Inferenzstatistik per Simulation: Bootstrap-Konfidenzintervalle in der Sekundarstufe II mit Excel

ANDREAS KAUFMANN, ECHALLENS, JOACHIM ENGEL, LUDWIGSBURG

Zusammenfassung: *Dieser Aufsatz enthält einen Vorschlag für die Einführung von Bootstrap-Konfidenzintervallen in der Sekundarstufe II. Zuerst wird der Nutzen dieser Methoden für die angewandte Statistik und für den Mathematikunterricht erläutert. Anschließend wird das Verfahren an zwei Beispielen präsentiert, seine Eigenschaften werden anschaulich begründet und Fehlerquellen und Konvergenzfragen werden diskutiert. Schließlich werden Vorschläge zum Erweitern und Verfeinern präsentiert.*

1 Einleitung

Erheben von Daten, Erstellen von Grafiken wie z. B. Histogramme, Streudiagramme oder Errechnen numerischer Werte wie Mittelwerte, Mediane oder Streumaße sind sehr konkrete Operationen. Dagegen sind Konzepte, die auf der Zufallsvariabilität dieser Statistiken beruhen wie Stichprobenverteilungen, Standardfehler, Konfidenzintervalle, Hypothesentests und P-Werte abstrakt und schwer für Schüler zu verstehen. Bootstrap-Methoden nehmen die konkreten Konzepte, mit denen Schüler vom Arbeiten mit Daten vertraut sind, und wenden sie zur Herleitung von Stichprobenverteilungen an. Der Einsatz geeigneter Software hilft nicht nur den für Simulationen nötigen hohen Rechenaufwand zu bewältigen, sondern unterstützt vor allem auch konzeptionelles Verstehen. Wir illustrieren die Bootstrap-Methode zur Herleitung von (approximativen) Konfidenzintervallen für Mittelwerte und Korrelationskoeffizienten.

2 Bootstrap: Recycling von Daten

Die grundlegende Idee des Bootstraps ist wie folgt: Angenommen, man könnte beliebig oft Stichproben vom vorgegebenen Umfang *n* aus der Population ziehen und ließe sich dann aus jeder Stichprobe die Wer-

Stochastik in der Schule 35 (2015) 2, S. 14–20

te der gewünschten Statistik (Mittelwert, Median, Varianz etc.) errechnen. Dann ließe sich aus zahlreichen solchen Stichprobenstatistiken die (empirische) Stichprobenverteilung erstellen. Nun kosten bei konkreten Anwendungen Stichproben Ressourcen (Zeit, Geld etc.) und das Ziehen von sehr vielen Stichproben gegebenen Umfangs ist aus praktischen Gründen indiskutabel. Bei statistischen Untersuchungen ist in der Regel die Population nicht verfügbar, sondern man hat nur eine Stichprobe. Hier setzt die Bootstrap-Idee an: Anstelle aus der nicht-verfügbaren Population werden weitere Stichproben – Bootstrap-Stichproben – aus der vorliegenden Stichprobe gezogen. Durch dieses Recycling der Daten (die Fachliteratur spricht von *Resampling*) können Schätzer zwar nicht verbessert werden, da der ursprünglich vorliegende Stichprobenumfang fest ist. Es ist jedoch möglich die Verteilung von Stichprobenstatistiken durch die Verteilung der Bootstrap-Stichproben zu approximieren. Diese Approximation ist umso besser, je mehr die vorliegende Stichprobe der Population ähnelt. Diese Idee geht auf Brad Efron (1979) zurück.

Bootstrap-Konfidenzintervalle erhält man dann, indem dann z. B. das 2,5 % und 97,5 % Perzentil der Bootstrap-Verteilung als Konfidenzgrenzen genommen wird. Bootstrap-Konfidenzintervalle sind viel flexibler als die klassischen Konfidenzintervalle, da die Berechnung analytisch unzugänglicher Stichprobenverteilungen – wie z. B. des Medians oder von Kennzahlen für robuste explorative Datenanalysen – beim Bootstrap durch Simulationen ersetzt werden, und weil man nicht auf Hypothesen hinsichtlich der Verteilung der Daten angewiesen ist (Engel & Grübel 2008). Die in den meisten Verfahren der klassischen Statistik vorausgesetzten Normalverteilungsannahmen sind hier nicht erforderlich. Das begründet die enorme Bedeutung dieser Methode für die moderne angewandte Statistik. Verschiedene Autoren (z. B. Cobb 2007, Engel 2010, Hesterberg 2014, Pfannkuch & Budgett 2014) erklären, dass der Bootstrap neben seiner Nützlichkeit als flexible Methode für viele Anwendungssituationen auch ein erhebliches didaktisches Potenzial besitzt. Die Rechenwege zur Gewinnung von Referenzverteilungen in der klassischen Inferenzstatistik sind in vielen Ausbildungsstufen – wie zum Beispiel in der Sekundarstufe II – nicht realisierbar. Simulationen bieten hier einen Ausweg, die formale Mathematik auf ein Minimum zu reduzieren und den Fokus auf konzeptionelles Verstehen zu richten.

Das zentrale Konzept der Inferenzstatistik ist die Stichprobenverteilung. Resampling Methoden erlauben dem Lernenden durch Simulationen zu erfahren, wie sich eine Statistik von Stichprobe zu Stichprobe unterscheidet. Anhand von Bootstrap-Konfidenzintervallen kann die beschreibende Statistik in der Sekundarstufe II zur Inferenzstatistik weitergeführt werden. Wenn Schüler die Idee des Bootstraps an einfachen Beispielen wie etwa dem Mittelwert verstanden haben, können sie das Verfahren leicht auf andere Kennwerte übertragen. Damit lassen sich dann vielfältige Projektarbeiten in der Sekundarstufe II durchführen.

Johnson (2001) zeigt in seinem Aufsatz, wie man Bootstrap-Konfidenzintervalle Studenten nahe bringen kann, die bereits mit der klassischen Inferenzstatistik vertraut sind. Christie (2004) erklärt wie man Bootstrap-Konfidenzintervalle mit Excel berechnen kann. Mit Excel lassen sich die einzelnen Schritte sowie das gesamte Verfahren gut visualisieren, das Programmieren ist relativ einfach und viele Schüler haben bereits Erfahrung mit diesem Programm. Im Rahmen dieses Aufsatzes gehen wir nicht näher auf die Gestaltung der Excel-Arbeitsblätter ein, die jedoch auf Anfrage von den Autoren zur Verfügung gestellt werden. Alternativ lässt sich der Bootstrap auch verständnisfördernd mit didaktisch konzipierter Stochastik-Software wie Fathom (Erickson 2001) oder Tinkerplots (Watson 2013) implementieren.

Auf begrifflicher Ebene verwenden wir den Begriff des *approximativen Modells der Population*, wenn wir von den Daten sprechen, dies um das Simulieren anhand der Daten zu veranschaulichen.

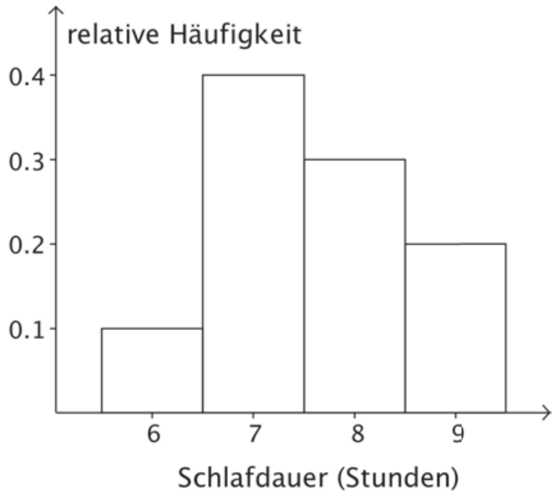
3 Bootstrap-Konfidenzintervalle für das arithmetische Mittel

Als einführendes Beispiel betrachten wir Konfidenzintervalle für den Mittelwert μ einer Population. Die Stichprobenverteilung von (empirischen) Mittelwerten ist annähernd symmetrisch verteilt (das sagt uns der zentrale Grenzwertsatz), wenn der Stichprobenumfang groß genug ist. Somit wird auch das Konfidenzintervall für den Populationsmittelwert annähernd symmetrisch sein. Im Folgenden soll die Idee von Bootstrap-Konfidenzintervallen illustriert werden. Um der Klarheit willen ist das konstruierte Beispiel bewusst einfach gewählt worden und mit einem wenig realistischen Kontext versehen. Es geht uns hierbei nicht um Authentizität einer Anwendung, sondern um ein Verständnis für das methodische Vorgehen beim Bootstrap. Die folgenden Überlegungen lassen sich leicht verallgemeinern auf Bootstrap-Konfidenzintervalle für Kennwerte, die durch Schätzer mit symmetrischer Stichprobenverteilung geschätzt werden.

3.1 Eine bekannte Population

Die Schulleitung einer Fachhochschule kennt die Schlafgewohnheiten aller Studenten, da bei der Immatrikulation alle Studenten in einem Fragebogen angeben mussten, wie lange sie täglich schlafen. Zunächst gehen wir also von der Annahme aus, dass die Population schon bekannt sei. Das ist zugegebenermaßen sehr unrealistisch, und – da die Population bekannt ist – erfordert es auch keinerlei Inferenzstatistik. Diese Annahme soll auch nur als illustrativer Zwischenschritt dienen. Durch Anwenden verschiedener Modelle (exaktes Modell und approximatives Modell) wird die Stimmigkeit der Verfahren dadurch plausibel, dass die ermittelten Werte zum Populationswert passen. Die Populationsdaten sind in Tabelle 1 zusammengefasst. Der Mittelwert der Population beträgt $\mu = 7{,}6$ Stunden.

Schlafdauer [in Stunden]	Relative Häufigkeit
6	10 %
7	40 %
8	30 %
9	20 %



Tab. 1: Relative Häufigkeiten der täglichen Schlafdauer der Population der Studenten der Fachhochschule und Histogramm.

3.2 Exaktes Modell der Population und 95 % Schwankungsbereich

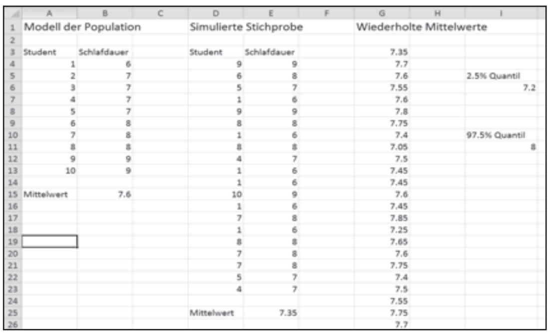
Lena und Hasso studieren an der Fachhochschule. Im Rahmen einer Projektarbeit sollen sie unter anderem untersuchen wie lange ihre Mitstudenten täglich schlafen. Sie fragen gemäß Zufallsprinzip 20 Mitstudenten, wie lange sie täglich schlafen und berechnen das arithmetische Mittel. In welchem Bereich würde dieser Mittelwert in 95 % der „Fälle" liegen, vorausgesetzt sie könnten die Befragung von 20 zufällig ausgewählten Mitstudenten beliebig oft wiederholen?

Diese Frage beantworten wir durch Simulieren von Umfragewerten: Als exaktes Modell für die Population betrachten wir die folgenden 10 Studenten und deren Schlafdauer (Tabelle 2). Dieses Modell spiegelt die Population repräsentativ wider, da die relativen Häufigkeiten dieser Stichprobe den relativen Häufigkeiten der Population gleichen.

Student	Schlafdauer [in Stunden]
1	6
2	7
3	7
4	7
5	7
6	8
7	8
8	8
9	9
10	9

Tab. 2: Exaktes Modell für die Population der Studenten der Fachhochschule.

Um eine Zufallsstichprobe mit 20 Umfragewerten zu erzeugen gehen wir so vor: Wir erzeugen eine Zufallszahl zwischen 1 und 10, sie entspricht dem Student aus dem Modell, der befragt wird, und wir schreiben die entsprechende Schlafdauer auf. Das machen wir 20 Mal und erhalten so 20 Umfragewerte, welche eine simulierte Zufallsstichprobe bilden (Spalte E in Tabelle 3), aus der sich ein simulierter Mittelwert errechnen lässt (Zelle E25 in Tabelle 3). Das wiederholen wir 1000 Mal und erhalten so 1000 simulierte Mittelwerte (Spalte G in Tabelle 3).



	A	B	C	D	E	F	G	H	I
1	Modell der Population			Simulierte Stichprobe			Wiederholte Mittelwerte		
2									
3	Student	Schlafdauer		Student	Schlafdauer		7.35		
4	1	6		9	9		7.7		
5	2	7		6	8		7.6		2.5% Quantil
6	3	7		5	7		7.55		7.2
7	4	7		1	6		7.6		
8	5	7		9	9		7.8		
9	6	8		8	8		7.75		
10	7	8		1	6		7.4		97.5% Quantil
11	8	8		8	8		7.05		8
12	9	9		4	7		7.5		
13	10	9		1	6		7.45		
14				1	6		7.45		
15	Mittelwert	7.6		10	9		7.6		
16				1	6		7.45		
17				7	8		7.85		
18				1	6		7.25		
19				8	8		7.65		
20				7	8		7.6		
21				7	8		7.75		
22				5	7		7.4		
23				4	7		7.5		
24							7.55		
25				Mittelwert	7.35		7.75		
26							7.7		

Tab. 3 : Anhand des exakten Modells der Population (Spalte B) wird eine Stichprobe simuliert (Spalte E) und deren Mittelwert berechnet (Zelle E16). Das wird 1000 Mal wiederholt und die Mittelwerte in Spalte G gespeichert.

Sei $q(0{,}025)$ das 2,5 % Perzentil dieser simulierten Mittelwerte und $q(0{,}975)$ das 97,5 % Perzentil. Bei vielen Versuchen (z. B. bei 1000 Wiederholungen) liegt in ca. 95 % der Fälle der simulierte Mittelwert im Bereich $[q(0{,}025; q(0{,}975)] = [7{,}2; 8]$. Wir nennen dieses Intervall 95 %-Schwankungsbereich der Mittelwerte der Stichproben. Die halbe Breite des 95 %-Schwankungsbereiches ist $b = [q(0{,}975) - q(0{,}025)]/2 = 0{,}4$. Allgemein betrachtet gilt: Ist μ der Populationsmittelwert und $b = q(0{,}975) - \mu = \mu - q(0{,}025)$, dann ist der 95 %-Schwankungsbereich $[\mu - b, \mu + b]$.

3.3 Approximatives Modell der Population und Bootstrap-Konfidenzintervall für den Mittelwert

Lena und Hasso sind die Population nicht bekannt. Sie befragen gemäß Zufallsprinzip 20 Mitstudenten, wie lange sie täglich schlafen und erhalten 20 Umfragewerte (Tabelle 4), für die sich als arithmetisches Mittel $\bar{x} = 7{,}5$ errechnet. Lena sagt: „Hätten wir 20 andere Personen befragt, dann hätten wir einen anderen Mittelwert erhalten, vielleicht $\bar{x} = 7{,}1$“. Hasso sagt: „Oder vielleicht $\bar{x} = 7{,}8$. Wie sicher können wir uns bezüglich des Populationsmittelwerts sein?“ Gesucht ist ein um das Stichprobenmittel $\bar{x}$ symmetrisches Intervall, das den Populationsmittelwert μ in 95 % der Fälle enthält, d. h. in 100 hypothetisch durchgeführten Umfragen sollte das jeweilige Intervall in ca. 95 Fällen den Populationsmittelwert μ enthalten.

Student	Schlafdauer [in Stunden]
1	7
2	9
3	7
4	7
5	9
6	8
7	6
8	7
9	8
10	7
11	7
12	9
13	8
14	9
15	6
16	8
17	7
18	7
19	6
20	8

Tab. 4: Umfragewerte von Lena und Hasso. Sie bilden das approximative Modell der Population.

Fall 1:

Hasso ruft bei der Schulleitung an und erhält den Tipp $b = 0{,}4$. Anhand des Tipps ist die Lösung das Intervall $7{,}5 \pm 0{,}4 = [7{,}1; 7{,}9]$. In 95 % der Fälle liegt der Populationsmittelwert μ im Intervall $[\bar{x} - b; \bar{x} + b]$. Denn: In 95 % der Fälle liegt der Umfragemittelwert $\bar{x}$ im 95 %-Schwankungsbereich und das Intervall $[\bar{x} - b, \bar{x} + b]$ enthält μ in diesen Fällen (siehe Abbildung 1).

$$I = [\bar{x} - b; \bar{x} + b] = [7{,}1; 7{,}9]$$

bezeichnen wir deshalb als genaues 95 %-Konfidenzintervall für den Populationsmittelwert μ.

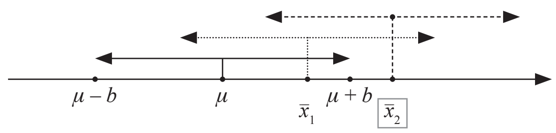


Abb. 1: Das als durchgezogene Linie dargestellte Intervall ist der 95 %-Schwankungsbereich $[\mu - b; \mu + b]$ der Mittelwerte der Stichproben. Die gepunktete Linie stellt das 95 %-Konfidenzintervall $[\bar{x}_1 - b; \bar{x}_1 + b]$ dar und enthält μ, da $\bar{x}_1$ im 95 %-Schwankungsbereich von μ liegt. Die gestrichelte Linie ist das 95 %-Konfidenzintervall $[\bar{x}_2 - b; \bar{x}_2 + b]$; es enthält nicht μ, da $\bar{x}_2$ nicht im 95 %-Schwankungsbereich von μ liegt.

Fall 2:

Lena und Hasso erhalten keinen Tipp von der Schulleitung. Da das Modell der Population nicht bekannt ist, ist auch b nicht bekannt. Deshalb verwenden sie die Umfragewerte, d. h. ihre Stichprobe (Tabelle 4) als approximatives Modell der Population an Stelle des unbekannten exakten Modells der Population (Tabelle 2). Sie ziehen per Simulation wiederholt Stichproben *(Bootstrap-Stichproben)* vom Umfang $n = 20$ aus der vorhandenen Stichprobe um b zu approximieren, d. h.

$$b = [q(0{,}975) - q(0{,}025)]/2$$

wird geschätzt durch

$$\hat{b} = [\hat{q}(0{,}975) - \hat{q}(0{,}025)]/2\,.$$

Dabei ist $\hat{q}(0{,}975)$ bzw. $\hat{q}(0{,}025)$ das 97,5 % bzw. 2,5 % Quantil der Verteilung der Mittelwerte, die anhand des approximativen Modells der Population simuliert wurden (siehe Tabelle 5). Die beiden unbekannten Quantile sowie der Populationsmittelwert werden somit ersetzt durch die entsprechenden Quantile der Bootstrap-Verteilung und den Mittelwert der vorliegenden Daten.

In unserem Fall ist

$$\hat{b} = \frac{7{,}9 - 7{,}15}{2} = 0{,}375$$

und wir erhalten das approximative 95 % Konfidenzintervall $\tilde{I} = [\bar{x} - \hat{b}; \bar{x} + \hat{b}]$ = [7,125; 7,875] für den Populationsmittelwert. Alternativ hätte man – unter Verzicht auf die Annahme der Symmetrieannahme – auch die entsprechenden Perzentile der Bootstrap-Verteilung als approximatives Konfidenzintervall nehmen können, d. h.

$$\tilde{I} = [\hat{q}(0{,}025); \hat{q}(0{,}975)] = [7{,}15; 7{,}9].$$

In Abgrenzung zu weiteren im Zusammenhang mit dem Bootstrap stehenden Varianten von Intervallen wird in der Fachliteratur dieses Intervall als Bootstrap-Perzentilintervall *(bootstrap percentile interval)* bezeichnet (Engel 2010).

	A	B	C	D	E	F	G	H	I
1	Daten			Resampling			Wiederholte Mittelwerte		
2									
3	Nummer	Schlafdauer		Nummer	Schlafdauer		7,35		
4	1	7		4	7		7,95		
5	2	9		10	7		7,45		
6	3	7		9	8		7,75		2,5% Quantil
7	4	7		4	7		7,45		7,15
8	5	9		5	9		7,6		
9	6	8		4	7		7,55		
10	7	6		10	7		7,15		
11	8	7		9	8		7,35		97,5% Quantil
12	9	8		7	6		7,35		7,9
13	10	7		10	7		7,65		
14	11	7		9	8		7,15		
15	12	9		10	7		7,4		
16	13	8		6	8		7,3		
17	14	9		3	7		7,75		
18	15	6		6	8		7,5		
19	16	8		4	7		7,7		
20	17	7		8	7		7,55		
21	18	7		3	7		7,4		
22	19	8		10	7		7,5		
23	20	6		6	8		7,4		
24							7,55		
25	Mittelwert	7,5		Mittelwert	7,35		7,25		
26							7,75		
27							7,6		
28							7,65		

Tab. 5: Als approximatives Modell der Population werden die Daten (Spalte B) verwendet. Daraus wird eine neue Stichprobe simuliert (Spalte E) und dessen Mittelwert berechnet (Zelle E25). Das wird 1000 Mal wiederholt und die Mittelwerte werden in Spalte G gespeichert.

4 Bootstrap-Konfidenzintervalle bei asymmetrischen Stichprobenverteilungen

Lena und Hasso untersuchen den Zusammenhang zwischen Motivation und Erfolg für das Studium an der Fachhochschule für Architektur. In einer Umfrage haben sie 12 Mitstudenten nach ihrer Motivation und ihrem Erfolg befragt. Die Motivationswerte können Werte zwischen 0 und 5 annehmen, und die Erfolgswerte haben Werte zwischen 0 und 10. Die Ergebnisse sind in Tabelle 6 zusammengefasst.

Der Pearsonsche Korrelationkoeffizient zwischen Motivation und Erfolg in der Stichprobe beträgt $r = 0{,}69$. Nun wollen Lena und Hasso ein 95 %-Konfidenzintervall für die Korrelation zwischen Motivation und Erfolg in der Population ermitteln.

Im jetzt vorliegenden Fall ist die Stichprobenverteilung asymmetrisch. Somit ist auch die Symmetrieannahme für das Konfidenzintervall nicht mehr haltbar. Im Fall des Pearsonschen Korrelationskoeffizienten ϱ erhält man einen 95 % Schwankungsbereich der Art $[\varrho - b_1; \varrho + b_2]$, der nicht um ϱ zentriert ist (siehe Abbildung 2). Dennoch kann man zunächst ganz analog zum vorangegangenen Beispiel vorgehen und Bootstrap-Stichproben (d. h. neue Stichproben aus der vorhandenen Stichprobe) ziehen. Ganz analog wie beim Konfidenzintervall für den Mittelwert erhält man das Bootstrap-Perzentilintervall durch die Perzentile der Bootstrap-Verteilung: $[\hat{q}(0{,}025); \hat{q}(0{,}975)] = [r - \hat{b}_1; r + \hat{b}_2]$, wobei $\hat{b}_1 = r - \hat{q}(0{,}025)$; $\hat{b}_2 = \hat{q}(0{,}975) - r$ ist.

Um eine Zufallsstichprobe mit 12 Umfragewerten zu erzeugen, gehen Lena und Hasso daher wie folgt vor: Sie erzeugen eine Zufallszahl zwischen 1 und 12. Sie entspricht dem Studenten aus dem approximativen Modell der Population (die Daten), der befragt wird und schreiben seine Motivation und seinen Erfolg auf. Das wird 12 Mal wiederholt und resultiert somit in 12 gepaarten Umfragewerten, die eine simulierte Zufallsstichprobe bilden (Spalten F und G in Tabelle 7), aus der sich ein Korrelationskoeffizient berechnen lässt (Zelle G17 in Tabelle 7). Dieser Vorgang wird sehr oft (z. B. 1000 mal) wiederholt, wodurch man 1000 anhand der Daten simulierte Korrelationskoeffizienten erhält (Spalte I in Tabelle 7). Im vorliegenden Fall ergibt eine Simulation mit Excel $\hat{q}(0{,}025) = 0{,}28$ und $\hat{q}(0{,}975) = 0{,}92$ und somit das Intervall [0,28; 0,92] als Bootstrap-Perzentil-Intervall

Student	Motivation	Erfolg
1	2	5
2	4	7
3	3	7
4	5	8
5	2	5
6	4	9
7	4	8
8	3	4
9	1	3
10	3	5
11	3	8
12	5	6

Tab. 6: Umfragewerte von Lena und Hasso. Sie bilden das approximative Modell der Population.

	A	B	C	D	E	F	G	H	I	J	K
1	Daten				Resampling				Wiederholte Korrelationen		
2											
3	Student	Motivation	Erfolg		Nummer	Motivation	Erfolg				
4	1	2	5		6	4	9		0.44		2.5% Quantil
5	2	4	7		12	5	6		0.7		0.28
6	3	3	7		4	5	8		0.66		
7	4	5	8		7	4	8		0.68		
8	5	2	5		7	4	8		0.55		97.5% Quantil
9	6	4	9		10	3	5		0.86		0.92
10	7	4	8		10	3	5		0.67		
11	8	3	4		11	3	8		0.62		
12	9	1	3		12	5	6		0.69		
13	10	3	5		4	5	8		0.77		
14	11	3	8		2	4	7		0.68		
15	12	5	6		1	2	5		0.35		
16									0.61		
17		Korrelation	0.69			Korrelation	0.44		0.8		
18									0.53		
19									0.77		

Tab. 7: Als approximatives Modell der Population werden die gepaarten Daten (Spalten B und C) verwendet. Daraus wird eine neue Stichprobe (Spalten F und G) und Korrelation (Zelle G17) simuliert. Das wird 1000 Mal wiederholt und die Korrelationen werden in Spalte I gespeichert.

für den Korrelationskoeffizienten zwischen Motivation und Studienerfolg.

Allerdings ist im Fall asymmetrisch verteilter Schätzer das Bootstrap-Perzentilintervall oft verzerrt, was zu falschen Überdeckungswahrscheinlichkeiten führt (siehe Efron and Tibshirani, S. 178 und Abbildung 2). Eine plausible Alternative im asymmetrischen Fall mit weniger Verzerrung ist das sogenannte klassische Bootstrap-Intervall *(basic bootstrap interval)*. Ausgangspunkt für das klassische Bootstrap-Intervall ist die Differenz zwischen Schätzwert r und Populationsparameter ϱ. Angenommen wir hätten kritische Werte c_1 und c_2, so dass gilt: $P(c_1 \leq r - \varrho \leq c_2) = 95\ \%$. Wegen

$$95\ \% = P(q(0{,}025) \leq r \leq q(0{,}975))$$

$$= P(q(0{,}025) - \varrho \leq r - \varrho \leq q(0{,}975) - \varrho))$$

sind die um ϱ verminderten Perzentile der Stichprobenverteilung von r die exakten Werte für b_1 und b_2. Einfaches Umstellen ergibt $P(r - c_2 \leq \varrho \leq r - c_1) = 95\ \%$, d. h. $[r - c_2; r - c_1]$ ist ein 95 % Konfidenzintervall für ϱ. Da die Verteilung von r aber nicht bekannt ist, ersetzen wir sie durch die Bootstrap-Verteilung, und die entsprechenden Perzentile werden – da ϱ nicht bekannt ist – um r vermindert, d. h. c_1 und c_2 werden ersetzt durch

$$\hat{c}_1 = \hat{q}(0{,}025) - r,\ \hat{c}_2 = \hat{q}(0{,}975) - r.$$

Dies führt schließlich zu

$$95\ \% = P(r - c_2 \leq \varrho \leq r - c_1)$$

$$\approx P(2r - \hat{q}(0{,}975) \leq \varrho \leq 2r - \hat{q}(0{,}0025))$$

$$= P(r - \hat{c}_2 \leq \varrho \leq r - \hat{c}_1).$$

Demnach ist das approximative 95 %-Konfidenzintervall $\tilde{I} = [r - \hat{c}_2; r - \hat{c}_1]$.

Im obigen Beispiel mit den Perzentilen der Bootstrapverteilung von 0,28 bzw. 0,92 ergibt sich für den Korrelationskoeffizienten zwischen Motivation und Studienerfolg [0,46; 1] als klassisches Bootstrap-Konfidenzintervall, wenn man noch berücksichtigt, dass Korrelationskoeffizienten nie größer als 1 werden können.

Visuell lassen sich die Überdeckungswahrscheinlichkeiten folgendermaßen veranschaulichen:

Ist ϱ der Populationskorrelationskoeffizient, dann ist der 95 % Schwankungsbereich $[\varrho - b_1, \varrho + b_2]$, wobei $b_1 = \varrho - q(0{,}025)$, $b_2 = q(0{,}975) - \varrho$.

Ersetzen wir ϱ durch r und die Perzentile durch die entsprechenden Perzentile der Bootstrap-Verteilung, so erhalten wir das Bootstrap-Perzentilintervall, d. h. $\hat{b}_1 = r - \hat{q}(0{,}025)$, $\hat{b}_2 = \hat{q}(0{,}975) - r$. Wie oben erläutert, ist das Intervall $[r - \hat{b}_1, r + \hat{b}_2]$ jedoch verzerrt, d. h. es hat nicht die korrekten Überdeckungswahrscheinlichkeiten. Ein Vergleich mit dem oben hergeleiteten klassischen Bootstrap-Intervall ergibt $\hat{b}_1 = -\hat{c}_1$, $\hat{b}_2 = \hat{c}_2$. Somit ist $[r - \hat{b}_2, r + \hat{b}_1]$ das klassische Bootstrap-Intervall.

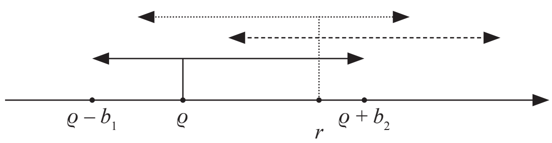


Abb. 2: Das durchgezogene Intervall ist der 95 % Schwankungsbereich $[\varrho - b_1; \varrho + b_2]$ der asymmetrisch verteilten Kennzahl ϱ. Eine Stichprobe liefert den Schätzwert r, welcher im 95 % Schwankungsbereich $[\varrho - b_1; \varrho + b_2]$ liegt. Das gestrichelte 95 % Bootstrap-Perzentilintervall $[r - b_1; r + b_2]$ enthält ϱ nicht, obwohl r im 95 % Schwankungsbereich $[\varrho - b_1; \varrho + b_2]$ liegt. Das gepunktete 95 % klassische Bootstrap-Intervall hingegen enthält ϱ.

5 Konvergenz der Methode

Bei der Bestimmung des 95 %-Schwankungsbereiches wurden die Perzentile durch Simulieren von 1000 Zufallsstichproben ermittelt. Dadurch ist eigentlich das hier eingeführte exakte 95 %-Konfidenzintervall – welches anhand der halben Breite $b = q(0{,}975) - \mu$ bestimmt wird – nur näherungsweise exakt. Jedoch hängt die Genauigkeit von b nur von dem Rechenaufwand ab, der durch Erhöhung der Anzahl der Simulationen beliebig erhöht werden kann. Beim Bestimmen des approximativen 95 %-Konfidenzintervalls erscheint eine weitere Fehlerquelle:

Das exakte Modell der Population wird angenähert durch das approximative Modell der Population, also die Daten. Die Qualität dieser Approximation hängt davon ab, wie weit die Daten repräsentativ für die Grundgesamtheit sind.

Die Daten bilden eine approximative Abbildung der Population. Mit zunehmendem Stichprobenumfang streben die relativen Häufigkeiten der Umfragewerte in der Zufallsstichprobe gegen die relativen Häufigkeiten der Population.

Dadurch strebt $\hat{b} = \hat{q}(0{,}975) - \bar{x}$ gegen $b = q(0{,}975) - \mu$ und somit strebt die Überdeckungshäufigkeit des approximativen 95 %-Konfidenzintervalls gegen 95 %.

Man kann zeigen, dass die Überdeckungshäufigkeit des approximativen $(100 - \alpha)$%-Konfidenzintervalls $1 - \alpha + C/\sqrt{n}$ beträgt, wobei C eine verteilungsabhängige Konstante ist und n den Stichprobenumfang bezeichnet (Hall 1992). Das Bootstrap-Verfahren ist nur so gut, wie es die Repräsentativität der vorliegenden Stichprobe erlaubt, weshalb es bei kleinen Stichproben weniger geeignet ist. Die vorangegangenen Beispiele mit $n = 20$ bzw. $n = 12$ dienten rein illustrativen Zwecken.

6 Verfeinerungen und Erweiterungen

Neben dem Bootstrap-Perzentilintervall und dem klassischen Bootstrap-Intervall gibt es noch viele weitere Varianten (z. B. studentized bootstrap, parametric bootstrap, bias corrected bootstrap, accelerated bootstrap) mit zum Teil deutlich verbesserten Konvergenzeigensschaften. Für Details muss auf die Literatur (z. B. Davison & Hinkely 1997, DiCiccio & Efron1996) verwiesen werden.

Es ist instruktiv, die Breite des Schwankungsbereiches und des Konfidenzintervalls in Abhängigkeit des Stichprobenumfangs und in Abhängigkeit des Konfidenzniveaus (z. B. mit Hilfe von Excel) zu untersuchen und die Ergebnisse dann anschaulich zu deuten. Die Idee, dass das Konfidenzintervall mit zunehmendem Konfidenzniveau breiter und mit zunehmendem Stichprobenumfang feiner wird, kann dann anschaulich interpretiert werden.

Im Rahmen eines statistischen Projektes können Studenten vielfältige Fragestellungen bearbeiten, indem sie zuerst eine Datenerhebung durchführen (z. B. durch eine Umfrage), anhand der beschreibenden Statistik verschiedene Kennwerte ausrechnen und anschließend anhand der hier präsentierten Methoden Konfidenzintervalle für die untersuchten Kennwerte bestimmen.

Literatur

Christie, D. (2004): Resampling mit Excel. In: *Stochastik in der Schule* 24(3), S. 22–27.

Cobb, G. (2007): The introductory statistics course: A Ptolemaic curriculum? In: *Technology Innovations in Statistics Education*, 1(1), S. 1–15.

Davison, A. C. & Hinkley, D. V. (1997): Bootstrap Methods and their Applicatins. Cambridge University Press.

DiCiccio, T. & Efron, B. (1996): Statistical Science 11 (3), 189–228.

Efron, B. (1979): Bootstrap methods: another look at the jackknife. In: *Annals of Statistics* 7(1), S. 1–26.

Efron, B. & Tibshirani, R. (1993): An Introduction to the Bootstrap. Chapman & Hall: London.

Engel, J. (2010): On teaching bootstrap confidence intervals. In: C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Engel, J. & Grübel, R. (2008): Bootstrap – oder die Kunst, sich selbst aus dem Sumpf zu ziehen. In: *Mathematische Semesterberichte*, 55, S. 113–130.

Erickson, T. (2002): Fifty Fathoms. Eeps-media: Oakland

Hall, P. (1992): The Bootstrap and Edgworth Expansion. Springer-Verlag: London, 1992.

Hesterberg, T. (2014): Bootstraping for learning statistics. In: K. Makar; B. de Sousa & R. Gould (Eds.). *Proceedings of the Ninth International Conference on Teaching Statistics*, Voorburg, The Netherlands: International Statistical Institute.

Johnson, R. W. (2001): An Introduction to the Bootstrap. In: *Teaching Statistics* 23(2), S. 49–54.

Pfannkuch, M. & Budgett, S. (2014): Constructing inferential concepts through bootstrap and randomizationtest simulations: a case study. In: K. Makar; B. de Sousa & R. Gould (Eds.). *Proceedings of the Ninth International Conference on Teaching Statistics*, Voorburg, The Netherlands: International Statistical Institute.

Watson, J. (2013): Resampling with Tinkerplots. In: *Teaching Statistics* 35(1), S. 32–36.

Anschrift der Verfasser

Andreas Kaufmann
Institut für Allgemeinbildende Fächer
Fachhochschule Westschweiz
CH-1040 Echallens
andreas.kaufmannn@telkel.ch

Joachim Engel
Institut für Mathematik und Informatik
Pädagogische Hochschule Ludwigsburg
Reuteallee 36
D-71634 Ludwigsburg
engel@ph-ludwigsburg.de