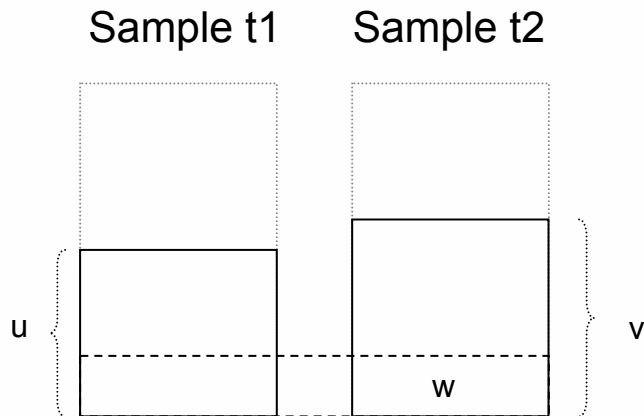


Capture – Recapture Technik

Technik zur Bestimmung der Größe von Populationen



u Anzahl der Zielpersonen in 1. Stichprobe (d.h. Zahl der Personen mit gesuchter Eigenschaft)

v Anzahl der Zielpersonen in 2. Stichprobe

w Anzahl der Zielpersonen aus 2. Stichprobe, die zuvor auch in 1. Stichprobe enthalten waren.

x Schätzung der gesuchten Anzahl aller Zielpersonen

$$\frac{w}{u} = \frac{v}{x} \quad | \cdot x$$

$$x \cdot \frac{w}{u} = v \quad | \cdot u$$

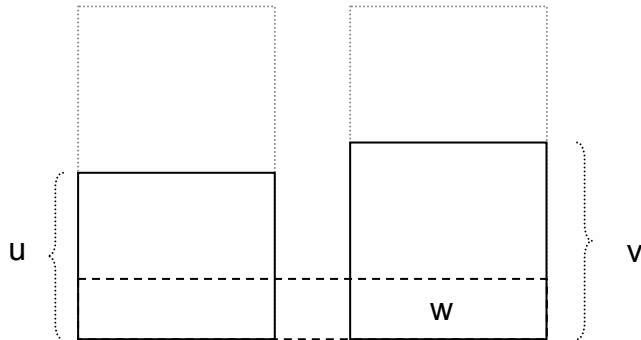
$$x \cdot w = u \cdot v \quad | \div w$$

$$x = \frac{u \cdot v}{w}$$

Annahmen

1. „Re-Capture“ – Wahrscheinlichkeit w/u ist gleich der „Capture“ - Wahrscheinlichkeit insgesamt v/x
2. Voneinander unabhängige Zufallsstichproben
3. Population bleibt unverändert zwischen 1. und 2. Stichprobe

Capture – Recapture Technik (Beispiel)



u Anzahl der Zielpersonen in 1. Stichprobe
(d.h. Zahl der Personen mit gesuchter
Eigenschaft)

v Anzahl der Zielpersonen in 2. Stichprobe

w Anzahl der Zielpersonen aus 2. Stichprobe,
die zuvor auch in 1. Stichprobe enthalten
waren.

x Schätzung der gesuchten Anzahl aller Zielpersonen

$$\frac{w}{u} = \frac{v}{x}$$

$$\frac{10}{30} = \frac{40}{x}$$

$$x = \frac{u \cdot v}{w}$$

$$x = \frac{30 \cdot 40}{10} = 120$$

Gesucht: Anzahl x der „Jogger“ eines Wohngebietes, die ein angrenzendes Naherholungsgebiet nutzen.

1. Stichprobe (an einem 1. Wochenende) förderte u=30 Jogger zutage.
2. Stichprobe (am Wochenende darauf) förderte v=40 Jogger zutage, von denen 10 auch am Wochenende zuvor registriert worden waren.

„Recapture“-Wahrscheinlichkeit $w/u = 10/30 = 0,333 = \frac{1}{3}$
Bei Berechnung implizit angenommene „Capture“-Wahrscheinlichkeit ist nun ebenfalls $\frac{1}{3}$, d.h.: 40 von x stellen ein Drittel der „Population“ der Jogger dieses Gebietes dar:

$$40 = 0,333 \cdot x \quad x = \frac{40}{0,333} = 120$$

Kriterien zur Beurteilung eines Sample Designs

- (1) MSE: Mittlerer quadrierter Fehler
(mean squared error) der Schätzungen**
- (2) Kosten der Realisierung eines Sample Designs
- (3) Kombination von (1) und (2)
- (4) Durchführbarkeit des Sample Designs

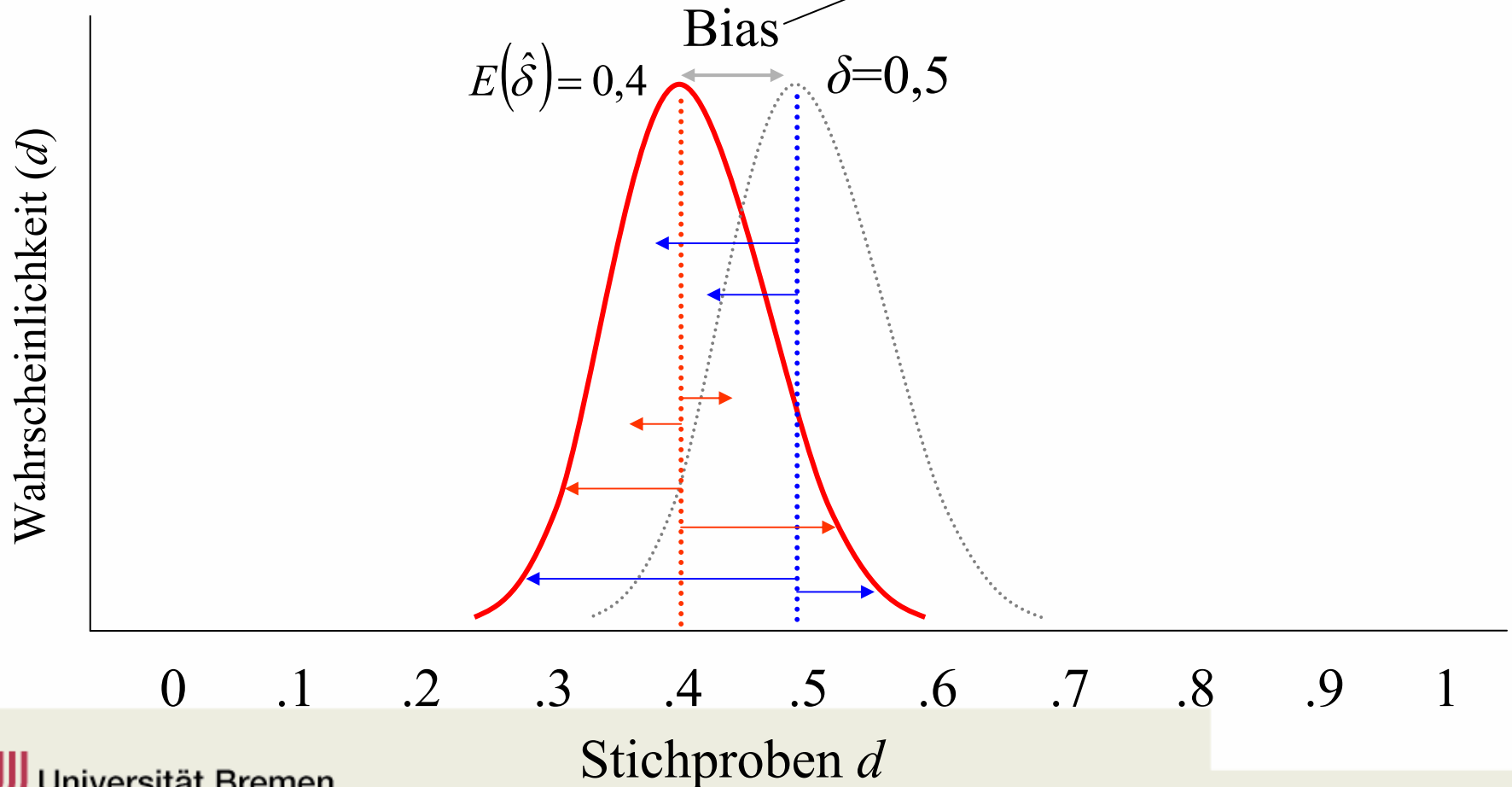
$$MSE(\hat{\delta}) = \sum_{i=1}^C (\hat{\delta}_i - \delta)^2 \pi_i$$

Der MSE ist definiert als Summe der quadrierten Abweichungen der mit der relativen Häufigkeit π_i ihres Auftretens gewichteten $i=1, \dots, C$ Stichprobenschätzungen vom wahren Wert δ des geschätzten Parameters.

Bezug: Samplingverteilung des Schätzers

$$MSE(\hat{\delta}) = Var(\hat{\delta}) + B^2(\hat{\delta}) \quad [1]$$

$$B(\hat{\delta}) = E(\hat{\delta}) - \delta$$



$$\textit{Nonresponse Bias} = p_{NR}(\bar{x}_R - \bar{x}_{NR}) \quad [2]$$

$$\bar{x} = (1 - p_{NR})\bar{x}_R + p_{NR}\bar{x}_{NR} \quad [3]$$

$$\bar{x} = \bar{x}_R - p_{NR}\bar{x}_R + p_{NR}\bar{x}_{NR} \quad [4]$$

$$\bar{x} + p_{NR}\bar{x}_R - p_{NR}\bar{x}_{NR} = \bar{x}_R \quad [5]$$

$$\bar{x} + p_{NR}(\bar{x}_R - \bar{x}_{NR}) = \bar{x}_R \quad [6]$$

$$p_{NR}(\bar{x}_R - \bar{x}_{NR}) = \bar{x}_R - \bar{x} \quad [7]$$

Zur Rolle des Ausschöpfungsgrades ..

bekannte Größen

$$\bar{x} = (1 - p_{NR})\bar{x}_R + p_{NR}\bar{x}_{NR} \quad [3]$$

*Mögl. Minimum des wahren
Samplewertes via ..*

$$\bar{x} = (1 - p_{NR})\bar{x}_R + p_{NR} \times 0 = (1 - p_{NR})\bar{x}_R$$

*Mögl. Maximum des wahren
Samplewertes via ..*

$$\bar{x} = (1 - p_{NR})\bar{x}_R + p_{NR} \times 1 = (1 - p_{NR})\bar{x}_R + p_{NR}$$


.. am Beispiel eines Anteilswertes

Beispiel:

$$\bar{x}_R = 0,6$$

„wahrer“ Samplewert

$\underbrace{\hspace{10em}}$

	$1-p_{NR}$	p_{NR}	min	max	Diff.
Antwortrate 	0,40	0,60	0,24	0,84	0,6
	0,50	0,50	0,30	0,80	0,5
	0,60	0,40	0,36	0,76	0,4
	0,70	0,30	0,42	0,72	0,3
	0,80	0,20	0,48	0,68	0,2
	0,90	0,10	0,54	0,64	0,1

$$\bar{x}_{\min} = (1 - p_{NR}) \bar{x}_R$$

$$\bar{x}_{\max} = (1 - p_{NR}) \bar{x}_R + p_{NR}$$