

Stichprobenverteilung bei unterschiedlichen Auswahlätzen

Begleitende Unterlagen zur Übung *Induktive Statistik*

Michael Westermann

Universität Essen

Inhaltsverzeichnis

1	Einleitung	2
2	Stichprobenfunktion	2
3	Stichprobenverteilung	2
4	Numerisches Beispiel	3

1 Einleitung

Gegenstand der induktiven Statistik ist u.a. der Repräsentativschluß oder auch der indirekte Schluß. D.h. es wird mittels der Daten, welche in einer Stichprobe enthalten sind, auf die Verteilung der Merkmale in der Grundgesamtheit geschlossen.

Diese Vorgehensweise ist notwendig, da es nicht immer möglich, bzw. nur unter extremem Zeit- und Kostenaufwand möglich ist, die relevante Grundgesamtheit zu befragen, bzw. zu untersuchen. Eine Situation, in der die Untersuchung der Grundgesamtheit zwangsläufig unmöglich ist, findet man bei der statistischen Qualitätskontrolle vor. Soll bei der Überprüfung des Produktionsprozesses von Glühlampen die durchschnittliche Brenndauer der Produktion ermittelt werden, so ist dies nur mit einer Stichprobe möglich. Alles andere würde bedeuten, daß die gesamte Produktion zur Überprüfung herangezogen würde und somit nicht mehr verkauft werden könnte.

2 Stichprobenfunktion

Es wird eine Stichprobe vom Umfang n betrachtet. Diese wurde aus einer Grundgesamtheit vom Umfang N entnommen. Der Quotient $\frac{n}{N}$ heißt *Auswahlsatz*. Die Daten der Stichprobe werden in eine Stichprobenfunktion eingesetzt. Diese Stichprobenfunktion ist eine Rechenvorschrift, welche vorgibt, wie mit den konkret ermittelten Daten umzugehen ist. Allgemein handelt es sich bei einer Stichprobenfunktion um eine Funktion der Stichprobenrealisationen, also

$$\hat{\Theta} = g(X_1, X_2, \dots, X_n). \quad (1)$$

Wegen der Zufälligkeit der tatsächlichen Stichprobenrealisationen (bei den einzelnen X_i 's handelt es sich um Zufallsvariablen) handelt es sich bei der Schätz- bzw. Stichprobenfunktion um eine Zufallsvariable mit entsprechenden Eigenschaften. Zufallsvariablen werden durch ihre Wahrscheinlichkeitsfunktion und ihre Parameter hinreichend beschrieben.

Auf diesen Seiten kann (zumindest gedanklich bzw. beispielhaft) gezeigt werden, welche Eigenschaften solch eine Stichprobenfunktion besitzt.

3 Stichprobenverteilung

Alle möglichen Realisationen einer Stichprobenfunktion¹ mit ihren entsprechenden Wahrscheinlichkeiten bilden zusammen die *Stichprobenverteilung*. Theoretisch bedeutet dies, daß man alle möglichen Stichproben betrachtet und zu jeder Stichprobe die entsprechende Realisation der Stichprobenfunktion berechnet. Tut man dies für alle $\binom{N}{n}$ möglichen Stichproben² und erstellt eine Häufigkeitsverteilung für diese Stichprobenrealisationen, so gelangt man zur Stichprobenverteilung. Diese Verteilung wird durch die Verteilungsfunktion und durch die Angabe des Erwartungswertes und der Varianz genau beschrieben.

Die Varianz einer Stichprobenverteilung wird *Schätz- bzw. Stichprobenfehler* genannt.

Eine im statistischen Sinne gute Schätzfunktion zeichnet sich dadurch aus, daß ihr Erwartungswert mit dem (unbekannten) Erwartungswert der Grundgesamtheit übereinstimmt

$$E(\hat{\Theta}) = \Theta.$$

Trifft dies zu, so handelt es sich um einen *erwartungstreuen* Schätzer.

¹ Eine konkrete Realisation der Schätzfunktion lautet dann beispielsweise $\hat{\Theta}_{(1)} = g(x_1, x_2, \dots, x_n)$, d.h. es handelt sich dabei um einen Funktionswert von Gl. 1. Die allgemein beschriebene Stichprobenfunktion konkretisiert sich durch die tatsächlich eingetretenen Realisationen der Zufallsvariablen (aus X_i wird dann x_i) zu einem bestimmten Wert.

² Ziehen ohne Zurücklegen, bei Ziehen mit Zurücklegen existieren $\binom{N+n-1}{n}$ mögliche Stichproben.

Wenn die Schätzfunktion zusätzlich noch so konstruiert ist, daß der Stichprobenfehler mit zunehmendem Stichprobenumfang verschwindet, so nennt man diese Schätzfunktion *konsistent*

$$\lim_{n \rightarrow \infty} V(\hat{\Theta}) = 0.$$

Theoretisch kann gezeigt werden, daß es sich bei der Schätzfunktion

$$\hat{\Theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

um einen erwartungstreuen und konsistenten Schätzer für den Erwartungswert der Grundgesamtheit handelt.

Erwartungstreue

Der Erwartungswert des Schätzers errechnet sich als

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n \cdot \mu = \mu.$$

Das arithmetische Mittel ist somit ein erwartungstreuer Schätzer.

Konsistenz

Die Varianz der Schätzfunktion ermittelt sich für den Fall *zmz* als

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

bzw.

$$V(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

für den Fall *zoz*. Beide Ausdrücke werden mit zunehmendem Stichprobenumfang immer kleiner. Somit handelt es sich um eine konsistente Schätzfunktion.

4 Numerisches Beispiel

Das nun folgende Beispiel soll die oben erwähnten Zusammenhänge verdeutlichen. Es wird angenommen, die Grundgesamtheit bestehe aus $N = 10$ Elementen:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Es handelt sich hierbei um eine Gleichverteilung nach Art eines Würfels, jedoch mit 10 Ausprägungen. Diese Grundgesamtheit weist einen Erwartungswert von $\mu = 5,5$ und eine Varianz von $\sigma^2 = 8,25$ auf.

Es wird nun gezeigt wie sich die Stichprobenverteilung des arithmetischen Mittels für unterschiedliche Stichprobenumfänge entwickelt. Es werden Stichproben vom Umfang $n = 1, 2, \dots, 9, 10$ gezogen, wobei es sich im letzten Fall nicht um eine Stichprobe handelt sondern um eine Totalerhebung, da Stichprobe und Grundgesamtheit sich nicht mehr unterscheiden. Wird das Zufallsexperiment *ohne Zurücklegen durchgeführt*, ermittelt sich der Stichprobenfehler wie oben bereits erwähnt als

$$V(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1}.$$

Die Anzahl aller möglichen Stichproben ermittelt man im Fall zoz mittels des Binomialkoeffizienten $\binom{N}{n}$. Es ergeben sich für $N = 10$ und $n = 1, 2, \dots, 10$ folgende Anzahlen möglicher Stichproben:

Stichprobenumfang	mögliche Stichproben
1	10
2	45
3	120
4	210
5	252
6	210
7	120
8	45
9	10
10	1

Es wird nun beispielhaft für den Fall $n = 2$ die Vorgehensweise zur Ermittlung der Stichprobenverteilung dargestellt. Wie gerade gezeigt, existieren $\binom{10}{2} = 45$ Möglichkeiten, eine Stichprobe vom Umfang 2 aus der oben beschriebenen Grundgesamtheit zu entnehmen. Alle diese Stichproben führen zu einem ganz bestimmten arithmetischen Mittel. Diese möglichen Stichproben und ihr jeweiliger Stichprobenmittelwert sind in der folgenden Tabelle dargestellt:

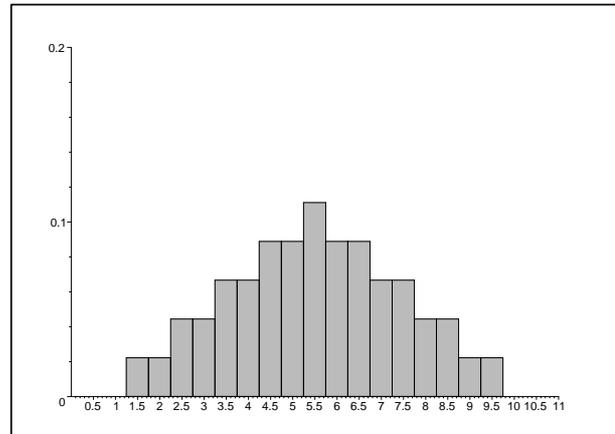
Tabelle 1 Mögliche Stichproben und Stichprobenmittelwerte im Falle $n = 2$

i	Stichprobe	\bar{x}	i	Stichprobe	\bar{x}	i	Stichprobe	\bar{x}
1	(1, 2)	1,5	16	(2, 9)	5,5	31	(5, 6)	5,5
2	(1, 3)	2	17	(2, 10)	6	32	(5, 7)	6
3	(1, 4)	2,5	18	(3, 4)	3,5	33	(5, 8)	6,5
4	(1, 5)	3	19	(3, 5)	4	34	(5, 9)	7
5	(1, 6)	3,5	20	(3, 6)	4,5	35	(5, 10)	7,5
6	(1, 7)	4	21	(3, 7)	5	36	(6, 7)	6,5
7	(1, 8)	4,5	22	(3, 8)	5,5	37	(6, 8)	7
8	(1, 9)	5	23	(3, 9)	6	38	(6, 9)	7,5
9	(1, 10)	5,5	24	(3, 10)	6,5	39	(6, 10)	8
10	(2, 3)	2,5	25	(4, 5)	4,5	40	(7, 8)	7,5
11	(2, 4)	3	26	(4, 6)	5	41	(7, 9)	8
12	(2, 5)	3,5	27	(4, 7)	5,5	42	(7, 10)	8,5
13	(2, 6)	4	28	(4, 8)	6	43	(8, 9)	8,5
14	(2, 7)	4,5	29	(4, 9)	6,5	44	(8, 10)	9
15	(2, 8)	5	30	(4, 10)	7	45	(9, 10)	9,5

Aus diesen möglichen Stichproben kann man die Wahrscheinlichkeitsverteilung des arithmetischen Mittels – also die Stichprobenverteilung – herleiten.

Sie lautet

\bar{x}_i	$f(\bar{x}_i)$	\bar{x}_i	$f(\bar{x}_i)$
1,5	0,0222	6	0,0889
2	0,0222	6,5	0,0889
2,5	0,0444	7	0,0667
3	0,0444	7,5	0,0667
3,5	0,0667	8	0,0444
4	0,0667	8,5	0,0444
4,5	0,0889	9	0,0222
5	0,0889	9,5	0,0222
5,5	0,1111		



Stichprobenverteilung von \bar{x} für $n = 2$

Erwartungswert und Stichprobenfehler lauten:

$$E(\bar{x}) = 5,5 \quad \text{und} \quad V(\bar{x}) = \frac{8,25}{2} \cdot \frac{8}{9} = 3,6667.$$

Zur Ermittlung der übrigen Stichprobenverteilungen und ihrer Parameter wird analog vorgegangen. Die folgende Tabelle stellt die daraus resultierenden Größen zusammen:

Tabelle 2 Stichprobenverteilung für unterschiedliche Stichprobenumfänge

Stichprobenumfang	Erwartungswert	Stichprobenfehler
1	5.5	8.25
2	5.5	3.6667
3	5.5	2.1389
4	5.5	1.375
5	5.5	0.9167
6	5.5	0.6111
7	5.5	0.3929
8	5.5	0.2292
9	5.5	0.1019
10	5.5	0

Anhand dieser Übersicht wird deutlich, was Konsistenz bedeutet. Die Vergrößerung des Stichprobenumfangs hat zur Folge, daß die Spannweite der möglichen Stichprobenergebnisse kontinuierlich abnimmt. Diese Abnahme der Spannweite drückt sich durch den Stichprobenfehler aus, welcher im Extremfall ($n = 10$) den Wert Null annimmt, was auch nur logisch ist. Wenn die Grundgesamtheit bezüglich eines Merkmals befragt wird, so begeht man bei der Schätzung des Mittelwertes keinen Fehler, weil es sich ja eben nicht um eine Schätzung sondern um eine Totalerhebung handelt und folglich nur ein einziger Wert – der Parameter der Grundgesamtheit – ermittelt wird. Dieser Sachverhalt wird abschließend auch noch in der folgenden Abbildung dargestellt.

Abb. 1 Stichprobenverteilung für unterschiedliche Stichprobenumfänge

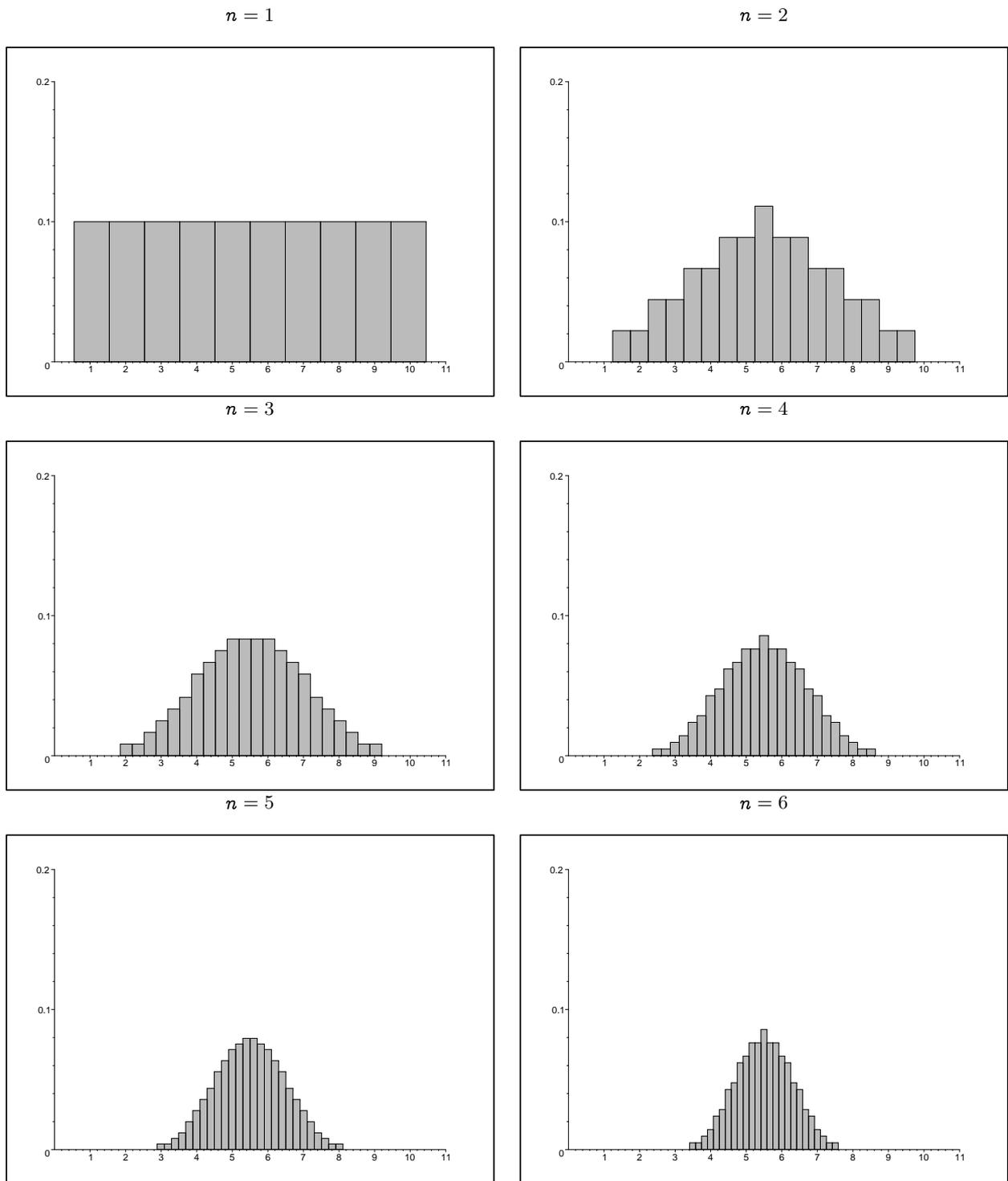
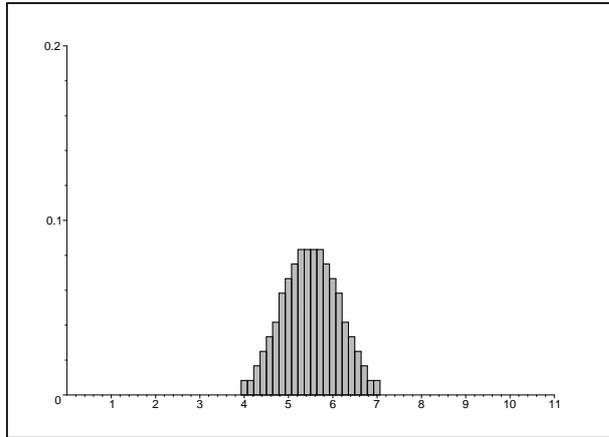
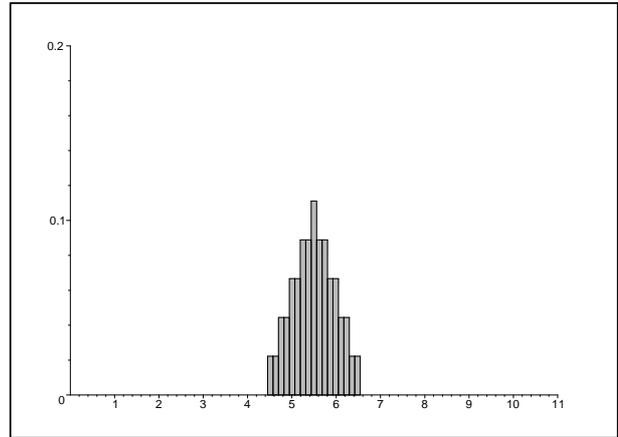


Abb. 1 (Fortsetzung) Stichprobenverteilung für unterschiedliche Stichprobenumfänge

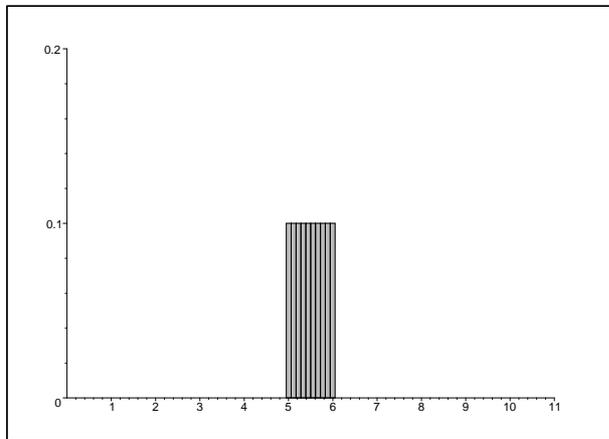
$n = 7$



$n = 8$



$n = 9$



$n = 10$

