person, Objekt oder Paar zwei zusammengehörige Werte. Aus diesen beiden Werten wird jeweils eine Differenz gebildet.

- Hat sich der Blutdruck nach einer Behandlung verändert?
- Sind Personen nach einem Training schneller als vorher?
- Unterscheiden sich zwei Messmethoden bei denselben Objekten?

Wichtig: gepaarte Stichproben

Dieses Dokument behandelt den Zwei-Stichproben-t-Test für **gepaarte** oder **abhängige** Stichproben. Die beiden Werte gehören jeweils zusammen. Bei unabhängigen Gruppen wird stattdessen der Zwei-Stichproben-t-Test für unabhängige Stichproben verwendet.

Situation	**Passender Test**	**Beispiel**
dieselben Personen zweimal	gepaarter t-Test	vor und nach Training
zugeordnete Paare	gepaarter t-Test	Zwillinge oder Matching
zwei unabhängige Gruppen	unabhängiger t-Test	Klasse A vs. Klasse B

2 Differenzen

Der gepaarte t-Test wird über die Differenzen berechnet. Wichtig ist, dass die Richtung der Differenz vorher festgelegt wird.

$$d_i = x_{i,2} - x_{i,1}$$

Symbol	Bedeutung	Beispiel
$x_{i,1}$	erster Wert des Paares	vor dem Training
$x_{i,2}$	zweiter Wert des Paares	nach dem Training
d_i	Differenz des Paares	nachher minus vorher

Danach wird geprüft, ob der Mittelwert der Differenzen von 0 abweicht. Mathematisch ist der gepaarte t-Test deshalb ein Ein-Stichproben-t-Test auf den Differenzen.

3 Hypothesen

Die Nullhypothese besagt, dass der mittlere Unterschied 0 ist:

$$H_0 : \mu_d = 0$$

Die Alternativhypothese hängt von der Fragestellung und der gewählten Differenzrichtung ab:

zweiseitig:	$H_1 : \mu_d \neq 0$	Unterschied in beide Richtungen
rechtsseitig:	$H_1 : \mu_d > 0$	die zweite Messung ist im Mittel grösser
linksseitig:	$H_1 : \mu_d < 0$	die zweite Messung ist im Mittel kleiner

4 Teststatistik

Zuerst werden aus allen Paaren die Differenzen d_i berechnet. Danach benötigt man den Mittelwert der Differenzen $\bar{d}$, die Standardabweichung der Differenzen s_d und die Anzahl der Paare n.

$$\bar{d} = \frac{d_1 + d_2 + \cdots + d_n}{n}$$

$$s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$$

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$

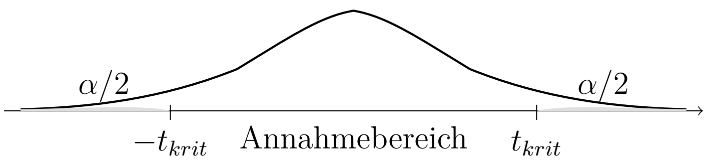
Symbol	Bedeutung
d_i	Differenz innerhalb eines Paares
$\bar{d}$	Mittelwert der Differenzen
s_d	Standardabweichung der Differenzen
n	Anzahl der Paare
$df = n - 1$	Freiheitsgrade

Je weiter der berechnete t-Wert von 0 entfernt ist, desto stärker spricht die Stichprobe gegen die Nullhypothese.

5 Entscheidungsregel

Die Entscheidung erfolgt ausschliesslich durch den Vergleich mit dem kritischen t-Wert t_{krit}.

Testart	**H_0 verwerfen, wenn ...**	**Ablehnungsbereich**
zweiseitig	$\|t\| > t_{krit}$	links und rechts
rechtsseitig	$t > t_{krit}$	rechts
linksseitig	$t < -t_{krit}$	links



6 Voraussetzungen

- Die abhängige Variable ist metrisch.
- Die Werte liegen als zusammengehörige Paare vor.
- Die Paare sind voneinander unabhängig.
- Die Differenzen d_i sind annähernd normalverteilt.
- Bei grösseren Stichproben ist der Test relativ robust gegenüber moderaten Abweichungen von der Normalverteilung.

7 Kritische t-Werte

Der kritische t-Wert hängt von drei Informationen ab:

- dem Signifikanzniveau α,
- der Testart,
- den Freiheitsgraden $df = n - 1$.

df	**zweiseitig 5%**	**einseitig 5%**	**zweiseitig 1%**
7	2.365	1.895	3.499
9	2.262	1.833	3.250
11	2.201	1.796	3.106
14	2.145	1.761	2.977

8 Beispiel

Bei zehn Personen wird die Punktzahl vor und nach einem Training gemessen. Die Differenz wird definiert als:

$$d = \text{nachher} - \text{vorher}$$

Die beobachteten Differenzen sind:

$$3,\ 4,\ 2,\ 5,\ 1,\ 3,\ 4,\ 2,\ 6,\ 3$$

Es wird zweiseitig mit $\alpha = 0.05$ getestet, ob sich die Punktzahl im Mittel verändert hat.

$$H_0 : \mu_d = 0 \qquad H_1 : \mu_d \neq 0$$

$$\bar{d} = 3.3 \qquad s_d = 1.49$$

$$t = \frac{3.3}{1.49/\sqrt{10}} = \frac{3.3}{0.47} = 6.99$$

Die Freiheitsgrade sind:

$$df = n - 1 = 10 - 1 = 9$$

Für einen zweiseitigen Test mit $df = 9$ und $\alpha = 0.05$ gilt ungefähr $t_{krit} = 2.262$. Da $|6.99| > 2.262$, wird H_0 verworfen.

Interpretation: Die Stichprobe spricht dafür, dass sich die Punktzahl nach dem Training im Durchschnitt verändert hat.

9 Zusammenfassung

Schritt	Frage	Formel oder Entscheidung
1	Differenzen	$d_i = x_{i,2} - x_{i,1}$
2	Hypothesen	$H_0 : \mu_d = 0$
3	Standardfehler	$SE = s_d/\sqrt{n}$
4	Teststatistik	$t = \bar{d}/SE$
5	Freiheitsgrade	$df = n - 1$
6	Entscheidung	Vergleich mit t_{krit}

Aufgaben: Gepaarter Zwei-Stichproben-t-Test

1 – Blutdruck

Bei acht Personen wird der Blutdruck vor und nach einer Behandlung gemessen. Die Differenz wird definiert als $d = \text{nachher} - \text{vorher}$. Die Differenzen lauten:

$$-4,\ -6,\ -3,\ -5,\ -2,\ -7,\ -4,\ -5$$

Testen Sie linksseitig mit $\alpha = 0.05$, ob der Blutdruck nach der Behandlung im Durchschnitt kleiner ist.

- Formulieren Sie H_0 und H_1.
- Berechnen Sie $\bar{d}$, s_d und den t-Wert.
- Bestimmen Sie t_{krit} und treffen Sie eine Entscheidung.

2 – Reaktionszeit

Eine Gruppe von $n = 12$ Personen absolviert einen Reaktionstest vor und nach einem Training. Die Differenz wird definiert als $d = \text{nachher} - \text{vorher}$. Es gilt $\bar{d} = -6.4$ Millisekunden und $s_d = 8.2$ Millisekunden. Testen Sie linksseitig mit $\alpha = 0.05$, ob die Reaktionszeit nach dem Training kleiner ist.

- Formulieren Sie H_0 und H_1.
- Berechnen Sie den t-Wert.
- Interpretieren Sie das Ergebnis.

3 – Zwei Messmethoden

Bei $n = 15$ Objekten werden zwei Messmethoden verglichen. Die Differenz wird definiert als $d = \text{Methode B} - \text{Methode A}$. Es gilt $\bar{d} = 2.1$ und $s_d = 5.6$. Testen Sie zweiseitig mit $\alpha = 0.05$, ob sich die beiden Messmethoden im Durchschnitt unterscheiden.

- Formulieren Sie H_0 und H_1.
- Berechnen Sie t und df.
- Treffen Sie eine Entscheidung mit t_{krit}.

4 – Multiple Choice

Welche Aussagen sind korrekt? (Mehrfachantworten möglich)

- ❑ Beim gepaarten t-Test werden zuerst Differenzen gebildet.
- ❑ Die Freiheitsgrade betragen $n - 1$, wobei n die Anzahl der Paare ist.
- ❑ Der gepaarte t-Test setzt voraus, dass die beiden Gruppen unabhängig sind.
- ❑ Die Richtung der Differenz sollte vor der Berechnung festgelegt werden.
- ❑ Mathematisch ist der gepaarte t-Test ein Ein-Stichproben-t-Test auf den Differenzen.

Lösungen: Gepaarter Zwei-Stichproben-t-Test

Lösung zu Aufgabe 1

$$H_0 : \mu_d = 0 \qquad H_1 : \mu_d < 0$$

$$\bar{d} = -4.5 \qquad s_d = 1.60$$

$$t = \frac{-4.5}{1.60/\sqrt{8}} = \frac{-4.5}{0.57} = -7.94$$

Die Freiheitsgrade sind $df = 8 - 1 = 7$. Für einen linksseitigen Test mit $\alpha = 0.05$ gilt ungefähr $t_{krit} = 1.895$.

$$-7.94 < -1.895$$

Entscheidung: H_0 wird verworfen. Die Stichprobe spricht dafür, dass der Blutdruck nach der Behandlung im Durchschnitt kleiner ist.

Lösung zu Aufgabe 2

$$H_0 : \mu_d = 0 \qquad H_1 : \mu_d < 0$$

$$t = \frac{-6.4}{8.2/\sqrt{12}} = \frac{-6.4}{2.37} = -2.70$$

Die Freiheitsgrade sind $df = 12 - 1 = 11$. Für einen linksseitigen Test mit $\alpha = 0.05$ gilt ungefähr $t_{krit} = 1.796$.

$$-2.70 < -1.796$$

Entscheidung: H_0 wird verworfen. Es gibt einen Hinweis darauf, dass die Reaktionszeit nach dem Training im Durchschnitt kleiner ist.

Lösung zu Aufgabe 3

$$H_0 : \mu_d = 0 \qquad H_1 : \mu_d \neq 0$$

$$t = \frac{2.1}{5.6/\sqrt{15}} = \frac{2.1}{1.45} = 1.45$$

Die Freiheitsgrade sind $df = 15 - 1 = 14$. Für einen zweiseitigen Test mit $\alpha = 0.05$ gilt ungefähr $t_{krit} = 2.145$.

$$|1.45| < 2.145$$

Entscheidung: H_0 wird nicht verworfen. Die Stichprobe liefert keinen ausreichenden Hinweis darauf, dass sich die beiden Messmethoden im Durchschnitt unterscheiden.

Lösung zu Aufgabe 4

✓ Richtig – Beim gepaarten t-Test werden zuerst Differenzen gebildet.

✓ Richtig – Die Freiheitsgrade betragen $n - 1$, wobei n die Anzahl der Paare ist.

✗ Falsch – Die Werte sind gepaart oder abhängig, nicht unabhängig.

✓ Richtig – Die Richtung der Differenz bestimmt die Interpretation des Vorzeichens.

✓ Richtig – Der Test entspricht einem Ein-Stichproben-t-Test auf den Differenzen.