Vergleich zweier Anteilswerte aus abhängigen Stichproben

$$
\begin{array}{cc|cc|c}
 & & \multicolumn{2}{c}{t_1} & \\
 & & 1 & 2 & \\
\hline
t_0 & 1 & p_{11} & p_{12} & p_{1+} \\
 & 2 & p_{21} & p_{22} & p_{2+} \\
\hline
 & & p_{+1} & p_{+2} & 1.0
\end{array}
$$

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

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

(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 Anteilswerte aus abhängigen Stichproben

$$\begin{array}{cc|cc|c} & & \multicolumn{2}{c}{t_1} & \\ & & 1 & 2 & \\ \hline t_0 & 1 & p_{11} & p_{12} & p_{1+} \\ & 2 & p_{21} & p_{22} & p_{2+} \\ \hline & & p_{+1} & p_{+2} & 1.0 \end{array}$$

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

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

(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) $\neq$ 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

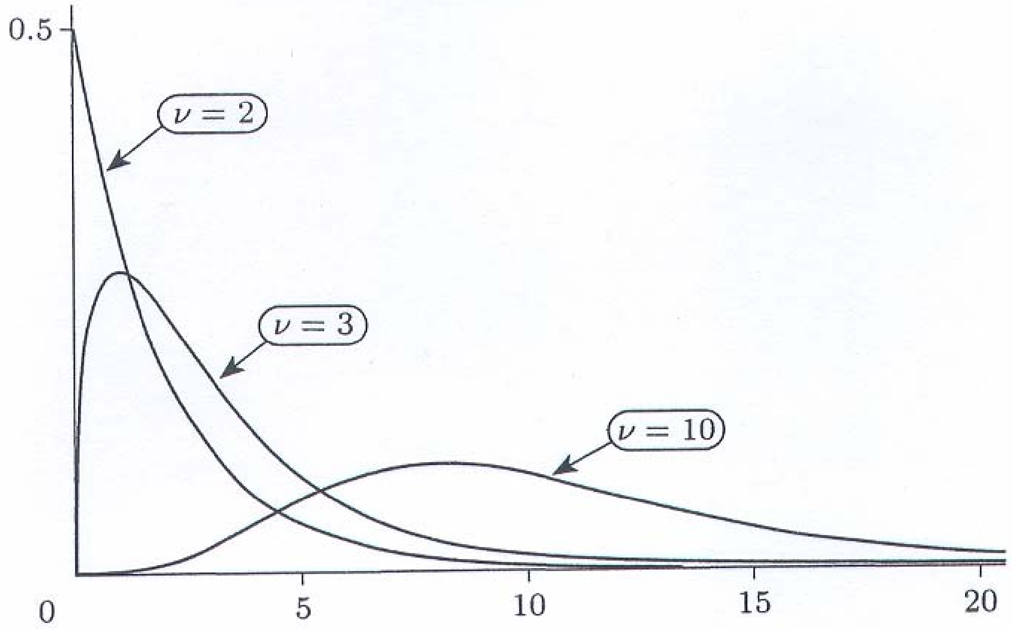
df = Anzahl Freiheitsgrade (Anzahl der Summanden)

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

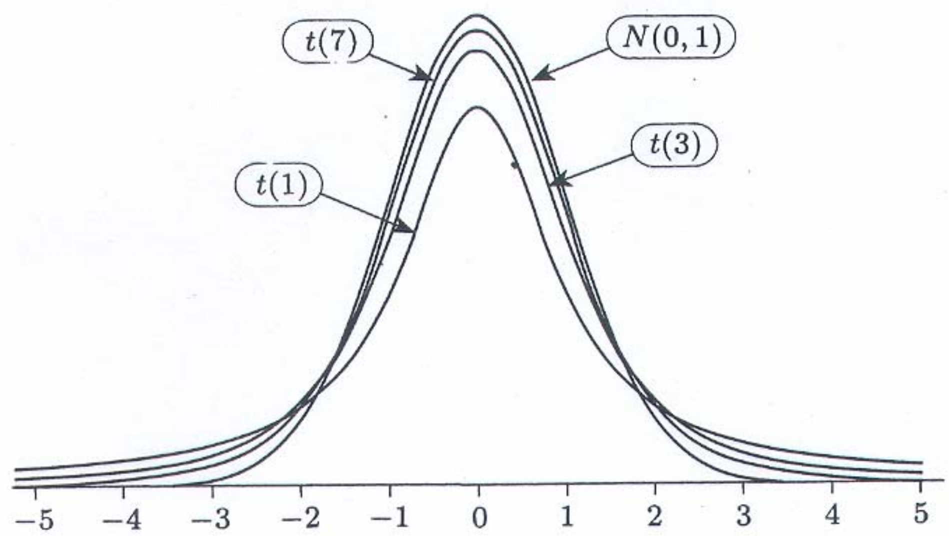


Quelle: Daly, F. et al. (1999) Elements of Statistics. Harlow: Addison-Wesley, p. 288

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

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

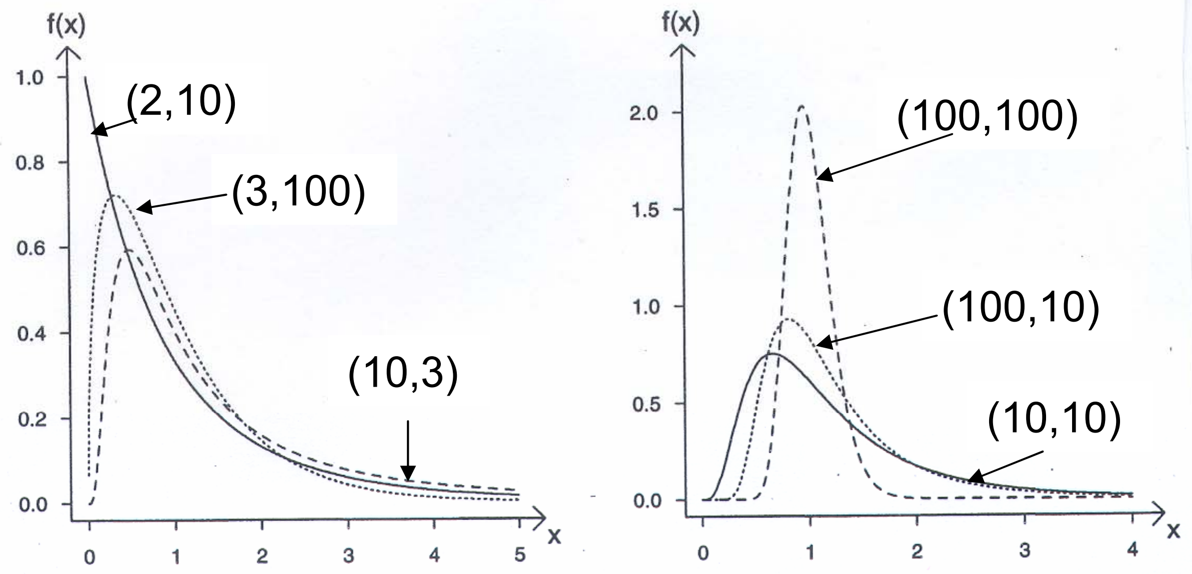
Ausgewählte Quantile der t-Verteilung

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

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



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 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)

	B1	B2	
A1	*a*	*b*	a+b(N_1)
A2	*c*	*d*	c+d
	a+c	b+d	N

Exakte Analyse: Fisher-Yates Test

Randhäufigkeiten als Parameter fixiert.

Gesucht: Wahrscheinlichkeit, dass sich von N_1 Beobachtungen *a* unter B1 und *b* unter B2 befinden, wenn insgesamt *a+c* Beobachtungen unter B1 und *b+d* Beobachtungen unter B2 gemacht wurden.

Quelle: Bortz/Lienert/Boehnke (2008) Verteilungsfreie Methoden in der Biostatistik. Berlin: Springer, S. 110-112

Wahrscheinlichkeitsverteilung der Zufallsvariablen *a* bei wiederholter Entnahme von N1 aus N Beobachtungen „ohne Zurücklegen“: hypergeometrische Verteilung:

$$p(a) = \frac{\binom{a+c}{a} \times \binom{b+d}{b}}{\binom{N}{a+b}}$$

$$= \frac{(a+b)! \times (c+d)! \times (a+c)! \times (b+d)!}{N! \times a! \times b! \times c! \times d!}$$

Kombinationsregel
Wählt man aus *n* verschiedenen Objekten *r* zufällig aus und lässt hierbei die Reihenfolge außer acht, ergeben sich

für die *r* Objekte $\binom{n}{r} = \frac{n!}{r! \times (n-r)!}$. verschiedene Möglichkeiten.