Hypothesen über einen Populationsanteil

Teststatistik

$$Z = \frac{p_1 - \pi_1}{\sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n}}}$$

(mit Zurücklegen)

$$Z = \frac{p_1 - \pi_1}{\sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n} \cdot \frac{N - n}{N - 1}}}$$

(ohne Zurücklegen*)

p_1 Stichprobenanteil π_1 in H_0 unterstellter Wert des Populationsanteils

---

*Korrekturfaktor vernachlässigbar, wenn $\frac{N}{n} > 20$

NV als Kennwerteverteilung, wenn

$$n \cdot \frac{\pi_1}{1 - \pi_1} > 9 \quad \text{und} \quad n \cdot \frac{1 - \pi_1}{\pi_1} > 9$$

(z.B.: wenn *n* im Bereich von 25-30 und 0,25 ≤ π1 ≤ 0,75);

Wird π1 durch p1 geschätzt, dann sollte n > 60

Vergleich zweier Anteilswerte aus unabhängigen Stichproben

$$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1 \cdot (1 - p_1)}{n_1} + \frac{p_2 \cdot (1 - p_2)}{n_2}}}$$

Für H_0: $\pi_1 - \pi_2 = 0$

$$\hat{\sigma}_{p_1 - p_2} = \sqrt{\hat{\pi}_{pooled} \cdot (1 - \hat{\pi}_{pooled}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\hat{\pi}_{pooled} = p_1 \cdot \frac{n_1}{n_1 + n_2} + p_2 \cdot \frac{n_2}{n_1 + n_2}$$

Orgqual\info	0	1	Gesamt
0	105 70,0%	251 47,5%	356 52,5%
1	45 30,0%	277 52,5%	322 47,5%
n	150	528	678

$$z = \frac{(0{,}3 - 0{,}525) - 0}{\sqrt{\frac{0{,}21}{150} + \frac{0{,}249}{528}}} = \frac{-0{,}225}{\sqrt{0{,}0014 + 0{,}00047}} = \frac{-0{,}225}{\sqrt{0{,}00187}} = \frac{-0{,}225}{0{,}043} = -5{,}2$$

1

$$\hat{\pi}_{pooled} = 0{,}3 \cdot \frac{150}{150+528} + 0{,}525 \cdot \frac{528}{150+528} = 0{,}475$$

$$\hat{\sigma}_{p_1-p_2} = \sqrt{0{,}475 \cdot (1-0{,}475) \cdot \left(\frac{1}{150} + \frac{1}{528}\right)} = 0{,}046$$

Hypothesen über einen Populationsmittelwert

Teststatistik

$$T = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}_x}{\sqrt{n}}} = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n-1}}}$$

(Wenn X in der Population normalverteilt und die Populationsvarianz unbekannt ist; bei df=n-1 Freiheitsgraden)

$$Z = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}_x}{\sqrt{n}}} = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n-1}}}$$

(Wenn X in der Population nicht normalverteilt und die Populationsvarianz unbekannt ist; falls n>30)

Vergleich zweier Mittelwerte aus unabhängigen Stichproben

$$H_0:\ \mu_1 - \mu_2 = 0 \qquad H_1:\ \mu_1 - \mu_2 \neq 0$$

Standardfehler der Differenz zweier Mittelwerte

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Bei **gleichen** Populationsvarianzen $\sigma_1^2 = \sigma_2^2 = \sigma^2$

resultiert:

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\sigma^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Schätzung der gemeinsamen Populationsvarianz

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}\left(x_{i1}-\bar{x}_1\right)^2 + \sum_{i=1}^{n_2}\left(x_{i2}-\bar{x}_2\right)^2}{\left(n_1-1\right)+\left(n_2-1\right)}$$

Schätzung des Standardfehlers der Differenz zweier Mittelwerte

$$\hat{\sigma}_{\left(\bar{x}_1-\bar{x}_2\right)} = \sqrt{\frac{\sum_{i=1}^{n_1}\left(x_{i1}-\bar{x}_1\right)^2 + \sum_{i=1}^{n_2}\left(x_{i2}-\bar{x}_2\right)^2}{\left(n_1-1\right)+\left(n_2-1\right)}} \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$\bar{x}_1$ $\bar{x}_2$ $(\bar{x}_1 - \bar{x}_2)$ **normal**verteilt, wenn $n_1 \geq 30$ und $n_2 \geq 30$

t-verteilt, mit *df* = $n_1 + n_2 - 2$

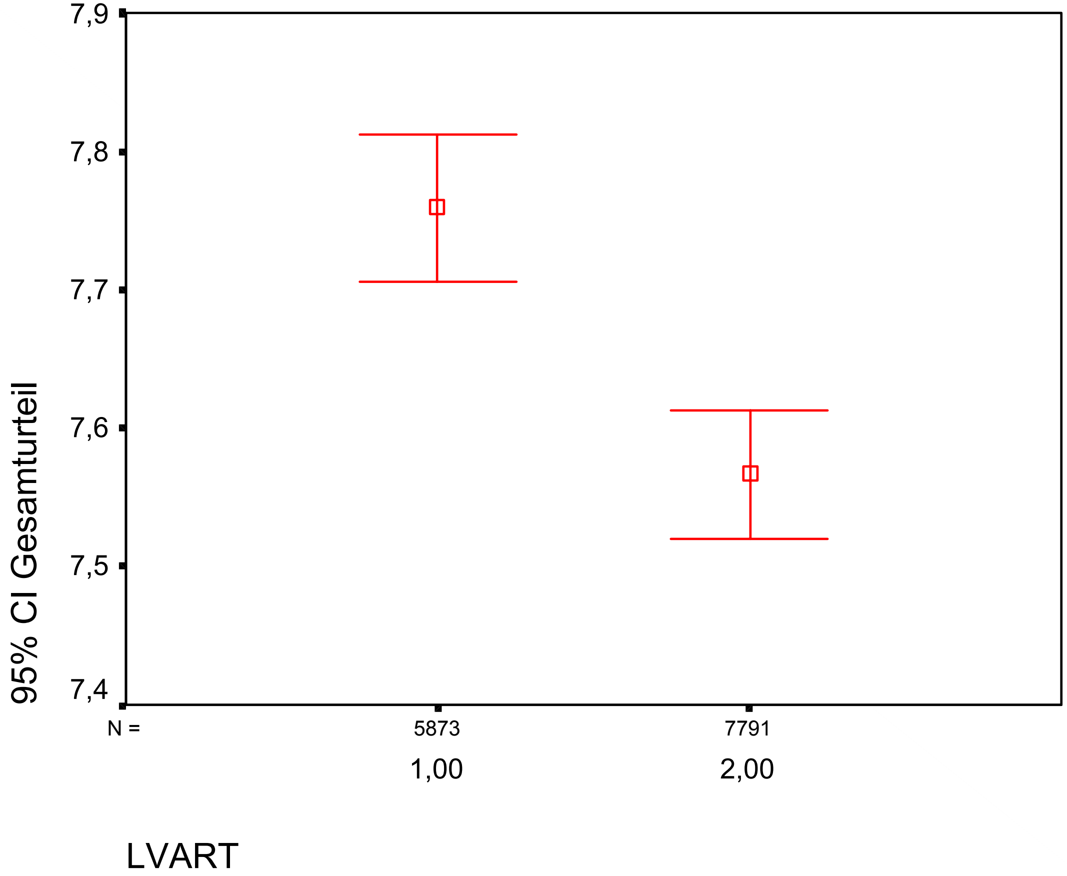
wenn

gleiche Populationsvarianzen und das Merkmal in den verglichenen Populationen normalverteilt ist

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$$

bzw. $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$ für H_0: $\mu_1 - \mu_2 = 0$

Gesamtbeurteilung der Qualität von Lehrveranstaltungen: 1=sehr schlecht, .., 11=sehr gut



$$\hat{\sigma}_{\bar{x}_1} = 0{,}0269$$

$$\hat{\sigma}_{\bar{x}_2} = 0{,}0236$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = 0{,}036$$

$$\bar{x}_1 = 7{,}76 \qquad \bar{x}_2 = 7{,}57 \qquad \Delta = 0{,}19$$

$$n_1 = 5.873 \qquad n_2 = 7.791 \qquad n = 13.664$$

Im Beispiel:

t=5,35; df=13.662; Sig.(2-seitig)=0,00

95% Konfidenzintervall

$$0{,}19 \pm 1{,}96 \times 0{,}036 = 0{,}19 \pm 0{,}07$$

Untere Grenze: 0,12

Obere Grenze = 0,26

Korrektur der Freiheitsgrade bei **heterogenen** Varianzen

$$df_{korr} = \frac{1}{\dfrac{c^2}{n_1 - 1} + \dfrac{(1-c)^2}{n_2 - 1}}$$

$$c = \frac{\hat{\sigma}^2_{\bar{x}_1}}{\hat{\sigma}^2_{\bar{x}_1} + \hat{\sigma}^2_{\bar{x}_2}}$$

Im Beispiel:

$$c = \frac{\hat{\sigma}^2_{\bar{x}_1}}{\hat{\sigma}^2_{\bar{x}_1} + \hat{\sigma}^2_{\bar{x}_2}} = \frac{0{,}0269^2}{0{,}0269^2 + 0{,}0236^2} = 0{,}5651$$

$$df_{korr} = \frac{1}{\dfrac{0{,}3193}{5872} + \dfrac{0{,}1891}{7790}} = 12.713{,}54$$

Nach Bortz (1999) Statistik für Sozialwissenschaftler. Berlin, S. 137f.

Gruppenstatistiken					
	V2	N	Mittel-wert	Standard-abweichung	Standardfehler des Mittelwertes
V1	0	150	5,980	2,3473	,1917
	1	528	7,263	2,3093	,1005

Wahrgenommene Organisationsqualität (V1 [=„studorga“ auf nächster Folie]) und Verfügbarkeit von Informationsmaterialien (V2) (N=678)

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									5% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
STUDORG	Varianzen sind gleich	,002	,961	-5,984	676	,000	-1,28326	,21444	-1,70431	-,86220
	Varianzen sind nicht gleich			-5,930	237,139	,000	-1,28326	,21641	-1,70958	-,85694

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit	
		F	Signifikanz
STUDORGA	Varianzen sind gleich	,002	,961

STUDORGA = V1 (in vorheriger Folie)

Test bei unabhängigen Stichproben											
				T-Test für die Mittelwertgleichheit*							
				T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
				-5,984	676	,000	-1,283	,2144	-1,7043	-,8622

*unter der Annahme gleicher Varianzen

Test bei unabhängigen Stichproben											
				T-Test für die Mittelwertgleichheit							
				T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
				-5,930	237,139	,000	-1,283	,2164	-1,7096	-,8569

*unter der Annahme *un*gleicher Varianzen

Korrektur der Freiheitsgrade bei **heterogenen** Varianzen

$$df_{korr} = \frac{1}{\dfrac{c^2}{n_1 - 1} + \dfrac{(1-c)^2}{n_2 - 1}}$$

Im Beispiel:

$$c = \frac{\hat{\sigma}^2_{\bar{x}_1}}{\hat{\sigma}^2_{\bar{x}_1} + \hat{\sigma}^2_{\bar{x}_2}} = \frac{0{,}1917^2}{0{,}1917^2 + 0{,}1005^2} = 0{,}7844$$

$$df_{korr} = \frac{1}{\dfrac{0{,}6153}{149} + \dfrac{0{,}0465}{527}} = 237{,}09$$

Vergleich zweier Anteilswerte aus abhängigen Stichproben

		t_1: 1	t_1: 2	
t_0	1	p_{11}	p_{12}	p_{1+}
	2	p_{21}	p_{22}	p_{2+}
		p_{+1}	p_{+2}	1.0

$$d = p_{+1} - p_{1+}$$

$$H_0 : \pi_{+1} = \pi_{1+}$$

(Marginale Homogenität)

$$z = d / s.e_{(d)}$$

$$H_0 : \pi_{12} = \pi_{21}$$

(Symmetrie)

Abhängige Samples $(n_0 = n_1 = n)$

$$s.e_{(d)} = \sqrt{\frac{p_{+1}(1-p_{+1}) + p_{1+}(1-p_{1+}) - 2(p_{11}p_{22} - p_{12}p_{21})}{n}}$$

In unabhängigen Samples $$s.e_{(d)} = \sqrt{\frac{p_{+1}(1-p_{+1})}{n_1} + \frac{p_{1+}(1-p_{1+})}{n_0}}$$

Cf. Agresti (1990) Categorical Data Analysis. New York, pp. 348-350

X2 \ X1	+	-	Σ
+	n_a	$\mathbf{n_b}$	n_{a+b}
-	$\mathbf{n_c}$	n_d	n_{c+d}
Σ	n_{a+c}	n_{b+d}	n

H_0: π(Feld c) = π(Feld b)

H_1: π(Feld c) ≠ π(Feld b)

$$Z = \frac{p(Feld\_c) - 0{,}5}{\sqrt{\frac{0{,}5 \cdot (1-0{,}5)}{n_b + n_c}}}$$

$$Z = \frac{p(Feld\_b) - 0{,}5}{\sqrt{\frac{0{,}5 \cdot (1-0{,}5)}{n_b + n_c}}}$$

Vergleich zweier Mittelwerte aus abhängigen Stichproben

Prüfung einer Differenz von Messwertpaaren über die Prüfung einer Hypothese über den Populations**mittelwert** der **Differenz**variablen **D**

$$D = X_1 - X_2$$

Test,
wie bei einer Hypothese über **einen** Populationsmittelwert

Chi²-verteilte Zufallsvariable

resultiert, wenn eine standardnormalverteilte Zufallsvariable quadriert wird oder wenn mehrere, voneinander unabhängige standardnormalverteilte Zufallsvariablen quadriert und aufsummiert werden

$$\chi^2_{df=\nu} = \chi^2_{\nu} = Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_{\nu}^2$$

Chi²-Verteilung mit *v* Freiheitsgraden

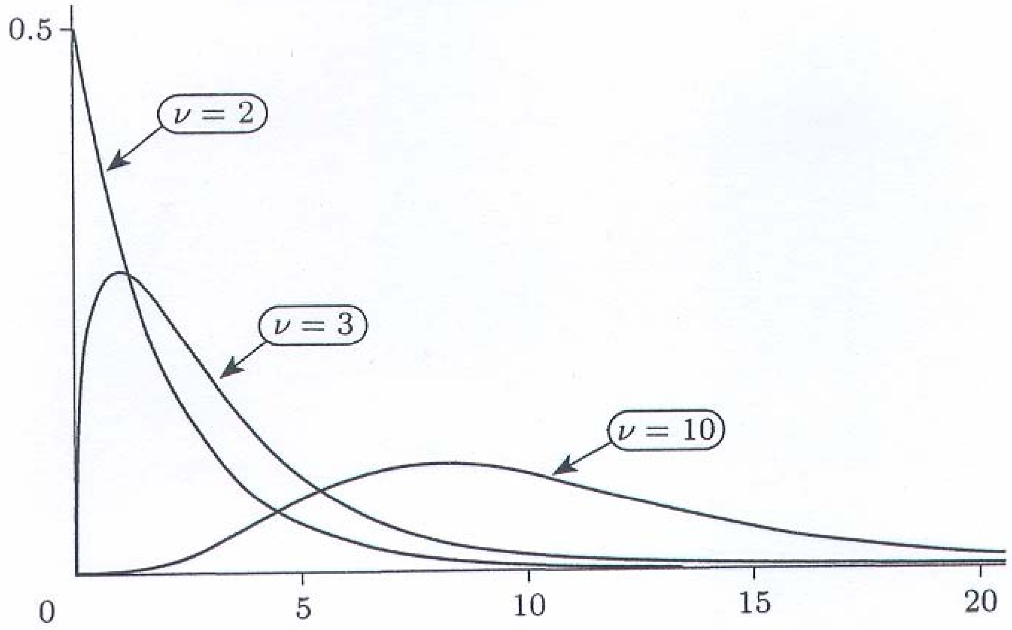
df = Anzahl Freiheitsgrade (Anzahl der Summanden)

Chi²-Verteilung

Kennwerteverteilung von Variationen und Varianzen (df=n-1)

Siehe Basislektüre (Kühnel/Krebs), Kap. 7.4

Chi²-Verteilung bei 2, 3 und 10 Freiheitsgraden

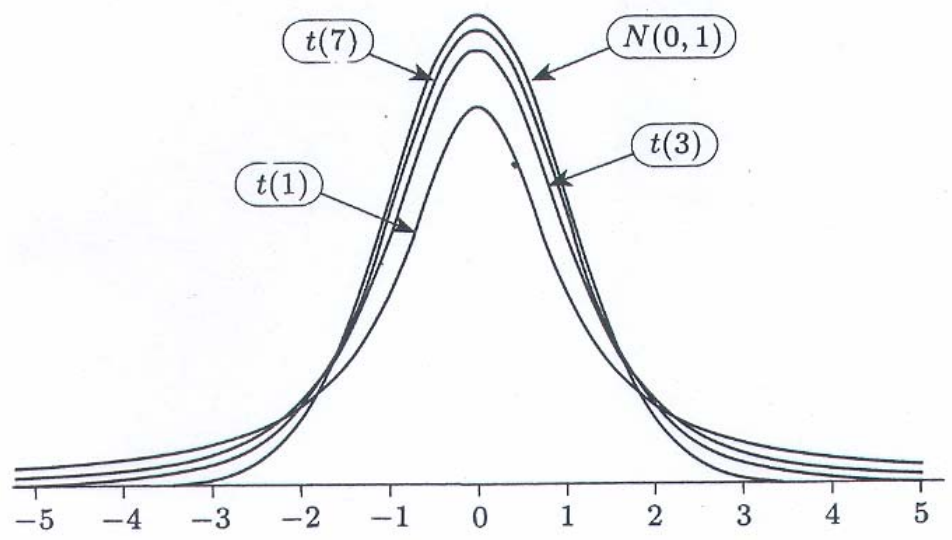


Quelle: Daly, F. et al. (1999) Elements of Statistics. Harlow: Addison-Wesley, p. 288

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{df}}}$$

T = t-verteilte Zufallsvariable T mit df Freiheitsgraden
Z = standardnormalverteilte Zufallsvariable
χ^2= von Z unabhängige, mit df Freiheitsgraden
chi^2-verteilte Zufallsvariable
df = Zahl der Freiheitsgrade

t-Verteilung bei 1, 3 und 7 Freiheitsgraden
Standardnormalverteilung *N*(0,1)



Quelle: Daly, F. et al. (1999) Elements of Statistics. Harlow: Addison-Wesley, p. 283

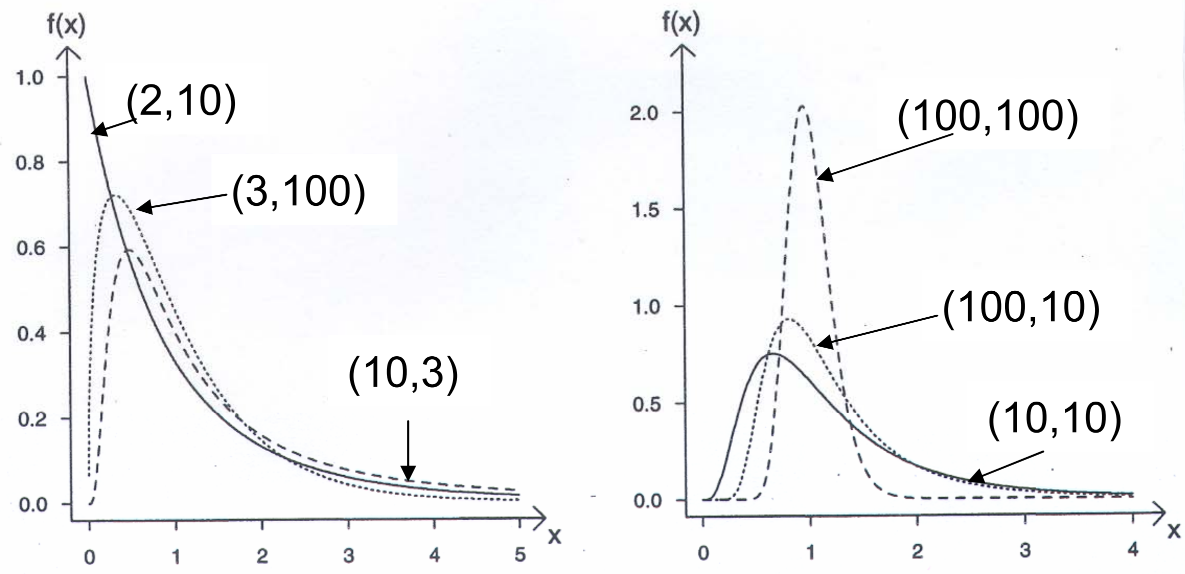
Ausgewählte Quantile der t-Verteilung

df	…	95,0%	97,5%	…
.				
.				
5	…	2,015	2,571	…
.				
10	…	1,812	2,228	…
.				
15	…	1,753	2,131	…
.				
30	…	1,697	2,042	…
.				
120	…	1,658	1,980	…
∞	…	1,645	1,960	…

F-Verteilung

Kennwerteverteilung von Varianzverhältnissen

$$F = \frac{\dfrac{\chi_1^2}{df_1}}{\dfrac{\chi_2^2}{df_2}}$$



F-Verteilung bei (m,n) Freiheitsgraden

Quelle: Fahrmeir et al. (1999) Statistik. Berlin: Springer, S. 302

	Anteil		Mittelwert		Korrelation und Regression			Varianz	
Stich- probe	Anteils- wert	Anteils- wertdif- ferenz	Mittel- wert	Mittel- wertsdif- ferenz	Phi, Cr. V	r	b	Varianz	Varianz- verhält- nisse
groß	NV	NV	NV T	NV T	Chi^2	NV	T	NV Chi^2	F
klein	HG, BI								
Abhängig		NV							

Test bei asymptotischer Verteilung
Test, wenn „Finite-Sampling-Verteilung“, inclusive
„Exakter Tests“ (ohne Bezug auf NV oder daraus abgeleiteter Verteilung, Bezug: Kombinatorik)