Population und Stichprobe

Inhaltsverzeichnis

Population und Stichprobe

Lernhinweise

Im folgenden Lernschritt werden Sie erfahren, wie anhand einer Stichprobe auf die unbekannten Kennwerte der zugrundeliegenden Population geschlossen werden kann. Weil eine Stichprobe im Gegensatz zu einer Vollerhebung per Definition immer nur eine Teilmenge aller interessierenden Untersuchungseinheiten darstellt, ist dieser Rückschluss auf die Population mit Unsicherheit behaftet. Die Verfahren der schliessenden Statistik erlauben allerdings die Quantifizierung dieser Unsicherheit. Dadurch ist es möglich, anhand von Stichprobendaten Aussagen über die Grundgesamtheit zu machen.

Benötigte Vorkenntnisse
Um diese Lektion erfolgreich bearbeiten zu können, sollten Sie Kenntnisse über die folgenden Themengebiete besitzen:

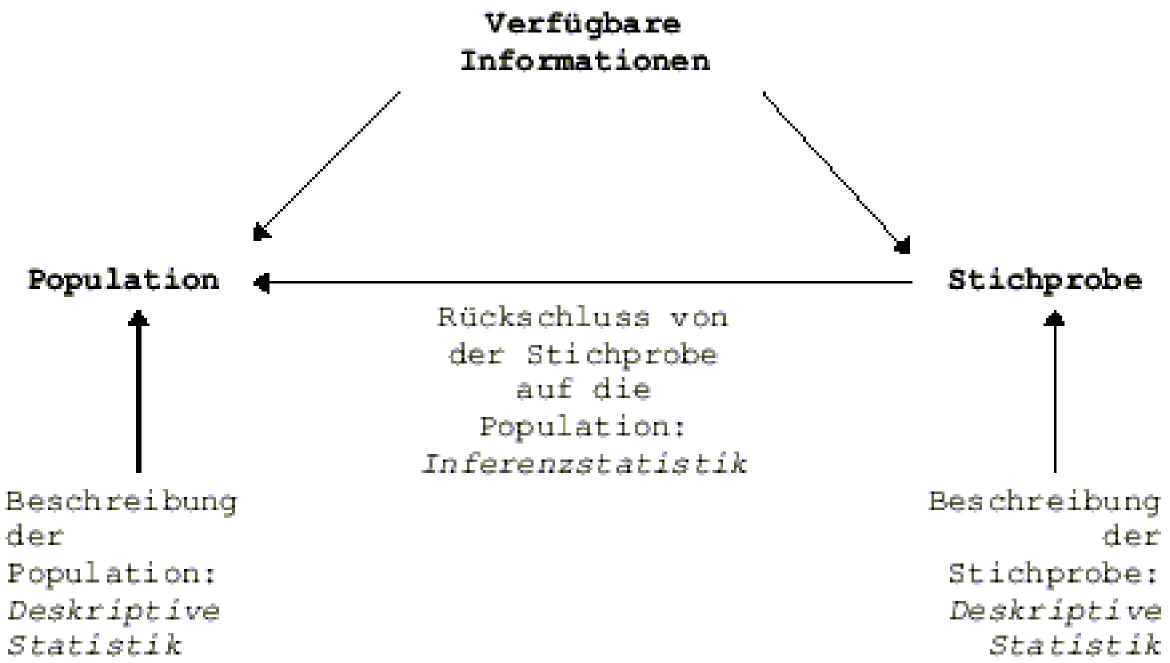
- Verfahren der deskriptiven Statistik (univariate Kennwerte)
- Die Grundlagen der Wahrscheinlichkeitstheorie.
- Kenntnisse über Zufallsvariablen und deren Parameter (insbesondere über standardnormalverteilte Zufallsvariablen).

Lernziele

- Sie verstehen die Logik hinter dem Rückschluss von der Stichprobe auf die Grundgesamtheit.
- Sie wissen, wie man unbekannte Populationsparameter aus Stichprobendaten schätzen kann und wie man deren Präzision quantifizieren kann.

Einführung

Nur sehr selten werden in sozialwissenschaftlichen Untersuchungen Daten zu sämtlichen interessierenden Untersuchungseinheiten erhoben. Zumeist beschränkt man sich in der Praxis darauf, die Informationen nur für einen Teil der Beobachtungen tatsächlich zu erheben. Diese Teilmenge an Untersuchungseinheiten bezeichnet man als eine Stichprobe.
Dennoch ist es zumeist das erklärte Ziel von empirischen Untersuchungen, die Ergebnisse über die konkrete Stichprobe hinaus auf die zugrundeliegende Gesamtheit an Beobachtungen zu verallgemeinern. Da über die Stichprobe immer nur ein Teil der interessierenden Elemente beobachtet wird, besteht Unsicherheit im Rückschluss von der Stichprobe auf die Grundgesamtheit (siehe Grafik unten).



In dieser Lektion wird die zentrale Frage der schliessenden Statistik behandelt: Wie kann anhand von Stichprobendaten auf die umfassendere Gesamtheit an Untersuchungseinheiten geschlossen werden? In dieser Lektion werden Sie die zugrundeliegende Logik hinter dem Rückschluss von der Stichprobe auf die Grundgesamtheit kennen lernen. Zudem werden Sie erfahren, wie sich die Unsicherheit der Stichprobenziehung quantifizieren lässt.

Theorie (1-13)

1. Fragestellung

Die zentrale Fragestellung, welche in dieser Lektion behandelt wird, lässt sich am besten anhand eines Beispieles illustrieren.
Stellen Sie sich die folgende Situation vor:
Sie werden als ForscherIn von der Universitätsleitung beauftragt, die durchschnittliche Wochenarbeitszeit von Studierenden an der Universität Zürich zu bestimmen. Zur Beantwortung der Fragestellung erhalten Sie allerdings aus Datenschutzgründen nur Zugang zu 500 zufällig ausgewählten Studierenden. Die Stichprobe, welche Ihnen zur Verfügung steht, umfasst somit gerade einmal 2,24 % aller Studierenden an der Universität Zürich (Im Wintersemester 2002/03 waren insgesamt 22'362 Studierende an der Universität Zürich immatrikuliert). Den Wert, welchen Sie aus der vorliegenden Stichprobe ermitteln können, ist offensichtlich von der Zusammensetzung der Stichprobe abhängig und ist nicht notwendigerweise mit dem Wert der Population identisch.
Es stellt sich somit die zentrale Frage, ob und mit welcher Genauigkeit anhand der Stichprobendaten auf die Wochenarbeitszeit aller Studierenden an der Universität insgesamt geschlossen werden kann: Man spricht auch vom Rückschluss von der Stichprobe auf die Grundgesamtheit, welcher der schliessenden Statistik gerade auch ihren Namen gibt.
In dieser Lektion werden Sie die Logik hinter dem Rückschluss von der Stichprobe auf die Grundgesamtheit kennen lernen. Sie werden lernen, wie man die mit der Stichprobenziehung verbundene Unsicherheit quantifizieren kann und damit ausgehend von einer Stichprobe Aussagen über die zugrundeliegende Gesamtheit an Beobachtungen machen kann.
Die enorme praktische Bedeutung dieser Fragestellung lässt sich an einigen wenigen Beispielen illustrieren:

- Im Vorfeld von Abstimmungen und Wahlen werden häufig Prognosen zum Ausgang präsentiert. Auch wenn Meinungsforschungsinstitute nur eine kleine Stichprobe aller Wahl- und Stimmberechtigten befragen, ist es offensichtlich das Ziel solcher Umfragen, Aussagen über die Gesamtheit aller Wahl- und Stimmberechtigten zu machen. Man ist am tatsächlichen Ausgang von Wahlen oder Abstimmungen interessiert, und diese hängen von den Entscheidungen aller Urnengänger ab.
- Zur Evaluation von wirtschafts- und sozialpolitischen Massnahmen werden immer häufiger empirische Studien durchgeführt, welche ebenfalls in aller Regel anhand von Stichprobendaten erarbeitet werden. Auch in diesem Falle spielt die Verallgemeinerung von Ergebnissen über die konkrete Stichprobe hinaus eine zentrale Rolle: Aussagen über die Wirksamkeit oder Unwirksamheit von politischen Massnahmen sollten sich auf die Gesamtheit der betroffenen Personen beziehen, und nicht alleine auf die untersuchte Stichprobe.

2. Definitionen und Notation

Bevor wir die Lösung des gestellten Problemes angehen, erscheint es als sinnvoll, zunächst einige wichtige Definitionen sowie die entsprechende Notation festzuhalten.

- **Statistische Einheit / Statistisches Element / Beobachtung:**
  Beobachtungen sind die "Träger" der letztlich interessierenden Grössen, der Merkmale / Variablen. In aller Regel handelt es sich dabei in sozialwissenschaftlichen Zusammenhängen um Individuen. Es kann sich dabei allerdings auch um Haushalte, Firmen, Gemeinden, etc. handeln
- **Grundgesamtheit / Population / Universum:**

Als Grundgesamtheit bezeichnet man die Menge aller im Hinblick auf die Fragestellung der Studie relevanten Beobachtungen. Je nach Fragestellung variiert somit die zugrundeliegende Population. Es ist deshalb notwendig, die Population nach verschiedenen und insbesondere nach inhaltlichen Abgrenzungskriterien eindeutig zu definieren.

Werden zu sämtlichen Beobachtungen einer Population Daten erhoben, spricht man von einer **Voll- oder Totalerhebung.**

Wird nur für eine Teilmenge der Grundgesamtheit Daten erhoben, spricht man von einer **Teil- oder Stichprobenerhebung**

- **Stichprobe:**

  Als Stichprobe bezeichnet man eine Teilmenge aus der Studienpopulation. Im Gegensatz zu einer Vollerhebung werden somit bei einer Stichprobenerhebung nicht für sämtliche Beobachtungen Daten erhoben.

  Im folgenden gehen wir von der Annahme aus, dass eine sogenannte **einfache Zufallsstichprobe** gezogen wird. Bei dieser Art der Stichprobenziehung hat jedes Element der Population dieselbe positive Wahrscheinlichkeit, Eingang in die konkrete Stichprobe zu finden. Das Ziehen einer einfachen Zufallsstichprobe entspricht dem "Urnenmodell" aus der Wahrscheinlichkeitstheorie: Für jedes einzelne Element liegt jeweils eine Kugel in der Urne, aus welcher zufällig die Beobachtungen gezogen werden.

Anmerkungen zur Notation:

Zur Unterscheidung von Populations- und Stichprobenkennwerten werden die ersten mit griechischen und die zweiten mit lateinischen Buchstaben bezeichnet. In statistischen Untersuchungen interessiert man sich sehr häufig für einzelne Momente der Verteilung eines bestimmten Merkmales. Die folgenden Populationskennwerte (arithmetisches Mittel, Varianz, Anteilswert), welche uns im folgenden weiter beschäftigen werden, haben Sie bereits im Rahmen der deskriptiven Statistik kennen gelernt:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

$$\pi = Pr(x_i = 1) = \frac{\sum_{i=1}^{N} x_i}{N}$$

In Analogie zu den Populationkennwerten lassen sich mit Stichprobendaten die folgenden Stichprobenkennwerte berechnen:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$p = Pr(x_i = 1) = \frac{\sum_{i=1}^{n} x_i}{n}$$

Zentral für das Verständnis dieser Lektion ist der folgende Sachverhalt: Populationskennwerte sind feste (d.h. nicht variable) Grössen, sie sind in aller Regel jedoch unbekannt, weil vollständige Informationen über die Population selten bis nie vorhanden sind. Stichprobenkennwerte hingegen sind als zufällige Grössen aufzufassen, da sie von Stichprobe zu Stichprobe unterschiedlich ausfallen können. Stichprobenkennwerte können damit als das Ergebnis eines Zufallsexperimentes (die Stichprobenziehung) aufgefasst werden. Stichprobenkennwerte sind somit nicht als feste Grössen zu behandeln, sondern als Realisierungen einer Zufallsvariablen.

3. "Dilemma" der Stichprobenziehung

Das "Dilemma" der Stichprobenziehung besteht darin, dass nicht nur die einhergehende Unsicherheit (aufgrund der Unvollständigkeit der Inormationen), sondern auch die Kosten relativ zu einer Vollerhebung gegeneinander abgewogen werden müssen. In aller Regel wird man in der Praxis zum Schluss kommen, dass die Informationsgewinnung mittels einer Stichprobe effizienter ist als die Durchführung einer Totalerhebung. Zum einen dürfte es offensichtlich sein, dass Stichprobenerhebungen, bei welchen in aller Regel nur einige hundert Personen befragt werden, erheblich schneller und billiger durchzuführen sind als Vollerhebungen. Die geringeren Kosten stellen den enormen Vorteil von Stichproben dar.
Demgegenüber besitzen nun aber Stichproben den Nachteil, dass sie immer nur unvollständiger Informationen über die Population liefern können, weil ja per Definition jeweils nur eine Teilmenge der Grundgesamtheit erhoben wird. Der Nachteil von Stichprobendaten ist somit ein Problem unvollständiger Daten. Daran schliesst sich nun die Frage an, welche Aussagen bezüglich der Population anhand von Stichprobendaten überhaupt möglich sind

Zusammenfassung: Vollerhebung versus Stichprobenerhebung

	Vorteil	Nachteil
Vollerhebung	Keine Informationsunsicherhei	Hohe Kosten
Stichprobenerhebung	Geringe Kosten	Informationsunsicherheit

Um diese zentrale Frage der schliessenden Statistik zu beantworten, bedienen wir uns eines Gedankenexperimentes. Wir werden sehen, dass glücklicherweise unter bestimmten Annahmen selbst bei kleinen Stichproben relativ präzise Aussagen über die Population möglich sind.

4. Gedankenexperiment

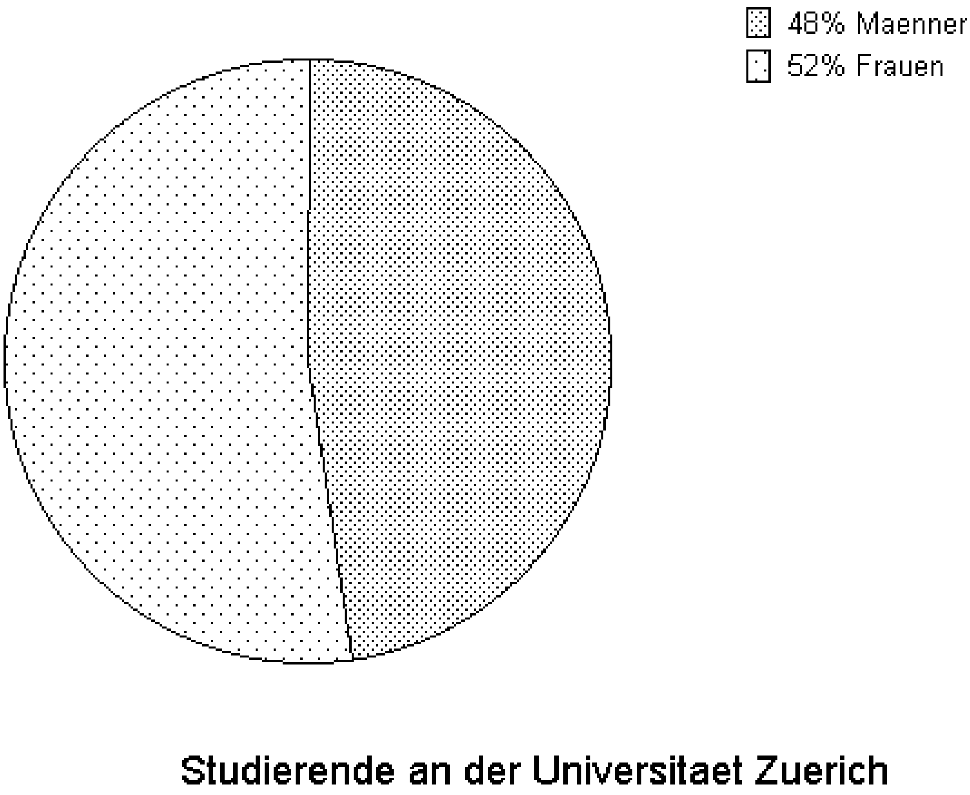
Welche Rückschlüsse lassen sich nun anhand von Stichprobendaten auf die zugrundeliegende Population ziehen? Der Schlüssel zum Verständnis dieses Rückschlüsses liegt in einem Gedankenexperiment.
Wir gehen zunächst von der kontrafaktischen Annahme aus, dass wir die Verteilung eines Merkmales (und damit auch alle seine Momente der Verteilung wie beispielsweise das arithmetische Mittel oder die Varianz) in der Population vom Umfang N kennen. Wir stellen uns nun die folgende Frage: Was passiert, wenn man aus dieser Population nun sehr viele Stichproben derselben Grösse zieht und für jede Stichprobe denselben Kennwert berechnet?
Wir ziehen nun also theoretisch unendlich viele Stichproben vom jeweils gleichen Umfang n. Für jede einzelne Stichprobe bildet man den interessierenden Kennwert (z.B. das arithmetische Mittel). Da sich bei jedem erneuten Ziehen einer Stichprobe die Zusammensetzung der Beobachtungen in aller Regel unterscheidet, erhält man für jede Stichprobe einen unterschiedlichen Stichprobenkennwert. Alle diese unterschiedlichen Stichprobenkennwerte bilden nun ihrerseits selbst eine Verteilung, die sogenannte Stichprobenkennwerteverteilung. Diese Verteilung bildet ab, wie der interessierende Kennwert nach der Zusammensetzung der Stichprobe variiert.
Falls sich nun diese Stichprobenkennwerteverteilung über eine Verteilung mit einer bekannten Wahrscheinlichkeitsdichtefunktion beschreiben lässt, lässt die Kenntnis dieser Verteilung folgendes zu:

- Es lässt sich eine Aussage darüber machen, mit welcher Wahrscheinlichkeit ein bestimmter Stichprobenkennwert anzutreffen ist.
- Die Kenntnis der Stichprobenkennwerteverteilung wird es erlauben, anhand eines einzelnen Stichprobenkennwertes Aussagen darüber zu machen, innerhalb welchen Wertebereiches der unbekannte Populationskennwert mit einer bestimmten Wahrscheinlichkeit zu erwarten ist.

Bevor wir also die praktisch relevante Frage beantworten können, welche Rückschlüsse sich anhand einer einzelnen Stichprobe auf die Population ziehen lassen, müssen wir die Frage nach der Stichprobenkennwerteverteilung beantworten.
Wir werden im folgenden den obigen Gedankengang für den Anteils- und den Mittelwert anhand von verschiedenen Computersimulationen illustrieren. Für andere Kennwerte existiert ebenfalls eine entsprechende Kennwerteverteilung, deren Herleitung auf demselben Gedanengang beruht.

5. Verteilung der Stichprobenanteilswerte

An dieser Stelle greifen wir noch einmal das Beispiel der Studierenden an der Universität Zürich auf, um das Konzept einer Stichprobenkennwerteverteilung anhand einer Computersimulation zu illustrieren.
Für die Population der Studierenden an der Universität Zürich ist die Geschlechterverteilung bekannt: Im Wintersemester 2002/03 waren von der insgesamt eingeschriebenen Studierenden 52,41% Frauen. Wir generieren also zunächst am Computer eine Modellpopulation, welche das entsprechende Geschlechterverhältnis an der Universität Zürich abbilden soll.



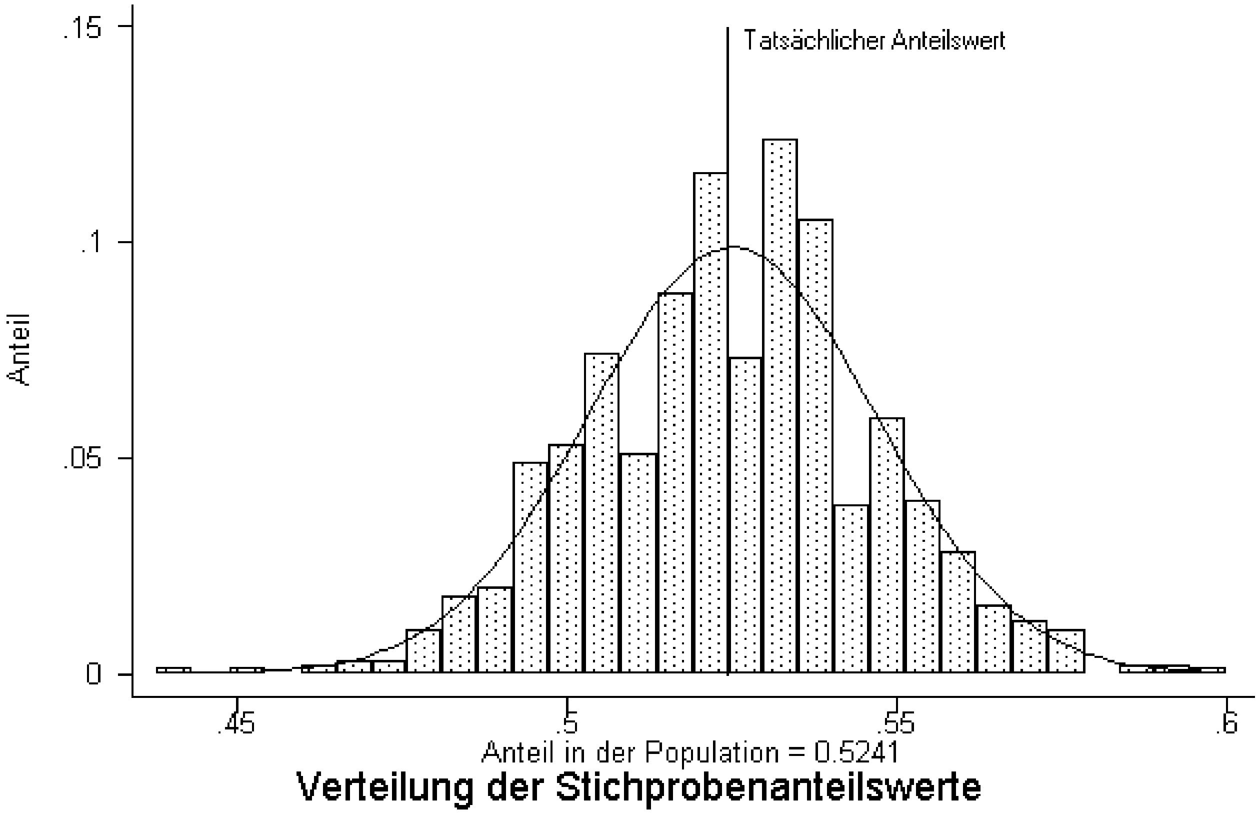
Zunächst soll gezeigt werden, dass aus unterschiedlichen Stichproben zum Teil sehr unterschiedliche Stichprobenanteilswerte resultieren. Aus den ersten fünf Zufallsstichproben vom Umfang n=500 resultierten die folgenden Werte für den geschätzten Frauenanteil:

Stichprobe	**Frauenanteil**
Nr.1	49,8 %
Nr.2	52,8 %
Nr.3	53,2 %
Nr.4	53,8 %
Nr.5	52,2 %

Zum Vergleich: Der tatsächliche Frauenanteil in der Population beträgt 52,41 %.

Stichprobe	**Frauenanteil**
Uni ZH	52,41 %

Aus der Tabelle ist ersichtlich, dass für die ersten fünf Zufallsstichproben jeweils ein anderer Anteilswert resultiert, und dass diese fünf Stichprobenanteilswerte unterschiedlich nahe am tatsächlichen Populationsanteilswert liegen.
Wir führen nun dieselbe Prozedur am Computer für 1'000 Zufallsstichproben mit jeweils einem Stichprobenumfang von n = 500 durch. Stellt man die Verteilung der 500 Stichprobenanteilswerte grafisch dar, erhält man die Stichprobenverteilung des Anteilswertes.



Aus der Grafik wird folgendes ersichtlich:

- Die Verteilung der unterschiedlichen Stichprobenanteilswerte weist annähernd die Form einer Normalverteilung auf, wie die hinterlegte Normalverteilungskurve andeutet.
- Die unterschiedlichen Stichprobenanteilswerte sind um den tatsächlichen Populationswert von 52,41 % zentriert.
- Stichprobenkennwerte, welche "nahe" am tatsächlichen Populationskennwert liegen, sind wahrscheinlicher als Stichprobenkennwerte, welche "nicht nahe" am wahren Kennwert liegen. Wie der Grafik zu entnehmen ist, treten Stichprobenkennwerte kleiner als 0.45 oder grösser als 0.60 beispielsweise nur sehr selten auf.

6. Verteilung von Stichprobenmittelwerten

In diesem Abschnitt führen wir dieselbe Simulation wie vorher für den Anteilswert für das arithmetische Mittel durch, und zwar für unterschiedliche Verteilungsformen des untersuchten Merkmales in der Population.
Dazu wurden wiederum am Computer fiktive Daten erstellt, welche unterschiedliche Verteilungsformen des Merkmales darstellen. Es wurde jeweils eine Modellpopulation von N = 100'000 erstellt. Betrachten Sie sich zuächst die unterschiedlichen Verteilungen des Merkmales in der Population.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Nun ziehen wir wiederum für jede Population jeweils 1'000 Zufallsstichproben vom Umfang n = 500 und berechnen für jede einzelne Stichprobe das arithemetische Mittel. So erhält man unterschiedliche Verteilungen der entsprechenden Stichprobenmittelwerte. Diese finden Sie in der unteren Grafik dargestellt.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Beachten Sie insbesondere, dass alle vier Stichprobenmittelwertverteilungen jeweils annähernd eine Normalverteilung aufweisen, und zwar erstaunlicherweise praktisch unabhängig von der Verteilungsform des Merkmales in der Population. Zudem gilt auch hier, dass die Verteilung der Stichprobenmittelwerte um den tatsächlichen Populationsmittelwert zentriert ist.
Diese Beobachtung führt uns nun zur Formulierung einer der bekanntesten und wichtigsten Gesetzmässigkeiten der Statistik.

7. Zentrales Grenzwerttheorem

Ohne mathematische Herleitung formulieren wir an dieser Stelle den bereits vorher beobachteten Sachverhalt: Die Stichprobenverteilung über theoretisch unendlich viele Stichproben von Mittelwerten als auch von Anteilswerten folgt einer Normalverteilung - und zwar praktisch unabhängig von der Verteilungsform des Merkmales in der Population - falls die Stichprobe genügend gross ist. Diese Gesetzmässigkeit nennt man auch **das zentrale Grenzwerttheorem.**

Arithmetisches Mittel:
Die Stichprobenverteilung der Mittelwerte weist den folgenden Erwartungswert auf:

$$E(\overline{x}_n) \quad = \quad \mu$$

Das heisst, dass der Erwartungswert der Stichprobenmittelwerte dem tatsächlichen Populationsmittelwert entspricht. Dies bedeutet, dass die Stichprobenmittelwerte am wahren Populationsmittelwert zentriert sind.
Für die Varianz und die Standardabweichung der Stichprobenmittelwerte ergeben sich die folgenden Formeln:

$$\begin{aligned} Var(\overline{x}_n) &= \frac{\sigma^2}{n} \\ SD(\overline{x}_n) &= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Die Standardabweichung der Stichprobenmittelwerte bezeichnet man auch als Standardfehler der Stichprobenmittelwerte. Der Standardfehler ist von zentraler Bedeutung, weil er ein Mass dafür ist, wie präzise ein Stichprobenmittelwert den unbekannten Populationsmittelwert schätzt. Aus obiger Formel ist ersichtlich, dass der Standardfehler lediglich von der Populationsvarianz sowie dem Stichprobenumfang abhängt. Insbesondere hat die Grösse der Population keinen Einfluss auf den Standardfehler.
Es gilt:

- Je grösser die Varianz des Merkmales in der Population, desto grösser ist der Standardfehler der Stichprobenmittelwerte. Bei konstanter Stichprobengrösse ist damit der Standardfehler umso grösser, je grösser die Populationsvarianz ist.
  Eine Stichprobe vom Umfang n schätzt damit den Populationsmittelwert umso besser, je weniger das Merkmal in der Population streut, d.h. je homogener die Beobachtungen in der Population sind.
- Je grösser der Stichprobenumfang, desto kleiner ist der Standardfehler. Auch dies sollte intuitiv sein, da mit steigendem Stichprobenumfang die Informationsunsicherheit über die Population reduziert wird. Im Extremfall von n = N wird der Standardfehler gleich Null, da in diesem Falle der Populationsmittelwert ohne Unsicherheit berechnet werden kann.

Anteilswert:
Für den Anteilswert erhält man analoge Ergebisse wie für das arithmetische Mittel. Der Erwartungswert der Stichprobenanteilswerte entspricht ebenfalls dem tatsächlichen Anteilswert in der Population:

$$E(p_n) = \pi$$

Die Varianz sowie die Standardabweichung der Stichprobenanteilswerte berechnen sich nach den folgenden Formeln:

$$Var(p_n) = \frac{\pi \cdot (1-\pi)}{n}$$
$$SD(p_n) = \sqrt{\frac{\pi \cdot (1-\pi)}{n}}$$

Auch hier gilt, dass der Standardfehler der Anteilswerte die Präzision umschreibt, mit welcher der Populationsanteilswert aus den Stichprobendaten geschätzt werden kann. Ebenso wie für den Mittelwert gilt auch hier, dass der Standardfehler nur von der Varianz der Merkmales in der Population sowie vom Stichprobenumfang abhängt.
Wiederum gilt:

- Je grösser die Varianz des Merkmales in der Population, desto grösser ist der Standardfehler des Anteilwertes. Dies bedeutet wiederum, dass eine Stichprobe vom Umfang n der Populationsanteilswert umso präziser schätzt, je homogener die Verteilung des Merkmales in der Population ist.
- Der Standardfehler sinkt mit steigendem Stichprobenumfang. Auch für den Anteilswert gilt somit, dass mit steigendem Stichprobenumfang der unbekannte Populationsanteilswert präziser geschätzt werden kann.

Exakte Verteilungsform der Stichprobenkennwerte:
Erwartungswert und Varianz bzw. Standardabweichung charakterisieren nun allerdings noch nicht die exakte Form der Verteilung der Stichprobenanteilswerte respektive der Stichprobenmittelwerte.
Gemäss dem zentralen Grenzwerttheorem weisen nun allerdings beide Verteilungen bei genügend grossem Stichprobenumfang eine Normalverteilung auf, welche durch die beiden besprochenen Grössen determiniert wird.
Damit kennen wir nun aber die exakte Verteilungsform sowohl der Stichprobenanteilswerte als auch der Stichprobenmittelwerte:

$$\begin{aligned} \overline{x}_n &\sim N\left(\mu, \frac{\sigma^2}{n}\right) \\ p_n &\sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right) \end{aligned}$$

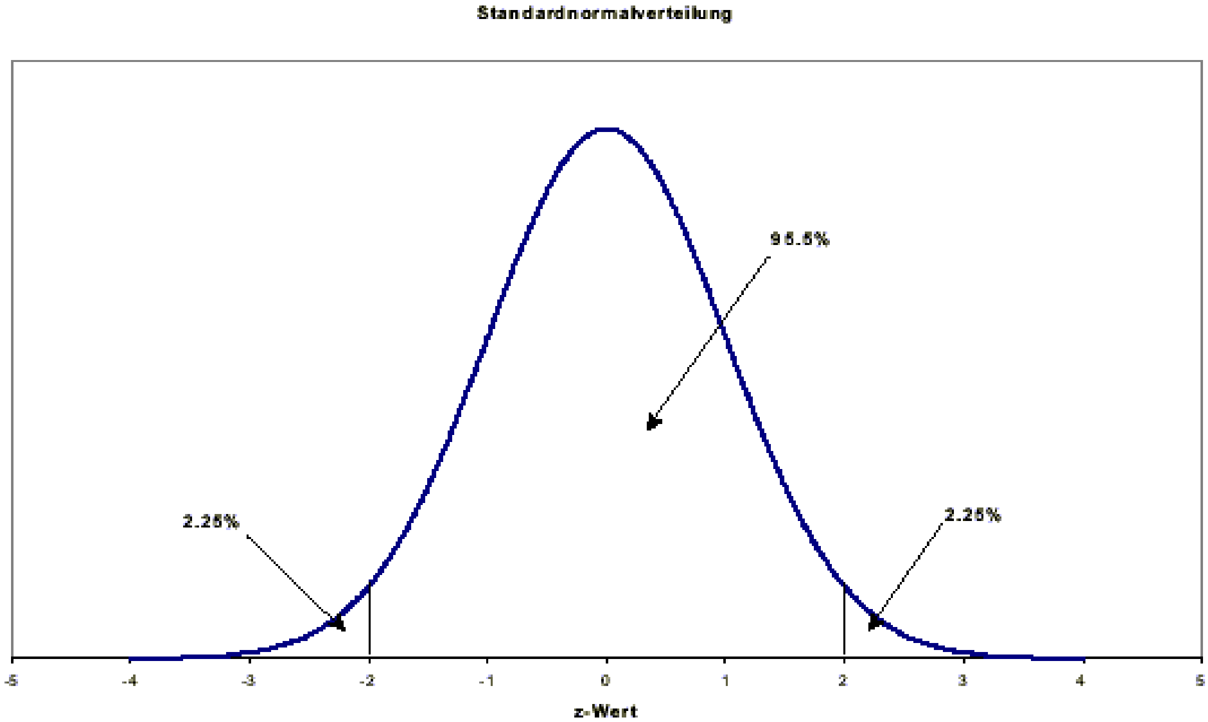
Inwiefern hilft uns nun aber diese Gesetzmässigkeit bei der Lösung des gestellten Problemes?
Wir machen uns zunächst die Tatsache zunutze, dass sich jede beliebige Normalverteilung in die Standardnormalverteilung überführen lässt. Dazu bilden wir die z-Werte unserer Stichprobenmittelwerte:

$$\frac{\overline{x}_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$$

Beziehungsweise für die Stichprobenanteilswerte:

$$\frac{p_n - \pi}{\left(\sqrt{\frac{\pi(1-\pi)}{n}}\right)} \sim N(0,1)$$

Die standardisierten Werte bringen somit zum Ausdruck, um wieviele Standardabweichungen ein Stichprobenmittelwert beziehungsweise ein Stichprobenanteilswert vom entsprechenden Populationskennwert entfernt ist.
Da wir nun die exakte Verteilungsform kennen, können wir eine Aussage darüber machen, wie wahrscheinlich oder eben unwahrscheinlich es ist, einen Stichprobenkennwert einer bestimmten Grösse zu beobachten.
Um beispielsweise die Wahrscheinlichkeit zu bestimmen, einen Stichprobenmittelwert zu beobachten, welcher 2 Standardabweichungen vom tatsächlichen Populationsmittelwert entfernt ist, muss man die Flächen links von z = -2 und die Fläche rechts von z = 2 unter der Standardnormalverteilung addieren.

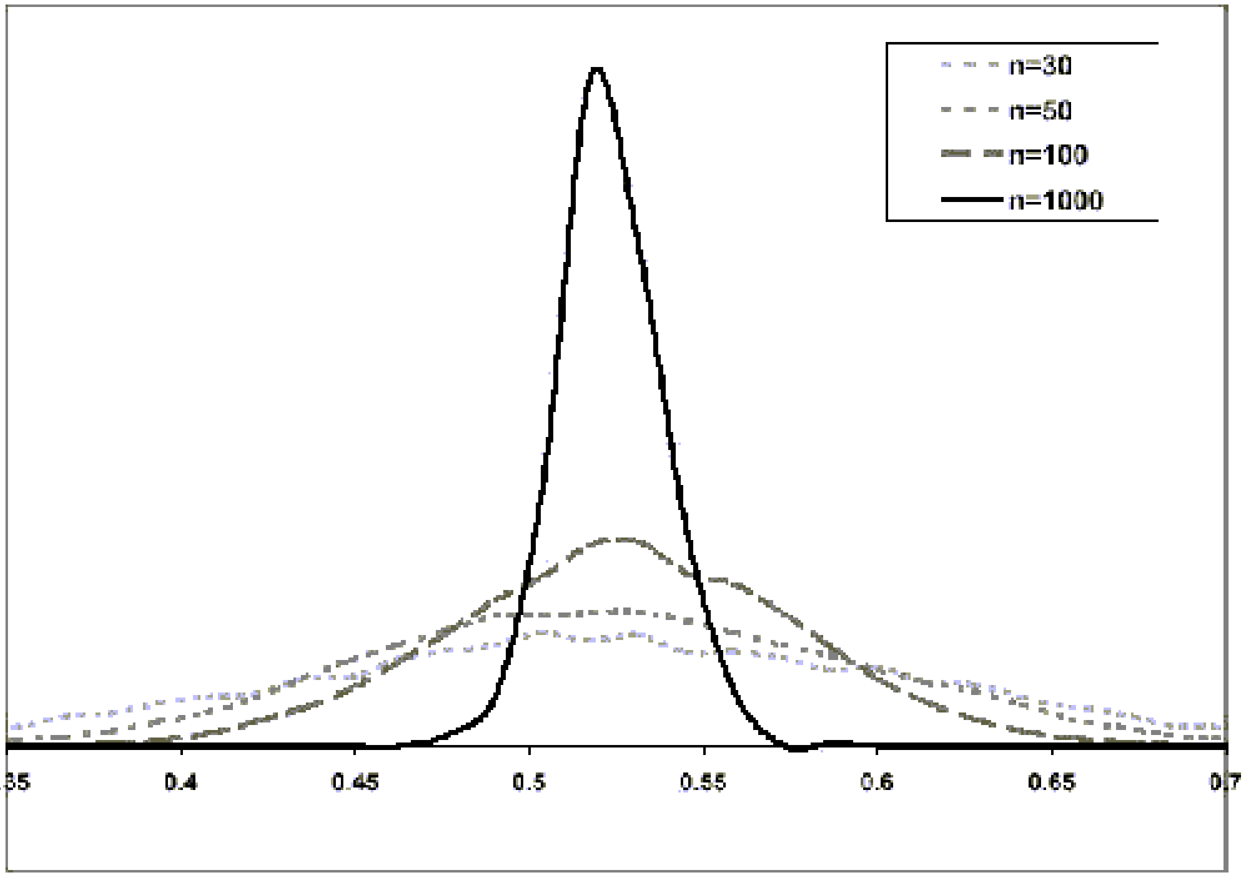


Sowohl die Fläche links als auch die Fläche rechts umfassen aufgrund der Symmetrie der Standardnormalverteilung jeweils ungefähr die Wahrscheinlichkeit von 2,5 %. Das heisst, die Wahrscheinlichkeit, dass sie einen Stichprobenmittelwert beobachten, welcher 2 oder mehr Standardabweichungen vom tatsächlichen Populationsmittelwert entfernt ist, beträgt nur ungefähr 5 %.

8. Einfluss der Stichprobengrösse auf den Standardfehler

Wie haben bereits gesehen, dass die Stichprobengrösse entscheidenden Einfluss auf den Standardfehler sowohl des Mittelwerts als auch des Anteilwertes ausübt. Der Einfluss der Stichprobengrösse soll nun anhand des Anteilwertes noch genauer illustriert werden, für den Mittelwert gelten allerdings analoge Ergebnisse.

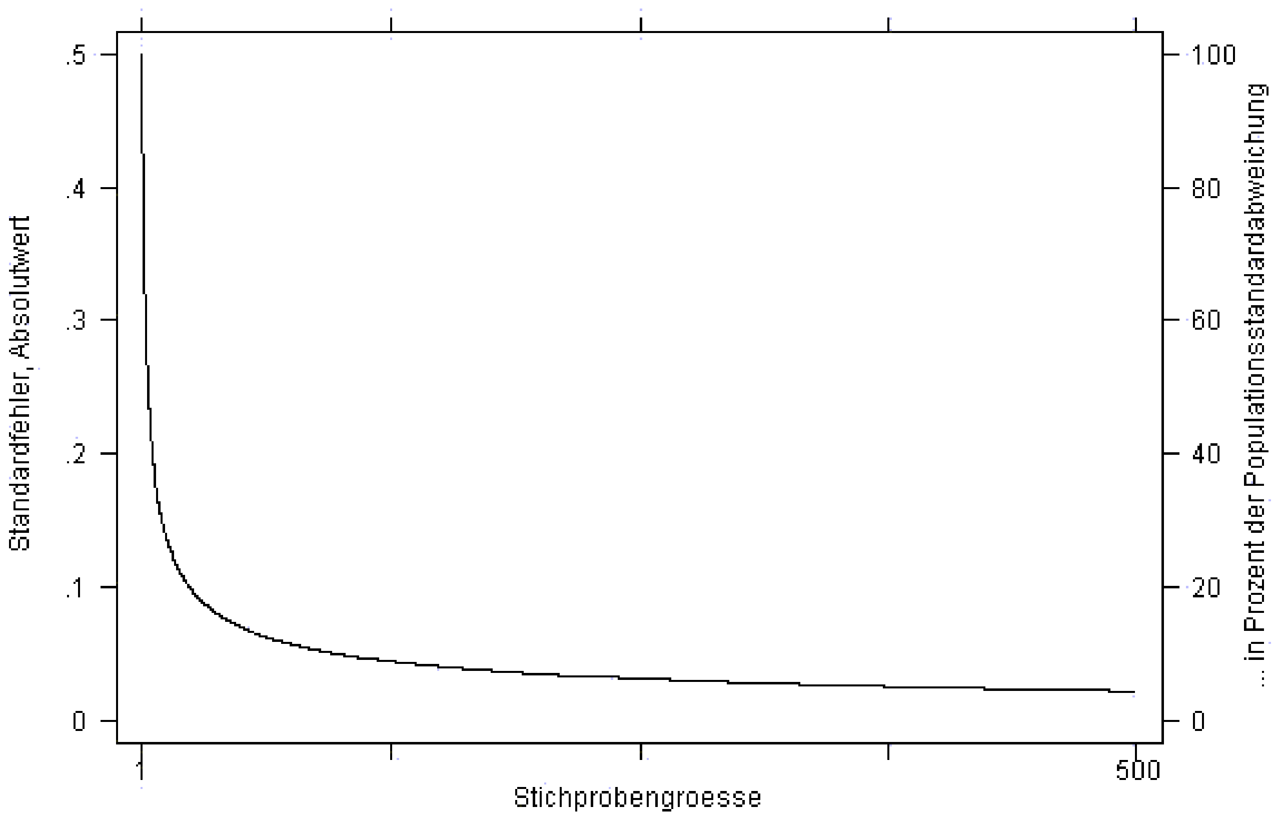
Der Einfluss der Stichprobengrösse auf den Standardfehler des Anteilwertes lässt sich wiederum am einfachsten über eine Computersimulation zeigen. Dazu ziehen wir aus der Modellpopulation, welche die Geschlechterverteilung der Studierenden an der Universität Zürich abbildet, jeweils 1'000 Zufallsstichproben für unterschiedliche Stichprobengrössen.



Anhand der Grafik ist deutlich zu erkennen, dass zwar alle Stichprobenverteilungen unabhängig vom Stichprobenumfang um den tatsächlichen Populationswert zentriert sind, dass aber mit steigendem Stichprobenumfang die Streuung der Stichprobenanteilswerte geringer wird. Mit steigendem n konzentriert sich somit die Stichprobenverteilung immer näher am wahren Anteilswert.
Der Zusammenhang zwischen der Stichprobengrösse und dem Standardfehler lässt sich auch noch auf eine andere Art und Weise illustrieren. Der Standardfehler des Anteilwertes ist gegeben durch die folgenden Grösse:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \frac{\sigma_\pi}{\sqrt{n}}$$

Die Populationsvarianz bzw. die Populationsstandardabweichung ist eine Konstante und lässt sich durch den Forscher in aller Regel nicht beeinflussen. Der Standardfehler variiert somit umgekehrt proportional zur Wurzel aus n. Der Zusammenhang zwischen dem Standardfehler und der Stichprobengrösse für das Beispiel der Geschlechterverteilung ist unten grafisch dargestellt:



Zum einen ist der negative Zusammenhang zwischen der Stichprobengrösse und dem Standardfehler klar ersichtlich. Zum anderen zeigt sich, dass der Zusammenhang zwar monoton, aber nicht linear ist: Der Standardfehler "reagiert" zunächst sehr stark auf eine Vergrösserung der Stichprobe. Ist die Stichprobe bereits relativ gross, lässt sich durch eine weitere Vergrösserung der Stichprobe der Standardfehler nicht mehr markant verkleinern.
Aus diesem Ergebnis lässt sich der folgende Schluss ziehen: Will man als sozialwissenschafliche ForscherIn den Standardfehler reduzieren (und damit die Unsicherheit in Bezug auf den gesuchten Populationskennwert), sollte man um eine möglichst grosse Stichprobe besorgt sein, da man auf die Varianz des Merkmales in der Population keinen Einfluss nehmen kann.
Allerdings ist eine grosse Stichprobe alleine noch keine Garantie für eine präzise Schätzung des Populationskennwertes. Diese Präzision hängt ebenfalls entscheidend von der Art und Weise ab, wie die Stichprobe gezogen wurde. Weitere Informationen dazu finden Sie in einem separaten Lernschritt zur Stichprobenziehung.

9. Mutungsintervall für Stichprobenmittelwerte und -anteilswerte

Das zentrale Grenzwerttheorem ermöglicht es uns nun, Aussagen darüber zu formulieren, mit welcher Wahrscheinlichkeit ein Stichprobenkennwert innerhalb eines zentralen Wertebereiches zu erwarten ist, wenn aus der Population eine einzelne Stichprobe der Grösse n gezogen wird. Wir interessieren uns nun für die Wahrscheinlichkeit, dass ein (standardisierter) Stichprobenkennwert einer Zufallsstichprobe um + / - z Standardabweichungen vom tatsächlichen Populationskennwert entfernt ist. Wir stellen uns hier also die Frage,

wie wahrscheinlich bzw. wie unwahrscheinlich konkrete Realisierungen von Stichprobenkennwerten sind, wenn wir eine einzelne Zufallsstichprobe aus einer bekannten Population ziehen (d.h. die Verteilung des Merkmales in der Population ist bekannt).
Einen solchen zentralen Wertebereich, innerhalb dessen der Stichprobenkennwert mit einer bestimmten Wahrscheinlichkeit zu erwarten ist, nennt man Mutungs- oder Prognoseintervall.
Ein Mutungsintervall für Stichprobenmittelwerte berechnet sich als:

$$Pr\left(-z_{1-\alpha/2} \leq \frac{\overline{x}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}\right) = 1-\alpha$$

Ein Mutungsintervall für Stichprobenanteilswerte berechnet sich als:

$$Pr\left(-z_{1-\alpha/2} \leq \frac{p_n - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \leq z_{1-\alpha/2}\right) = 1-\alpha$$

Mutungsintervalle besitzen somit folgende Interpretation: Beim Ziehen einer Zufallsstichprobe aus der Population beträgt die Wahrscheinlichkeit, dass der daraus berechnete Stichprobenkennwert innerhalb des Intervalles auf der linken Seite der Gleichung liegt, 1-alpha.

Vorgehensweise zur Bestimmung von Mutungsintervallen:

- Bestimmen Sie die gewünschte Wahrscheinlichkeit, mit welcher der Stichprobenkennwert innerhalb des Intervalles zu erwarten ist. Möchte man beispielsweise wissen, wie gross das Intervall ist, innerhalb dessen der Stichprobenkennwert mit einer Wahrscheinlichkeit von 90 % zu erwarten ist, ergeben sich die folgenden Grössen:

$$(1-\alpha) = 0.9$$
$$\alpha = 0.1$$

- Bestimmen Sie nun:

$$(1-\alpha/2) = 0.95$$

- Suchen Sie für diesen Wert den entsprechenden z-Wert der Standardnormalverteilung aus einer statistischen Tabelle oder bestimmen Sie diesen Wert am Computer.
  Für unser Beispiel sollten Sie den Wert 1,645 erhalten.
- Sie kennen nun bereits das zentrale Mutungsintervall.
  Für das Beispiel: 90 % aller z-transformierten Stichprobenkennwerte beim wiederholten zufälligen Ziehen einer Stichprobe werden sich im Intervall von + / 1,645 Standardabweichungen vom tatsächlichen Populationskennwert entfernt befinden.

Die z-Werte zu den wichtigsten Mutungsintervallen:

- 90 % aller Beobachtungen einer standardnormalverteilten Zufallsvariable liegen im Wertebereich von + / - 1,645.
- 95 % aller Beobachtungen einer standardnormalverteilten Zufallsvariable liegen im Wertebereich von + / - 1,96.
- 99 % aller Beobachtungen einer standardnormalverteilten Zufallsvariable liegen im Wertebereich von + / - 2,576

Grafische Illustration von Mutungsintervallen:
Das Konzept des Mutungsintervalles lässt sich grafisch illustrieren. Dazu wurde wiederum eine standardnormalverteilte Verteilung für eine Population von N = 10'000 am Computer generiert. Anschliessend wurden 100 Zufallsstichproben vom Umfang n = 500 aus dieser Population gezogen und der jeweilige Stichprobenmittelwert berechnet. In der folgenden Grafik sind die 100 Stichprobenmittelwerte und verschiedene Mutungsintervalle dargestellt.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

10 Punktschätzung für unbekannte Populationskennwerte

Da wir nun wissen, wie sich die Stichprobenkennwerte um den wahren Populationskennwert verteilen, können wir uns der praktisch relevanten Frage zuwenden, wie anhand einer einzelnen Stichprobe auf den Populationkennwert geschlossen werden kann, wenn dieser unbekannt ist (Zur Erinnerung: Bei der Berechnung von Mutungsintervallen geht man davon aus, dass die Populationskennwerte bekannt sind).
Zunächst stellt sich die Frage, wie der gewünschte Populationskennwert am besten aus den vorliegenden Stichprobendaten geschätzt werden kann. Solche Schätzungen für den Populationskennwert selbst nennt man auch Punktschätzungen.
Aus der Formulierung des zentralen Grenzwerttheorems wissen wir bereits, dass die Erwartungswerte der Stichprobenverteilung sowohl des Anteilwertes als auch des arithmetischen Mittels dem jeweiligen Populationskennwert entsprechen. Die besten Punktschätzer für diese beiden Kennwerte sind somit gegeben durch:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

$$\hat{\pi} = \frac{\sum_{i=1}^{n} x_i}{n} = p$$

Die Varianz und die Standardabweichung des interessierenden Merkmales lässt sich anhand der folgenden Formeln schätzen:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = s^2$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}(x_i - \overline{x})^2} = \sqrt{s^2} = s$$

Punktschätzungen liefern uns zwar die besten Schätzungen für die unbekannten Populationskennwerte, sie geben uns aber keine Auskunft darüber, wie präzise diese Schätzungen sind.

11. Vertrauensintervalle, falls die Populationsvarianz bekannt ist

Punktschätzungen für Populationsparameter liefern uns die bestmögliche Schätzung für den unbekannten Populationskennwert, wenn dieser aus Stichprobendaten geschätzt werden soll. Das Problem an Punktschätzungen ist allerdings, dass die mit der Stichprobenziehung verbundene Unsicherheit nicht berücksichtigt wird.
Wir wissen allerdings bereits, dass der Standardfehler eines Stichprobenkennwertes Auskunft über die Präzision gibt, mit welcher dieser Kennwert geschätzt werden kann. Mit dieser zusätzlichen Information lassen sich nun sogenannte Konfidenz- oder Vertrauensintervalle berechnen, welche im Gegensatz zu Punktschätzungen Auskunft über die Unsicherheit bezüglich dem Populationskennwert zum Ausdruck bringen. Vertrauensintervalle lassen sich durch einfaches Umformen aus den entsprechenden Mutungsintervallen herleiten. Für den Stichprobenmittelwert ergibt sich folgende Formel für das Vertrauensintervall:

$$\begin{aligned} 1-\alpha &= Pr\left(-z_{1-\alpha/2} \leq \frac{\overline{x}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}\right) \\ &= Pr\left(-z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{x}_n - \mu \leq z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \\ &= Pr\left(\overline{x}_n - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x}_n + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

Für den Stichprobenanteilswert ergibt sich analog folgende Berechnung des Vertrauensintervalles:

$$\begin{aligned}
1-\alpha &= Pr\left(-z_{1-\alpha/2} \leq \frac{p_n - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \leq z_{1-\alpha/2}\right) \\
&= Pr\left(-z_{1-\alpha/2} \cdot \sqrt{\frac{\pi(1-\pi)}{n}} \leq p_n - \pi \leq z_{1-\alpha/2} \cdot \sqrt{\frac{\pi(1-\pi)}{n}}\right) \\
&= Pr\left(p_n - z_{1-\alpha/2} \cdot \sqrt{\frac{\pi(1-\pi)}{n}} \leq \pi \leq p_n + z_{1-\alpha/2} \cdot \sqrt{\frac{\pi(1-\pi)}{n}}\right)
\end{aligned}$$

Wie lassen sich nun Vertrauensintervalle interpretieren? Vertrauensintervalle sind so konstruiert, dass sie mit einer vorgegebenen Wahrscheinlichkeit den tatsächlichen Populationskennwert umschliessen. Anders formuliert: Der unbekannte Populationskennwert befindet sich mit der vorgegebenen Wahrscheinlichkeit innerhalb des berechneten Vertrauensintervalles.
Die Breite des Vertrauensintervalles drückt aus, mit welcher Genauigkeit anhand der Stichprobe auf den Populationskennwert geschlossen werden kann. Die Breite des Konfidenzintervalles für den Stichprobenmittelwert bzw. -anteilwert ist gegeben durch den folgenden Ausdruck:

$$\begin{aligned}
b_{VI} &= 2z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\
b_{VI} &= 2z_{1-\alpha/2} \cdot \sqrt{\frac{\pi(1-\pi)}{n}}
\end{aligned}$$

Anhand der obigen Formeln ist ersichtlich, welche Grössen die Breite des Vertrauensintervalles beeinflussen:

- Die Wahrscheinlichkeit, mit welcher das Vertrauensintervall den unbekannten Populationskennwert umschliessen soll. Je grösser diese Wahrscheinlichkeit sein soll, desto breiter wird das entsprechende Vertrauensintervall. Wollte man etwa ein Vertrauensintervall bilden, welches den wahren Populationskennwert mit 100% Wahrscheinlichkeit umschliesst, erhält man ein Vertrauensintervall von unendlicher Breite!
- Der Standardfehler des Stichprobenkennwertes. Je kleiner der Standardfehler des Stichprobenkennwertes, desto kürzer wird das Vertrauensintervall. Hier fliesst somit direkt die Präzision ein, mit welcher der Kennwert geschätzt werden kann.

12. Vertrauensintervalle, falls die Populationsvarianz unbekannt ist

Es besteht nun allerdings das Problem, dass in aller Regel die Populationsstandardabweichung, welche zur Berechnung des Vertrauensintervalles notwendig ist, unbekannt ist und deshalb ebenfalls anhand der Stichprobendaten geschätzt werden muss.
Man ersetzt deshalb die Populationsstandardabweichung durch die entsprechende Stichprobenstandardabweichung:

$$\hat{\sigma} \quad = \quad \sqrt{\frac{1}{n-1}(x_i - \overline{x})^2} = \sqrt{s^2} = s$$

Aufgrund der Tatsache, dass die Populationsstandardabweichung aus den vorliegenden Stichprobendaten geschätzt werden muss, gilt, dass die Stichprobenkennwerte nun nicht mehr exakt, sondern nur noch approximativ normalverteilt sind. Diese Approximation gilt ebenfalls nur bei genügend grossem Stichprobenumfang (was als "genügend grosser" Stichprobenumfang gilt, hängt von der Verteilung des Merkmales in der Population ab. Für einen Stichprobenumfang von mehr als 30 Beobachtungen ist die Approximation für praktische Zwecke häufig genügend):

$$\begin{aligned} \overline{x}_n &\approx N\left(\mu, \frac{s^2}{n}\right) \\ p_n &\approx N\left(\pi, \frac{p(1-p)}{n}\right) \end{aligned}$$

Weil bei grossem Stichprobenumfang die Normalverteilung nach wie vor eine gute Approximation für die Verteilung der Stichprobenkennwerte darstellt, ändert sich nichts Grundlegendes an der Berechnung von Vertrauensintervallen. Man ersetzt nun lediglich die Populationsstandardabweichung durch die entsprechende Schätzung aus den Stichprobendaten. Für den Stichprobenmittelwert ergibt sich somit:

$$1-\alpha \quad \approx \quad Pr\left(\overline{x}_n - z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \overline{x}_n + z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

Und für den Stichprobenanteilswert:

$$1-\alpha \quad \approx \quad Pr\left(p_n - z_{1-\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p_n + z_{1-\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right)$$

Vertrauensintervalle lassen sich ebenfalls grafisch darstellen. Dazu wurden in der unteren Darstellung dieselben Stichproben verwendet wie zur Darstellung von Mutungsintervallen.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

13. Zusammenfassung zum Lernschritt

In dieser Lektion wurden die folgenden Punkte besprochen:

- Die praktische Notwendigkeit von Stichproben und deren Problem der Informationsunsicherheit bezüglich dem Rückschluss auf die Population.
- Die Logik hinter dem Rückschluss von der Stichprobe auf die Population:

- Stichprobenkennwerteverteilungen geben Auskunft darüber, welche Stichprobenkennwerte beim wiederholten Ziehen einer Stichprobe aus derselben Population zu erwarten sind.
- Das zentrale Grenzwerttheorem beschreibt die Verteilung von Stichprobenkennwerten. Bei genügend grossem Stichprobenumfang lassen sich die Stichprobenverteilungen von Mittelwert und Anteilswert über die Normalverteilung beschreiben.
- Darauf aufbauend geben Mutungsintervalle darüber Auskunft, wie wahrscheinlich bzw. unwahrscheinlich bestimmte Realisierungen von Stichprobenkennwerten sind, wenn die Population bekannt ist.
- Basierend auf dem zentralen Grenzwerttheorem haben wir Punktschätzungen für Populationskennwerte kennen gelernt.
- Vertrauensintervalle berücksichtigen die mit der Stichprobenziehung verbundene Unsicherheit. Sie geben Auskunft darüber, in welchem Wertebereich der unbekannte Populationskennwert mit einer bestimmten Wahrscheinlichkeit zu erwarten ist.

Fallbeispiel

Das GfS-Forschungsinstitut führt im Auftrag des Schweizer Fernsehens vor den anstehenden Nationalratswahlen 2003 das "SRG SSR Wahlbarometer 03" durch
Aus der Population aller Wahlberechtigten wurden in der fünften Welle des Wahlbarometers im Juni rund 2000 Personen nach ihrer Wahlabsicht befragt: "Wenn am nächsten Sonntag schon Nationalratswalen wären, welcher Partei würden Sie heute Ihre Stimme hauptsächliche abgeben?"
Aus den Angaben der 1015 Personen, welche auch tatsächlich an der Wahl teilzunehmen gedenken, ergeben sich die folgenden Parteienstärken (kleinere Parteien sind nicht aufgeführt):

Partei	Stimmenanteil
SVP	26%
SP	23%
FDP	19%
CVP	15%

In der obigen Tabelle sind bereits die Punktschätzungen für die vier Bundesratsparteien dargestellt. Wir wissen nun aber, dass diese Schätzungen mit Unsicherheit behaftet sind und dass die tatsächlichen Anteilswerte in der Population von diesen Schätzungen abweichen können. Wir wollen uns insbesondere fragen, welchen Einfluss diese Unsicherheit auf die Rangfolge der Parteienstärke haben könnte.
Um diese Frage zu beantworten, bilden wir für alle vier Parteien ein 95%-Vertrauensintervall für den geschätzten Stimmenanteil. Da eine grosse Stichprobe (n = 1015) vorliegt, können wir davon ausgehen, dass die Stichprobenanteilswerte approximativ normalverteilt sind:

$$\frac{p_n - \pi}{\left(\sqrt{\frac{\pi(1-\pi)}{n}}\right)} \sim N(0,1)$$

Ein 95%-Vertrauensintervall ergibt sich demzufolge über:

$$1-\alpha \;\approx\; Pr\left(p_n - z_{1-\alpha/2}\cdot\sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p_n + z_{1-\alpha/2}\cdot\sqrt{\frac{p(1-p)}{n}}\right)$$

Nach dieser Berechnung ergeben sich somit die folgenden Vertrauensintervalle für die vier Parteien:

Partei	**95%-Vertrauensintervall**
SVP	[.2330, .2870]
SP	[.2041, .2559]
FDP	[.1658, .2142]
CVP	[.1280, .1720]

Während die Punktschätzungen eine klare Rangfolge der Parteien suggerieren, zeigen die entsprechenden Intervallschätzungen, dass die Rangfolge durchaus auch anders aussehen könnte bei der gegebenen Stichprobengrösse:

- Die SP könnte einen höheren Stimmenanteil aufweisen als die SVP, die FDP einen höheren als die SP und die CVP einen höheren als die FDP
- Eher auszuschliessen ist allerdings, dass die FDP einen höheren Stimmenanteil gewinnt als die SVP und dass die CVP einen höheren Stimmenanteil gewinnt als die SP oder die SVP.