Kapitel 1

Stichprobenverteilung

D. 1. 1 (*Grundgesamtheit, Stichprobe*)
Eine *Grundgesamtheit* (*Population)* ist die Menge aller Elemente, die Gegenstand einer statistischen Untersuchung sind.
Eine *Stichprobe* ist eine Untermenge der Grundgesamtheit.

BS. 1. 1.

Grundgesamtheit:	Volkzählung
Stichprobe:	Mikrozensus

<u>B. 1. 1.</u> (*Auswahlmethoden der Stichprobenziehung*)

1. **Auswahlmethoden der Stichprobenziehung nach dem Zufallsprinzip**

Auswahl der Stichprobe nach dem Zufallsprinzip		**Beispiel einer Erhebung**
Einfache Zufallsauswahl	Alle Elemente der Grundgesamtheit müssen die gleiche Chance haben, in die Stichprobe aufgenommen zu werden.	Aus der Grundgesamtheit "Hundebesitzer in Berlin" (feststellbar wegen der Hundesteuer) wird eine Stichprobe gezogen.
Geschichtete Auswahl	Die Grundgesamtheit wird entsprechend definierter Kriterien in verschiedene Schichten aufgeteilt. Aus diesen Teilgesamtheiten werden dann nach dem Zufallsprinzip Stichproben entnommen.	Die Grundgesamtheit von Handelsunternehmen wird unterteilt in Warenhäuser, Fachgeschäfte, Billigmärkte etc. Innerhalb jeder dieser Schichten wird per Zufallsprinzip bestimmt, welche Geschäfte genau an der Studie teilnehmen.
Klumpenauswahl	Die Grundgesamtheit wird in Klumpen unterteilt, aus denen eine entsprechende Auswahl nach dem Zufallsprinzip getroffen wird. Alle gezogenen Elemente nehmen dann an der Untersuchung teil.	Die Grundgesamtheit (das Land Berlin) hat z.B. 100 Grundschulen (=Klumpen), aus denen 20 Grundschulen nach dem Zufallsprinzip ausgesucht werden und dann **komplett** an einer Studie als Stichprobe teilnehmen.
Mehrstufige Auswahl	Die Grundgesamtheit wird in Klumpen bzw. Schichten unterteilt, aus denen eine entsprechende Auswahl nach dem Zufallsprinzip getroffen wird. Aus dieser getroffenen Auswahl wird dann wiederum eine begrenzte Anzahl nach dem Zufallsprinzip ausgewählt, die an der Studie teilnimmt.	Es soll untersucht werden, ob in einer Stadt genügend Kindergartenplätze existieren oder ob das Betreuungsangebot ausgebaut werden muss. Die Grundgesamtheit (eine Stadt) wird in 10 Klumpen (einzelne Bezirke) eingeteilt. Aus diesen Bezirken werden dann z.B. fünf Bezirke per Zufall ausgewählt, aus diesen fünf Bezirken werden dann alle Familien mit kleinen Kindern eruiert, aus dieser Gruppe werden dann per Zufall die eigentlichen Stichprobenmitglieder (z.B. 200 Familien mit kleinen Kindern) gezogen.

2. Auswahlmethoden der nicht-zufälligen Stichprobenziehung

Auswahl der nicht zufallsorientierten Stichprobe		**Beispiel einer Erhebung**
Willkürliche Auswahl	Die Stichprobe wird "aufs Geratewohl" bestimmt.	In bestimmten Fast-Food Restaurants werden morgens zwischen 10 und 11 Uhr diejenigen Menschen befragt, die sich "gerade zufällig" im Restaurant aufhalten.
Konzentrationsauswahl	Bei der Zusammensetzung der Stichprobe wird bewusst/absichtlich auf eine Konzentration geachtet.	Bei Bundestagswahlen werden die Hochrechnungen von Wahlergebnissen in Ballungszentren (z.B. großen Städten) erstellt, deren Bevölkerungszusammensetzung typisch für das ganze Land ist.
Quotenauswahl	Man kennt die Verteilung relevanter Merkmale in der Grundgesamtheit und setzt die Stichprobe entsprechend dieser Quote zusammen.	Wenn der Anteil von Frauen in Deutschland, die Spitzenpositionen (= Jahresgehalt von > 130 TEuro) besetzten, bei 5 % (=Quote) liegt und man eine Stichprobe von Menschen in Spitzenpositionen untersucht, liegt der Anteil der Frauen in dieser Stichprobe bei 5 %.

D. 1. 2. (*Statistik, Parameter*)
Eine *Statistik* ist eine Maßzahl einer Stichprobe; *ein Parameter* eine Maßzahl der Grundgesamtheit:

Maßzahl	Statistik (Stichprobe)	Parameter (Grundgesamtheit)
Arithmetisches Mittel	$\bar{x}$	μ
Standardabweichung	s	σ
Anzahl der Elemente	n	N
Anteil	p	P

D. 1. 3. (*Stichprobenverteilung*)
Eine *Stichprobenverteilung* ist die Verteilung der Statistiken einer Stichprobe.

B. 1. 2.
Es gilt:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \begin{cases} \dfrac{\sigma^2}{n} & \text{mit Zurücklegen} \\ \dfrac{\sigma^2}{n} \cdot \left(\dfrac{N-n}{N-1} \right) & \text{ohne Zurücklegen} \end{cases}$$

Hier sind:

μ :	Mittelwert der Grundgesamtheit
σ^2 :	Varianz der Grundgesamtheit
$\mu_{\bar{x}}$:	Mittelwert der Stichprobenverteilung
$\sigma_{\bar{x}}^2$:	Varianz der Stichprobenverteilung

BS. 1. 2.
Eine Grundgesamtheit bestehe aus den Zahlen

2, 3, 6, 8, 11

Betrachtet seien alle möglichen Stichproben des Umfangs 2 *mit Zurücklegen*.
Bestimmen Sie

1. den Mittelwert der Grundgesamtheit
2. die Standardabweichung der Grundgesamtheit
3. den Mittelwert der Stichprobenverteilung
4. die Standardabweichung der Stichprobenverteilung

Lösung:
1.

$$\mu = \frac{2+3+6+8+11}{5} = 6.0 .$$

2.

$$\sigma^2 = \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = 10.8 ,$$

$$\sigma = 3.29.$$

3.
Es gibt 25 Stichproben des Umfangs 2 mit Zurücklegen:

(2,2)	(2,3)	(2,6)	(2,8)	(2,11)
(3,2)	(3,3)	(3,6)	(3,8)	(3,11)
(6,2)	(6,3)	(6,6)	(6,8)	(6,11)
(8,2)	(8,3)	(8,6)	(8,8)	(8,11)
(11,2)	(11,3)	(11,6)	(11,8)	(11,11)

Die entsprechenden Stichprobenmittelwerte lauten:

2.0	2.5	4.0	5.0	6.5
2.5	3.0	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

Es gilt:

$$\mu_{\bar{x}} = \frac{150}{25} = 6.0 .$$

Damit gilt: $\mu_{\bar{x}} = \mu$.

4.

$$\sigma_{\bar{x}}^2 = \frac{(2-6)^2 + (2.5-6)^2 + ... + (11.0-6)^2}{25} = \frac{135}{25} = 5.40$$

d.h.

$$\sigma_{\bar{x}} = \sqrt{5.40} = 2.32.$$

Damit wurde gezeigt: $\sigma_{\bar{x}}^2 = 5.40 = \frac{10.8}{2} = \frac{\sigma^2}{n}$

BS. 1. 3.

Lösen Sie das Beispiel BS 1 für den Fall *ohne Zurücklegen*.

Lösung:

1. – 2.

Wie oben:

$$\mu = 6, \quad \sigma^2 = 10.8, \quad \sigma = 3.29 .$$

3.

Es gibt 10 Stichproben des Umfangs 2 *ohne Zurücklegen*:

(2,3), (2,6), (2,8), (2,11), (3,6), (3,8), (3,11), (6,8), (6,11), (8,11).

Die entsprechenden Mittelwerte lauten:

2.5, 4.0, 5.0, 6.5, 4.5, 5.5, 7.0, 7.0, 8.5, 9.5

Damit hat man

$$\mu_{\bar{x}} = \frac{2.5+4.0+5.0+6.5+4.5+5.5+7.0+7.0+8.5+9.5}{10} = 6.0$$

d. h. $\mu_{\bar{x}} = \mu$.

4.

$$\sigma^2_{\bar{x}} = \frac{(2.5-6.0)^2+(4.0-6.0)^2+(5.0-6.0)^2+...+(9.5-6.0)^2}{10} = 4.05$$

d.h. $\sigma_{\bar{x}} = 2.01$.

Damit wurde gezeigt: $\frac{\sigma^2}{n}\cdot\left(\frac{N-n}{N-1}\right) = \frac{10.8}{2}\cdot\left(\frac{5-2}{5-1}\right) = 4.05 = \sigma^2_{\bar{x}}$.

B. 1. 3. (*Form der Stichprobenverteilung des Mittelwertes*)

1.

Für eine *normalverteilte Grundgesamtheit* gilt: Die Verteilung des Stichprobenmittelwertes ist *normal* mit

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

2.

Für eine „große" Stichprobe ($n \geq 30$) gewählt aus einer *beliebig verteilten Grundgesamtheit* gilt: Die Verteilung des Stichprobenmittelwertes ist *näherungsweise normal* mit

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

BS. 1. 4.

Die jährlichen Ersparnisse der reichen Leute über 50 in den USA sind normalverteilt mit einem Durchschnitt von $125000 und einer Standardabweichung von $25000.

1. Beschreiben Sie die Verteilung der Stichprobenmittelwerte des Umfangs 16
2. Wie groß ist die Wahrscheinlichkeit, dass die durchschnittlichen Ersparnisse von 16 reichen Personen über 50 in den USA größer sind als $135000?
3. Wie groß ist die Wahrscheinlichkeit, , dass die durchschnittlichen Ersparnisse von 16 reichen Personen über 50 in den USA zwischen $124000 und $129000 liegen?

Lösung:

1.

Die Verteilung der Stichprobenmittelwerte des Umfangs 16 ist normal mit

$$\mu_{\bar{x}} = \$125000, \qquad \sigma_{\bar{x}} = \frac{25000}{\sqrt{16}} = \$6250.$$

2.

$$P(\bar{x} > 135000) = 1 - P(\bar{x} \leq 135000) \approx 1 - P(\bar{x} < 135000)$$

$$= 1 - \Phi\left(\frac{135000-125000}{6250}\right) = 1 - \Phi(1.6)$$

$$= 1 - 0.945201 = 0.054799\ .$$

3.

$$P(124000 \le \bar{x} < 129000) = F(129000) - F(124000)$$

$$= \Phi\left(\frac{129000 - 125000}{6250}\right) - \Phi\left(\frac{124000 - 125000}{6250}\right)$$

$$= \Phi(0.64) - \Phi(-0.16) = \Phi(0.64) - \left(1 - \Phi(-0.16)\right)$$

$$= \Phi(0.64) - 1 + \Phi(0.16) = 0.7389 - 1 + 0.5636 = 0.3025.$$

(Letzte Aktualisierung: 04.01.2015)