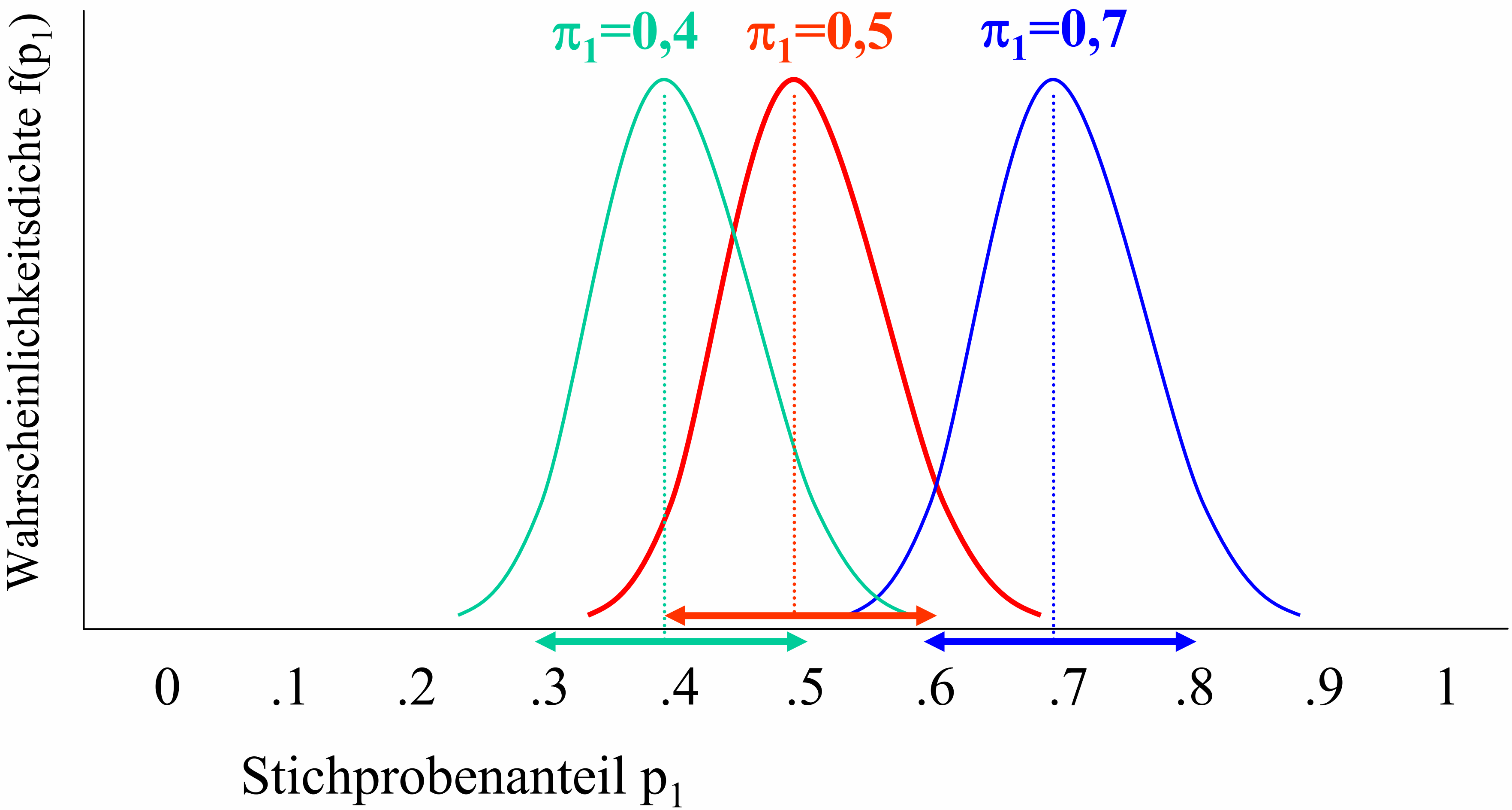
Bei verschiedenen Populationsanteilen π_1 zu erwartende Stichprobenanteile p_1



» **Null**hypothese $\mathbf{H_0}$

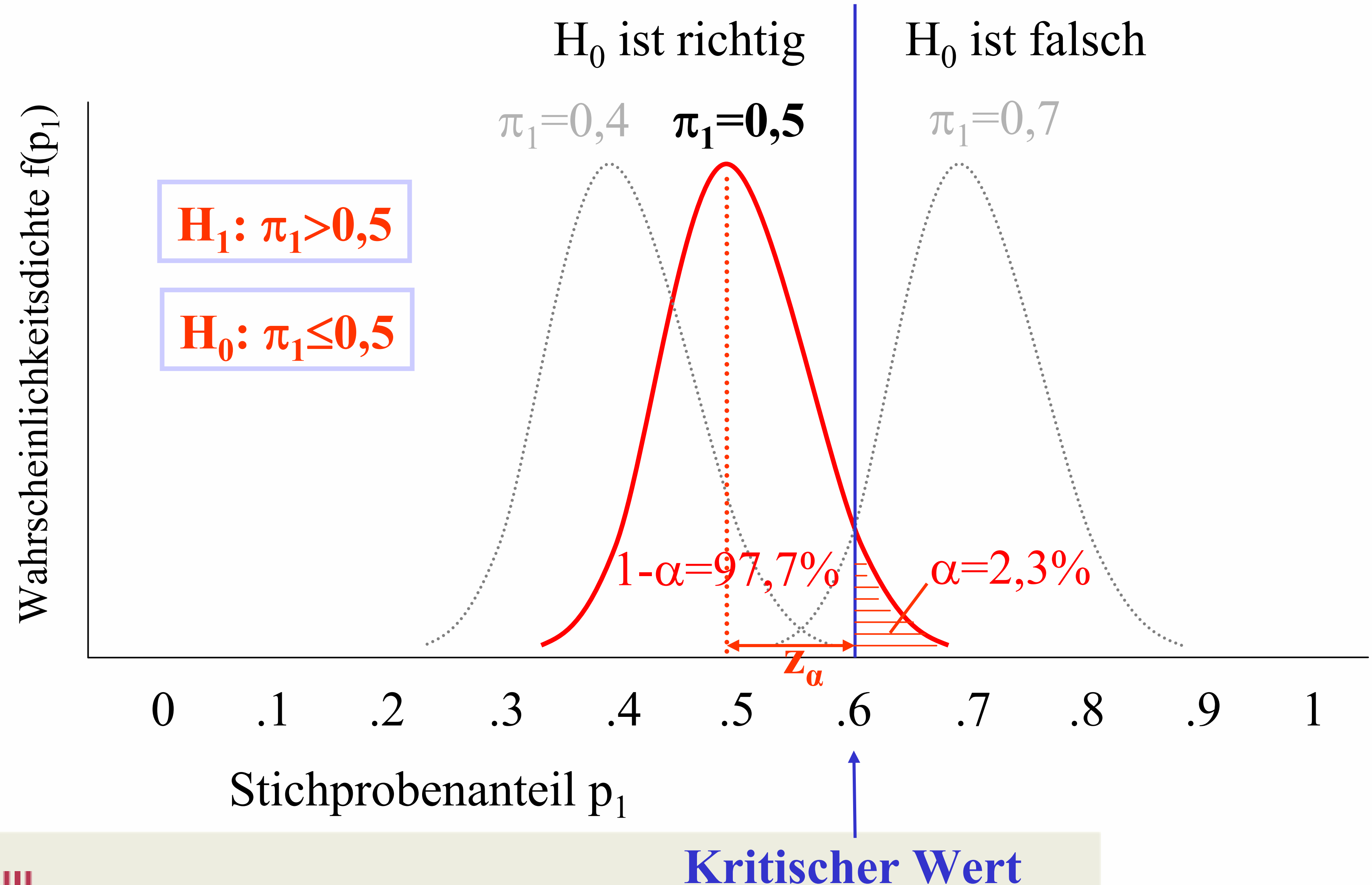
» **Forschungs**- bzw. Alternativhypothese $\mathbf{H_1}$

» Bestimmung der Wahrscheinlichkeit des Stichprobenbefundes unter der Annahme der Gültigkeit von $\mathbf{H_0}$

» Festlegung einer Irrtumswahrscheinlichkeit $\boldsymbol{\alpha}$

α- und β-Fehler bei statistischen Entscheidungen

	In der Grundgesamtheit gilt die ...	
Entscheidung aufgrund der Stichprobe zugunsten der ...	Nullhypothese H_0	Alternativhypothese H_1
H_0	Richtige Entscheidung	**β-Fehler**
H_1	**α-Fehler**	Richtige Entscheidung



Beispiel nach Kühnel/Krebs (2001) Statistik für die Sozialwissenschaften. Reinbek bei Hamburg, S. 256ff

Schritte beim Test statistischer Hypothesen

» Formulierung von Null- und Alternativhypothese

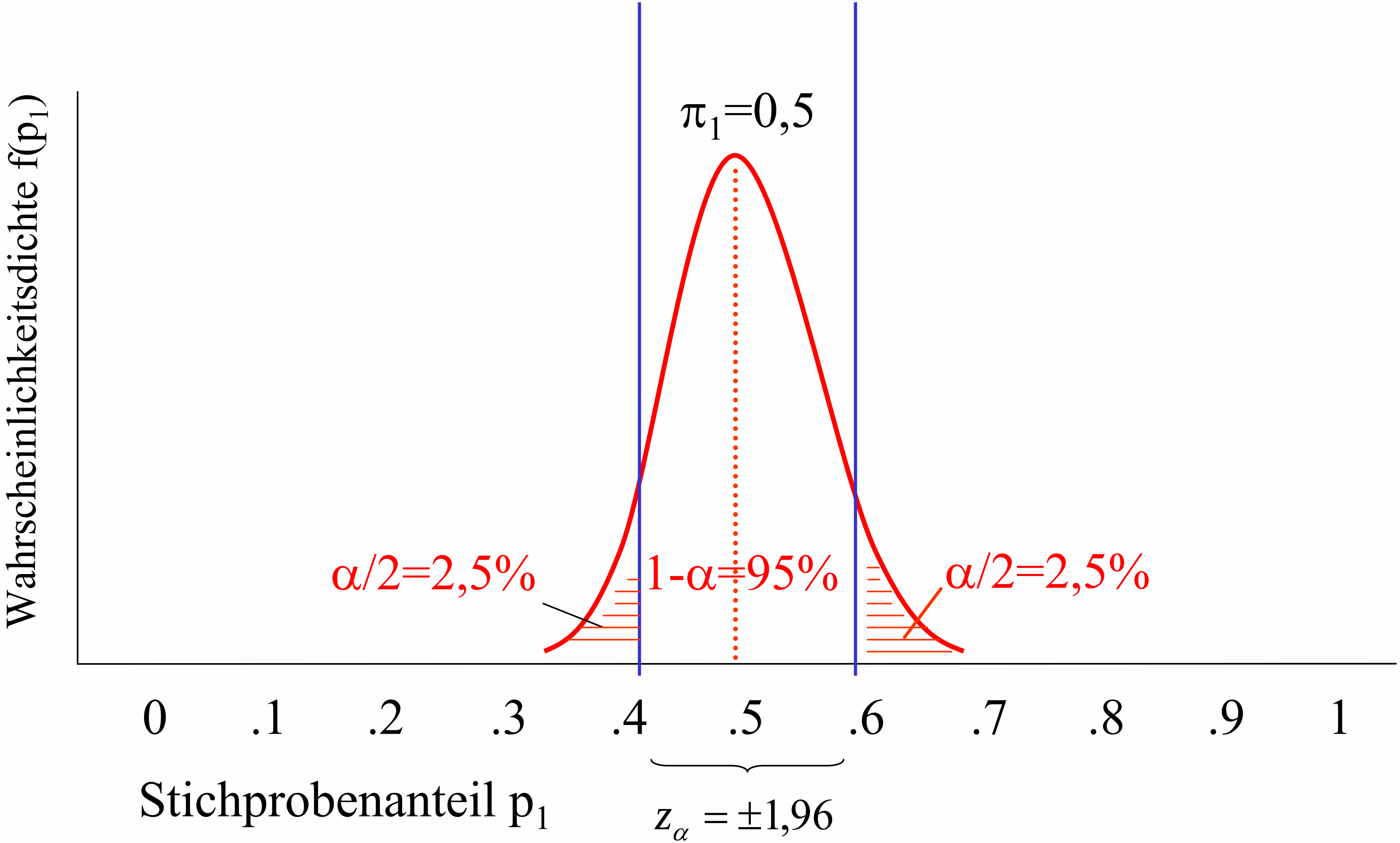
» Festlegung der Irrtumswahrscheinlichkeit α

» Teststatistik berechnen

$$z_\alpha = \frac{p_1 - \pi_{1|H_0}}{\sqrt{\frac{\pi_{1|H_0} \cdot \left(1 - \pi_{1|H_0}\right)}{n}}} = \frac{0{,}6 - 0{,}5}{\sqrt{\frac{0{,}5 \cdot 0{,}5}{100}}} = \frac{0{,}1}{0{,}05} = 2$$

Im Beispiel aus Kühnel/Krebs mit n=100

» Entscheidung treffen



Beispiel nach Kühnel/Krebs (2001) Statistik für die Sozialwissenschaften. Reinbek bei Hamburg, S. 256ff

Quantile der Standardnormalverteilung (Struktur der NV-Tabelle)

Quantilwahrscheinlichkeiten

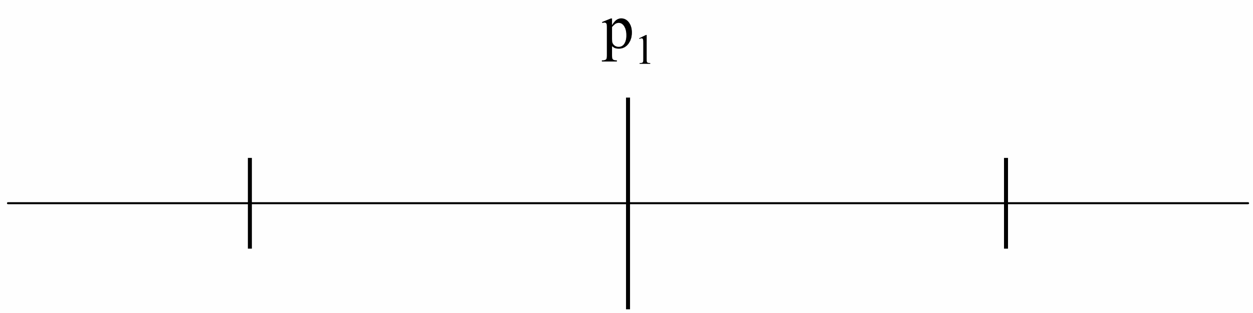
Quantilwerte

α	z_α		α	z_α
0,000	$-\infty$		0,950	1,645
0,005	-2,576		…	…
…	…		0,975	1,960
0,025	-1,960		…	…
…	…		0,995	2,576
0,050	-1,645		1,00	∞
…	…			
0,500	0,000			
…	…			

Bestimmung des 95%-Konfidenzintervalls:

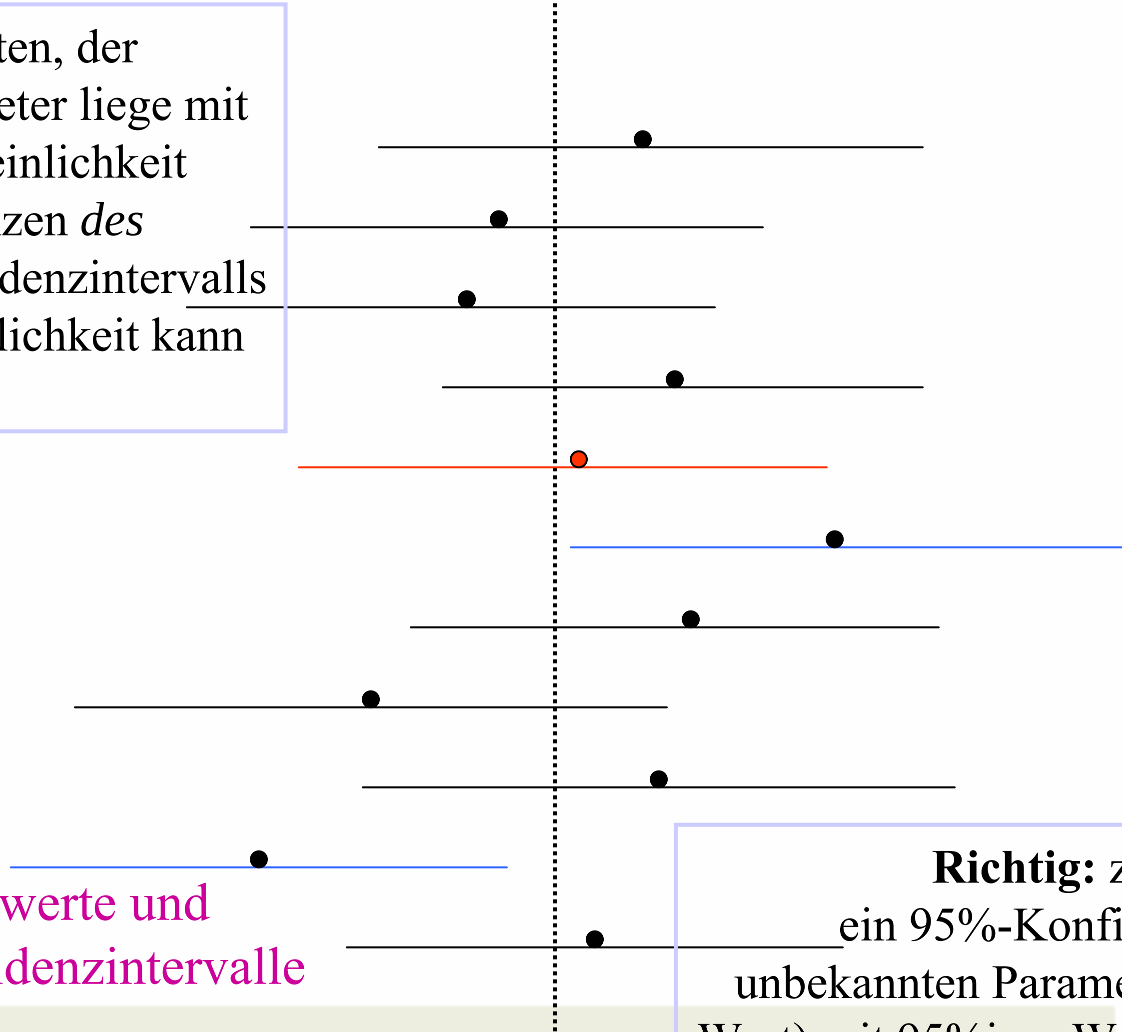
$$c.i(\pi_1) = p_1 \pm z_{1-\alpha/2} \cdot \hat{\sigma}(p_1)$$

$$c.i(\pi_1) = p_1 \pm z_{1-\alpha/2} \cdot \sqrt{\frac{p_1 \cdot (1 - p_1)}{n}}$$



Parameter in der Grundgesamtheit

Falsch: zu behaupten, der unbekannte Parameter liege mit 95%iger Wahrscheinlichkeit innerhalb der Grenzen *des berechneten* Konfidenzintervalls (diese Wahrscheinlichkeit kann nur 1 oder 0 sein)



Stichprobenkennwerte und deren 95% Konfidenzintervalle

Richtig: zu behaupten, daß ein 95%-Konfidenzintervall den unbekannten Parameter (Populations-Wert) mit 95%iger Wahrscheinlichkeit enthält

Vgl. z.B. Kühnel/Krebs (2001) Statistik für die Sozialwissenschaften. Reinbek bei Hamburg, S. 242 - 244

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 $0{,}25 \leq \pi_1 \leq 0{,}75$);

Wird π_1 durch p_1 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: $\pi_1 - \pi_2 = 0$

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

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