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**

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

Binomialverteilung

X	Anzahl Erfolge (weiße Kugeln)
π	Anteil von Erfolgen (weißen Kugeln) in Population
ρ	Anteil von Misserfolgen in Population
n	Stichprobenumfang

$$\rho = 1 - \pi$$

Binomial-expansion

$$(\rho + \pi)^n$$

	Wahrscheinlichkeit ς (X)			
X	n=1	n=2	n=3	n=4
0	ρ	ρ^2	ρ^3	ρ^4
1	π	$2\rho\pi$	$3\rho^2\pi$	$4\rho^3\pi$
2		π^2	$3\rho\pi^2$	$6\rho^2\pi^2$
3			π^3	$4\rho\pi^3$
4				π^4
Σ	$(\rho+\pi)^1$	$(\rho+\pi)^2$	$(\rho+\pi)^3$	$(\rho+\pi)^4$

Beispiel

- Sampling-Verteilung vom **Umfang n = 3**

- 4 mögliche Ergebnisse: 0, 1, 2, 3 weiße Kugeln

- Zu bestimmen: Wahrscheinlichkeit jedes dieser möglichen Ereignisse

---

X=0 $\rho\rho\rho = \rho^3$ Prob(3 schwarze Kugeln zu ziehen)

X=1 $\pi\rho\rho + \rho\pi\rho + \rho\rho\pi$

$= \rho^2\pi + \rho^2\pi + \rho^2\pi$

$= 3\rho^2\pi$

Prob(1 weiße und 2 schwarze Kugeln in einer beliebigen Sequenz zu ziehen); 3 Sequenzen

X=2

$$\pi\pi\rho + \pi\rho\pi + \rho\pi\pi$$
$$= \rho\pi^2 + \rho\pi^2 + \rho\pi^2$$
$$= 3\rho\pi^2$$

Prob(2 weiße und 1 schwarze Kugel in einer beliebigen Sequenz zu ziehen); 3 Sequenzen

X=3

$$\pi\pi\pi = \pi^3$$

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

$\boxed{\pi = 0{,}7}$ $\boxed{\rho = 1 - \pi = 0{,}3}$

Umfang n = 4 Mögliche Ergebnisse X = 0, 1, 2, 3

X = 0 ρ^4 $0{,}3^4 = 0{,}008$

X = 1 $4\rho^3\pi$ $4 \cdot 0{,}3^3 \cdot 0{,}7 = 0{,}076$

Sequenz	1	**s**	f	f	f
	2	f	**s**	f	f
	3	f	f	**s**	f
	4	f	f	f	**s**

$X = 2$ $\qquad 6\rho^2\pi^2 \qquad 6 \cdot 0{,}3^2 \cdot 0{,}7^2 = 0{,}265$

Sequenz 1	s	s	f	f
2	s	f	s	f
3	f	s	f	s
4	f	f	s	s
5	f	s	s	f
6	s	f	f	s

$X = 3$ $\qquad 4\rho\pi^3 \qquad 4 \cdot 0{,}3 \cdot 0{,}7^3 = 0{,}412$

Sequenz 1	s	s	s	f
2	s	s	f	s
3	s	f	s	s
4	f	s	s	s

$\mathrm{X} = 4$ π^4 $0{,}7^4 = 0{,}24$

Anzahl der Anordnungen (Sequenzen)

$$\binom{n}{n_s} = \frac{n!}{(n - n_s)! \cdot n_s!}$$

n Stichprobenumfang

n_s Anzahl Erfolge

$$P(X = n_s) = \binom{n}{n_s} \cdot \rho^{n-n_s} \cdot \pi^{n_s}$$

Permutation

$N! = N \cdot (N-1) \cdot (N-2) \cdot (N - n) \cdot \ \ldots \ \cdot 3 \cdot 2 \ \cdot 1$

N! = Anzahl der Anordnungen der N Elemente mit Beachtung der Reihenfolge

Variationen

$$_N V_n = \frac{N!}{(N-n)!}$$

Anzahl der Möglichkeiten, n Elemente aus N Elementen
ohne Zurücklegen **mit Berücksichtigung der Reihenfolge** auszuwählen

Kombinationen

$$_N K_n = \frac{N!}{(N-n)! \cdot n!} = \frac{N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot (N-n+1)}{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1} = \binom{N}{n}$$

Anzahl der Möglichkeiten, n Elemente aus N Elementen ohne Zurücklegen **ohne Berücksichtigung der Reihenfolge** der Elemente auszuwählen

Hypergeometrische Verteilung

Wahrscheinlichkeit, dass eine Stichprobe
n_1 Elemente mit den Ausprägungen 1 und
n_0 Elemente mit den Ausprägungen 0 enthält

Zu ermitteln:

Wie oft können n_1 Elemente aus der Teilpopulation N_1,
wie oft können n_0 Elemente aus der Teilpopulation N_0 und
wie oft können n Elemente aus der Population N gezogen werden

Die Anzahl der jeweiligen Kombinationen ohne Zurücklegen wird ermittelt nach dem Binomialkoeffizienten:

$$_{M}K_{m} = \frac{M!}{(M-m)! \cdot m!} = \binom{M}{m}$$

(Anzahl der Möglichkeiten, m aus M Elementen ohne Zurücklegen auszuwählen)

$$_{N}K_{n} = \frac{N!}{(N-n)! \cdot n!} = \binom{N}{n}$$

.. n aus N

$$_{N_1}K_{n_1} = \frac{N_1!}{(N_1-n_1)! \cdot n_1!} = \binom{N_1}{n_1}$$

.. n_1 aus N_1

$$_{N_0}K_{n_0} = \frac{N_0!}{(N_0-n_0)! \cdot n_0!} = \binom{N_0}{n_0}$$

.. n_0 aus N_0

Wahrscheinlichkeit der Stichprobenzusammensetzung von n_1 und n_0 Elementen:

$$P(X = n_1) = \frac{\binom{N_1}{n_1} \cdot \binom{N_0}{n_0}}{\binom{N}{n}}$$

<u>Für (z.B.:)</u>
N=130
$\pi = 0{,}7$
$N_1 = 91$
$(= 0{,}7 \cdot 130)$
$N_0 = 39$

$n_1 = 11$
$n_0 = 5$

$$P(X = n_1) = \frac{\binom{91}{11} \cdot \binom{39}{5}}{\binom{130}{16}}$$

$$= \frac{47325339895743 \cdot 575757}{121515484051014339000} = \underline{\underline{0{,}2242}}$$

$$\binom{16}{11} \cdot 0{,}3^5 \cdot 0{,}7^{11} = \underline{\underline{0{,}2098}} \quad \text{(nach Binomialverteilung)}$$

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**

Fortsetzung

n=16

Erwartungswert der Zufallsvariablen X:

$$\mu_X = n \cdot \frac{N_1}{N} = n \cdot \pi_1$$

(identisch für hypergeometrische Verteilung und Binomialverteilung)

Varianz der Zufallsvariablen X:

$$\sigma_X^2 = n \cdot \pi_1 \cdot \pi_0 \cdot \frac{N-n}{N-1} = n \cdot \pi_1 \cdot (1-\pi_1) \cdot \frac{N-n}{N-1}$$

(hypergeometrische Verteilung)

$$\sigma_X^2 = n \cdot \pi_1 \cdot \pi_0 = n \cdot \pi_1 \cdot (1-\pi_1)$$

(Binomialverteilung)

Varianz der hypergeometrischen Verteilung ist um den Quotienten

$$\frac{N-n}{N-1}$$

geringer als die Varianz der Binomialverteilung

Wenn jedoch N relativ zu n sehr groß ist,
nähert sich dieser Quotient 1 an

Unterschiede zwischen hypergeometrischer und Binomialverteilung
für praktische Anwendungen vernachlässigbar, wenn

$$\frac{N}{n} > 20$$

NV (statt Binomialverteilung) 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