Grundgesamtheit (Population)

Verteilung des Merkmals in der Grundgesamtheit (mit ihren Parametern [**Populations**kennwerten])

Stichprobenkennwerteverteilung
im Zufallsexperiment
wiederholter Stichprobenziehung
(Sampling distribution; Samplingverteilung)

Stichprobe (Sample)

Verteilung des Merkmals in der Stichprobe
(Berechnung von **Stichproben**kennwerten)

Sampling als Zufallsexperiment

Beispiel: Populationsumfang N=130
Anteilswert $\pi = 0{,}7$
„Erfolg“ = Ziehung einer weißen Kugel

Experiment 1
Versuch 1 und 2

Sample-Umfang n=4

Anzahl der gezogenen Samples, jeweils **k=100**

Versuche 1 und 2 zusammengelegt zu **k=200**

Experiment 2
Versuch 1 und 2

Sample-Umfang n=16

Anzahl der gezogenen Samples, jeweils **k=100**

Experiment 1

Erfolge		Experimental				Theoretisch
		1. Versuch		2. Versuch		
Anzahl	Anteil	Häufigkeit		Häufigkeit		Häufigkeit
X	p	Abs.	Rel.	Abs.	Rel.	Rel.
0	**0,00**	1	0,01	1	0,01	**0,01**
1	**0,25**	4	0,04	12	0,12	**0,07**
2	**0,50**	26	0,26	24	0,24	**0.26**
3	**0,75**	42	0,42	47	0,47	**0,41**
4	**1,00**	27	0,27	16	0,16	**0,24**
	Σ	100	1,00	100	1,00	1,00
k=100	mean	**0,7250**		**0,6625**		**0,7**
k=100	s.d.	**0,2196**		**0,2300**		**0,2291**

k=200	mean	**0,6938**
k=200	s.d.	**0,2272**

n=4

„bias“ (Δ)

k=100	mean	0,025	-0,0375	0,7
k=100	s.d.	-0,0095	0,0009	0,2291
k=200	mean	-0,0062		
k=200	s.d.	-0,0019		
n=16	mean	0,0		0,7
n=16	s.d.	-0,0086		0,1146

Experiment 2

Erfolge		Experimental			Theoretisch
		1. Versuch		2. Versuch*	
Anzahl	Anteil	Häufigkeit		Häufigkeit	Häufigkeit
X	p	Abs.	Rel.	Rel.	Rel.
0	**0,0000**	0	0,00	0,00	**0,00**
1	**0,0625**	0	0,00	0,00	**0,00**
2	**0,1250**	0	0,00	0,00	**0,00**
3	**0,1875**	0	0,00	0,00	**0,00**
4	**0,2500**	0	0,00	0,00	**0,00**
5	**0,3125**	0	0,00	0,00	**0,00**
6	**0,3750**	0	0,00	0,00	**0,01**
7	**0,4375**	1	0,01	0,04	**0,02**
8	**0,5000**	4	0,04	0,04	**0,05**

Teil 1 der Tabelle

*Modifizierter Samplingprozeß

n=16

Anzahl	Anteil	Häufigkeit		Häufigkeit		Häufigkeit
X	**p**	Abs.	Rel.		Rel.	Rel.
9	**0,5625**	10	0,10		0,10	**0,10**
10	**0,6250**	19	0,19		0,20	**0,17**
11	**0,6875**	26	0,26		0,21	**0,21**
12	**0,7500**	18	0,18		0,22	**0,20**
13	**0,8125**	13	0,13		0,10	**0,15**
14	**0,8750**	5	0,05		0,07	**0,07**
15	**0,9375**	4	0,04		0,02	**0,02**
16	**1,0000**	0	0,00		0,00	**0,00**
	Σ	100	1,00	100	1,00	1,00
	Mean	0,7000		0,6894		**0,7**
	s.d.	0,1060		0,1106		**0,1146**

Eigenschaften von Schätzern

- » Erwartungstreue
- » Effizienz
- » Konsistenz

Faktoren, die sich auf den Standardfehler auswirken

[1] Populationsvarianz

[2] Stichprobenumfang

[3] Auswahlsatz: Verhältnis n/N

[4] Designeffekt, bei komplexen Auswahlverfahren

[1] Varianz des Merkmals in der Grundgesamtheit
[2] Stichprobenumfang *n*

$$s.e = \sqrt{\frac{Populationsvarianz}{Stichprobenumfang}}$$

Beispiele:

$$\sigma_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}}$$

s.e des arith. Mittels

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi \times (1 - \pi)}{n}}$$

s.e des Anteilswertes

Wenn Erhöhung des Stichprobenumfangs um das ℓ-fache,

$$n \to \lambda \times n$$

dann Reduktion des Standardfehlers um das

$$\sqrt{\lambda - fache}$$

auf

$$s.e \to 1/\sqrt{\lambda} \times s.e$$

[3] durch den Auswahlsatz ***f***
(Verhältnis Stichproben-***n*** zu Populations-***N***)

Korrekturfaktor für finite Populationen

$$k = 1 - \frac{n}{N} = 1 - f.$$

N = Zahl der Einheiten der Auswahlgrundlage
(„Populationsumfang“)
n = Zahl der Einheiten in der Stichprobe
(„Stichprobenumfang“)
n/N = *f* = Auswahlsatz

Beispiel: Standardfehler für den Anteilswert π

Schätzung von π durch p

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{p \times (1-p)}{n}}$$

Ergänzt um den Korrekturfaktor für finite Populationen

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{p \times (1-p)}{n}} \times \sqrt{1-\frac{n}{N}}$$

Eine Vernachlässigung der Korrektur für finite Populationen bewirkt in meisten Fällen nur eine marginale bis vernachlässigbare Überschätzung der Samplingvarianz.
Der Effekt auf den Standardfehler s.e beträgt annähernd $1 - f/2$:

$f = n/N$	0,10	0,04	0,01	0,001	0,0001
$\sqrt{1-f}$	0,9487	0,9798	0,9950	0,9995	0,99995
$1 - f/2$	0,9500	0,9800	0,9950	0,9995	0,99995

Cf. Leslie Kish (1965) Survey Sampling. New York, p.44

Designeffekt

$$deff = \frac{faktischeSamplingVarianz}{srsSamplingVarianz}$$

Die Annahme eines Designeffektes von 2 führt
zu einem um den Faktor 1,4 erhöhten Standardfehler

deff Korrekturfaktor: $\sqrt{2} = 1{,}414$

Im Falle des Standardfehlers eines Anteilswertes:

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{p \times (1-p)}{n}} \times \sqrt{1-\frac{n}{N}} \times \sqrt{2}$$