

Stichprobenziehung

Inhaltsverzeichnis

Stichprobenziehung	2
Lernhinweise	2
Einführung	2
Theorie (1-8)	3
1. Fragestellung	4
2. Vor- und Nachteile der Stichprobenziehung	4
3. Definition der Population	5
4. Kategorisierung von Stichprobenverfahren	6
5. Probabilistische Stichprobenverfahren	7
Einfache Zufallsstichprobe	
Systematische Zufallsstichprobe	
Geschichtete Zufallsstichprobe	
Klumpenstichprobe	
Zwei- und mehrstufige Stichprobenverfahren	
6. Nichtprobabilistische Stichprobenverfahren	12
Willkürliche Stichprobe	
Bewusste Stichprobe	
Quotenstichprobe	
7. Stichprobenziehung - Fehlerquellen	13
Nichtstichprobenfehler	
Stichprobenfehler	
Variabilität der Stichprobenmittelwerte	
Totaler Stichprobenfehler	
Nonresponse	
Nonresponse, unsystematisch	
Nonresponse, systematisch	
8. Zusammenfassung zum Lernschritt	19
Fallbeispiel	19

Stichprobenziehung



Hier können Sie eine PDF-Version dieser Lerneinheit herunterladen.

Lernhinweise

Im folgenden Lernschritt werden Sie verschiedene Verfahren der Stichprobenziehung und mögliche Fehlerquellen beim Rückschluss von der Stichprobe auf die Population kennenlernen

Lernziele

- Sie kennen unterschiedliche Verfahren der Stichprobenziehung, deren Vor- und Nachteile sowie deren Anwendungsmöglichkeiten.
- Sie kennen die potentiellen Fehlerquellen von Stichprobendaten im Hinblick auf inferenzstatistische Aussagen.

Benötigte Vorkenntnisse

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

- Stichprobe und Grundgesamtheit
- Was heisst "Repräsentativität"?

Hinweise zur Bearbeitung

Beim Anklicken des "next" - Buttons werden Sie nach der oben dargestellten Rubriken-Reihenfolge durch die Lektion geführt: (1) Lernhinweis, (2) Einführung, (3) Theorie, (4) Fallbeispiel, (5) Lernkontrolle.

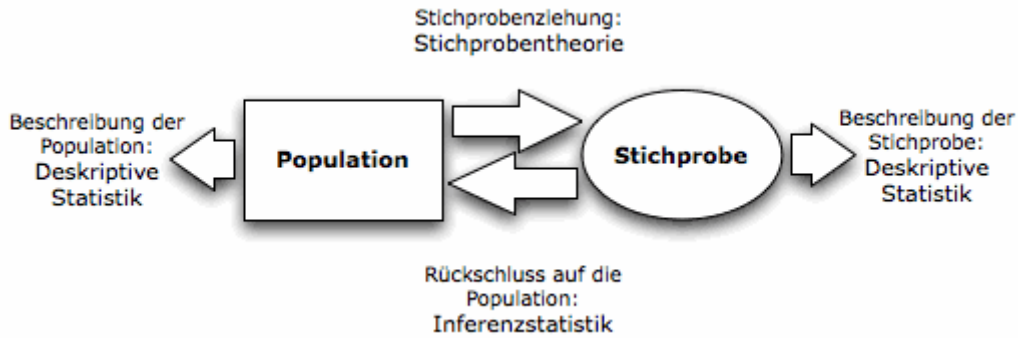
Einführung

Das Arbeiten mit Stichprobendaten ist zentraler Bestandteil sozialwissenschaftlicher Untersuchungen, da aus Zeit- sowie aus Kostengründen praktisch nie Vollerhebungen durchgeführt werden (können). Dies führt zu einem Informationsproblem, sofern die Aussagen, welche über die Stichprobendaten gewonnen werden können, auf die zugrundeliegende Grundgesamtheit an Beobachtungen verallgemeinert werden sollen.

In diesem Lernschritt lernen Sie unterschiedliche Verfahren der Stichprobenziehung kennen. Die Darstellung der Verfahren beschränkt sich allerdings auf eine einführende Beschreibung der verschiedenen Verfahren, ihre Vor- und Nachteile sowie ihre Anwendungsmöglichkeiten anhand verschiedener Beispiele.

Anschliessend werden verschiedene Fehlerquellen vorgestellt, welche den Rückschluss von der Stichprobe auf die Population beeinflussen oder gar verunmöglichen können.

Bevor wir uns verschiedenen Arten der Stichprobenziehung zuwenden, soll zunächst dargelegt werden, wie die Stichprobenziehung in die Verfahren der Statistik eingeordnet werden soll (siehe Grafik unten).



Aus der Grafik wird ersichtlich, dass es im Hinblick auf die Konzepte von Stichprobe und Population ausreicht, die Verfahren der Statistik grob in drei Kategorien einzuteilen:

- **Die beschreibende (deskriptive) Statistik** dient der Beschreibung sowohl von Populations- als auch von Stichprobendaten:
 - Liegt eine Vollerhebung vor, genügen die Verfahren der beschreibenden Statistik oftmals aus, da per Definition einer Vollerhebung keine Informationsunsicherheit besteht. Es sind in diesem Falle keine inferenzstatistischen Verfahren notwendig.
 - Liegt lediglich eine Stichprobe vor, so bilden die Verfahren der beschreibenden Statistik meistens den Ausgangspunkt einer empirischen Analyse. Diese Ergebnisse beziehen sich allerdings immer nur auf die konkret vorliegende Stichprobe. Die Verfahren der beschreibenden Statistik lassen keinen Rückschluss auf die Grundgesamtheit zu!
- **Die schliessende Statistik (Inferenzstatistik)** findet immer dann Anwendung, wenn zwar nur Informationen für eine Stichprobe vorliegen, die Aussagen allerdings auf die zugrundeliegende Population verallgemeinert werden sollen.
- **Die Stichprobentheorie** schliesslich befasst sich mit der vorgelagerten Frage, wie aus einer Population überhaupt eine Stichprobe gezogen werden soll, damit anschliessend der Rückschluss auf die Population möglich ist.

Im Rahmen dieses Lernschrittes werden wir uns darauf beschränken, eine Einführung in die verschiedenen Konzepte der Stichprobenziehung zu geben. Wer sich tiefer in das Gebiet der Stichprobentheorie einarbeiten möchte, dem seien die folgenden Bücher empfohlen:

- Henry, Gary T. (1990). "Practical Sampling". Newbury Park, London, New Delhi: Sage Publications.
- Kalton, Graham (1983). "Introduction to Survey Sampling". Series on Quantitative Applications in the Social Sciences, 07-035. Beverly Hills, London, New Delhi: Sage Publications.
- Kish, Leslie (1995). "Survey Sampling". New York, u.a.: Wiley & Sons.
- Levy, Paul S. und Stanley Lemeshow (1999). "Sampling of Populations -- Methods and Applications". New York: Wiley & Sons.

Theorie (1-8)

Inhaltsübersicht

- **1. Fragestellung**
- **2. Vor- und Nachteile der Stichprobenziehung**
- **3. Definition der Population**

- **4. Kategorisierung von Stichprobenverfahren**
- **5. Probabilistische Stichprobenverfahren?**
- **6. Nichtprobabilistische Stichprobenverfahren**
- **7. Stichprobenziehung - Fehlerquellen**
- **8. Zusammenfassung zum Lernschritt**

1. Fragestellung

Für die allermeisten empirischen Untersuchungen in den Sozialwissenschaften liegen die notwendigen Informationen nur für einen Teil der interessierenden Beobachtungen - d.h. für eine Stichprobe - vor. Das Ziel der meisten empirischen Untersuchungen sind allerdings Aussagen über die der Stichprobe zugrundeliegende Population, also Aussagen über die konkrete Stichprobe hinaus. Empirische Studien zielen in der Regel auf allgemeine und nicht auf spezifische Aussagen ab. Aus diesem Grund sind inferenzstatistische Verfahren von so grosser Bedeutung, weil sie es - unter gewissen Voraussetzungen - erlauben, Ergebnisse über eine konkrete Stichprobe hinaus auf die zugrundeliegende Grundgesamtheit zu verallgemeinern.

Eine Stichprobe liefert definitionsgemäss immer nur unvollständige Informationen in Bezug auf die interessierende Grundgesamtheit. Es stellt sich damit die zentrale Frage, ob und wie sich Ergebnisse, welche anhand der Stichprobendaten gewonnen werden, auf die Population verallgemeinern lassen. Die Verfahren der Inferenzstatistik liefern das notwendige Instrumentarium für diesen Rückschluss und ermöglichen eine Quantifizierung der Unsicherheit, welche mit der Informationsunvollständigkeit der Stichprobe einhergeht.

In dieser Lektion geht es allerdings nicht um den Rückschluss von der Stichprobe auf die Population, sondern um den der Inferenzstatistik vorgelagerten Schritt der Stichprobenziehung und die damit verbundenen Probleme. Wir werden verschiedene Verfahren der Stichprobenziehung sowie ihre spezifischen Vor- und Nachteile kennenlernen. Wir werden in dieser Lektion ausserdem auf mögliche Fehlerquellen im Zusammenhang mit Stichprobenziehungen zu sprechen kommen. Schliesslich versuchen wir, dem Begriff der "Repräsentativität" auf die Spur zu kommen.

2. Vor- und Nachteile der Stichprobenziehung

Bei Stichprobenerhebungen werden die Informationen immer nur für einen Teil aller interessierenden Beobachtungen erhoben. Vor- und Nachteile von Stichprobenerhebungen beziehen sich somit auf den Vergleich mit einer Voll- oder Totalerhebung, bei welcher für sämtliche Beobachtungen der interessierenden Population Informationen erhoben werden.

Nachteil

- *Informationsunvollständigkeit:*
Der offensichtliche Nachteil einer Stichprobenerhebung besteht darin, dass eine Stichprobe immer nur unvollständige Informationen über die Population liefert, da per Definition eine Stichprobe immer nur eine Teilmenge der Population darstellt.

Da in den meisten Fällen Aussagen über die Grundgesamtheit erwünscht sind, stellt sich die Frage, weshalb man diese mit der Stichprobenziehung verbundene Informationsunsicherheit überhaupt in Kauf nehmen sollte. Dazu betrachten wir uns die Vorteile einer Stichprobenerhebung (= Nachteile einer Vollerhebung).

Vorteile

- *Kosten:*
Eine Totalerhebung ist - sofern die Population nicht (sehr) klein ist - ausserordentlich teuer. Die Kosten für die Schweizerische Volkszählung belaufen sich beispielsweise auf mehr als 100 Millionen!
- *Eine Vollerhebung ist nicht immer möglich:*
Insbesondere in der Qualitätsprüfung von Produkten ist eine Vollerhebung dann nicht möglich, wenn die Beobachtungen durch die Erhebung beschädigt oder zerstört werden.
- *Zeit:*
Die Durchführung einer Stichprobenerhebung erfordert wesentlich weniger Zeit als eine Totalerhebung. Insbesondere in der kommerziellen Markt- und Meinungsforschung, aber auch in der sozialwissenschaftlichen Forschung ist man daran interessiert, dass die Ergebnisse möglichst rasch vorliegen und Aktualität besitzen. Man denke etwa an das Beispiel von Wahl- oder Abstimmungsprognosen. Hier ist es etwa von zentraler Bedeutung, dass die Ergebnisse so rasch wie möglich vorliegen.
- *Präzision:*
Unter Umständen ist es sogar denkbar, dass eine Stichprobe präziser ist als eine Vollerhebung. Bei Vollerhebungen sind sehr viel mehr Personen in die Datenerhebung involviert, zudem stellt die Dateneingabe und -validierung ein grösseres Fehlerrisiko dar als bei einer Stichprobe. Der Vorteil einer genauen und fehlerlosen Datenerhebung und -eingabe einer Stichprobe kann unter Umständen sogar den Nachteil der Informationsunsicherheit wettmachen (sofern sich diese Unsicherheit quantifizieren lässt). Da die Qualität einer Stichprobe nicht vom Umfang der Population abhängig ist, fällt dieser Vorteil umso mehr ins Gewicht, je grösser die Population ist.

3. Definition der Population

Wesentlicher Bestandteil einer Stichprobenerhebung ist eine exakte und eindeutige Definition der interessierenden Grundgesamtheit, aus welcher die Stichprobe gezogen wird. Diese vollzieht sich notwendigerweise vor dem Hintergrund der interessierenden Fragestellung sowie vor der aktuellen Ziehung einer Stichprobe.

Zunächst ist zu definieren, welches die Beobachtungseinheiten (Individuen, Haushalte, etc.) sind.

Bei der Definition der Population geht es darum, eindeutige Abgrenzungskriterien zu formulieren, sodass für jede Beobachtung bestimmt werden kann, ob sie Teil der Grundgesamtheit ist oder nicht. Die Abgrenzungskriterien der Population müssen in dreifacher Hinsicht definiert werden:

1. *In geografischer Hinsicht:*
Hier ist zu definieren, welche Beobachtungen in geografischer Hinsicht zur Population gehören. Diese Abgrenzung ist im allgemeinen relativ unproblematisch.
2. *In zeitlicher Hinsicht:*
Hier ist ein bestimmter Zeitpunkt oder ein Zeitfenster zu bestimmen, nach welchem Beobachtungen zur Population gehören oder nicht.
3. *In sachlicher Hinsicht:*
Die Bestimmung der inhaltlichen Kriterien, nach welchen Beobachtungen zur Grundgesamtheit gehören oder nicht, ist häufig nicht einfach. Bei der Festlegung der sachlichen Abgrenzungskriterien tauchen nicht selten Definitionsprobleme auf.

Beispiele für die Definition einer Population

Insbesondere die sachliche Abgrenzung der Population ist oftmals schwierig, wie die folgenden beiden Beispiele verdeutlichen sollen:

1. Stellen Sie sich vor, Sie möchten eine Umfrage unter allen EinwohnerInnen der Stadt Zürich durchführen, welche zu einem gegebenen Zeitpunkt älter als 18 Jahre sind.

Die interessierende Population sind somit alle Personen über 18 Jahren mit einem Wohnsitz in Zürich. Für eine eindeutige Definition der Grundgesamtheit sind allerdings noch weitere Aspekte zu berücksichtigen. Beispielsweise:

- Wie behandelt man Wochenaufenthalter?
- Wie geht man mit Personen um, welche zwar ihren Wohnsitz in der Stadt Zürich haben, sich aber vorübergehend abgemeldet haben?
- Gibt es einen theoretischen Grund, nur Personen zu befragen, welche seit einer bestimmten Minimaldauer ihren Wohnsitz in Zürich haben?
- Sollen sämtliche ausländischen EinwohnerInnen befragt werden oder nur solche mit einer Niederlassungs- oder Aufenthaltsbewilligung?
- Was geschieht mit Personen, welche in einem sogenannten Kollektivhaushalt (z.B. psychiatrische Klinik) leben?

2. Stellen Sie sich als zweites Beispiel vor, dass Sie eine Befragung unter den Studierenden an der Universität Zürich durchführen möchten.

Die Grundgesamtheit bilden offensichtlich alle Personen, welche zu einem bestimmten Zeitpunkt an der Universität immatrikuliert sind. Allerdings sind auch in diesem Beispiel weitere Aspekte zu berücksichtigen. Beispielsweise:

- Sollen alle Personen mit einer gültigen Immatrikulation berücksichtigt werden oder soll man sich auf eine spezifischere Gruppe beschränken (z.B. nur neuimmatrikulierte Personen)?
- Was geschieht mit Personen, welche zwar immatrikuliert sind, aber zu interessierenden Zeitpunkt nicht "aktiv" studieren (z.B. Auslandsaufenthalt, Krankheit, Praktikum)?
- Zählen Doktorierende zur Grundgesamtheit oder nicht?
- Sollen ausländische Studierende, welche im Rahmen eines Austauschprogrammes ein Semester an der Universität Zürich studieren, berücksichtigt werden oder nicht?

4. Kategorisierung von Stichprobenverfahren

Die verschiedenen Verfahren der Stichprobenziehung lassen sich grob in die folgenden beiden Kategorien einordnen:

- Probabilistische Stichprobenverfahren. Diese zeichnen sich durch die folgenden Eigenschaften aus:
 - Die Selektion der Beobachtungen basiert auf einem strikten Zufallsmechanismus (Urnenmodell, Zufallszahlentabelle, Zufallszahlengenerator).
 - Jedes Element der Population hat eine positive, d.h. eine von Null verschiedene, Wahrscheinlichkeit, Eingang in die Stichprobe zu finden. Die Auswahlwahrscheinlichkeit der Beobachtungen ist bekannt oder lässt sich berechnen.
 - Es ist allerdings nicht notwendig, dass alle Beobachtungen dieselbe Auswahlwahrscheinlichkeit aufweisen.

- Probabilistische Stichprobenverfahren erlauben den Rückschluss von der Stichprobe auf die Population.
- Nicht-probabilistische Stichprobenverfahren. Diese zeichnen sich durch die folgenden Eigenschaften aus:
 - Die Selektion der Beobachtungen basiert nicht auf einem Zufallsmechanismus, sondern erfolgt durch subjektive Entscheidungen (bewusste oder willkürliche Auswahl von Beobachtungen).
 - Dies führt im allgemeinen dazu, dass ein Teil der Beobachtungen eine Auswahlwahrscheinlichkeit von Null aufweist.
 - Die Auswahlwahrscheinlichkeit der einzelnen Beobachtungen ist weder bekannt noch lässt sich diese berechnen.
 - Bei nicht-probabilistischen Stichprobenverfahren ist kein Rückschluss von der Stichprobe auf die Population möglich. Bei diesen Verfahren der Stichprobenziehung ist oftmals völlig unklar, auf welche Grundgesamtheit sich die Stichprobe bezieht.

Eine probabilistische Stichprobenziehung ist eine notwendige Voraussetzung für den Rückschluss von der Stichprobe auf die Population: Alle inferenzstatistischen Verfahren gehen von der Voraussetzung einer probabilistischen Stichprobe aus! Liegt eine nicht-probabilistische Stichprobe vor, lassen sich keine inferenzstatistischen Verfahren anwenden.

5. Probabilistische Stichprobenverfahren

Probabilistische Stichprobenverfahren weisen die zentrale Eigenschaft auf, dass die Auswahl der Beobachtungen für die Stichprobe durch einen strikten Zufallsmechanismus zustande kommt. Dies garantiert zum einen, dass die Auswahlwahrscheinlichkeit für alle Beobachtungen der Population grösser als Null ist. Es werden somit keine Beobachtungen systematisch von der Stichprobenziehung ausgeschlossen.

Die Selektion durch einen objektiven Zufallsmechanismus erlaubt im weiteren den Rückschluss von der Stichprobe auf die Population. Bevor wir uns den einzelnen Verfahren der Stichprobenziehung zuwenden, soll zunächst erläutert werden, was unter einer zufälligen Auswahl von Beobachtungen gemeint ist:

Zufallsauswahl von Beobachtungen:

Vorauszuschicken ist, dass für alle Arten der Zufallsauswahl eine vollständige und eindeutige Auflistung aller Elemente notwendig ist, aus welcher die Stichprobe gezogen werden soll. Liegt eine solche Auflistung aller Elemente der Population vor, lässt sich über die folgenden Hilfsmittel eine zufällige Stichprobe aus dieser Grundgesamtheit ziehen:

- "Urnenmodell": Jedem Element der Population wird ein Los mit einer eindeutigen Nummer zugewiesen (d.h. alle Elemente der Population werden durchgehend numeriert). Alle Lose kommen in eine Urne und daraus werden sukzessive blind Lose gezogen, bis die gewünschte Grösse der Stichprobe erreicht ist.
- Zufallszahlentabelle / Zufallszahlengenerator: Das "Urnenmodell" ist in den seltensten Fällen praxistauglich, deshalb erfolgt die Zufallsauswahl entweder über eine Zufallszahlentabelle beziehungsweise heutzutage meist über einen Zufallszahlengenerator, wie er in vielen Statistikprogrammen vorhanden ist. Sowohl eine Zufallszahlentabelle als auch ein Zufallszahlengenerator erlauben es, aus einer vorgegebenen Population eine zufällige Stichprobe im Sinne des "Urnenmodelles" zu ziehen.

Stichprobenziehung

Voraussetzungen

- Alle Elemente der interessierenden Population müssen entweder physisch präsent sein oder es muss eine aktuelle, vollständige und eindeutige Liste aller Elemente vorhanden sein.
- Dies setzt voraus, dass die Population bekannt ist oder grundsätzlich identifizierbar ist.

Vorgehen

Bei einer einfachen Zufallsstichprobe wird aus einer Population vom Umfang N eine Stichprobe der Grösse n per Zufallsmechanismus gezogen.

Eigenschaften

- Die Auswahlwahrscheinlichkeit (d.h. die Wahrscheinlichkeit, Eingang in die Stichprobe zu finden) ist bei einer einfachen Zufallsstichprobe für alle Beobachtungen der Population gleich gross.
- Die Auswahlwahrscheinlichkeiten der Beobachtungen sind untereinander unabhängig.
- Bei einer einfachen Zufallsstichprobe der Grösse n aus einer Population vom Umfang N sind alle möglichen Stichproben gleich wahrscheinlich.

Anmerkungen

- Die einfache Zufallsstichprobe ist von ausserordentlicher theoretischer Bedeutung. Die Annahme einer einfachen Zufallsstichprobe erleichtert beispielsweise die Berechnung von Standardfehlern von Schätzfunktionen.
- Die meisten inferenzstatistischen Verfahren und insbesondere auch ihre Implementierung in Statistikprogrammen gehen von der Annahme einer einfachen Zufallsstichprobe aus.
- Die Voraussetzungen für eine einfache Zufallsstichprobe sind in der Praxis allerdings häufig nicht gegeben, d.h. es ist eher selten, dass für eine Population eine vollständige und aktuelle Liste vorhanden ist. Häufig ist nicht einmal die Grundgesamtheit bekannt oder identifizierbar.

Voraussetzungen

- Die Population muss bekannt oder grundsätzlich identifizierbar sein.
- Es muss eine aktuelle, vollständige und eindeutige Liste aller Elemente der Population vorhanden sein oder:
Die Population ist eine Sequenz von Elementen in einem fixen Zeitintervall, wobei zumindest die Grösse der Population (ungefähr) bekannt sein muss (z.B. Fließbandproduktion, Posteingang o.ä.).

Vorgehen

Bei einer systematischen Zufallsstichprobe vom Umfang n aus einer Population der Grösse N wird nur die erste Beobachtung per Zufallsmechanismus ausgewählt.

Die verbleibenden $(n-1)$ Beobachtungen werden - ausgehend von der ersten Beobachtung - in einem fixen Selektionsintervall k ausgewählt, wobei dieses gegeben ist durch den folgenden Ausdruck:

$$k = \left(\frac{N}{n} \right)$$

Es ist dabei notwendig, das Selektionsintervall auf eine gerade Zahl zu runden.

Anmerkungen

- Die Eigenschaften einer systematischen Zufallsstichprobe stimmen in der Regel mit denjenigen einer einfachen Zufallsstichprobe überein.
- Vorsicht ist geboten, wenn die Liste der Population ein zyklisches Muster aufweist. Fällt das Selektionsintervall mit diesem Zyklus zusammen, erhält man eine höchst selektive Stichprobe. Der Rückschluss von der Stichprobe auf die Population kann in diesem Fall mit einem grossen Fehler behaftet sein.

Beispiel

Aus der Grundgesamtheit aller im Wintersemester 2002/03 an der Universität Zürich immatrikulierten Studierenden (inkl. Doktorierende und KandidatInnen des Höheren Lehramtes) soll eine systematische Zufallsstichprobe von 10% gezogen werden.

- Die Population umfasst 22'362 Elemente. Wir gehen an dieser Stelle davon aus, dass wir Zugriff auf eine vollständige und eindeutige Liste aller Elemente haben.
- Um eine Stichprobe von 10% zu erhalten, müssen 2236 Personen befragt werden. Das Selektionsintervall beträgt in diesem Falle zehn, d.h. jede zehnte Person auf der Liste muss befragt werden.
- Die erste Person wird nun folgendermassen bestimmt: Aus den ersten zehn Personen auf der Liste der Population wird per Zufallsmechanismus eine einzelne Person ausgewählt. Dies sei beispielsweise die siebte Person auf der Liste.
- Die verbleibenden Personen werden ausgehend von dieser ersten Person ausgewählt, indem jeweils zur Nummer der vorangehenden Person das Selektionsintervall addiert wird. Die zweite Person, welche für die Stichprobe ausgewählt wird, ist somit die 17. Person, die dritte Person diejenige mit der Nummer 27. Dieses Vorgehen wird nun solange wiederholt, bis man die gesamte Liste der Population durchgegangen ist und damit die gewünschte Grösse der Stichprobe erreicht ist.

Voraussetzungen

- Die Population muss bekannt oder grundsätzlich identifizierbar sein.
- Es muss eine aktuelle, vollständige und eindeutige Liste aller Elemente der Population vorhanden sein.
- Es muss zusätzlich die Information über ein Gruppierungsmerkmal (Schichtungsmerkmal) vorhanden sein:
 - Die Information über das Schichtungsmerkmal muss für alle Beobachtungen der Population vorhanden sein.
 - Jedes Element der Population muss sich eindeutig einer Ausprägung des Schichtungsmerkmals zuordnen lassen.
 - Das Schichtungsmerkmal muss dergestalt sein, dass sich die Population vollständig in sich nicht überschneidende Subpopulationen aufteilen lässt.

Vorgehen

Das Schichtungsmerkmal unterteilt die Population in verschiedene Subpopulationen, wobei die Anzahl der Subpopulationen durch die Anzahl der Ausprägungen des Schichtungsmerkmals abhängig ist. Aus jeder Subpopulation wird nun per Zufallsmechanismus eine einfache Zufallsstichprobe gezogen.

Anmerkung

- Eine geschichtete Stichprobe bietet sich vor allem in zwei Situationen an:
 - Wenn garantiert werden soll, dass die relativen Häufigkeitsanteile der Subpopulationen ihren Anteilen in der Population entsprechen. Dies ist über eine geschichtete Stichprobe besser zu erreichen als über eine einfache Zufallsstichprobe. In diesem Fall wird aus jeder Subpopulation derselbe Anteil an Beobachtungen als Stichprobe gezogen. Man nennt dies proportionale Schichtung.
 - Möglicherweise ist man an spezifischen Subpopulationen interessiert, deren Anteil in der Population relativ klein ist. Sowohl eine einfache Zufallsstichprobe als auch eine proportional geschichtete Zufallsstichprobe realisierten dann wahrscheinlich eine zu geringe Fallzahl für die interessierende Subpopulation.
In diesem Fall wird nicht aus jeder Subpopulation derselbe Anteil an Personen gezogen. Bestimmte Subpopulationen werden bewusst über- bzw. unterproportional zu ihrem Anteil in der Population gezogen. Man nennt dies disproportionale Schichtung.
- Der Vorteil von geschichteten Zufallsstichproben ist der damit verbundene Gewinn an Präzision (geringere Standardfehler) relativ zur einfachen Zufallsstichprobe. Diese zusätzliche Präzision ist von zwei Faktoren abhängig:
 - Die Variabilität des interessierenden Merkmals zwischen den Subpopulationen sollte möglichst gross sein: Je grösser die Differenz der schichtspezifischen Mittelwerte des interessierenden Merkmals, desto grösser der Präzisionsgewinn. D.h. das Schichtungsmerkmal sollte möglichst eng mit dem interessierenden Merkmal zusammenhängen.
 - Die einzelnen Subpopulationen sollten jeweils in sich möglichst homogen sein bezüglich dem interessierenden Merkmal, d.h. die Variabilität des interessierenden Merkmals sollte innerhalb der Subpopulationen möglichst klein sein.

Beispiel

Als Beispiel dient wiederum die Population aller Personen, welche im Wintersemester 2002/03 an der Universität Zürich immatrikuliert sind. Wiederum soll eine Stichprobe von 10% aus dieser Grundgesamtheit gezogen werden. Wir gehen davon aus, dass es sich bei der geplanten Studie um geschlechtsspezifische Fragestellungen handelt.

Betrachtet man sich die Unterschiede im Frauenanteil zwischen den Fakultäten, so zeigt sich, dass hier das Ziehen einer geschichteten Zufallsstichprobe möglich ist, da für alle Personen der Population die Information zu einem Schichtungsmerkmal (Fakultät) vorhanden ist und dieses Schichtungsmerkmal ausserdem mit einem für die Studie relevanten Merkmal (Geschlecht) in einem Zusammenhang steht.

Um nun eine (proportional) geschichtete Zufallsstichprobe von 10% zu ziehen, wird aus jeder Subpopulation (Fakultät) eine einfache Zufallsstichprobe von 10% gezogen.

Voraussetzungen

- Die Population muss bekannt oder identifizierbar sein.
- Es existiert ein Gruppierungsmerkmal (Klumpenmerkmal), welches die folgenden Eigenschaften aufweist:
 - Die gesamte Population kann vollständig Subpopulationen (Klumpen) aufgeteilt werden.
 - Jedes Element der Population lässt sich eindeutig einem solchen Klumpen zuordnen.

Stichprobenziehung

- Es existiert eine aktuelle, vollständige und eindeutige Liste aller Klumpen (nicht aller Elemente!).

Vorgehen

Aus der Liste der Klumpen wird per Zufallsmechanismus eine einfache Zufallsstichprobe von Klumpen gezogen. Danach werden für alle Elemente der so bestimmten Klumpen die notwendigen Informationen erhoben. Im Gegensatz zu den vorher besprochenen Verfahren werden hier also nicht einzelne Elemente, sondern ganze Gruppierungen (Klumpen) von Elementen ausgewählt.

Anmerkungen

- Klumpen sind in der Regel "natürlich" auftretende Gruppierungen wie beispielsweise Schulen oder Haushalte. Geografische / politische Einheiten wie beispielsweise Gemeinden stellen ebenfalls geeignete Klumpen dar.
- Klumpenstichproben sind in der Praxis sehr häufig, da oftmals keine Liste aller Populationselemente, wohl aber eine Liste von Klumpen vorhanden ist. Zudem sind Klumpenstichproben einfacher und billiger zu realisieren als andere Arten der Stichprobenziehung, da a priori sehr viel weniger Informationen zum Ziehen der Stichprobe erforderlich sind
- Diesem "ökonomischen" Vorteil steht allerdings der gewichtige Nachteil gegenüber, dass Klumpenstichproben in der Regel eine - zum Teil sehr viel - geringere Präzision (grössere Standardfehler) aufweisen als eine einfache Zufallsstichprobe. Der Präzisionsverlust hängt von drei Faktoren ab:
 - Die Differenz zwischen den einzelnen Klumpenmittelwerten und dem Populationsmittelwert des interessierenden Merkmales sollte möglichst gering sein. Je heterogener die Klumpen untereinander sind in Bezug auf das interessierende Merkmal, desto grösser ist der Verlust an Präzision.
 - Die einzelnen Klumpen sollten jeweils in sich möglichst grosse Variabilität des interessierenden Merkmales aufweisen. Je geringer die Streuung des Merkmales innerhalb der Klumpen, desto grösser ist der Verlust an Präzision.
 - Je weniger Klumpen erhoben werden, desto grösser ist der Verlust an Präzision. Dies fällt vor allem dann ins Gewicht, wenn nur wenige aber grosse Klumpen vorliegen.

Beispiel

Es soll eine Befragung unter aller PrimarschülerInnen des Kantons Zürich durchgeführt werden. Sie stellen fest, dass jedoch keine vollständige Liste aller PrimarschülerInnen zugänglich oder vorhanden ist.

Sie stellen jedoch fest, dass es im Vergleich dazu relativ einfach, eine Liste aller Schulhäuser im Kanton Zürich zu erhalten. Aus dieser Liste der Schulhäuser ziehen Sie eine einfache Zufallsstichprobe. Alle SchülerInnen an den so ausgewählten Schulhäusern werden schliesslich für Ihre Studie befragt.

Voraussetzungen

- Siehe dazu die Voraussetzungen zu den jeweiligen Stichprobenverfahren.

Vorgehen

In der Praxis werden häufig verschiedene Stichprobenverfahren auf zwei oder mehr Stufen miteinander kombiniert. Ein praktisch relevanter Fall ist etwa die zweistufige Klumpenstichprobe:

1. In einer ersten Stufe wird eine Zufallsstichprobe von Klumpen gezogen. In der Praxis ist nun allerdings der Umfang der einzelnen Klumpen oft viel zu gross, als dass sämtliche Elemente der ausgewählten Klumpen erhoben werden könnten.
2. In einer zweiten Stufe wird deshalb aus jedem Klumpen eine einfache oder geschichtete Zufallsstichprobe gezogen.

Ein Beispiel für ein mehrstufiges Stichprobenverfahren ist das zwei- oder mehrmalige Ziehen einer Klumpenstichprobe mit einer abschliessenden einfachen oder geschichteten Zufallsstichprobe. Diese wird aus Gründen der Praktikabilität häufig verwendet.

Anmerkungen

- Zwei- und mehrstufige Stichprobenverfahren sind in der Praxis häufig, da dies oft die einzige Möglichkeit der Stichprobenziehung überhaupt darstellt (weil etwa häufig keine Liste aller Elemente der Grundgesamtheit zugänglich ist).
- Zwei- und mehrstufige Stichprobenverfahren sind aus Kostengründen oftmals die einzige Möglichkeit einer Stichprobenziehung.
- Mehrstufige Stichprobenverfahren gehen oft mit einem Verlust an Präzision einher, zudem werden kompliziertere Verfahren zur Berechnung von Standardfehlern benötigt.

Beispiel

Sie möchten eine Studie durchführen, welcher die Population der ständigen Wohnbevölkerung in der Schweiz zugrunde liegt. Da keine Liste zu dieser Population vorliegt bzw. Sie keinen Zugang zu einer solchen Liste haben, entscheiden Sie sich aus praktischen sowie ökonomischen Gründen für ein zweistufiges Stichprobenverfahren:

1. In der ersten Stufe erstellen Sie eine Liste aller Gemeinden der Schweiz. Aus dieser Liste werden zufällig Gemeinden ausgewählt. Für die so ausgewählten Gemeinden erhalten Sie die entsprechenden Einwohnerregister.
2. In der zweiten Stufe wird nun aus den vorhandenen Einwohnerregistern jeweils eine einfache oder eine geschichtete Zufallsstichprobe an Personen gezogen.

6. Nichtprobabilistische Stichprobenverfahren

Der Rückschluss von der Stichprobe auf die Population ist an die Voraussetzung gebunden, dass die Selektion der Stichprobe über einen objektiven Zufallsmechanismus geschieht.

Nicht-probabilistische Stichprobenverfahren weisen diese Eigenschaft nicht auf, da die Selektion der Stichprobe nach subjektiven Kriterien erfolgt. Somit lässt sich bei solchen Stichprobenverfahren kein Rückschluss auf die Population ziehen. Dennoch sind nicht-probabilistische Verfahren in bestimmten Situationen sinnvoll anzuwenden:

- Die Population selbst ist unbekannt und lässt sich grundsätzlich nicht identifizieren.
- Das Interesse der Studie richtet sich explizit auf ganz spezifische Beobachtungen (z.B. Fallstudien) und nicht auf Verallgemeinerungen darüber hinaus.
- Eine Studie dient rein deskriptiven oder explorativen Zwecken.

Bei einer willkürlichen / pragamtischen Stichprobe werden die Beobachtungen primär aufgrund ihrer einfachen Verfügbarkeit oder ihrer raschen oder einfachen Erreichbarkeit für die Datenerhebung ausgewählt.

- Strassenbefragungen
- BesucherInnen einer Veranstaltung.
- Leserumfragen in Zeitungen und Zeitschriften.

Bei solchen Stichproben werden bewusst einzelnen oder mehrere Elemente der Population der Population in die Stichprobe einbezogen oder weggelassen. Verschiedene Möglichkeiten für bewusste Stichproben sind beispielsweise:

- Die Auswahl von möglichst ähnlichen oder möglichst unähnlichen Fällen: Die Beobachtungen werden so für die Stichprobe ausgewählt, dass sie möglichst ähnliche / unähnliche Eigenschaften aufweisen. In Fallstudien (welche per Definition keine Zufallsstichproben darstellen) werden beispielsweise häufig zwei Beobachtungen mit möglichst gegensätzlichen Eigenschaften untersucht.
- Die Auswahl von typischen Fällen: Hier werden diejenigen Beobachtungen für die Stichprobe ausgewählt, von denen man weiss oder annimmt, dass sie typische, durchschnittliche und eben keine extremen Eigenschaften aufweisen.
- Die Auswahl von kritischen Fällen: Es werden insbesondere oder ausschliesslich diejenigen Beobachtungen ausgewählt, von denen man weiss, dass deren Einschluss in die Stichprobe für die Glaubwürdigkeit oder Akzeptanz einer Studie ausschlaggebend sind.

Bei sogenannten Quotenstichproben wird die Population zunächst wie bei einer geschichteten Zufallsstichprobe nach einem oder meistens mehreren Schichtungsmerkmalen in verschiedene Subpopulationen unterteilt. Basierend auf dieser Einteilung der Population erhalten die Interviewer entsprechende Vorgaben ("Quoten"), wieviele Beobachtungen mit welchen bestimmten Merkmalskombinationen zu erheben sind. Die Quoten werden zumeist so gewählt, dass die Stichprobenanteile der Subpopulationen möglichst den Populationsanteilen entsprechen.

Im Gegensatz zu einer geschichteten Zufallsstichprobe werden nun allerdings die Beobachtungen aus den Subpopulationen nicht per Zufallsmechanismus bestimmt, sondern es bleibt dem jeweiligen Interviewer überlassen, wie er seine vorgegebenen Quoten erfüllt. Die Auswahl der einzelnen Beobachtungen der Interviewer erfolgen damit über subjektive Kriterien und nicht per Zufallsverfahren.

Viele kommerzielle Markt- und Meinungsforschungsinstitute arbeiten nach wie vor mit Quotenverfahren. Allerdings ist an dieser Stelle noch einmal festzuhalten, dass auch Quotenstichproben aus einem nicht-probabilistischen Stichprobenverfahren resultieren und damit inferenzstatistische Verfahren nicht anzuwenden sind.

7. Stichprobenziehung - Fehlerquellen

Sie wissen bereits, dass sich die Präzision einer Schätzfunktion über ihre Varianz oder über ihren Standardfehler quantifizieren lässt. Unter der Annahme einer einfachen Zufallsstichprobe ist beispielsweise der Standardfehler des Stichprobenmittelwertes durch den folgenden Ausdruck gegeben:

$$se(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Der Standardfehler des Stichprobenmittelwertes hängt somit ausschliesslich von zwei Grössen ab:

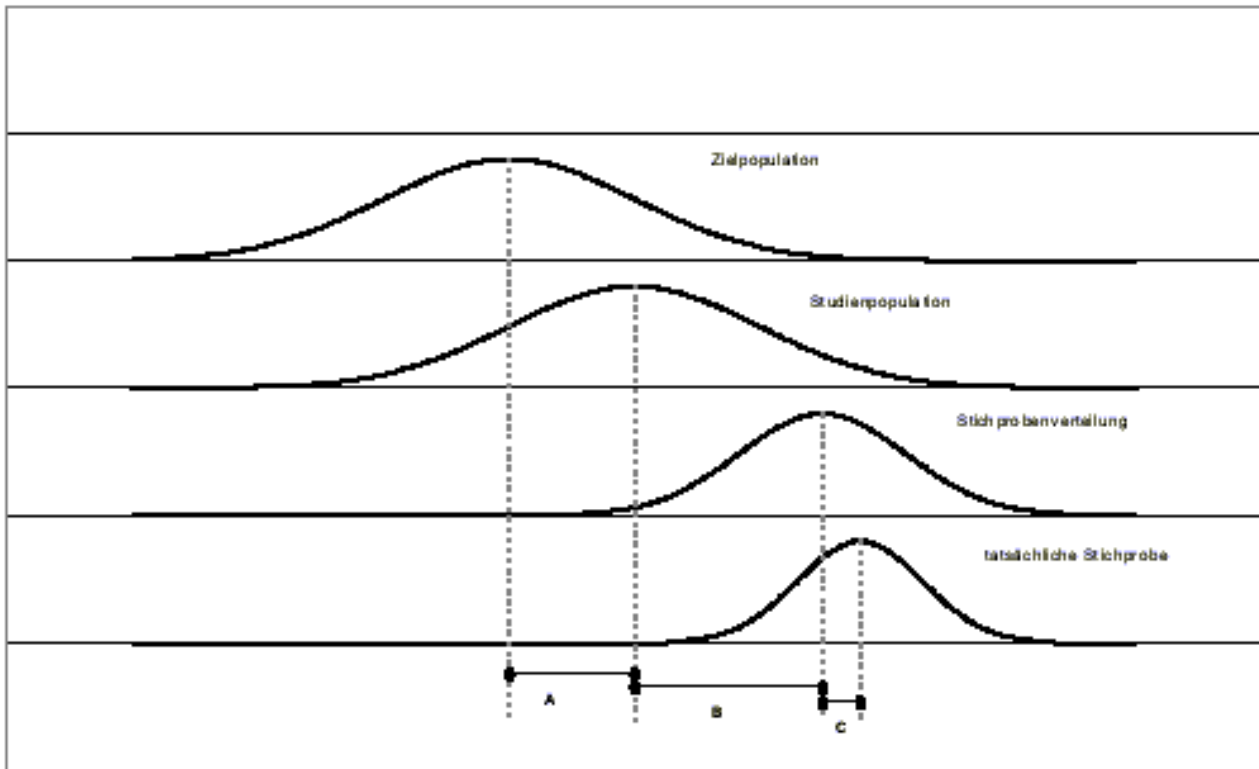
- Je grösser die Varianz des Merkmales in der Population, desto grösser ist der Standardfehler.
- Je kleiner die Stichprobe ist, desto grösser ist der Standardfehler.

Stichprobenziehung

Da die Populationsvarianz offensichtlich ausserhalb der Beeinflussbarkeit liegt, entscheidet offenbar einzig und allein die Stichprobengrösse über die Präzision von Populationschätzungen?

Leider nein, beziehungsweise nur unter bestimmten Voraussetzungen. Es existiert eine ganze Reihe von Faktoren neben der Stichprobengrösse, welchen die Präzision von Populationsschätzungen beeinflussen können. Wir wollen uns im folgenden anhand des Beispiels des Stichprobenmittelwertes mit möglichen Fehlerquellen bei der Stichprobenziehung befassen.

In untenstehender Grafik sind die möglichen Fehlerquellen beim Rückschluss von der Stichprobe auf die Population zusammenfassend grafisch dargestellt.



Es lassen sich drei Arten von Fehlern im Zusammenhang mit der Stichprobenziehung benennen:

- A: Nichtstichprobenfehler.
- B: Stichprobenfehler.
- C: Variabilität der Stichprobenmittelwerte.

Im folgenden werden die einzelnen Fehlerquellen und deren Ursachen im einzelnen diskutiert.

Der sogenannte Nichtstichprobenfehler ist gegeben durch die Differenz zwischen dem Mittelwert der ursprünglichen Zielpopulation und dem Mittelwert der tatsächlichen Studienpopulation:

$$\mu_T - \mu_S$$

Die Zielpopulation ist diejenige Population, auf welche sich die Studie beziehen möchte und aus welcher somit die Stichprobe gezogen werden sollte. Aus den folgenden Gründen kann es allerdings passieren, dass sich die ursprünglich vorgesehene Zielpopulation von der faktisch erreichbaren Studienpopulation unterscheidet:

- "listing und framing": Darunter versteht man das Problem, dass keine aktuelle, keine vollständiger oder aber keine eindeutige Liste aller Elemente der Population vorhanden oder konstruierbar ist. Fehlende oder mehrfache Einträge von einzelnen Beobachtungen auf der Liste können zu einer Diskrepanz zwischen Ziel- und Studienpopulation führen.
- "nonresponse": Unter "nonresponse" versteht man das vollständige (unit-nonresponse) oder das teilweise (item-nonresponse) Fehlen von Informationen zu einzelnen Beobachtungen. Steht dieses Verhalten in einem systematischen Zusammenhang mit dem interessierenden Merkmal oder einem anderen Merkmal, welches seinerseits mit dem interessierenden Merkmal in Zusammenhang steht, dann führt dies ebenfalls zu einem Nichtstichprobenfehler. "Nonresponse" unsystematischer Art führt zwar nicht zu einem Nichtstichprobenfehler, reduziert aber dennoch die Präzision von Populationsschätzungen, weil der Stichprobenumfang reduziert wird.
- Messfehler: Können die interessierenden Merkmale nur mit einem systematischen Messfehler erfasst werden, führt auch dies zu einem Nichtstichprobenfehler. Ein unsystematischer Messfehler bewirkt zwar keinen Fehler, reduziert aber wiederum die Präzision von Populationsschätzungen.

Auf den potentiellen Einfluss von "nonresponse" kommen wir später noch im Detail zurück.

Der sogenannte Stichprobenfehler ist gegeben durch die Differenz zwischen dem Mittelwert der Studienpopulation und dem Erwartungswert der Stichprobenmittelwerte:

$$E(\bar{x}_n) - \mu_S$$

Für diese Diskrepanz gibt es primär zwei mögliche Ursachen:

- Selektionsfehler: Ein Selektionsfehler liegt im allgemeinen dann vor, wenn nicht alle Beobachtungen der Population dieselbe Auswahlwahrscheinlichkeit aufweisen und dies in der Auswertung nicht entsprechend berücksichtigt wird.
- Schätzfehler: Ein Schätzfehler resultiert aus der Verwendung eines ungeeigneten Schätzers (verzerrte oder inkonsistente Schätzfunktion) für den interessierenden Populationskennwert.

Die Variabilität der Stichprobenmittelwerte lässt sich durch den Standardfehler des Mittelwertes quantifizieren:

$$se(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Wird nur eine einzige Stichprobe realisiert, weicht dieser Stichprobenmittelwert in der Regel vom Erwartungswert der Stichprobenmittelwerte ab (wobei diese Diskrepanz abhängig ist von der Grösse des Standardfehlers):

$$\bar{x}_n - E(\bar{x}_n)$$

Wie bereits erwähnt hängt diese Abweichung im Falle des Stichprobenmittelwertes von zwei Grössen ab:

- Der Populationsvarianz: Je heterogener das Merkmal in der Population verteilt ist, desto grösser ist der Standardfehler des Stichprobenmittelwertes und desto grösser ist die zu erwartende Abweichung zwischen dem realisierten Stichprobenmittelwert und dem Erwartungswert der Stichprobenmittelwerte.
- Der Stichprobengrösse: Je grösser die Stichprobe, desto kleiner fällt der Standardfehler der Stichprobenmittelwerte aus. Damit wird mit steigendem Stichprobenumfang auch die zu erwartende Differenz zwischen Erwartungswert und tatsächlichem Stichprobenmittelwert kleiner.

Der gesamte oder totale Stichprobenfehler ist somit die Summe aus Nichtstichprobenfehler, Stichprobenfehler und Stichprobenvariabilität:

$$(\mu_t - \mu_S) + (E(\bar{x}_n) - \mu_S) + (\bar{x}_n - E(\bar{x}_n))$$

Zusammenfassend lässt sich somit an dieser Stelle festhalten:

- Nichtstichproben- und Stichprobenfehler können zu einer selektiven Stichprobe führen. Im Fall einer starken Selektion lässt auch eine noch so grosse Stichprobe keine präzise Schätzung von Populationskennwerten zu.
- Ceteris paribus ist eine grössere Stichprobe immer besser, da die Stichprobenvariabilität in jedem Fall reduziert wird. Allerdings ist damit noch keine Präzision der Populationsschätzungen garantiert.
- Die Stichprobengrösse ist dann ausschlaggebend für die Präzision von Populationsschätzungen, wenn weder ein Nichtstichproben- noch ein Stichprobenfehler vorliegt.

Nonresponse: Das Problem von fehlenden Informationen auf der Ebene der Beobachtungen

Eines der häufigsten Probleme insbesondere von Umfragedaten ist nonresponse. Unter nonresponse versteht man, dass zu einem, zu mehreren oder zu allen interessierenden Merkmalen einer Studie die entsprechenden Angaben für eine Beobachtung fehlen.

Es sind vor allem zwei Arten von Nonresponse zu unterscheiden:

- **Unit-Nonresponse** : Für eine bestimmte Beobachtung liegen überhaupt keine Informationen vor.
- **Item-Nonresponse** : Für eine bestimmte Beobachtung liegen für eines oder für mehrere (allerdings nicht für alle) Merkmale keine Informationen vor.

Nonresponse ist somit prinzipiell ein Problem unvollständiger Informationen.

Die entscheidende Frage in Zusammenhang mit Nonresponse ist, ob es einen systematischen Selektionsmechanismus gibt oder ob das Fehlen der Angaben zufälliger Natur ist.

Da Nonresponse nun aber ein Problem fehlender Daten ist, ist es in aller Regel sehr schwierig wenn nicht gar unmöglich, auf den zugrundeliegenden Selektionsmechanismus zu schliessen. Falls dieser Mechanismus nicht bestimmt werden kann, ist es nur schwer möglich, dem Problem des Nonresponse beizukommen.

Wir werden uns im folgenden zwei Arten von Nonresponse und deren Einfluss auf die Schätzung eines Populationsmittelwertes betrachten:

- Unsystematischer (zufälliger) Nonresponse.
- Systematischer Nonresponse.

Wir werden den Einfluss von Nonresponse wiederum anhand einer Computersimulation illustrieren. Es wird sich zeigen, dass die beiden Arten von Nonresponse einen sehr unterschiedlichen Einfluss auf die Schätzung des Populationsmittelwertes ausüben.

Der Einfluss von unsystematischem Nonresponse auf die Schätzung eines Populationsmittelwertes soll anhand der folgenden Computersimulation illustriert werden:

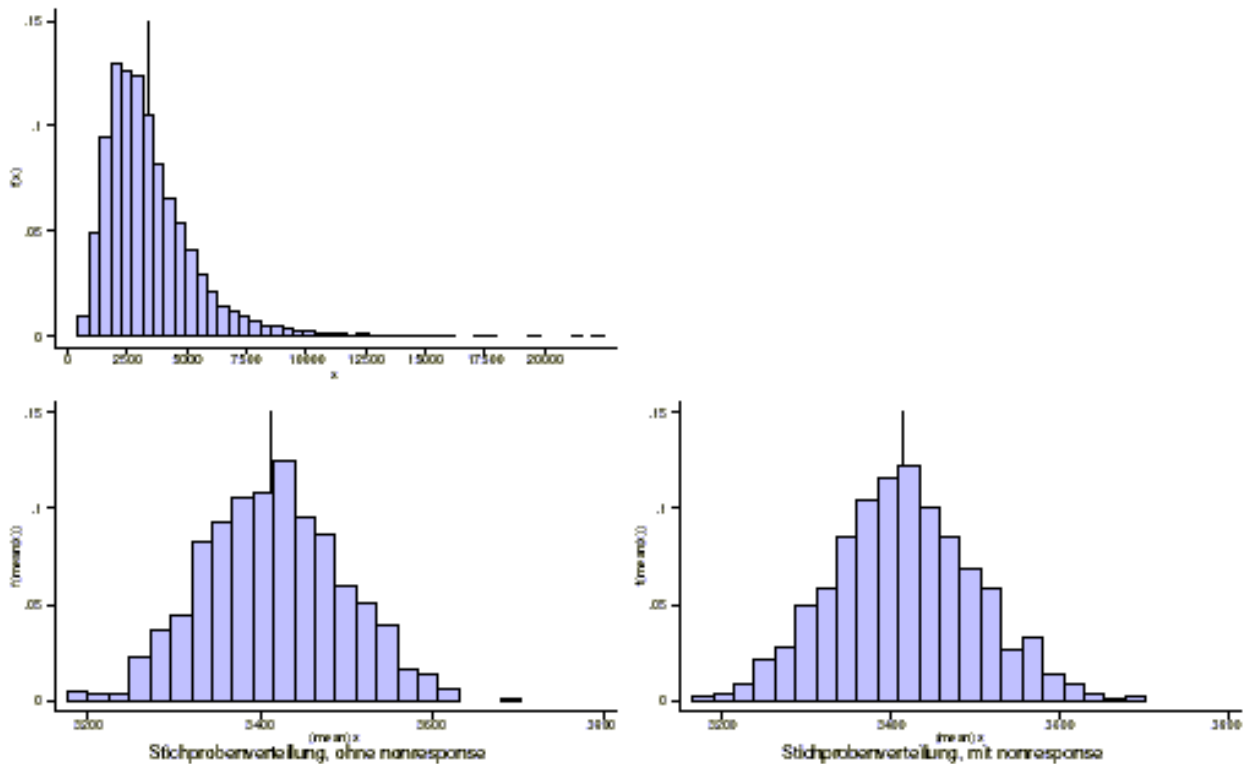
- Es wird eine Verteilung mit bekannten Parametern für eine Population vom Umfang $N = 10'000$ generiert.
- Die Verteilung von X stelle eine fiktive Einkommensverteilung in der Population dar.

Der Populationsmittelwert von X sei bekannt:

$$E(X) = \mu = 3'408$$

- Wir ziehen nun zunächst 1'000 einfache Zufallsstichproben vom Umfang $n = 500$ aus der Population und berechnen für jede Stichprobe den Mittelwert von X .
Auf diese Weise erhält man die Stichprobenkennwerteverteilung, falls kein Nonresponse vorliegt.
- In einem zweiten Schritt wird Nonresponse von zufälliger Natur simuliert. Wie gehen davon aus, dass die Nonresponse-Rate 10% betrage, d.h. für 10% der Beobachtungen liegen keine Informationen bezüglich X vor.
D.h. wir ziehen nun ebenfalls 1'000 einfache Zufallsstichproben aus der Population, allerdings nur mit dem Umfang $n = 450$.
Auf diese Weise erhält man die Stichprobenverteilung, falls zufälliger Nonresponse vorliegt.

In der unteren Grafik ist zum einen die Verteilung von X in der Population dargestellt (links oben). Die unteren beiden Grafiken stellen die Verteilung der Stichprobenmittelwerte ohne beziehungsweise mit Nonreponse dar.



In Bezug auf zufälligen Nonresponse lässt sich das folgende Fazit ziehen:

- Bei zufälligem Nonresponse lässt sich der Populationsmittelwert nach wie vor unverzerrt schätzen.
- Allerdings ist bei zufälligem Nonresponse der Standardfehler der Schätzung grösser, als wenn kein Nonresponse vorliegt.

Der Effekt von Nonresponse auf die Schätzung des Populationsmittelwertes ist gänzlich anders, wenn Nonresponse nicht zufällig, sondern durch einen systematischen Selektionsmechanismus erfolgt.

Zur Illustration wird ebenfalls eine Computersimulation durchgeführt:

- Es wird eine Verteilung mit bekannten Parametern für eine Population vom Umfang $N = 10'000$ generiert.

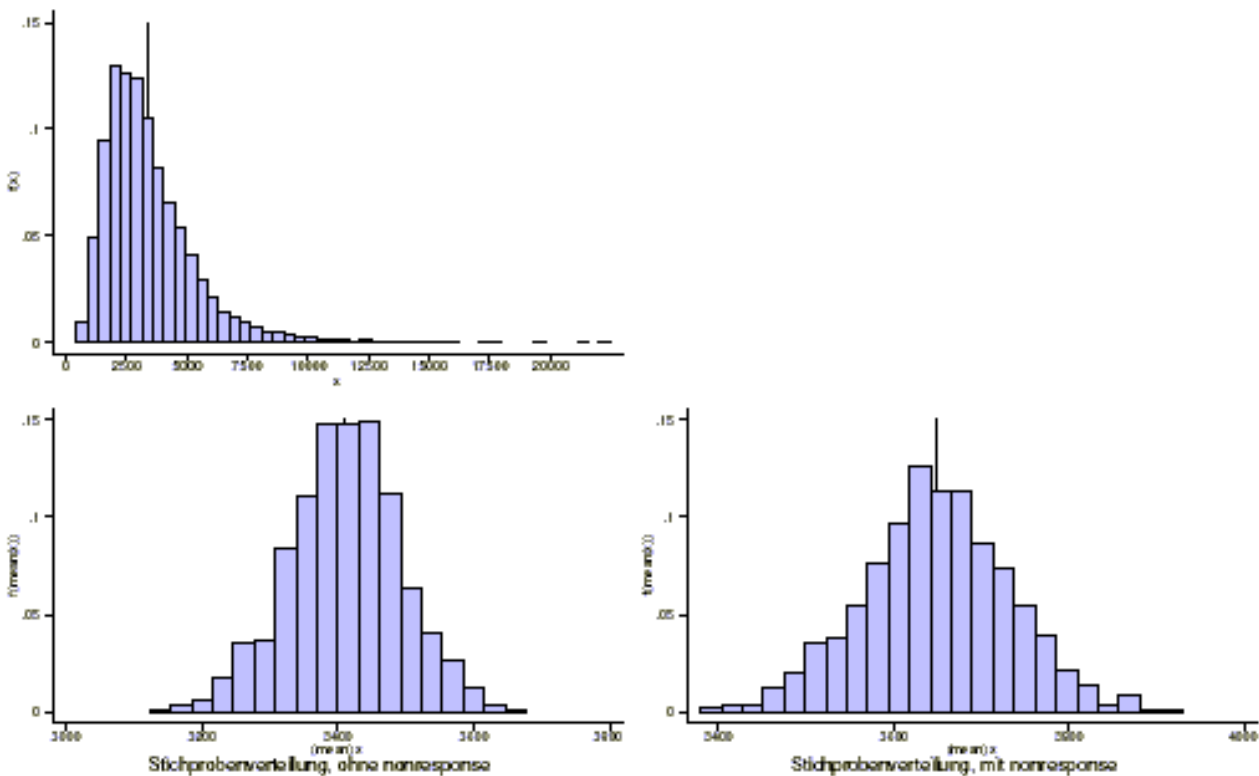
Stichprobenziehung

- Zunächst wird der Fall simuliert, dass kein Nonresponse vorliegt: Wie ziehen 1'000 einfache Zufallsstichproben vom Umfang $n = 500$ aus der Population und berechnen für jede Stichprobe den Mittelwert. So erhält man die Stichprobenverteilung ohne Nonresponse.
- Um den Fall von systematischem Nonresponse zu simulieren gehen wir davon aus, dass der folgende Selektionsmechanismus vorliegt. Die Indikatorfunktion I gibt an, ob die Ausprägung von X vorliegt ($I=1$) oder nicht ($I=0$).

$$I = \begin{cases} = 0 & \text{falls } X < 1'567 \\ = 1 & \text{falls } X \geq 1'567 \end{cases}$$

Um nun systematischen Nonresponse zu simulieren, ziehen wir wiederum 1'000 einfache Zufallsstichproben vom Umfang $n = 500$ aus der Population. Allerdings werden alle Beobachtungen entfernt, welche auf X eine Ausprägung kleiner als 1'567 aufweisen. Danach wird für jede Stichprobe der Mittelwert berechnet. So erhält man die Stichprobenverteilung des Mittelwertes mit Nonresponse.

In der unteren Grafik ist wiederum links oben die Verteilung von X in der Population dargestellt. Die unteren beiden Grafiken stellen die Stichprobenverteilung des Mittelwertes ohne beziehungsweise mit Nonresponse dar.



Fazit im Falle von systematischem Nonresponse (s. Grafik):

- Wie im Falle des unsystematischen Nonresponse ist der Standardfehler der Schätzung grösser. Dies ist auf die durch den Nonresponse bedingt Verkleinerung der Stichprobe zurückzuführen.
- Der Populationsmittelwert kann in diesem Fall jedoch ohne Korrektur nicht unverzerrt geschätzt werden.
- Und: Diese Verzerrung ist nicht abhängig von der Stichprobengrösse!

8. Zusammenfassung zum Lernschritt

In diesem Lernschritt wurden die folgenden Punkte besprochen:

- Es wurden verschiedene Arten der probabilistischen als auch der nicht probabilistischen Stichprobenziehung vorgestellt. Zudem wurden die Voraussetzungen sowie die Vor- und Nachteile der unterschiedlichen Stichprobenziehungen besprochen.
- Es wurden mögliche Fehlerquellen bei der Stichprobenziehung behandelt. Wir haben dabei insbesondere festgestellt, dass die Stichprobengrösse alleine noch keine präzisen Populationsschätzungen garantiert.

Fallbeispiel

Franklin D. Roosevelt beendete 1936 seine erste Amtszeit als Präsident der Vereinigten Staaten von Amerika. Sein republikanischer Herausforderer bei den Wahlen hiess Alfred Landon, Gouverneur aus Kansas. Zwar waren die Aufrüstung Deutschlands unter den Nationalsozialisten und der Spanische Bürgerkrieg Leitthemen in den grossen Tageszeitungen wie etwa der New York Times. Den Wahlkampf dominierten allerdings wirtschaftspolitische Themen. Die USA kämpften noch immer mit den Folgen der Grossen Depression: Mehr als neun Millionen Menschen waren arbeitslos, und die realen Einkommen waren zwischen 1929 und 1933 etwa um einen Drittel gesunken.

Die meisten Beobachter des Wahlkampfes sagten Roosevelt einen einfachen Sieg über seinen Herausforderer Landon voraus, nicht aber die bekannte Zeitschrift "Literary Digest". Diese prognostizierte Roosevelt einen Stimmenanteil von lediglich 43% und damit einen klaren Sieg für Landon. Die Zeitschrift basierte ihre Voraussage auf der grössten Stichprobe an Personen, welche je auf eine kommerzielle Umfrage geantwortet hatten: 2,4 Millionen! Zudem hatte das "Literary Digest" seit 1916 jeweils den richtigen Gewinner der Präsidentschaftswahlen vorausgesagt.

Wie wir allerdings wissen, hat dennoch Franklin Roosevelt die Wahl gewonnen, und zwar mit einer überwältigenden Mehrheit von 62% der Stimmen. Alfred Landon erzielte lediglich einen Stimmenanteil von 38% für seine Person! Selbst die riesige Stichprobe konnte somit nicht verhindern, dass das "Literary Digest" den wahrscheinlich grössten Prognosefehler in der Geschichte der Umfrageforschung publizierte: ganze 19 Prozentpunkte! Nicht zuletzt diese krasse Fehlprognose dürfte dazu geführt haben, dass die Zeitschrift wenig später (im darauffolgenden Jahr) eingestellt wurde.

Im Gegensatz zum "Literary Digest" sagte Georg Gallup, dessen eigenes Umfrageinstitut gerade im Aufbau war, korrekterweise Roosevelt als Sieger voraus, auch wenn seine eigenen Prognose ebenfalls um 6 Prozentpunkte vom tatsächlichen Ergebnis abwich. Erstaunlicherweise gelang es Gallup allerdings, das falsche Ergebnis des "Literary Digest" vor dessen Veröffentlichung mit einer Abweichung von lediglich einem Prozentpunkt korrekt vorauszusagen!

	Stimmenanteil (Roosevelt)	Abweichung (in Prozentpunkten)
Tatsächliches Ergebnis	62%	0
Prognose des "Literary Digest"	43%	-19
Gallups Prognose	56%	-6
Gallups Prognose des "Literary Digest"	44%	+1

Stichprobenziehung

Es stellt sich somit die Frage, weshalb die Umfrage des "Literary Digest" trotz der riesigen Stichprobe ein so falsches Resultat liefern konnte. Wir wollen uns zunächst die (kontrafaktische, wie wir unten sehen werden) Situation betrachten, dass tatsächlich eine korrekte Zufallsstichprobe gezogen wurde. Mit den vorhandenen Angaben ergibt sich das folgende 99%-Vertrauensintervall für den Stimmenanteil von Roosevelt:

$$\left[p - z_{(1-\alpha/2)} \sqrt{\frac{p(1-p)}{n}}, p + z_{(1-\alpha/2)} \sqrt{\frac{p(1-p)}{n}} \right]$$

Setzt man die obigen Angaben in diese Formel ein, erhält man das folgende 99%-Vertrauensintervall für den Stimmenanteil für Roosevelt:

$$[0.4294, 0.4306]$$

Unter der Annahme, dass das "Literary Digest" seine Stichprobe korrekt gezogen hätte, wäre die Wahrscheinlichkeit für einen Stimmenanteil von 62% für Roosevelt mehr als unwahrscheinlich gewesen!

Dies lässt nun die Feststellung zu, dass wohl eine sehr selektive Stichprobe vorgelegen haben muss:

- Das "Literary Digest" versandte insgesamt rund 10 Millionen Fragebögen. Allerdings wurden die angeschriebenen Personen nicht zufällig ausgewählt, sondern die meisten dieser Personen wurden aus vorhandenen Listen von Zeitschriftenabonnenten, Automobilbesitzern sowie aus Telefonverzeichnissen gezogen. Zu jener Zeit (1936) waren nun allerdings wohlhabendere Personen in solchen Listen klar übervertreten. Die Stichprobe war damit klarerweise verzerrt, indem reichere Personen mit einer grösseren Wahrscheinlichkeit Eingang in die Stichprobe fanden als ärmere Personen. Es waren aber gerade reichere Personen, welche Alfred Landon ihre Stimme gaben. Der erste Fehler des "Literary Digest" bestand somit in einem Selektionsfehler.
- Das zweite - und wahrscheinlich ausschlaggebende - Problem der Stichprobe war die enorme Zahl an Personen, welche nicht auf die Umfrage antworteten (nur gerade 24% aller angeschriebenen Personen schickten den Fragebogen zurück). Offensichtlich antworteten primär Personen mit einem starken Interesse am Ausgang der Umfrage. Und dies waren vor allem Personen, welche sich einen Wandel an der Spitze der USA wünschten und somit Alfred Landon ihre Stimme gaben. Wähler, welche Landon ihre Stimme gaben, antworteten weitaus häufiger auf die Umfrage der Zeitschrift als die Anhänger Roosevelts.

Die Stichprobe des "Literary Digest" war somit sowohl durch einen Nichtstichprobenfehler (nonresponse) als auch durch einen Stichprobenfehler (Selektionsfehler) verzerrt. Wie wir wissen, hilft bei einer solch selektiven Stichprobe auch eine noch so grosse Stichprobengrösse nicht, um Populationskennwerte präzise zu schätzen.