Beiträge
zur
Wirtschaftsgeographie
und
Regionalentwicklung

Nr. 1-2012

BAND 1-2012

Stichproben, Stichprobenauswahlverfahren und Berechnung des minimal erforderlichen Stichprobenumfangs

Ivo Mossig

Universität Bremen
Institut für Geographie

Institut für Geographie
Universität Bremen

Beiträge zur Wirtschaftsgeographie und Regionalentwicklung

Stichproben, Stichprobenauswahlverfahren und Berechnung des minimal erforderlichen Stichprobenumfangs

Ivo Mossig

Nr. 1-2012

Erscheinungsort: Bremen

Herausgeber: Prof. Dr. Ivo Mossig

Schriftleitung: Matthias Scheibner

Adresse: Universität Bremen
Institut für Geographie
Prof. Dr. Ivo Mossig
Bibliothekstraße 1
28359 Bremen

Tel.: 0421 / 218 67019
Fax: 0421 / 218 7183

E-Mail: mossig@uni-bremen.de

www.regionalentwicklung.uni-bremen.de

ISSN: 2191-124X

Bremen, Januar 2012

Vorwort

Die erste Fassung des vorliegenden Beitrags wurde bereits 1996 geschrieben. Es handelte sich damals um ein unveröffentlichtes Manuskript, das den Studierenden einer Lehrveranstaltung zur empirischen Sozialforschung am Institut für Geographie der Universität Gießen zur Verfügung gestellt wurde. In vielen der gängigen Lehrbücher zur empirischen Sozialforschung ist die Bestimmung des minimalen erforderlichen Stichprobenumfangs bei einer Zufallsstichprobe bis heute nicht enthalten. Vor diesem Hintergrund kam das Manuskript in den Folgejahren nicht nur in Gießen, sondern auch in Lehrveranstaltungen an den Universitäten in Heidelberg und in Bremen zum Einsatz. Auch wenn es bislang nie offiziell publiziert wurde, so wurde das Manuskript bereits mehrfach auch in wissenschaftlichen Arbeiten zitiert (vgl. http://scholar.google.de). An dieser Stelle möchte ich nicht versäumen darauf hinzuweisen, dass mich damals Harald Bathelt darauf aufmerksam gemacht hat, dass sich der minimal erforderliche Stichprobenumfang errechnen lässt und mir diesbezüglich zentrale Hinweise gegeben hat.

Nun gab es erneut einen guten Anlass, auf dieses Manuskript zu verweisen und ich entschied mich, das Manuskript zu überarbeiten und in der Reihe „Beiträge zur Wirtschaftsgeographie und Regionalentwicklung“ zu veröffentlichen. Dadurch kann es nun jede Person offiziell herunterladen und bei Bedarf zitieren.

Bremen, im Januar 2012

1 Grundbegriffe: Grundgesamtheit, Stichproben und Repräsentativität

Eine empirische Untersuchung verfolgt das Ziel, eine oder mehrere Aussagen über den Untersuchungsgegenstand der Erhebung zu machen. In der Sozialforschung ist der Untersuchungsgegenstand in der Regel eine Gruppe von Individuen (z.B. die Bewohner einer Stadt, die Studenten einer Universität, die Wahlberechtigten eines Landes) oder es sind Gemeinschaften von Individuen (z.B. alle Haushalte einer Gemeinde), die sich durch gemeinsame Merkmale auszeichnen. Je nach Fragestellung sind es ganz bestimmte Merkmale, welche den betreffenden Untersuchungsgegenstand kennzeichnen. Alle Individuen oder Gemeinschaften, welche diese Merkmale besitzen, bilden die Grundgesamtheit (oder Population) der Untersuchung. Als **Grundgesamtheit** wird die Menge aller Untersuchungselemente bezeichnet.

Die genaue Größe einer Grundgesamtheit ist nicht immer leicht feststellbar und setzt eine genaue Formulierung der Fragestellung voraus. So ist beispielsweise die Menge aller Kunden eines Ladengeschäfts endlich, doch ohne genaue Festlegung eines Untersuchungszeitraums unbestimmbar. Ist ein Zeitraum festgelegt (z.B. ein Jahr), so ist die Kundenzahl zwar feststellbar, aber mit einem enormen Erhebungsaufwand verbunden (vgl. Bahrenberg et al. 2010, S. 19).

Da in den meisten Fällen empirischer Sozialforschung eine Betrachtung jedes einzelnen Elements der Grundgesamtheit aus Zeit- und Kostengründen nicht erfolgen kann, wird eine Teilmenge aus der Grundgesamtheit ausgewählt. Diese Teilmenge wird in der Statistik als **Stichprobe** (bzw. Sample) bezeichnet. Man beachte, dass Grundgesamtheit und Stichprobe bei verschiedenen Fragestellungen variieren und somit keine festen Größen sind. So können die Studenten eines einzelnen Fachbereichs die Stichprobe einer Untersuchung über alle Studenten dieser Universität bilden. Im Rahmen einer anderen Fragestellung, die lediglich den Fachbereich betrifft, bilden diese Studenten dann selbst die Grundgesamtheit (vgl. Atteslander 1971, S. 206 f.).

Ziel einer jeden Stichprobenauswahl ist es, Rückschlüsse von den Ergebnissen der Stichprobe auf die Eigenschaften der Grundgesamtheit ziehen zu können. Ist dies gewährleistet, so bezeichnet man die Stichprobe als **repräsentativ** (Bahrenberg et al. 2010, S. 19). Ferner wird versucht, Aussagen über die Genauigkeit des Stichprobenergebnisses im Vergleich zur Befragung aller Individuen der Grundgesamtheit treffen zu können. Eine solche Untersuchung der kompletten Grundgesamtheit wird als Vollerhebung bezeichnet.

In den folgenden Kapiteln sollen nun verschiedene Auswahlverfahren zur Bildung einer Stichprobe und ein Berechnungsverfahren zur Bestimmung eines repräsentativen Stichprobenumfangs vorgestellt werden.

2 Stichprobenauswahlverfahren

Die Festlegung einer Stichprobe kann grundsätzlich nach zwei unterschiedlichen Prinzipien erfolgen. Einerseits nach dem Zufallsprinzip und andererseits durch bewusste Auswahlmethoden (Willkürstichprobe). Den zufallsgesteuerten Auswahlverfahren kommt dabei die besondere Bedeutung zu, dass eine Abschätzung des abweichenden „Fehlers“ der Stichprobe gegenüber der Grundgesamtheit möglich ist, sofern für jedes Element der Grundgesamtheit die Wahrscheinlichkeit bekannt ist, in die Stichprobe zu gelangen (vgl. Kap. 4). Eine solche Fehlerabschätzung kann für bewusste Auswahlverfahren nicht vorgenommen werden.

Tab. 1: Stichprobenauswahlverfahren

Zufällige Stichprobenauswahl	**Bewusste Auswahlverfahren** (Willkürstichprobe)
a) Reine Zufallsstichprobe	e) Quotenverfahren
b) Systematische Stichprobe	f) Konzentrationsprinzip
c) Geschichtete Zufallsstichprobe	
d) Klumpenstichprobe	

Quelle: Eigene Darstellung nach Bahrenberg et al. 2010, S. 20 ff., Reuber/Pfaffenbach 2005, S 52 ff., Atteslander 1971, S. 207 ff.

a) Reine Zufallsstichprobe (Randomverfahren):
Um eine reine Zufallsstichprobe zu bilden, muss die Grundgesamtheit endlich und lückenlos bekannt sein. Die Auswahl der Stichprobenelemente erfolgt dann anhand des sogenannten Urnenmodells (bzw. Lotterieverfahrens). Dabei wird für jedes Element der Grundgesamtheit eine Kugel oder ein entsprechend beschrifteter Zettel in eine Urne gelegt. Der Inhalt der Urne wird gut gemischt. Anschließend werden entsprechend des zuvor festgelegten Stichprobenumfangs Kugeln oder Zettel aus der Urne gezogen. Dadurch ist gewährleistet, dass jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit besitzt, in die Stichprobe zu gelangen (vgl. Atteslander 1971, S. 207). Heutzutage erfolgt die Auswahl einer Zufallsstichprobe in der Regel computergestützt durch die Anwendung einer entsprechenden Funktion in gängigen Softwareprogrammen.

b) Systematische Stichprobe:
Eine systematische Stichprobe ist eine Modifikation der reinen Zufallsstichprobe: Erneut besitzt jedes Element der Grundgesamtheit die gleiche Chance in die Stichprobe zu gelangen. Zuerst werden die Elemente der Grundgesamtheit in eine Rangordnung gebracht, die zufällig sein kann, oder den Ausprägungen einer bestimmten Variablen folgt, die unabhängig vom Untersuchungsgegenstand ist (z.B. die alphabetische Auflistung von Namen). Dann wird jedes x-te Element als Stichprobenelement ausgewählt, beginnend mit einem zufällig bestimmten Element, das zwischen 0 und x in der Rangordnung steht.

Zum Beispiel entsteht eine systematische Stichprobe, wenn aus dem alphabetisch geordneten Einwohnerverzeichnis einer Gemeinde jeder 20. Bewohner auswählt wird, beginnend mit einer zufällig bestimmten Person, die unter den ersten 20 Einträgen der alphabetischen Liste steht (vgl. Bahrenberg et al. 2010, S. 22).
Liegen Adressdaten vor, so kann z.B. die Grundgesamtheit nach Postleitzahlen sortiert werden. Eine systematische Stichprobe würde dann automatisch die räumliche Verteilung der Untersuchungselemente bei der Zusammensetzung der Stichprobe berücksichtigen.

c) Geschichtete Stichprobe:
Eine geschichtete Stichprobe ist vor allem dann sinnvoll, wenn die Grundgesamtheit sehr heterogen ist, also die Merkmalsausprägungen der Grundgesamtheit starke Unterschiede zeigen. Damit sämtliche Schattierungen der Grundgesamtheit in der Stichprobe ausreichend vertreten sind, müsste eine nach dem Prinzip der reinen Zufallsauswahl gebildete Stichprobe sehr groß sein, um Repräsentativität zu gewährleisten. In einem solchen Fall bietet sich eine geschichtete Stichprobe an.
Man erhält eine geschichtete Stichprobe durch Aufteilung der Grundgesamtheit in disjunkte[1] Klassen (Schichten): Die Klasseneinteilung (Schichtung) soll dabei so erfolgen, dass sich die Elemente einer jeden Klasse bezüglich der untersuchten Frage möglichst ähnlich verhalten, jedoch die Elemente aus unterschiedlichen Klassen verschiedene Eigenschaften besitzen. Anschließend wird dann aus jeder Klasse eine reine Zufallsstichprobe nach den Vorgaben zur Bestimmung der Fehlergenauigkeit (vgl. Kap 4) gezogen. Die Stichproben der einzelnen Schichten werden zur Gesamtstichprobe vereinigt. Die einzelnen Elemente besitzen durch dieses Vorgehen zwar nicht mehr die gleiche Wahrscheinlichkeit, in die Gesamtstichprobe zu gelangen, doch ist für jedes Element die individuelle Wahrscheinlichkeit bekannt. Die Einhaltung des Zufallsprinzips bleibt ausreichend gewährleistet.
Soll beispielsweise das Freizeitverhalten der Bevölkerung einer Stadt ermittelt werden, so wäre eine Einteilung der Bevölkerung in verschiedene Altersklassen sinnvoll, aus denen dann die jeweiligen Stichproben gezogen werden. Dieser Vorgehensweise würde die Vermutung zugrunde liegen, dass sich die Freizeitgestaltung mit zunehmendem Alter der Personen ändert. Eine mehrfache Schichtung durch mehrere Variablen kann natürlich auch vorgenommen werden, indem beispielsweise die oben erwähnten Altersklassen zudem nach dem Geschlecht der Personen ausdifferenziert (geschichtet) werden (vgl. Bahrenberg et al. 2010, S. 22 f.).

d) Klumpenstichprobe:
Von der Grundidee entspricht die Klumpenstichprobe einem Fallbeispiel, bei dem auf Grund einer detaillierten Einzelbetrachtung verallgemeinerbare Erkenntnisse gewonnen werden sollen. Eine Klumpenstichprobe ist nur möglich, wenn die Grundgesamtheit bereits in mehrere Gruppen bzw. „Klumpen“ (Cluster) unterteilt ist. Dieser Unterteilung sollte keine besondere Merkmalsausprägung zugrunde liegen, die für das Untersuchungsziel von besonderer Relevanz ist. Als Stichprobe wird dann eine Gruppe (ein Klumpen) zufällig gewählt und untersucht. Die Ergebnisse des Klumpens werden dann auf die anderen Gruppen übertragen (vgl. Hantschel/ Tharun 1980, S. 65).

[1] Disjunkt = überschneidungsfrei, d.h. jedes Element kann genau einer Klasse zugeordnet werden und nicht mehreren.

So kann beispielsweise das Freizeitverhalten der deutschen Großstadtbewohner anhand einer einzelnen Großstadt (z.B. den Bewohnern Münchens) untersucht werden, um auf das Freizeitverhalten aller Großstadtbewohner zu schließen. Dieses Beispiel zeigt offensichtlich, dass die Klumpenstichprobe am stärksten von der reinen Zufallsstichprobe abweicht und oftmals nicht repräsentativ ist. Es liegt in der Praxis sehr selten der Fall vor, dass die „natürliche" Gruppeneinteilung in keiner Wechselwirkung zum Untersuchungsziel steht. So wird in dem obigen Beispiel das Freizeitverhalten der Bewohner Münchens von den spezifischen Angeboten vor Ort geprägt sein, die sich z.B. aus der geographischen Lage Münchens (z.B. durch die Nähe zu den Alpen) ergeben, während sich in Hamburg aufgrund der naturräumlichen Lage ganz andere Freizeitaktivitäten anbieten. Folglich sind weder Hamburg noch München repräsentativ, um das Freizeitverhalten der deutschen Großstadtbewohner zu ermitteln. Zumal in der Praxis der untersuchte Klumpen normalerweise nicht zufällig ausgelost, sondern eine bewusst vorgenommene Auswahl argumentativ begründet wird. Es ist bei Klumpenstichproben daher genau zu prüfen, ob der gewählte Klumpen sich als repräsentative Stichprobe eignet oder ob er bezüglich der Fragestellung Eigenheiten aufweist, die das allgemeine Ergebnis verzerren (vgl. Bahrenberg et al. 2010, S. 23). Trotz dieser Mängel besitzt die Klumpenstichprobe in einigen Sonderfällen zwei erwähnenswerte Vorteile:

1. Die Untersuchung kann auf bestimmte, geographisch begrenzte Gebiete konzentriert werden. Dadurch können oft Wegkosten und Zeit gespart werden.
2. Durch die Konzentration auf einen Klumpen ist eine Vollerhebung dieses Klumpens leichter möglich. Dies ist von Bedeutung, wenn keine Listen über die Elemente der Grundgesamtheit existieren bzw. fehlerhaft sind und so ein anderes Auswahlverfahren verhindern. Möchte man z.B. die Studenten mit Wohnsitz in Bremen befragen und besitzt keine entsprechende Adressenliste, dann könnte so vorgegangen werden, dass man bestimmte Straßen oder Stadtteile als Klumpen auswählt. Auf Grund dieser räumlichen Eingrenzung ist es nun leichter möglich, alle dort wohnenden Studenten zu ermitteln und zu befragen (vgl. Atteslander 1971, 209 f.).

e) Das Quotenverfahren:

Das Quotenverfahren gehört zu den bewussten Auswahlverfahren und unterliegt nicht mehr dem reinen Zufallsprinzip. An dieser Stelle sei nochmals darauf hingewiesen, dass die statistischen Aussagen über die Fehlergenauigkeit bei einem willkürlichen Auswahlverfahren nicht gelten. Die Solidität der Stichprobe beruht in einem solchen Fall vornehmlich auf der Erfahrung des Forschers, d.h. vor allem auf seinen Kenntnissen über die Grundgesamtheit.

Das Quotenverfahren ist in der Markt- und Meinungsforschung stark verbreitet. Oberflächlich betrachtet besitzt es eine gewisse Ähnlichkeit mit der geschichteten Stichprobe, da wieder von einer stark heterogenen Grundgesamtheit ausgegangen wird. Die Grundgesamtheit wird erneut in solche Gruppen aufteilt, die eine möglichst homogene Struktur innerhalb jeder Gruppe bezüglich der Fragestellung erwarten lässt. Aufteilungsmerkmale können z.B. das Geschlecht, die Altersklasse oder der Bildungsgrad sein. Für jede dieser Gruppen wird nun eine Quote festgelegt, mit welchem Prozentanteil diese Gruppe in der Stichprobe vertreten sein soll. Bei der Festlegung der Quoten sind statistische Unterlagen wie statistische Jahrbücher häufig sehr nützlich, da sie für viele Merkmalsausprägungen die entsprechenden Quotenanteile an der Grundgesamtheit enthalten. Der Interviewer erhält dann entsprechend der festgelegten Quoten

genaue Anweisungen, wie viele Personen er je Gruppe befragen soll. Ein solcher Auftrag kann beispielsweise wie folgt lauten: Befragen sie bitte 50 weibliche und 40 männliche Studenten in den Bachelorstudiengängen und 30 weibliche und 25 männliche Studenten im Masterstudium. Innerhalb dieser Vorgaben kann der Interviewer nun nach verschiedenen Methoden die Interviewpartner auswählen. Man beachte, dass am Ende die Summe der Befragungen aller Interviewer mit den zuvor festgelegten Quoten übereinstimmen muss (vgl. Atteslander 1971, S. 210 f., Reuber/Pfaffenbach 2005, S. 59).
Der zentrale Unterschied zur geschichteten Stichprobe besteht darin, dass sich die Anzahl der ausgewählten Elemente in den einzelnen Schichten aus der Zusammensetzung der Grundgesamtheit ergibt. Beim Quotenverfahren findet eine bewusst willkürliche Festlegung statt, welche durchaus erheblich von den Anteilen in der Grundgesamtheit abweichen kann. Daher zählt das Quotenverfahren zu den Willkürstichproben.

f) Konzentrationsprinzip

Das Konzentrationsprinzip sieht die bewusste Ausklammerung bestimmter Einheiten aus der Grundgesamtheit vor, wenn die Fragestellung dies sinnvoll erscheinen lässt. Will man z.B. die Studierenden über die Qualität der Lehre in ihrem Studienfach befragen, so könnte es sinnvoll sein, a priori die Studierenden des ersten Semesters auszuklammern. Diesem Beispiel liegt die Annahme zugrunde, dass Erstsemester noch nicht genügend Erfahrungen gesammelt haben, um die Lehrqualität zu beurteilen. Als Nebenaspekt des Konzentrationsprinzips ist ein geringerer Erhebungsaufwand zu nennen.

3 Räumliche Perspektiven einer Stichprobe

Stichproben können auch unter räumlichen Gesichtspunkten diskutiert werden. An dieser Stelle soll auf zwei Aspekte eingegangen werden: Erstens die räumliche Dimension im Zuge der genauen Abgrenzung der Grundgesamtheit, aus der die Stichprobe gezogen wird und zweitens die räumlichen Implikationen im Hinblick auf die Auswahlmethode zur Bildung der Stichprobe.

Durch die jeweilige Definition der Grundgesamtheit ist es möglich, den räumlichen Bezugsrahmen der Ergebnisse zu verändern. Soll beispielsweise im Rahmen der Zentralitätsforschung der Kundeneinzugsbereich einer Stadt analysiert werden, so könnten sowohl die Passanten des innerstädtischen Einkaufsbereichs als auch die Haushalte in den umliegenden Gemeinden als Grundgesamtheit betrachtet und befragt werden. Bei der ersten Variante erhält der Forscher einen Überblick, woher die Passanten kommen, die den innerstädtischen Einkaufsbereich besuchen. Für die umliegenden Herkunftsorte ist jedoch nicht feststellbar, ob sich die dortige Bevölkerung überwiegend in dem fraglichen Einkaufsbereich versorgt oder ob es sich nur um Einzelfälle handelt, die in der untersuchten Stadt einkaufen. Man bekommt also keine Informationen darüber, welche Bedeutung diese Stadt im Vergleich zu anderen Versorgungszentren für die umliegenden Gemeinden hat. Dafür erhält man den maximalen Einzugsradius sowie die Anteile verschiedener Ortschaften an der Gesamtkundschaft des Einkaufsbereiches. Die alternative Möglichkeit, die Haushalte der umliegenden Stadtteile und Gemeinden zu befragen, entspricht einer anderen räumlichen Perspektive. Man könnte dadurch herausfinden, welche Bedeutung der betrachtete Einkaufsbereich in den verschiedenen Ortschaften des Einzugsgebietes besitzt und wo sich konkurrierende Standorte befinden. Eine absolute Außengrenze wäre durch dieses Vorgehen jedoch wesentlich schwieriger zu ermitteln.

Die räumliche Perspektive, die sich aus der gewählten Vorgehensweise zur Bildung der Stichprobe ergibt, wurde bereits im Rahmen der Darstellung der Klumpenstichprobe angedeutet (vgl. Kap. 2). Hier ergab sich ein enger Bezug zwischen der geographischen Lage einzelner Städte und der zu untersuchenden Fragestellung, die eine Auswahl als Klumpen erschwert bzw. die Repräsentativität in Frage stellt. Ebenso wurde darauf verwiesen, dass die systematische Stichprobe nach räumlichen Kriterien gebildet werden kann und durch eine entsprechende Sortierung die räumliche Verteilung der Stichprobenelemente mit der Grundgesamtheit übereinstimmt. Auch die geschichtete Stichprobe und das Quotenverfahren können unter räumlichen Aspekten angewendet werden, indem die erforderliche Gruppenbildung anhand der Lage im Raum vorgenommen wird. Ebenso kann das Konzentrationsprinzip in einigen Fällen anhand räumlicher Überlegungen durchgeführt werden. Bei einer reinen Zufallsstichprobe wäre eine solche räumliche Differenzierung jedoch nicht möglich. Es bleibt festzuhalten, dass die verschiedenen Stichprobenauswahlverfahren unterschiedliche Raumbezüge ermöglichen. Entscheidend für eine entsprechende Auswahl und Anwendung bleibt, ob dies zur Klärung der gegebenen Forschungsfrage sinnvoll oder gar notwendig ist.

4 Bestimmung des minimal erforderlichen Stichprobenumfangs

4.1 Allgemeine Vorüberlegungen

Die folgenden Ausführungen beschränken sich nicht auf sozialgeographische Befragungen, sondern finden in allen Gebieten der Geographie Anwendung, bei der durch empirische Versuche Forschungsergebnisse erzielt werden (z.B. bei Klimamessreihen). Jedes Mal hängt vom Umfang der Stichprobe (bzw. der Anzahl der Messungen) sowohl die Genauigkeit des Ergebnisses als auch der Erhebungsaufwand ab. Entsprechend soll ein Kalkül entwickelt werden, durch das bei einer vorgegebenen Genauigkeit der minimal erforderliche Stichprobenumfang bestimmt werden kann, um den Erhebungsaufwand so gering wie nötig zu halten.

Die folgenden allgemeinen Grundregeln sollten unabhängig von den mathematischen Berechnungen des Stichprobenumfangs beachtet werden (vgl. Bahrenberg et al. 2010, S. 20 f.):

1. Generell gilt das „Gesetz der großen Zahl". Dieses besagt, dass sich die Eigenschaften der Stichprobe mit wachsendem Stichprobenumfang den Eigenschaften der Grundgesamtheit annähern.
2. Je stärker die Werte der untersuchten Variablen streuen, desto größer sollte der Stichprobenumfang sein.
3. Stichprobenumfänge von weniger als 30 gelten allgemein als zu klein, um Repräsentativität zu erreichen.
4. Für die Repräsentativität ist weniger der relative Anteil der Stichprobe an der Größe der Grundgesamtheit von Bedeutung, sondern vielmehr die absolute Größe der Stichprobe.

Für eine mathematische Bestimmung des minimal erforderlichen Stichprobenumfangs muss die Untersuchung nach dem Zufallsprinzip organisiert sein, d.h.:

1. Jedes Element der Grundgesamtheit muss die theoretische Chance besitzen, in die Stichprobe zu gelangen. Die Wahrscheinlichkeit dafür muss bekannt sein.
2. Jedes Element repräsentiert die Grundgesamtheit zu gleichen Teilen.

Grundgesamtheiten können in endliche und in unendliche Grundgesamtheiten unterschieden werden. **Endliche Grundgesamtheiten** bestehen aus einer bestimmbaren, festen (also endlichen) Anzahl von Einzelelementen. Zur Ermittlung einer Stichprobe aus einer endlichen Grundgesamtheit kann das Urnenmodell (Lotterieverfahren) herangezogen werden, wobei ein herausgezogenes Element nicht wieder zurückgelegt werden darf. Beispiele aus der Praxis sind Befragungen von Personen und Haushalten. **Unendliche Grundgesamtheiten** entsprechen dem Urnenmodell mit Zurücklegen der gezogenen Elemente. Dadurch können unendlich viele Ziehungen vorgenommen werden. Unendliche Grundgesamtheiten liegen auch bei Messreihen von Klimadaten wie der Temperatur zugrunde, denn theoretisch kann an beliebig vielen Zeitpunkten (d.h. unendlich oft) ein Messwert abgelesen werden. Folglich existieren sowohl für eine endliche als auch für eine unendliche Grundgesamtheit jeweils eigene Berechnungsformeln für den minimal erforderlichen Stichprobenumfang.

4.2 Die Dichtefunktion der Normalverteilung

Beim Wurf eines Würfels kann nicht vorhergesagt werden, welche der sechs Zahlen geworfen wird. Jedoch lässt sich die Wahrscheinlichkeit W bestimmen, dass z.B. eine „Eins“ geworfen wird. Ob dieses Ereignis dann tatsächlich eintrifft, bleibt aber dem Zufall überlassen. Die Wahrscheinlichkeit W, dass ein bestimmtes Ereignis $X = a$ eintritt, ist allgemein wie folgt definiert:

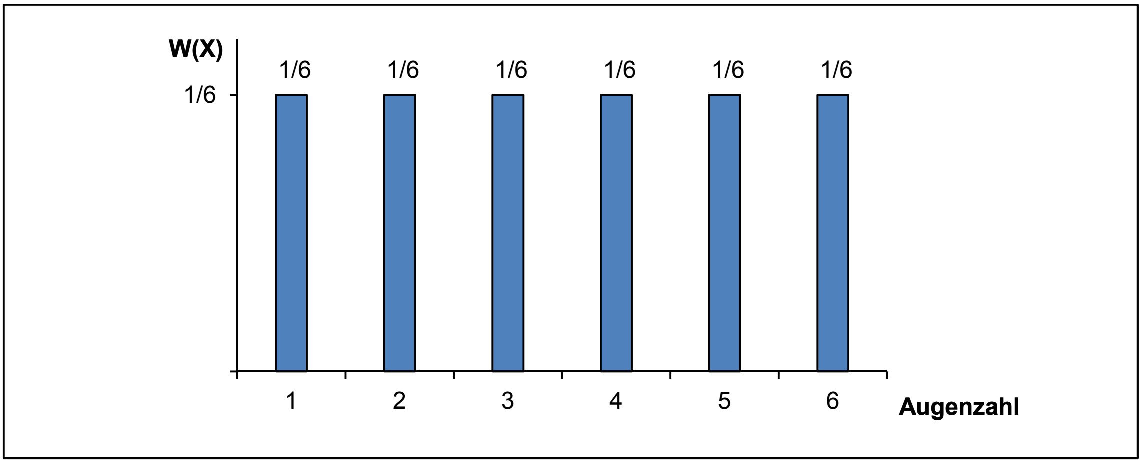
$$W(X = a) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.$$

Die Wahrscheinlichkeit z.B. eine „Eins“ zu würfeln ist demnach W(X = 1) = 1/6, denn von den sechs möglichen Ereignissen (Fällen) wird ein Fall als günstig betrachtet, nämlich wenn tatsächlich eine „Eins“ oben liegt. Entsprechend sind die Wahrscheinlichkeiten für den Wurf einer „Zwei“, „Drei“, „Vier“, „Fünf“ oder „Sechs“ ebenfalls jeweils 1/6. Ebenso berechnet man z.B. die Wahrscheinlichkeit, dass eine gerade Zahl geworfen wird:

$$W(X = 2{,}4{,}6) = \frac{3}{6} = \frac{1}{2} = 0{,}5.$$

Weiterhin existieren sechs mögliche Wurfereignisse (Fälle), die Anzahl der günstigen Fälle hat sich aber auf drei erhöht. Die Wahrscheinlichkeiten für die möglichen Wurfereignisse beim einmaligen Würfeln lassen sich als Stabdiagramm darstellen (vgl. Abb. 1).

Abb. 1: Wahrscheinlichkeiten der Ereignisse beim einmaligen Würfelwurf



Quelle: Eigene Darstellung

Allgemein sind folgende Eigenschaften von Wahrscheinlichkeiten festzuhalten:

1. Die Wahrscheinlichkeit ist immer eine Zahl zwischen 0 und 1, also $0 \leq W \leq 1$.[2]
2. Die Summe der Einzelwahrscheinlichkeiten ergibt immer 1.
   So gilt z.B. beim einmaligen Wurf eines Würfels:
   W(X=1)+W(X=2)+...+W(X=6) = 1/6 + 1/6 +...+ 1/6 = W(X=1,2,3,4,5,6) = 6/6 = 1.
   Dies wird auch als <u>sicheres Ereignis</u> bezeichnet.
3. Ein <u>unmögliches Ereignis</u> besitzt die Wahrscheinlichkeit 0.
   Der Wurf einer „Sieben“ ist beispielsweise ein unmögliches Ereignis, also W(X=7) = 0.

---

[2] Wahrscheinlichkeiten werden häufig in Prozentwert ausgedrückt. So entspricht der Wahrscheinlichkeit eines geraden Würfelwurfes W(X = 2,4,6) = 0,5 die Prozentangabe W(X = 2,4,6) = 50%.

Der tatsächliche Mittelwert der Zahlen 1,2,3,4,5,6 ist bekannt. Er beträgt $\mu = 3{,}5$. Im Folgenden soll aber angenommen werden, der Mittelwert sei unbekannt und soll durch ein Würfelexperiment bestimmt werden. Wir beginnen unser Experiment mit einmaligem Würfeln. Ein einzelner Wurf entspricht einer Stichprobe vom Umfang n=1. Die Wahrscheinlichkeit für jedes mögliche Ergebnis wurde bereits berechnet und in Abb. 1 dargestellt. Mit einer Wahrscheinlichkeit von jeweils 1/6 würde demnach der gesuchte Mittelwert eine der Zahlen von 1 bis 6 sein. Unser empirischer Versuch ist also bei einem Stichprobenumfang von n=1 sehr ungenau. Folglich muss der Stichprobenumfang erhöht werden.

Betrachten wir nun den Fall des zweifachen Wurfes (zwei Würfe nacheinander = Stichprobe vom Umfang n=2), wobei die jeweiligen Augenzahlen addiert und durch zwei (die Anzahl der Würfe) dividiert werden. Für jedes mögliche Ergebnis (Ereignis) ist nun zu berechnen, wie groß die Wahrscheinlichkeit ist, dass dieses Ereignis eintritt. Betrachten wir zunächst die Anzahl der möglichen Fälle: Beim zweimaligen Würfeln sind 36 Zahlenkombinationen möglich, denn jedes der sechs möglichen Ergebnisse des ersten Wurfes kann mit den sechs möglichen Ergebnissen des zweiten Wurfes kombiniert werden. Also gibt es $6 * 6 = 36$ Kombinationsmöglichkeiten. Nun muss für jedes mögliche Ereignis die Anzahl der günstigen Fälle ermittelt werden, um die Wahrscheinlichkeit zu bestimmen (vgl. Abb. 2):

- W(X=1) = 1/36, denn nur die eine Wurfkombination „1“ im ersten plus „1“ im zweiten Wurf ergibt (geteilt durch die Anzahl der Würfe) den Mittelwert 1.
- W(X=1,5) = 2/36 = 1/18. Die zwei günstigen Wurfkombinationen sind dabei „1“ im ersten und „2“ im zweiten Wurf sowie in umgekehrter Reihenfolge erst eine „2“ und dann eine „1“.

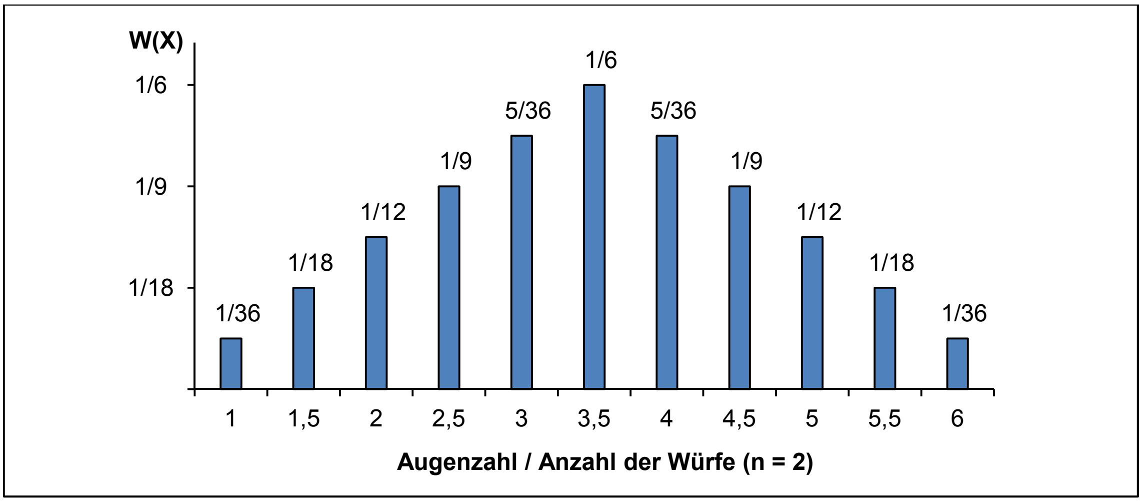
...

- W(X=3,5) = 6/36 = 1/6, mit den sechs günstigen Wurfpaaren (1;6), (2;5), (3;4), (4;3), (5;2) und (6;1).

...

- W(X=6) = 1/36. Die einzige günstige Wurfkombination ist (6;6).

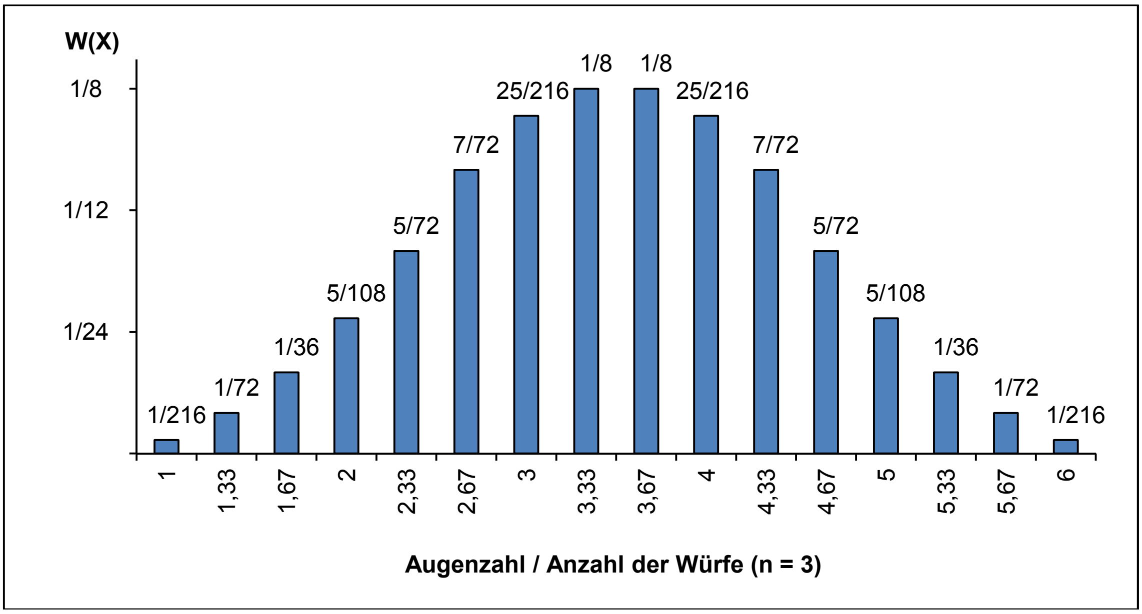
Abb. 2: Wahrscheinlichkeiten der Ereignisse bei zweimaligem Würfelwurf



Quelle: Eigene Darstellung

Wir sehen, dass sich die Wahrscheinlichkeiten für die möglichen Ereignisse verändert haben. In 1/6 der Fälle liegt bei zweimaligem Würfeln die durchschnittliche Augenzahl tatsächlich bei 3,5 und ist damit das Ereignis mit der höchsten Wahrscheinlichkeit, gefolgt von den beiden nächstliegenden mittleren Augenzahlen 3 und 4. Dennoch ist die Wahrscheinlichkeit, durch zweimaliges Würfeln den tatsächlichen Mittelwert zu erheben, mit 1/6 oder 16,67% ein sehr unbefriedigendes Ergebnis.

Abb. 3: Wahrscheinlichkeiten der Ereignisse bei dreifachem Würfelwurf



Quelle: Eigene Darstellung

Würde man das Würfelexperiment weiterführen und den Stichprobenumfang weiter erhöhen, zum Beispiel auf n=10 (zehnmaliges Würfeln), n=20 (zwanzigfaches Würfeln) usw., so würde man das Ergebnis laufend verbessern. Die Wahrscheinlichkeit, den tatsächlichen Mittelwert μ=3,5 zu treffen, würde sich sukzessiv erhöhen. Anhand des Würfelexperiments kann man also studieren, wie sich bei einer Zunahme des Stichprobenumfangs die Wahrscheinlichkeit erhöht, mit der Wurfkombination den tatsächlichen Mittelwert zu ermitteln.

In der Praxis wird nicht versucht die Wahrscheinlichkeit zu bestimmen, dass der tatsächliche Mittelwert genau getroffen wird. Dazu wäre letztlich eine Vollerhebung der Grundgesamtheit erforderlich, welche stets zum exakten Mittelwert führt. Gegenüber einer Stichprobe bedeutete aber eine Vollerhebung, dass keine Reduzierung des Aufwands und der Kosten erfolgt. Daher wird ein Ergebnisbereich um den gesuchten Mittelwert vorgegeben, der durch die Stichprobe erreicht werden soll. Dieser wird als „tolerierter Fehler ε“ bezeichnet. Der tolerierte Fehler ε gibt an, um welchen Betrag (um wie viel Prozentpunkte) der ermittelte Wert der Stichprobe höchstens vom tatsächlichen Wert der Grundgesamtheit abweichen darf:

$\varepsilon = |\mu - \bar{x}|$; mit μ = tatsächlicher Mittelwert und

$\bar{x}$ = Mittelwert des empirischen Versuches.

Für das Würfelbeispiel könnte man beispielsweise einen tolerierten Fehler von ε=0,5 festlegen. Wir würden somit bei unserem Würfelexperiment ein Ergebnis akzeptieren, das um maximal 0,5 vom tatsächlichen Mittelwert μ=3,5 abweicht. Folglich ist zu berechnen, mit welcher Wahrscheinlichkeit unser Würfelergebnis innerhalb des tolerierten Fehlers ε=0,5 um den tatsächlichen Mittelwert (also zwischen 3 und 4) liegt. Bei zweimaligem Würfeln sind drei Mittelwerte (Ereignisse) möglich, die in diesem Bereich liegen, nämlich X=3, X=3,5 sowie X=4.
Da die Gesamtwahrscheinlichkeit $W(3 \leq X \leq 4)$ die Summe der Einzelwahrscheinlichkeiten ist, gilt (vgl. Abb. 2):

$$W(3 \leq X \leq 4) = W(X=3) + W(X=3{,}5) + W(X=4) = 5/36 + 1/6 + 5/36 = 4/9 \approx 44{,}4\%.$$

Mit einer Wahrscheinlichkeit von 44,4% kann man also davon ausgehen, dass bei zweimaligem Würfeln unser Experiment ein Ergebnis innerhalb des gewählten tolerierten Bereichs um den Mittelwert liefert. Entsprechend niedriger ist die Wahrscheinlichkeit, mit nur einem Wurf innerhalb des tolerierten Fehlers von ε=0,5 zu liegen:

$$W(3 \leq X \leq 4) = W(X=3) + W(X=4) = 1/6 + 1/6 = 1/3 \approx 33{,}3\% \text{ (vgl. Abb. 1).}$$

Für einen Stichprobenumfang von n=3 (dreimaliges Würfeln) ergibt sich hingegen eine höhere Wahrscheinlichkeit (vgl. Abb. 3):

$$\begin{aligned} W(3 \leq X \leq 4) &= W(X=3) + W(X=3{,}33) + W(X=3{,}67) + W(X=4) \\ &= \frac{25}{216} + \frac{27}{216} + \frac{27}{216} + \frac{25}{216} = \frac{104}{216} = 48{,}15\%. \end{aligned}$$

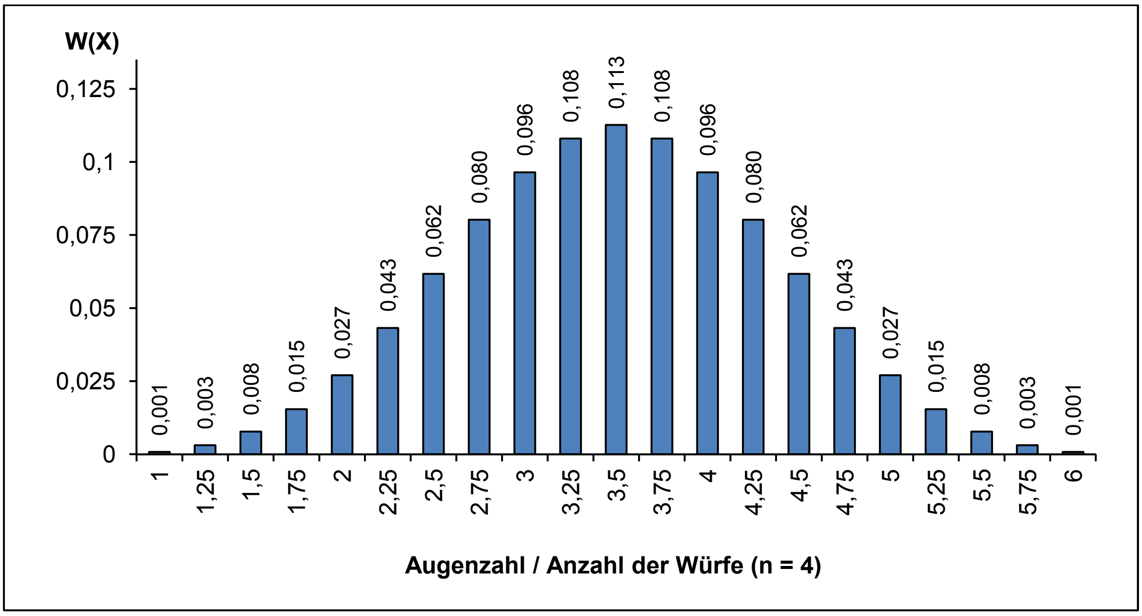
Mit zunehmendem Umfang der Stichprobe erhöht sich somit auch die Wahrscheinlichkeit, dass der gesuchte Mittelwert der Grundgesamtheit innerhalb des vorgegebenen tolerierten Fehlerbereichs von ε=0,5 liegt.

Würde man in einer Wahlprognose den tolerierten Fehler auf einen Prozentpunkt festlegen, so kann bei gegebenem Stichprobenumfang ausgerechnet werden, dass die empirisch ermittelte Prognose für die ABC-Partei mit x-prozentiger Wahrscheinlichkeit um maximal 1 Prozentpunkt vom tatsächlichen Wahlergebnis abweicht. Die Abweichung kann dabei natürlich sowohl nach oben als auch nach unten stattfinden.

Wie groß man den tolerierten Fehler wählt, hängt entscheidend von der Fragestellung der Untersuchung ab. So wird in empirischen Tests von Arzneimitteln ein wesentlich geringerer Fehler erlaubt sein als bei der Wahlprognose oder einer Haushaltsbefragung. Generell gilt, je kleiner der tolerierte Fehler gewählt wird, desto geringer ist die Wahrscheinlichkeit, dass die Stichprobe ein Ergebnis liefert, das innerhalb des tolerierten Bereichs liegt. Folglich ist der Stichprobenumfang mit kleiner werdendem toleriertem Fehler zu erhöhen, damit das Stichprobenergebnis mit genügend hoher Wahrscheinlichkeit in diesem Bereich liegen wird.

Zurück zu unserem Würfelexperiment. Aufgrund der Vergrößerung des Stichprobenumfangs von einem auf drei Würfe hat sich die Wahrscheinlichkeit von 33,3% (n=1) über 44,4% (n=2) auf 48,15% (n=3) erhöht, mit dem Experiment um maximal 0,5 vom tatsächlichen Mittelwert abzuweichen. Um ein Ergebnis im tolerierten Bereich mit möglichst hoher Wahrscheinlichkeit zu erhalten, muss der Stichprobenumfang weiter vergrößert werden. Die Berechnung der günstigen Fälle wird aber mit jedem weiteren Wurf zunehmend mühsamer, denn schon bei n=4 gibt es 1296 mögliche Wurfkombinationen (vgl. Abb. 4).

Abb. 4: Wahrscheinlichkeiten der Ereignisse bei einem viermaligen Würfelwurf



Quelle: Eigene Darstellung

Es kommt einem daher zugute, dass ab einem Stichprobenumfang von n=30 die Wahrscheinlichkeitsverteilung eines reinen Zufallsexperiments durch die Dichtefunktion der Normalverteilung („Gauß´sche Glockenkurve“) approximiert werden kann.[3]

Die Dichtefunktion der Normalverteilung hat folgenden Funktionsterm:

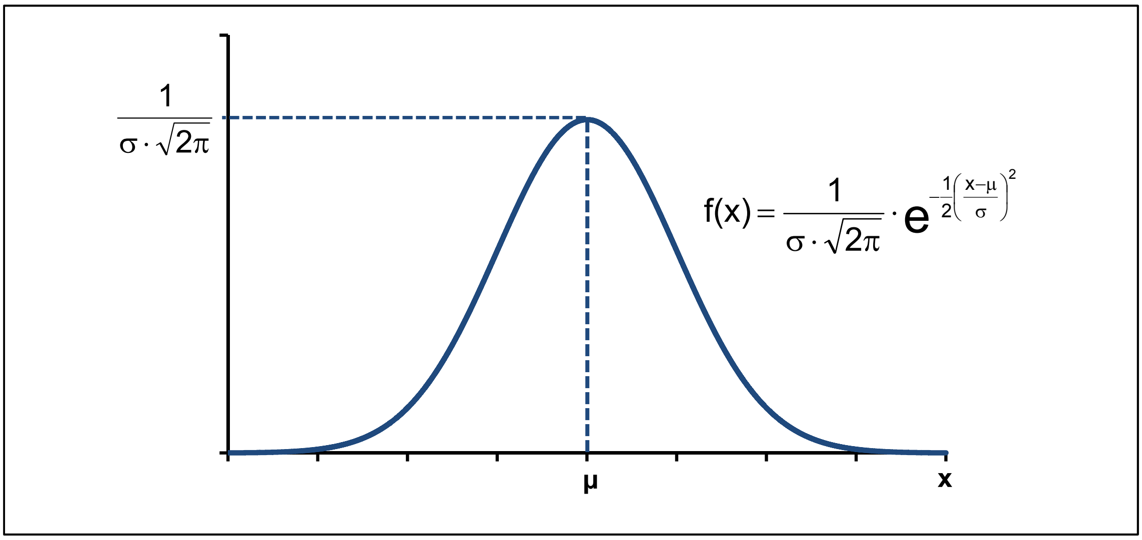
$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2};$$

mit σ = Standardabweichung und
μ = tatsächlicher Mittelwert.

Wie aus dem Funktionsterm erkennbar ist, hängt die Dichtefunktion der Normalverteilung von zwei Parametern ab: Der Standardabweichung σ und dem Mittelwert μ. Der generelle Funktionsverlauf ist in Abbildung 5 dargestellt.

[3] U.a. aus diesem Grund gelten Stichprobenumfänge von weniger als 30 im Allgemeinen als zu klein, um Repräsentativität zu erreichen.

Abb. 5: Die Dichtefunktion der Normalverteilung (Gauß´sche Glockenkurve)



Quelle: Eigene Darstellung

Die Gauß´sche Glockenkurve besitzt ein lokales Maximum an der Stelle x=μ mit

$$f(x = \mu) = \frac{1}{\sigma\sqrt{2\pi}}.$$

Sie ist symmetrisch zu der senkrechten Achse x=μ. Man sagt auch, sie ist symmetrisch um μ, den tatsächlichen Mittelwert. Ferner ist zu beachten, dass die Kurve Wahrscheinlichkeitsdichten darstellt. Das bedeutet, dass die Fläche unterhalb der Kurve - in den Grenzen eines gewählten Intervalls - die Wahrscheinlichkeit angibt, dass die Stichprobe ein Ergebnis innerhalb dieses Intervalls liefert. Diese Fläche kann durch das entsprechende Integral berechnet werden.

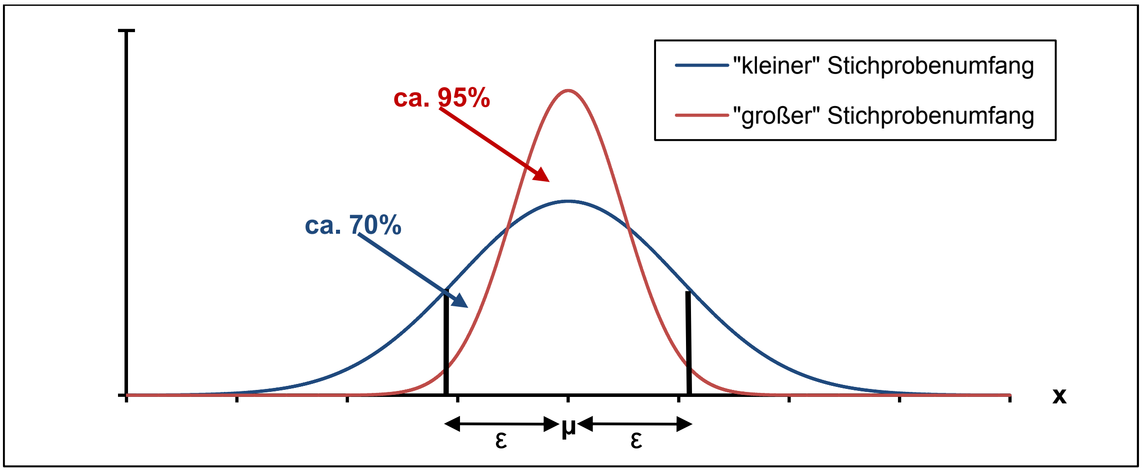
Ein Beispiel: Betrachtet wird wieder die durchschnittlich geworfene Augenzahl eines Würfels. Die Anzahl der Würfe (Stichprobenumfang) sei $n \geq 30$ und der tolerierte Fehler wie in den Beispielen zuvor $\varepsilon=0{,}5$. Die Wahrscheinlichkeit, mit der die durchschnittliche Augenzahl um maximal 0,5 vom tatsächlichen Mittelwert $\mu=3{,}5$ abweicht (also im tolerierten Bereich zwischen 3 bis 4 liegt), berechnet sich also durch:[4]

$$W(3 \leq X \leq 4) = \int_3^4 f(x)\,dx \quad ; \text{mit } f(x) = \text{Dichtefunktion der Normalverteilung.}$$

[4] Die Dichtefunktion der Normalverteilung bezieht sich nur auf stetige Variablen. Die durchschnittlichen Augenzahlen sind jedoch diskret. Ein exakteres mathematisches Vorgehen, z.B. die Mittelwerte der Würfe zuerst Klassen zuzuordnen und die Klassenbreite gegen den Grenzwert 0 gehen zu lassen, wird an dieser Stelle aus didaktischen Gründen vermieden.

Die Dichtefunktion der Normalverteilung unterliegt dem „Gesetz der großen Zahl". Für einen steigenden Stichprobenumfang (steigende Anzahl der Würfe) erhöht sich die „Treffergenauigkeit" (die Wahrscheinlichkeit), mit dem Ergebnis innerhalb des vom tolerierten Fehler ε vorgegebenen Intervalls zu liegen (vgl. Abb. 6). Der steilere Verlauf der Dichtefunktion für „große" Stichprobenumfänge entsteht dadurch, dass mit zunehmender Stichprobe die Standardabweichung σ kleiner wird.

Abb. 6: Verlauf der Dichtefunktion für „kleine" und für „große" Stichprobenumfänge
(Prozentangaben zum jeweiligen Flächeninhalt sind geschätzt)



Quelle: Eigene Darstellung

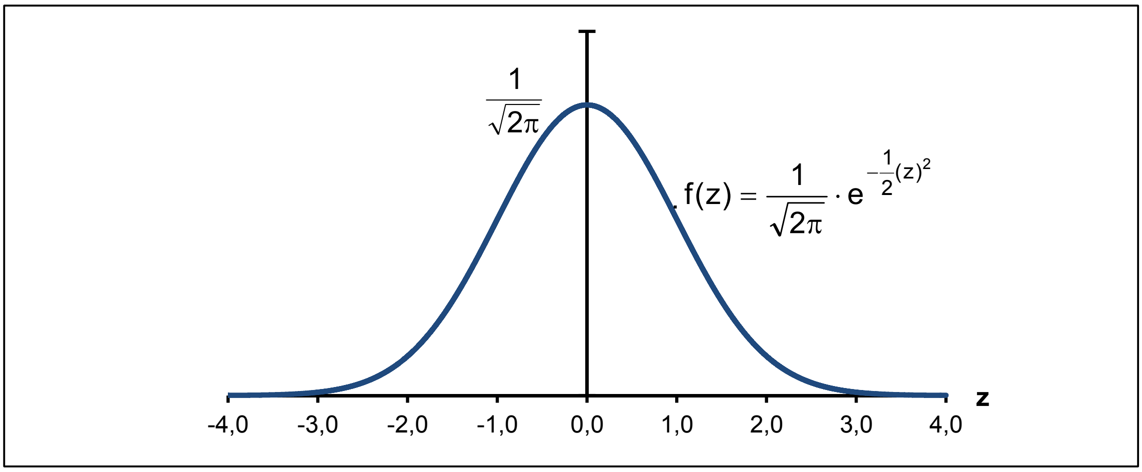
Die Berechnung des Integrals über die Dichtefunktion der Normalverteilung ist extrem schwierig und nur durch aufwendige numerische Verfahren näherungsweise möglich. Aus diesem Grund werden die Variablen in eine standardisierte Form gebracht. Die Variable X wird dabei durch die folgende Transformation in die standardisierte Variable Z überführt (vgl. Bahrenberg et al. 2010, S. 79 f.):

$$z = \frac{x - \mu}{\sigma}$$

Der besondere Vorteil ist, dass für standardisierte Variablen der Mittelwert stets μ=0 und die Standardabweichung σ=1 ist. Die Dichtefunktion der Normalverteilung geht dadurch in die sogenannte Standardnormalverteilungsfunktion über, die den folgenden Funktionsterm besitzt:

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(z)^2}$$

Abb. 7: Die Standardnormalverteilungsfunktion



Quelle: Eigene Darstellung

Die Standardnormalverteilung ist symmetrisch bezüglich der Y-Achse und besitzt an der Stelle z=0 ein lokales Maximum mit

$$f(z=0) = \frac{1}{\sqrt{2\pi}}.$$

Jeder Datensatz lässt sich standardisieren. Daher wurde die Arbeit geleistet, eine Tabelle für verschiedene standardisierte Intervallgrenzen zu errechnen, aus der die Werte der Integrale für verschiedene z-Werte abgelesen werden können (vgl. Tabelle im Anhang).

Die erste Spalte gibt die entsprechenden z-Werte der Abszisse vor.

Die zweite Spalte, ϕ (z), gibt die Wahrscheinlichkeit für alle Ereignisse an, die in standardisierter Form kleiner als z sind (vgl. Abb. 8). Dies ist gleichbedeutend mit dem Flächeninhalt unter der Standardnormalverteilungsfunktion von -∞ bis z. Oder mathematisch formuliert:

$$\phi(z) = \int_{-\infty}^{z} f(z)\,dz.$$

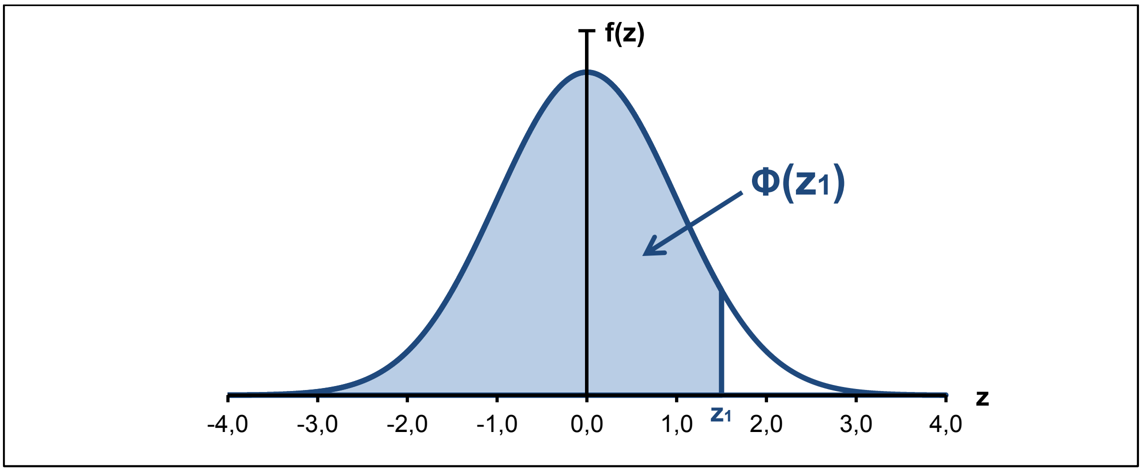
Man beachte, dass wegen der Achsensymmetrie die Flächen unterhalb des Funktionsgraphen links und rechts der Ordinate jeweils gleich groß sind. Da der gesamte Flächeninhalt von -∞ bis +∞ gleich der Gesamtwahrscheinlichkeit ist (also genau gleich 1), gilt entsprechend:

$$\phi(-z) = 1 - \phi(z).$$

Da die Wertetabelle keine Werte z<0 enthält, ist dieser Zusammenhang von Bedeutung, wenn der betrachtete z-Wert negativ ist. Die entsprechende Wahrscheinlichkeit für dieses negative z ist dann nach obiger Formel

$$1 - \phi(|z|).$$

Abb. 8: Linksseitige Wahrscheinlichkeit der Standardnormalverteilung



Quelle: Eigene Darstellung

<u>Beispiele für linksseitige Wahrscheinlichkeiten:</u>

a) ϕ (0,5) = W($Z \leq 0{,}5$) = 0,6915 = 69,15% (vgl. Tabelle im Anhang)

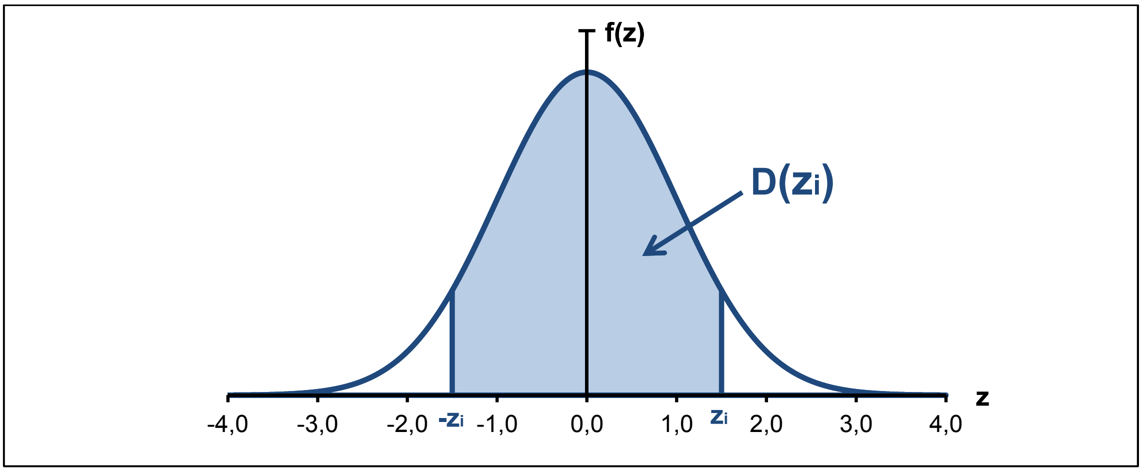
b) ϕ (2) = W($Z \leq 2$) = 0,9772 = 97,72%

c) ϕ (-1) = W($Z \leq -1$) = 1 - ϕ (1) = 1 - 0,8413 = 0,1587 = 15,87%.

Die dritte Spalte der Wertetabelle gibt zu den entsprechenden z-Werten die sogenannte zentrale Wahrscheinlichkeit D(z) an. Sie entspricht dem Flächeninhalt unter der Standardnormalverteilungsfunktion zwischen den Werten $-z_i$ bis z_i (vgl. Abb. 9). Zentrale Wahrscheinlichkeiten werden mathematisch exakt beschrieben durch:

$$D(z) = \int_{-z}^{z} f(z)\,dz\,.$$

Abb. 9: Zentrale Wahrscheinlichkeit der Standardnormalverteilung



Quelle: Eigene Darstellung

Beispiele für zentrale Wahrscheinlichkeiten:

a) $D(0,5) = W(-0,5 \leq Z \leq 0,5) = 0,3829 = 38,29\%$ (vgl. Tabelle im Anhang)
b) $D(1,0) = W(-1,0 \leq Z \leq 1,0) = 0,6827 = 68,27\%$.

Als Anwendungsbeispiel soll die Qualitätskontrolle von Schokoladentafeln betrachtet werden. Angenommen ein Hersteller verspricht seinen Kunden, seine Schokoladentafeln hätten ein durchschnittliches Gewicht von 100gr. pro Tafel. Durch umfangreiche Stichproben hat der Hersteller ermittelt, dass seine Tafeln tatsächlich ein mittleres Gewicht von μ=100gr. haben. Jedoch weichen einzelne Tafeln davon ab. Die ermittelte Standardabweichung beträgt σ=2,5gr. Anhand dieser Messwerte kann der Hersteller nun berechnen, wie groß die Wahrscheinlichkeit ist, dass ein Kunde zufälligerweise eine Schokoladentafel mit einem Gewicht unter x=95gr. kauft.

1. Standardisieren der Variablen $z = \frac{x - \mu}{\sigma}$:

$W(X \leq 95) = W(Z \leq [95-100]/2,5) = W(Z \leq -2,0)$.

2. Berechnen der linksseitigen Wahrscheinlichkeit für z=-2,0:
   $\phi(-2,0) = 1 - \phi(2,0) = 1 - 0,9772 = 0,0228 \Rightarrow 2,28\%$.
3. Antwort: Die Wahrscheinlichkeit, dass eine Tafel leichter als 95gr. ist, beträgt 2,28%.

4.3 Berechnung des minimal erforderlichen Stichprobenumfangs

Um die Dichtefunktion der Normalverteilung und die daraus resultierende Standardnormalverteilungsfunktion darzustellen, wurde bislang ein feststehender Stichprobenumfang unterstellt. Anhand des gewählten tolerierten Fehlers ε wurde ein Intervall festgelegt und berechnet, wie hoch der Prozentanteil der möglichen Ereignisse ist, die in den tolerierten Fehlerspielraum fallen.

Unser Ziel besteht nun darin, dass nicht nur der tolerierte Fehler, sondern auch die Wahrscheinlichkeit vorgegeben wird, mit der die empirische Untersuchung ein Ergebnis innerhalb des tolerierten Fehlerintervalls hervorbringt. Letzeres wird als Sicherheitswahrscheinlichkeit bezeichnet. Aus der Vorgabe des tolerierten Fehlers und der gewünschten Sicherheitswahrscheinlichkeit kann dann der minimal erforderlichen Stichprobenumfang errechnen werden. Dabei ist jedoch zu berücksichtigen, ob eine endliche oder unendliche Grundgesamtheit zugrunde liegt.

Eine mögliche Aufgabenstellung könnte wie folgt lauten: Wie groß muss bei einer Haushaltsbefragung mit einer Grundgesamtheit von N = 5000 Haushalten der Stichprobenumfang n sein, damit das Ergebnis mit einer Sicherheitswahrscheinlichkeit von 95% um nicht mehr als 5 Prozentpunkte (tolerierter Fehler ε = 0,05) vom tatsächlichen Ergebniswert abweicht.

Bei einer unendlichen Grundgesamtheit lässt sich der minimal erforderliche Stichprobenumfang nach der folgenden Formeln berechnen:[5]

Minimal erforderlicher Stichprobenumfang für eine unendliche Grundgesamtheit:

$$\boxed{n \geq z^2 \cdot \frac{P \cdot Q}{\varepsilon^2}} \qquad (1)$$

n = minimal erforderlicher Stichprobenumfang für eine unendliche Grundgesamtheit.

z = aus der zentralen Wahrscheinlichkeit der Standardnormalverteilung berechneter Wert der gewählten Sicherheitswahrscheinlichkeit.
Wurde beispielsweise eine Sicherheitswahrscheinlichkeit von 95% vorgegeben, so muss derjenige z-Wert aus der Wertetabelle abgelesen werden, für den gilt:
$D(z) = 95\ \% \Rightarrow z = 1{,}96$ (vgl. Tabelle im Anhang).[6]

ε = gewählter tolerierter Fehler.

P = tatsächlicher Mittelwert der Grundgesamtheit.
Oftmals entspricht dies dem prozentualen Anteilswert an der Grundgesamtheit. Soll beispielsweise der Stimmenanteil der ABC-Partei bei der nächsten Wahl ermittelt werden, so entspricht P dem Anteil der Wähler, die für diese Partei stimmen.
Beachte: P liegt zwischen 0% und 100%; dies entspricht $0 \leq P \leq 1$.

Q = 1-P.
Im Falle der Wahlprognose entspricht Q dem Anteil an der Wählerschaft, die nicht die ABC-Partei wählen.
Ebenfalls gilt $0 \leq Q \leq 1$.

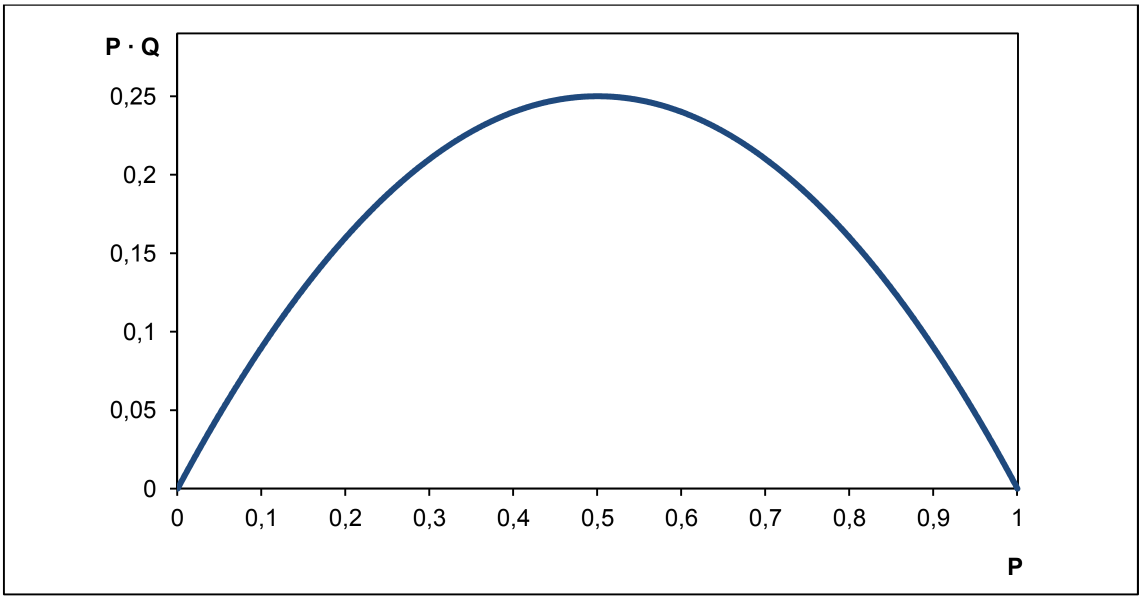
Wie man der Formel (1) entnehmen kann, wird der minimale Stichprobenumfang n in starkem Maße vom Produkt P·Q bestimmt. Mit wachsendem Produktwert P·Q erhöht sich auch der Stichprobenumfang n. Der tatsächliche Mittelwert P (und damit auch Q) ist unbekannt und soll erst durch die Befragung ermittelt werden. Man schätzt daher P so, dass das Produkt P·Q den größtmöglichen Wert annimmt, um selbst für den ungünstigsten Fall einen hinreichend großen Stichprobenumfang n zu errechnen.

Aus Abbildung 10 kann abgelesen werden, dass für P=0,5 das Produkt P·Q=0,25 maximal ist. Da Q = 1-P direkt von P abhängt, ist somit auch Q=0,5 zu wählen.

[5] Eine exakte mathematische Herleitung der Formeln würde an dieser Stelle zu weit führen. Es sei daher auf Rinne (1995, S. 370 f.) verwiesen.

[6] Für manche Sicherheitswahrscheinlichkeiten ist in der Tabelle kein z-Wert direkt ablesbar, wie z.B. für eine Sicherheitswahrscheinlichkeit von 99%, mit D(z=2,57)=98,98% und D(z=2,58)=99,01%. In einem solchen Fall ist der arithmetische Mittelwert zu bilden. Für eine gewählte Sicherheitswahrscheinlichkeit von 99% wäre also z=2,575 in obige Formel einzusetzen.

Abb. 10: Graphische Darstellung der Funktion P·Q mit Q = 1 - P und 0 ≤ P ≤ 1.
P·Q ist maximal für P=Q=0,5.



Quelle: Eigene Darstellung

Beispiel:

Es soll der minimal erforderliche Stichprobenumfang einer unendlichen Grundgesamtheit berechnet werden. Der tolerierte Fehler ε und die Sicherheitswahrscheinlichkeit sind wie folgt gewählt worden:

- tolerierter Fehler ε = 5%-Punkte $\Rightarrow$ ε = 0,05
- Sicherheitswahrscheinlichkeit 95% $\Rightarrow$ D(z) = 0,95 $\Rightarrow$ z = 1,96

$$n \geq z^2 \cdot \frac{P \cdot Q}{\varepsilon^2} = 1{,}96^2 \cdot \frac{0{,}5 \cdot 0{,}5}{0{,}05^2} = 384{,}16.$$

=> n = 385

Da als Stichprobenumfang nur ganze Zahlen in Frage kommen, wird stets auf die nächste größere ganze Zahl aufgerundet.

Minimal erforderlicher Stichprobenumfang für eine endliche Grundgesamtheit:

$$\boxed{n \geq \frac{N}{1+\frac{(N-1)\cdot\varepsilon^2}{z^2\cdot P\cdot Q}}} \qquad (2)$$

n = minimal erforderlicher Stichprobenumfang für eine endliche Grundgesamtheit.

N = Anzahl der Elemente in der Grundgesamtheit.

ε = gewählter tolerierter Fehler.

z = aus der zentralen Wahrscheinlichkeit der Standardnormalverteilung berechneter Wert der gewählten Sicherheitswahrscheinlichkeit (vgl. Formel (1)).

P = tatsächlicher Mittelwert der Grundgesamtheit bzw. prozentualer Anteilswert an der Grundgesamtheit (abzuschätzen wie nach Formel (1)).

Q = 1-P.

Beispiel: Minimal erforderlicher Stichprobenumfang für eine endliche Grundgesamtheiten mit N=4.548 Elementen (z.B. Haushalte in der Gemeinde XY) und den gewählten Vorgaben:
- tolerierter Fehler ε = 5%-Punkte $\Rightarrow \varepsilon = 0{,}05$
- Sicherheitswahrscheinlichkeit 99% $\Rightarrow D(z) = 0{,}99 \Rightarrow z = 2{,}575$:

$$n \geq \frac{N}{1+\frac{(N-1)\cdot\varepsilon^2}{z^2\cdot P\cdot Q}} = \frac{4548}{1+\frac{4547\cdot 0{,}05^2}{2{,}575^2\cdot 0{,}5\cdot 0{,}5}} = 578{,}8046.$$

=> n=579

Der Stichprobenumfang muss mindestens n=579 betragen, damit das Stichprobenergebnis bei einer Grundgesamtheit von 4.548 Elementen (Haushalten) mit einer Sicherheitswahrscheinlichkeit von 99% um nicht mehr als 5 Prozentpunkte vom tatsächlichen Ergebnis der Grundgesamtheit abweicht.

4.4 Einfluss der Parameter auf den Stichprobenumfang

In der Literatur zu Methoden der empirischen Sozialforschung wird häufig lediglich die Formel für unendliche Grundgesamtheiten angegeben (vgl. Atteslander 1971, Hantschel/Tharun 1980). Die Berechnungsformel für endliche Grundgesamtheiten, die in der empirischen Sozialforschung die wesentlich bedeutsamere Rolle spielt, ist weitaus seltener anzutreffen (z.B. Holm 1982, Rinne 1995). Dies ist vermutlich auf die wesentlich einfachere Handhabung der Formel für unendliche Grundgesamtheiten zurückzuführen.

Wird diese Formel auch bei einer endlichen Grundgesamtheit zur Berechnung des Stichprobenumfangs herangezogen, so bleiben die Aussagen bezüglich der Ergebnisgüte (maximal tolerierter Fehler, Sicherheitswahrscheinlichkeit) erhalten, denn der errechnete Stichprobenumfang für eine unendliche Grundgesamtheit ist stets größer als der Stichprobenumfang für eine endliche Grundgesamtheit. Das Ziel der Minimierung des Erhebungsaufwands wird in diesem Fall jedoch verfehlt. Tabelle 3 zeigt, dass für endliche und unendliche Grundgesamtheiten jeweils die entsprechende Formel Anwendung finden sollte, da sonst der Stichprobenumfang und damit der Aufwand um ein Vielfaches variieren kann.

Tab. 3: Stichprobenumfänge für endliche und unendliche Grundgesamtheiten bei gegebenen Sicherheitswahrscheinlichkeiten (S) und tolerierten Fehlern (ε)

gewählte Vorgaben:	Stichprobenumfang für endliche Grundgesamtheit (N=2700)	Stichprobenumfang für endliche Grundgesamtheit (N=6965)	Stichprobenumfang für unendliche Grundgesamtheit
S = 95% ε = 0,05	337	365	385
S = 99% ε = 0,05	533	606	664
S = 95 % ε = 0,01	2108	4038	9604

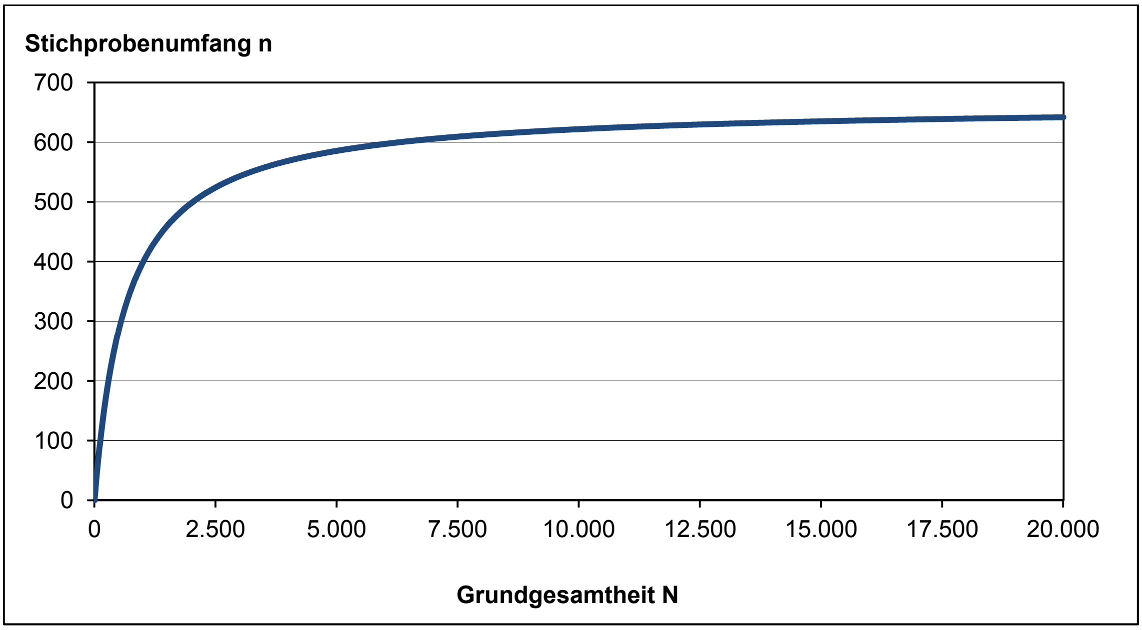
Quelle: Eigene Berechnungen

Nachfolgend soll untersucht werden, welchen Einfluss die drei Parameter Grundgesamtheit (N), tolerierter Fehler (ε) und Sicherheitswahrscheinlichkeit auf den minimal erforderlichen Stichprobenumfang ausüben. Beginnen wir mit der Frage, wie sich die Größe der Grundgesamtheit auf den minimal erforderlichen Stichprobenumfang auswirkt. In Abbildung 11 ist der minimal erforderliche Stichprobenumfang als Funktion in Abhängigkeit von der Größe der Grundgesamtheit N dargestellt. Ceteris paribus bleiben die Sicherheitswahrscheinlich S = 99% (=> z = 2,575) und der tolerierte Fehler ε=0,05 konstant.

Man erkennt, dass der minimal erforderliche Stichprobenumfang nicht proportional zur Größe der Grundgesamtheit zunimmt, sondern mit zunehmender Größe der Grundgesamtheit immer langsamer ansteigt. Dadurch kann erklärt werden, weshalb bei einer relativ großen Grundgesamtheit (z.B. bei einer Prognose für eine Bundestagswahl mit über 60 Mio. Wahlberechtigten) ein relativ kleiner Anteil der Grundgesamtheit für eine repräsentative Stichprobe ausreicht, während bei einer vergleichsweise kleinen Grundgesamtheit (z.B. eine Haushaltsbefragung in einem Dorf) der relative Anteil wesentlich größer sein muss. Dies bestätigt die Aussage, dass die Repräsentativität nicht durch den relativen Anteil der Stichprobe an der Grundgesamtheit erzeugt wird, sondern die absolute Größe der Stichprobe entscheidend ist.

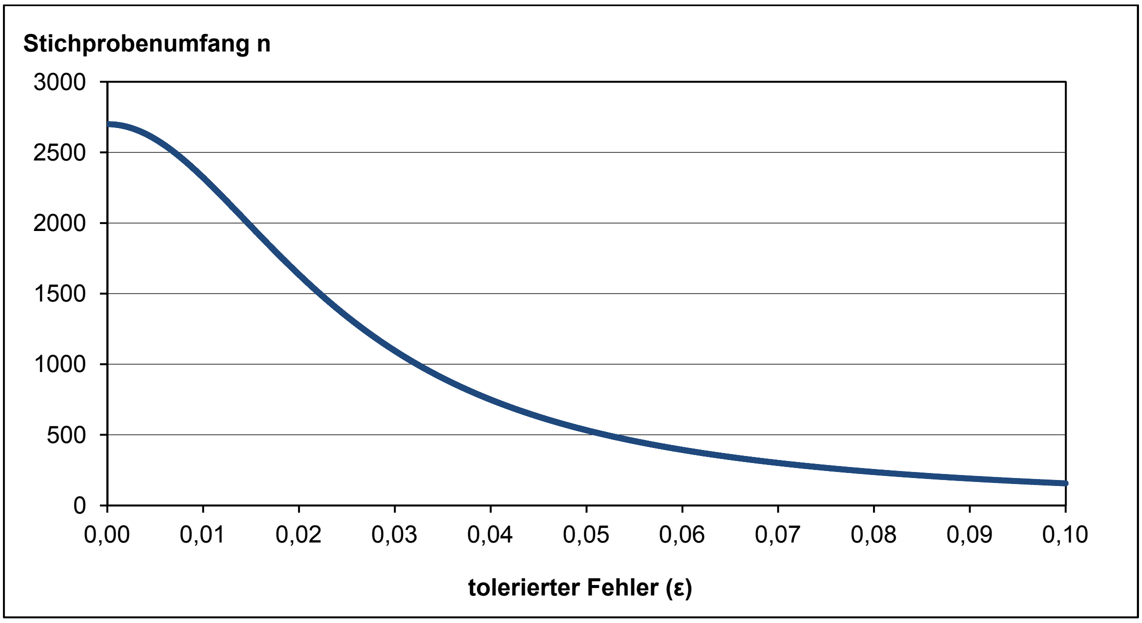
Entsprechend kann der Einfluss des tolerierten Fehlers ε auf die Größe des minimal erforderlichen Stichprobenumfangs analysiert werden, um die Festlegung auf den tolerierten Fehler ε zu erleichtern bzw. besser zwischen der gewünschten Ergebnisgenauigkeit und dem dazu erforderlichen Erhebungsaufwand abzuwägen. Dazu sei nun ceteris paribus die Grundgesamtheit auf N = 2700 und die Sicherheitswahrscheinlichkeit auf S=99% konstant festgelegt. Abbildung 12 gibt zu erkennen, dass die minimal erforderlichen Stichprobenumfänge n mit abnehmenden Fehlertoleranzen, etwa ab einem tolerierten Fehler von ε=0,05, rasch zunehmen.

Abb. 11: Minimal erforderlicher Stichprobenumfang n in Abhängigkeit von der Grundgesamtheit N bei konstant gewählter Sicherheitswahrscheinlichkeit von 99% (z = 2,575) und einem tolerierten Fehler ε = 0,05



Quelle: Eigene Darstellung

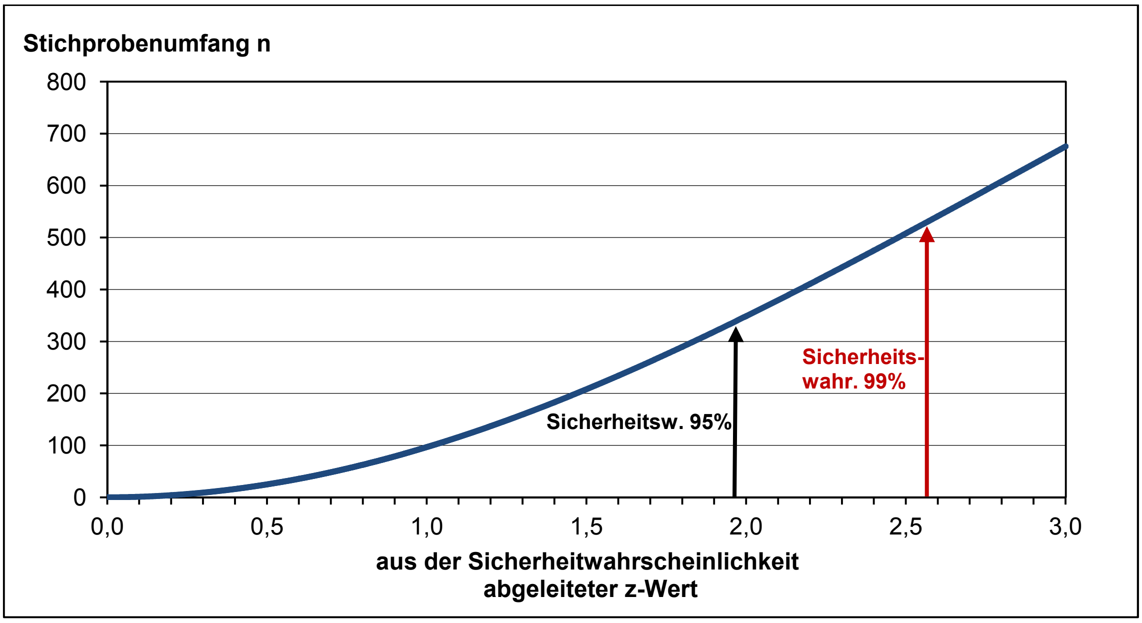
Abb. 12: Minimal erforderlicher Stichprobenumfang n in Abhängigkeit vom tolerierten Fehler ε bei konstant gewählter Grundgesamtheit N = 2700 mit einer Sicherheitswahrscheinlichkeit von 99% (z = 2,575)



Quelle: Eigene Darstellung

Eine entsprechende Darstellung ist auch für die Sicherheitswahrscheinlichkeit möglich. Dabei ist jedoch zu beachten, dass sich die Sicherheitswahrscheinlichkeit nicht gleichmäßig proportional zu den entsprechenden z-Werten verhält. Entsprechend ist bei konstant gehaltener Grundgesamtheit N=2700 und konstantem tolerierten Fehler ε=0,05 der minimal erforderlichen Stichprobenumfangs n nicht in direkter Abhängigkeit von der Sicherheitswahrscheinlichkeit dargestellt, sondern in Abhängigkeit vom standardisierten Wert z (vgl. Abb. 13).

Abb. 13: Minimal erforderlicher Stichprobenumfang n in Abhängigkeit von der Sicherheitswahrscheinlichkeit bei konstanter Grundgesamtheit N = 2700 und einem konstanten tolerierten Fehler ε = 0,05.



Quelle: Eigene Darstellung

5 Literatur

Atteslander, P. (1971): Methoden der empirischen Sozialforschung. Berlin.

Bahrenberg, G./Giese, E./Mevenklamp, N./Nipper, J. (2010): Statistische Methoden in der Geographie. Band 1: Univariate und bivariate Statistik. 5. Auflage. Stuttgart.

Bahrenberg, G./Giese, E./Nipper, J. (1999): Statistische Methoden in der Geographie. Band 1: Univariate und bivariate Statistik. 4. Auflage. Stuttgart.

Hantschel, R./Tharun, E. (1980): Anthropogeographische Arbeitsweisen. Braunschweig.

Holm, K. (1982): Die Befragung, Bd. 1: Der Fragebogen - Die Stichprobe. München.

Reuber, P./Pfaffenbach, C. (2005): Methoden der empirischen Humangeographie. Braunschweig.

Rinne, H. (1995): Taschenbuch der Statistik für Wirtschafts- und Sozialwissenschaften. Frankfurt a.M..

6 Anhang

(entnommen aus Bahrenberg et al. 1999, S. 224-225)

Tafel 2 Die Verteilungsfunktion $\phi(z)$ und die Funktion $D(z)$ der Standardnormalverteilung
Quelle: KREYSZIG 1968, S. 393–394

$\phi(-z) = 1 - \phi(z) \qquad D(z) = \phi(z) - \phi(-z)$

z	$\phi(z)$	$D(z)$	z	$\phi(z)$	$D(z)$	z	$\phi(z)$	$D(z)$	z	$\phi(z)$	$D(z)$
	0,	0,		0,	0,		0,	0,		0,	0,
0,01	5040	0080	0,43	6664	3328	0,85	8023	6047	1,27	8980	7959
0,02	5080	0160	0,44	6700	3401	0,86	8051	6102	1,28	8997	7995
0,03	5120	0239	0,45	6736	3473	0,87	8078	6157	1,29	9015	8029
0,04	5160	0319	0,46	6772	3545	0,88	8106	6211	1,30	9032	8064
0,05	5199	0399	0,47	6608	3616	0,89	8133	6265	1,31	9049	8098
0,06	5239	0478	0,48	6844	3688	0,90	8159	6319	1,32	9066	8132
0,07	5279	0558	0,49	6879	3759	0,91	8186	6372	1,33	9082	8165
0,08	5319	0638	0,50	6915	3829	0,92	8212	6424	1,34	9099	8198
0,09	5359	0717	0,51	6950	3899	0,93	8238	6476	1,35	9115	8230
0,10	5398	0797	0,52	6958	3969	0,94	8264	6528	1,36	9131	8262
0,11	5438	0876	0,53	7019	4039	0,95	8289	6579	1,37	9147	8293
0,12	5478	0955	0,54	7054	4108	0,96	8315	6629	1,38	9162	8324
0,13	5517	1034	0,55	7088	4177	0,97	8340	6680	1,39	9177	8355
0,14	5557	1113	0,56	7123	4245	0,98	8365	6729	1,40	9192	8385
0,15	5596	1192	0,57	7157	4313	0,99	8389	6778	1,41	9207	8415
0,16	5636	1271	0,58	7190	4381	1,00	8413	6827	1,42	9222	8444
0,17	5675	1350	0,59	7224	4448	1,01	8438	6875	1,43	9236	8473
0,18	5714	1428	0,60	7257	4515	1,02	8461	6923	1,44	9251	8501
0,19	5753	1507	0,61	7291	4581	1,03	8485	6970	1,45	9265	8529
0,20	5793	1585	0,62	7324	4647	1,04	8508	7017	1,46	9279	8557
0,21	5832	1663	0,63	7357	4713	1,05	8531	7063	1,47	9292	8584
0,22	5871	1741	0,64	7389	4778	1,06	8554	7109	1,48	9306	8611
0,23	5910	1819	0,65	7422	4843	1,07	8577	7154	1,49	9319	8638
0,24	5948	1897	0,66	7454	4907	1,08	8599	7199	1,50	9332	8664
0,25	5987	1974	0,67	7486	4971	1,09	8621	7243	1,51	9345	8690
0,26	6026	2051	0,68	7517	5035	1,10	8643	7287	1,52	9357	8715
0,27	6064	2128	0,69	7549	5098	1,11	8665	7330	1,53	9370	8740
0,28	6103	2205	0,70	7580	5161	1,12	8686	7373	1,54	9382	8764
0,29	6141	2282	0,71	7611	5223	1,13	8708	7415	1,55	9394	8789
0,30	6179	2358	0,72	7642	5285	1,14	8729	7457	1,56	9406	8812
0,31	6217	2434	0,73	7673	5346	1,15	8749	7499	1,57	9418	8836
0,32	6255	2510	0,74	7704	5407	1,16	8770	7540	1,58	9429	8859
0,33	9293	2586	0,75	7734	5467	1,17	8790	7580	1,59	9441	8882
0,34	6331	2661	0,76	7764	5527	1,18	8810	7620	1,60	9452	8904
0,35	6368	2737	0,77	7794	5587	1,19	8830	7660	1,61	9463	8926
0,36	6406	2812	0,78	7823	5646	1,20	8849	7699	1,62	9474	8948
0,37	6443	2886	0,79	7852	5705	1,21	8869	7737	1,63	9484	8969
0,38	6480	2961	0,80	7881	5763	1,22	8888	7775	1,64	9495	8990
0,39	6517	3035	0,81	7910	5821	1,23	8907	7813	1,65	9505	9011
0,40	6554	3108	0,82	7939	5878	1,24	8925	7850	1,66	9515	9031
0,41	6591	3182	0,83	7967	5935	1,25	8944	7887	1,67	9525	9051
0,42	6628	3255	0,84	7995	5991	1,26	8962	7923	1,68	9535	9070

Tafel 2 (Fortsetzung)

z	$\phi(z)$	$D(z)$	z	$\phi(z)$	$D(z)$	z	$\phi(z)$	$D(z)$
	0,	0,		0,	0,		0,	0,
1,69	9545	9090	2,13	9834	9668	2,57	9949	9898
1,70	9554	9109	2,14	9838	9675	2,58	9951	9901
1,71	9564	9127	2,15	9842	9684	2,59	9952	9904
1,72	9573	9146	2,16	9846	9692	2,60	9953	9907
1,73	9582	9164	2,17	9850	9700	2,61	9955	9909
1,74	9591	9181	2,18	9854	9707	2,62	9956	9912
1,75	9599	9199	2,19	9857	9715	2,63	9957	9915
1,76	9608	9216	2,20	9861	9722	2,64	9959	9917
1,77	9616	9233	2,21	9864	9729	2,65	9960	9920
1,78	9625	9249	2,22	9868	9736	2,66	9961	9922
1,79	9633	9265	2,23	9871	9743	2,67	9962	9924
1,80	9641	9281	2,24	9875	9749	2,68	9963	9926
1,81	9649	9297	2,25	9878	9756	2,69	9964	9929
1,82	9656	9312	2,26	9881	9762	2,70	9965	9931
1,83	9664	9328	2,27	9884	9768	2,71	9966	9933
1,84	9671	9342	2,28	9887	9774	2,72	9967	9935
1,85	9678	9357	2,29	9890	9780	2,73	9968	9937
1,86	9686	9371	2,30	9893	9786	2,74	9969	9939
1,87	9693	9385	2,31	9896	9791	2,75	9970	9940
1,88	9699	9399	2,32	9898	9797	2,76	9971	9942
1,89	9706	9412	2,33	9901	9802	2,77	9972	9944
1,90	9713	9426	2,34	9904	9807	2,78	9973	9946
1,91	9719	9439	2,35	9906	9812	2,79	9974	9947
1,92	9726	9451	2,36	9909	9817	2,80	9974	9949
1,93	9732	9464	2,37	9911	9822	2,81	9975	9950
1,94	9738	9476	2,38	9913	9827	2,82	9976	9952
1,95	9744	9488	2,39	9916	9832	2,83	9977	9953
1,96	9750	9500	2,40	9918	9836	2,84	9977	9955
1,97	9756	9512	2,41	9920	9840	2,85	9978	9956
1,98	9761	9623	2,42	9922	9845	2,86	9979	9958
1,99	9767	9534	2,43	9925	9849	2,87	9979	9959
2,00	9772	9545	2,44	9927	9853	2,88	9980	9960
2,01	9778	9556	2,45	9929	9857	2,89	9981	9961
2,02	9783	9566	2,46	9931	9861	2,90	9981	9963
2,03	9788	9576	2,47	9932	9865	2,91	9982	9964
2,04	9793	9586	2,48	9934	9869	2,92	9982	9965
2,05	9798	9596	2,49	9936	9872	2,93	9983	9966
2,06	9803	9606	2,50	9938	9876	2,94	9984	9967
2,07	9808	9615	2,51	9940	9879	2,95	9984	9968
2,08	9812	9625	2,52	9941	9883	2,96	9985	9969
2,09	9817	9634	2,53	9943	9886	2,97	9985	9970
2,10	9821	9643	2,54	9945	9889	2,98	9986	9971
2,11	9826	9651	2,55	9946	9892	2,99	9986	9972
2,12	9830	9660	2,56	9948	9895	3,00	9987	9973

Beiträge
zur
Wirtschaftsgeographie
und
Regionalentwicklung

Nr. 1 - 2012 - Ivo Mossig

Stichproben, Stichprobenauswahlverfahren und Berechnung des minimal erforderlichen Stichprobenumfangs

Kurzfassung

Wann ist eine Stichprobe repräsentativ? Wie viele Personen müssen befragt, wie viele Messwerte erhoben werden, um repräsentative Ergebnisse zu erzielen? Anhand welcher Verfahren kann die Auswahl der Stichprobenelemente vorgenommen werden? Welche Vor- und Nachteile sind mit dem jeweiligen Stichprobenauswahlverfahren verbunden?

Der vorliegende Beitrag widmet sich der Beantwortung dieser forschungspraktischen Fragen. Der minimal erforderliche Stichprobenumfang hängt von der anvisierten Fehlergenauigkeit (tolerierter Fehler ε), der gewählten Sicherheitswahrscheinlichkeit sowie der Größe der Grundgesamtheit ab. Es wird analysiert, wie sehr diese Parameter den Stichprobenumfang beeinflussen.

Bisher erschienen:

Nr. 3 - 2011 **Christian Kluck und Lars Schieber**
Die Kultur- und Kreativwirtschaft in Bremen. Standortfaktoren, Wissensquellen und Kooperationen in der Bremer Software- und Werbewirtschaft

Nr. 2 - 2011 **Lars Schieber und Ivo Mossig**
Clusterentwicklung und -politik in der Verpackungsmaschinenbau-Industrie Mittelhessens

Nr. 1 - 2011 **Lars Schieber und Ivo Mossig**
Clusterentwicklung und -politik in der Verpackungsmaschinenbau-Industrie Baden-Württembergs

Nr. 2 - 2010 **Ivo Mossig und Ansgar Dorenkamp**
Shopping-Malls und Business Improvement Districts als Instrumente zur Belebung innerstädtischer Geschäftszentren? Das Beispiel der Stadt Gießen.

Nr. 1 - 2010 **Ivo Mossig und Tobias Tkaczick**
Wohnsituation der Studentinnen und Studenten in Bremen