CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Statistik I

Einheit 5: Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

14.05.2025 | Prof. Dr. Stephan Goerigk

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wiederholung:

Inferenzstatistik:

- Umfasst alle statistischen Verfahren, die es erlauben, trotz der Informationsunvollständigkeit der Stichprobendaten Aussagen über eine Population zu treffen.

Population:

- Gesamtheit aller Merkmalsträger:innen, auf die eine Untersuchungsfrage gerichtet ist.

Stichprobe:

- Auswahl bestimmter Merkmalsträger:innen aus einer Population

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wiederholung:

Problem:

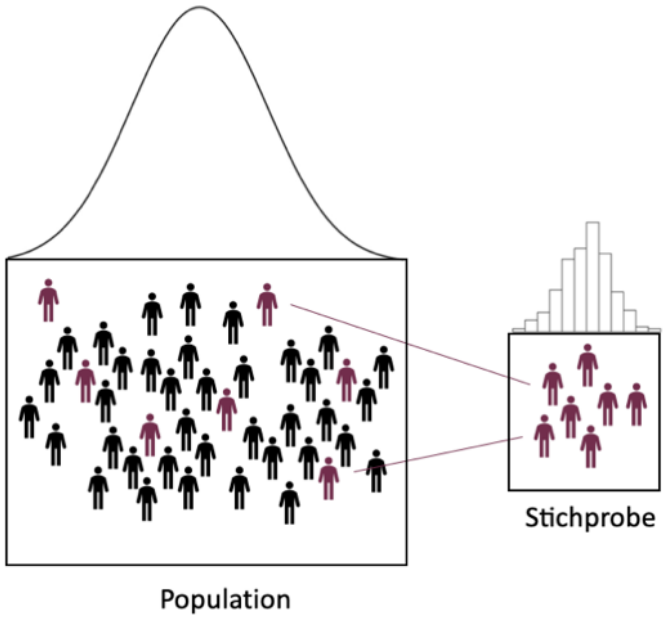
- Wenn nur ein Teil der Grundgesamtheit erfasst wird, z.B. 100 Personen, ist die **Informationslage** in Bezug auf die Untersuchungsfrage **unvollständig**. Wir können nicht einfach deskriptiv-statistische Methoden verwenden.
- Wie kann man trotzdem Aussagen treffen, die sich auf alle Personen der Grundgesamtheit beziehen, obwohl nur die Daten einer Stichprobe vorliegen?

Idee:

- Wir ziehen die Personen zufällig aus der Population in die Stichprobe.
- Wir greifen auf mathematische Methoden zur Formalisierung von Zufallsprozessen zurück $\rightarrow$ Wahrscheinlichkeitstheorie
- Aus diesen ergeben sich Methoden, die Rückschlüsse von der Stichprobe auf die Population erlauben $\rightarrow$ Inferenzstatistik

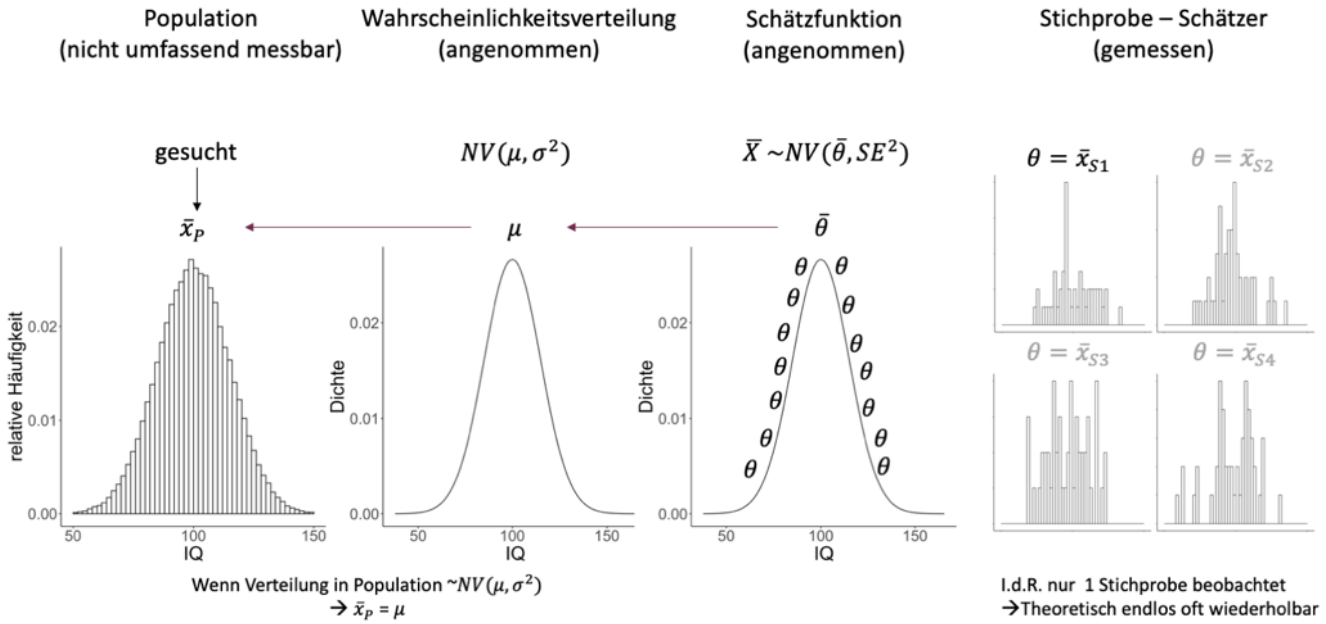
Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Logik des Schließens von Stichprobe auf Population (Einzelschritte folgen)



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Logik des Schließens von Stichprobe auf Population (Einzelschritte folgen)



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Inferenzstatistik:

- Schluss von Zufallsstichprobe auf Population
- Grundlage: Wahrscheinlichkeitsrechnung
- Zentral: Zufallsprozesse (Ausgang unsicher, nicht mit Sicherheit vorhersagbar)

Wahrscheinlichkeitsrechnung:

Mathematik ist der Versuch, alles zu bändigen, auch den Zufall.

Rudolf Taschner

- Statistischer Wahrscheinlichkeitsbegriff geht zurück auf 17. Jahrhundert (Frankreich)
- Im Jahr 1654 wandte sich der Glücksspieler Chevalier de Mere mit mehreren Fragen an den französischen Mathematiker Blaise Pascal

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Stochastik:

- Stochastik = die Kunst des Vermutens (altgriechisch)
- Mathematik setzt Vorstellung von Zufall voraus (= Modelle von Situationen, deren Ausgang unsicher ist)
- Keine Einzelereignisse vorhersagbar, aber:
- Erkennen von Regelmäßigkeiten bei Vorgängen, deren Ergebnisse vom Zufall abhängen.
- Zentraler Begriff: Zufallsexperiment

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Zufallsexperiment:

Im Prinzip beliebig oft wiederholbarer Vorgang, der nach bestimmter Vorschrift ausgeführt wird, wobei das Ergebnis vom Zufall abhängt, d.h. der Ausgang kann nicht eindeutig im voraus bestimmt werden.

- Folge von gleichartigen, voneinander unabhängigen Versuchen möglich.
- Entweder Folge voneinander unabhängiger Versuche mit einem Objekt oder jeweils einmaliger Versuche mit ”gleichartigen” (unabhängigen) Objekten.

Beispiele:

1. Ein Würfel wird wiederholte Male geworfen und es wird beobachtet, wie oft jede Zahl kommt.
2. Parteipräferenz bei weiblichen Jugendlichen zwischen 16 und 18 Jahren.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Zufallsexperiment - Nomenklatur:

- Die möglichen Ergebnisse eines Zufallsexperimentes heißen Elementarereignisse ω
- Die Menge aller möglichen Ergebnisse eines Zufallsexperimentes bezeichnet man als Ereignisraum Ω.
- Beispiel: 'Einmaliges Würfeln': Elementarereignisse sind {1}, {2}, {3}, {4}, {5}, {6}. Ereignisraum $\Omega = 1, 2, 3, 4, 5, 6$.
- Ereignis A: Teilmenge des Ereignisraums, z.B. alle geraden Augenzahlen beim Würfeln. Es gilt: $\omega \in A, A \subset \Omega$
- Sicheres Ereignis: Jenes Ereignis, welches unter gegebenen Bedingungen immer eintritt.
- Unmögliches Ereignis: Jenes Ereignis, welches unter gegebenen Bedingungen nie eintritt.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Definition der statistischen Wahrscheinlichkeit:

Die Wahrscheinlichkeit für das Auftreten eines Ereignisses A, $P_{(A)}$, ist jener Wert, bei dem sich die relative Häufigkeit $r_n(A)$ bei $n \to \infty$ Versuchen unter gleichen Bedingungen stabilisiert.

Die mathematische Formulierung:

$$P(A) = \lim_{n \to \infty} r_n(A)$$

In anderen Worten:

- Die Wahrscheinlichkeit eines Ereignisses gibt an, mit welcher relativen Häufigkeit das Ereignis einträte, wenn man den Versuch theoretisch unendlich oft wiederholen würde.
- Sie sagt jedoch nichts darüber aus, wie häufig das Ereignis bei einer kleinen Anzahl von Versuchen, z.B. $n = 5$, auftritt.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Laplace-Wahrscheinlichkeit

Bei Zufallsexperimenten, bei denen nur endlich viele, gleichwahrscheinliche Ergebnisse möglich sind, ergibt sich für ein beliebiges Ereignis A die Wahrscheinlichkeit $P(A)$:

$$P(A) = \frac{\text{Anzahl der } \omega \text{ in } A}{\text{Anzahl der } \omega \text{ in } \Omega} = \frac{\text{Anzahl der für A 'günstigen' Ereignisse}}{\text{Anzahl der möglichen Ereignisse}}$$

Beispiele: $P(K)$ bei Münzwürfen $= \frac{1}{2} = \lim_{n \to \infty} r_n(K)$

$P(1)$ beim Würfeln $= \frac{1}{6} = \lim_{n \to \infty} r_n(1)$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Axiome der Wahrscheinlichkeitsrechnung nach Kolmogoroff

Wahrscheinlichkeiten lassen sich durch drei Eigenschaften, die auch für relative Häufigkeiten gelten, und aus denen sich alle Rechenregeln für Wahrscheinlichkeiten ableiten lassen, charakterisieren:

1. Für die Wahrscheinlichkeit eines Ereignisses **gilt stets**:

$$0 \leq P_{(A)} \leq 1$$

2. Die Wahrscheinlichkeit eines **sicheren Ereignisses** beträgt

$$P_{(\Omega)} = 1$$

3. **Additionsregel der Wahrscheinlichkeit:** Die Wahrscheinlichkeit, dass eines von k einander ausschließenden Ereignissen auftritt, ist die Summe der einzelnen Wahrscheinlichkeiten $P_{(A_1)}, P_{(A_2)}, \ldots, P_{(A_k)}$.

$$P_{(A_1 \vee A_2 \vee \ldots \vee A_k)} = P_{(A_1)} + P_{(A_2)} + \ldots + P_{(A_k)}$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Rechenregeln: Unmögliches Ereignis

- Die Wahrscheinlichkeit des unmöglichen Ereignisses B beträgt

$$P_{(B)} = 0$$

- Wenn B ein unmögliches Ereignis ist, kann es nie eintreten:

$$rn_{(B)} = 0 \rightarrow P_{(B)} = 0$$

ACHTUNG:

- Aus $P_{(B)} = 0$ folgt nicht, dass B ein unmögliches Ereignis ist.
- Das bedeutet nur, dass der Grenzwert der relativen Häufigkeit bei $n \rightarrow \infty$ Null ist, woraus aber nicht folgt, dass B nie eintreten kann! (Analoges gilt für $P_{(A)} = 1$).

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Rechenregeln: Komplementärereignis

- $P_{(A)} + P_{(\bar{A})} = 1, P_{(\bar{A})} = 1 -_{P(A)}$
- $\bar{A}$ tritt immer dann ein, wenn A nicht eintritt → $r_n(A) + r_n(\bar{A}) = 1$

Beispiel - Münzwurf:

$P_{(K)} + P_{(Z)} = 0.5 + 0.5 = 1$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Stochastische Unabhängigkeit von Ereignissen

- Beim Ziehen mit Zurücklegen sind die einzelnen Wahrscheinlichkeiten gleich und die Ziehungen stochastisch unabhängig.
- Beim Ziehen ohne Zurücklegen ändern sich mit jeder Ziehung die Anteile der 'günstigen' ω_i , und daher auch die Wahrscheinlichkeiten. Die Ziehungen sind daher stochastisch abhängig.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Zufallsexperiment

- Jedes mögliche Ergebnis aus einem Zufallsexperiment nennen wir ein Elementarereignis ω
- Die Menge aller möglichen Ereignisse ist definiert als der Ereignisraum Ω
- Der Ereignisraum Ω heißt diskret, wenn er aus abzählbar vielen Elementarereignissen besteht
- Der Ereignisraum Ω heißt stetig, wenn er aus überabzählbar vielen Elementarereignissen besteht
- Zufallsexperiment ist ein allgemeiner Begriff, der Grundlage für die Inferenzstatistik ist

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Psychologische Fragestellungen:

- Praktisch alle psychologischen Theorien enthalten Aussagen über Populationen (nicht nur über isolierte Stichproben).
- Zu ihrer empirischen Überprüfung sind dann **immer** inferenzstatistische Methoden notwendig.

Beispiele für psychologische Fragestellungen:

- **Beispiel 1 (diskret):**
  - Wir interessieren uns für die relative Häufigkeit h_A der Personen in Europa, die an Angststörungen erkrankt sind.
- **Beispiel 2 (stetig):**
  - Wir interessieren uns für den Mittelwert $\bar{x}_{IQ}$ und die empirische Varianz s^2_{empIQ} des Intelligenzquotienten (IQ) von Personen in Europa.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Zufallsvariable

- Zufallsvariable lässt sich durch ihre Wahrscheinlichkeitsfunktion beschreiben, welche angibt, mit welcher Wahrscheinlichkeit die einzelnen Realisationen x_i auftreten.
- Es sei p_i die Wahrscheinlichkeit des Auftretens des Wertes x_i; dann ist

$$f(x_i) = P(X = x_i) = p_i; p_i \in [0, 1]$$

- Wenn alle möglichen Ausprägungen von X berücksichtigt wurden,ist die Summe aller möglichen Einzelwahrscheinlichkeiten p_i = 1

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Zufällige Ziehung einer einzelnen Person

Zufällige Ziehung einer **einzelnen** Person aus einer Population von N Personen:

Dieser Vorgang ist ein **Zufallsexperiment**:

- Wir wissen im Voraus nicht, welche Person gezogen wird.
- Die Ergebnismenge Ω ist die Menge aller Personen in der Population:

$$\Omega = Person_1, Person_2, \ldots, Person_i, \ldots, Person_N$$

- Wir setzen voraus, dass jede Person i in der Population die **gleiche Wahrscheinlichkeit** hat, gezogen zu werden.
- Alle Elementarereignisse haben die gleiche Wahrscheinlichkeit:

$$P(Person_i) = \frac{1}{N}$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Beispiel Angststörungen:

- Wir interessieren uns für die relative Häufigkeit h_A der Personen in Deutschland, die an Angststörungen erkrankt sind.
- Sei N_A die Anzahl der Angstpatienten in der Population und A_A die Menge der Angstpatienten in der Population:

$$A_A = Patient_1, Patient_2, \ldots, Patient_i, \ldots, Patient_N$$

- Die relative Häufigkeit der Angstpatienten in der Population ist also $h_A = \frac{N_A}{N}$
- Die Wahrscheinlichkeit, zufällig eine Angstpatient:in zu ziehen ist:

$$P(A_A) = P(Patient_1) + P(Patient_2) + P(Patient_{N_A})$$

- Die Wahrscheinlichkeit dafür, zufällig eine Angstpatient:in zu ziehen, entspricht also der relativen Häufigkeit der Angststörung in der Population:

$$P(A_A) = h_A$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Beispiel Angststörungen:

- Sei nun X eine Zufallsvariable, die den Wert 1 annimmt, falls die zufällig gezogene Person eine Angststörung hat, und 0, falls nicht.
- Diese Zufallsvariable ist eine Bernoulli-Variable und folgt somit einer Bernoulli-Verteilung.
- Der Parameter π der Bernoulli-Verteilung entspricht der Wahrscheinlichkeit, dass X den Wert 1 annimmt, also der Wahrscheinlichkeit, eine Angstpatient:in zu ziehen.
- Diese Wahrscheinlichkeit entspricht wiederum der relativen Häufigkeit der Angststörung in der Population (siehe letzte Folie).

Formal:

$$\pi = P(X = 1) = P(A_A) = h_A$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Beispiel Angststörungen:

Zusammengefasst: Unter der Voraussetzung, dass

- jede Person in der Population die gleiche Wahrscheinlichkeit hat, gezogen zu werden,
- X eine Zufallsvariable ist, die den Wert 1 annimmt, falls die gezogene Person eine Angststörung hat, und 0, falls nicht,

folgt X einer Bernoulli-Verteilung und der Wert des Parameters $\boldsymbol{\pi}$ dieser Bernoulli-Verteilung ist identisch mit dem Wert der relativen Häufigkeit h_A der Angststörung in der Population.

- Wenn wir herausfinden wollen, wie hoch die relative Häufigkeit der Angststörung in der Population ist, müssen wir lediglich herausfinden, welchen Wert der Parameter π hat.
- Wenn wir z.B. wüssten, dass $\pi = 0.3$ ist, wüssten wir auch, dass die relative Häufigkeit der Angststörung in der Population h_A = 0.3 ist.
- Da π ein Parameter einer Wahrscheinlichkeitsverteilung ist, können wir das Problem der Bestimmung einer deskriptivstatistischen Maßzahl in der Population (h_A) komplett in die Wahrscheinlichkeitstheorie verlagern und somit alle Mittel verwenden, die uns diese zur Verfügung stellt.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Dichtefunktion

- Eine stetige ZV X kann jeden Wert in einem Intervall [a, b] annehmen
- Die Wahrscheinlichkeiten der einzelnen Ausprägungen (Werte) einer stetigen ZV können (im Gegensatz zum diskreten Fall) nicht angegeben werden
- Es können nur Wahrscheinlichkeiten $f(x)dx$ angegeben werden, mit welchen die Werte innerhalb von Intervallen dx um die Werte x auftreten
- Beispielsweise fragt man nicht, wie viele Personen exakt 1.75 Meter groß sind, sondern z.B., wie viele Personen zwischen 1.75 und 1.76 Meter groß sind
- Die Funktion $f(x)$ heißt Dichtefunktion

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Dichtefunktion

- Die Wahrscheinlichkeit, dass die ZV Werte zwischen a und b annimmt, wird dann allgemein definiert als das Integral über die Dichtefunktion mit Integrationsgrenzen a und b.
- Analog zum diskreten Fall erhält man durch Integration die Verteilungsfunktion
- Die Wahrscheinlichkeit ist definiert als Fläche unter der Dichtefunktion

$$F(x) = P(X \leq x) = \int_{t \leq x} f(t)dt$$

Es gilt für alle $a < b$:

$$P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x)dx$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Erwartungswert

Beispiel: X ist die erhaltene Augenzahl bei einmaligem Würfeln; die Wahrscheinlichkeitsverteilung von X ist:

x_i	1	2	3	4	5	6
$f(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Welchen Wert ’erwarten’ wir, wenn wir dieses Zufallsexperiment sehr lange durchführen?
- Intuitiv erwarten wir $X = 1$ bei $\frac{1}{6}$ der Würfe, $X = 2$ bei $\frac{1}{6}$ bei der Würfe, usw.
- Der Durchschnitt von X auf lange Sicht ist der Erwartungswert von X
- Der Erwartungswert einer ZV ist ein Maß für das Zentrum der Verteilung

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Varianz der ZV

Die Varianz σ^2 ist ein Streuungsmaß der Verteilung

$$\sigma_X^2 = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Varianz der ZV

Beispiel: X ist die beobachtete Augenzahl bei einmaligem Würfeln; die Wahrscheinlichkeitsverteilung von X ist

x_i	1	2	3	4	5	6
$f(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\sigma^2 = E\left[X^2\right] - \underbrace{(E\left[X\right])^2}_{3.5^2} \text{ und } E\left[X^2\right] = \sum_{i=1}^{6} x_i^2 p(x_i^2)$$

$$E\left[X^2\right] = 1^2\,\frac{1}{6} + \cdots + 6^2\frac{1}{6} = 15.17,\ \sigma^2 = 2.92,\ \sigma = 1.71$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

α-Quantil

Als α-Quantil q_α wird ein Wert bezeichnet, unterhalb dessen ein vorgegebener Anteil α aller Fälle der Verteilung liegen

- Jeder Wert unterhalb von q_α unterschreitet den Anteil α, mit α als reelle Zahl zwischen 0 (gar kein Fall der Verteilung) und 1 (alle Fälle oder 100% der Verteilung)
- Für stetige ZV gilt:

$$F(q_\alpha) = P(X \leq q_\alpha) = \int_{t \leq q_\alpha} f(t)dt = \alpha$$

- Für diskrete ZV gilt (Aufrunden zur nächsten ganzzahligen Ausprägung):

$$F(q_\alpha) = P(X \leq q_\alpha) = \sum_{t \leq q_\alpha} P(X = t) \geq \alpha$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Diskrete Gleichverteilung

- Diese Verteilung beschreibt eine ZV, welche die Zahlen $1, 2, \cdots, m$ annehmen kann, und es gilt:

$$P(X = x) = \frac{1}{m} \text{ für alle } x = 1, 2, \cdots, m$$

$$E[X] = \frac{(m+1)}{2}$$

$$\sigma^2 = \frac{(m^2 - 1)}{12}$$

- Anwendung bei Zufallsexperimenten, deren Ergebnisse gleich häufig sind, also wenn angenommen wird, dass die m Elementarereignisse gleichwahrscheinlich sind

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Diskrete Gleichverteilung

Beispiel:

X = die erhaltene Augenzahl bei einmaligem Würfeln

$$E[X] = \frac{(6+1)}{2} = 3.5$$

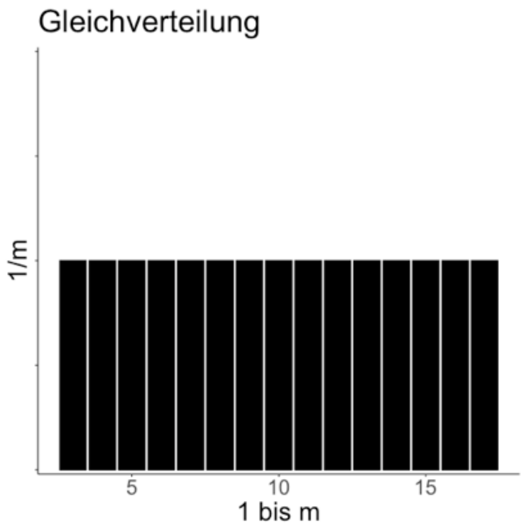
$$\sigma^2 = \frac{(6^2-1)}{12} = 2.92$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Diskrete Gleichverteilung



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Binomialverteilung

- Wir betrachten ein Zufallsexperiment mit 2 Ausgängen, ’Erfolg (2)’ und ’Misserfolg (1)’
- Die Wahrscheinlichkeit für Erfolg sei p, mit p zwischen 0 und 1
- Wir führen dieses Experiment n-mal durch, wobei zwischen den einzelnen Durchführungen Unabhängigkeit angenommen wird (’Ziehen mit Zurücklegen’)
- Die ZV X beschreibt die Anzahl der Erfolge und ist binomialverteilt mit Parametern n und p, $X \sim B(n, p)$

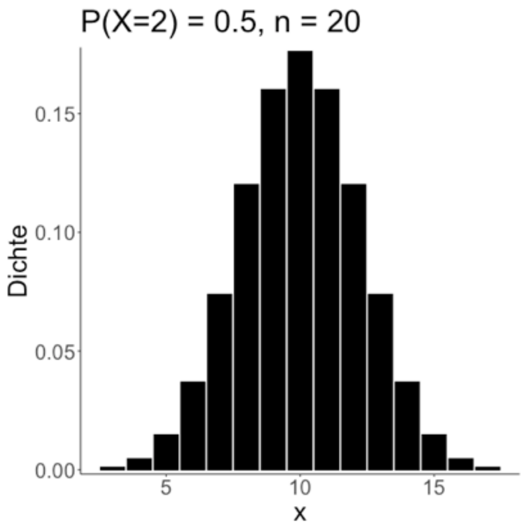
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ für } k = 0, 1, \cdots, n$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Binomialverteilung



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Binomialverteilung

- Beispiel: Ein Glücksrad besteht aus 20 Feldern, wobei 5 davon Gewinnfelder sind.
- Wie groß ist die Wahrscheinlichkeit, dass Sie zwei Mal gewinnen, wenn Sie das Glücksrad drei Mal drehen?
- Experiment mit 2 Ausgängen, Erfolg (5 Gewinnfelder) und Misserfolg
- $n = 3$, weil wir das Glücksrad drei Mal drehen
- $p = \frac{5}{20} = 0.25$ ist die Wahrscheinlichkeit zum Erfolg

$$P(X = 2) = \binom{3}{2} 0.25^2 (1 - 0.25)^1 = \frac{3!}{2!1!} 0.0625 \cdot 0.75 = 0.14$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle diskrete Verteilungen

Binomialverteilung

- Binomialverteilte ZV nimmt Werte zwischen 0 und n an
- Binomialverteilung ist symmetrisch für $p = 0.5$
- Je kleiner/größer p desto rechts/links-schiefer die Verteilung
- Summe mehrerer Bernoulli-Variablen

Erwartungswert und Varianz:

$$E[X] = np$$

$$\sigma^2 = np(1 - p)$$

- Für $n = 1$: $B(1, p)$ ist eine Bernoulli-ZV mit Erwartungswert p und Varianz $p(1 - p)$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle stetige Verteilungen

Normalverteilung (NV)

- Die NV ist eine stetige Verteilung, die durch 2 Parameter μ und σ charakterisiert ist
- Es sei X eine ZV die $N(\mu, \sigma^2)$ verteilt ist; X kann Werte zwischen $-\infty$ und $+\infty$ annehmen

Dichtefunktion $\varphi_{(x)}$:

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

- Geht x → $\pm\infty$ strebt $\varphi(x)$ gegen 0
- $\varphi(x)$ ist symmetrisch um μ

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle stetige Verteilungen

Normalverteilung (NV)

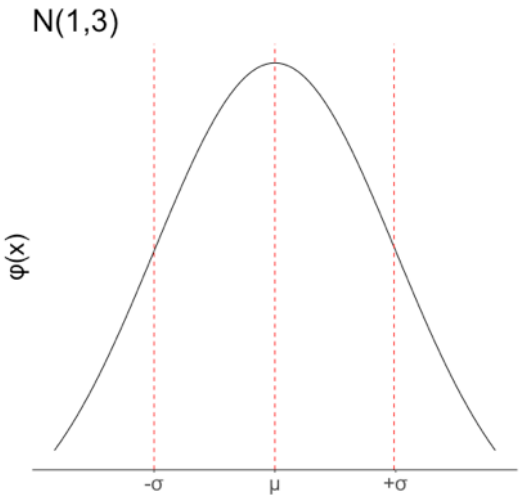
- σ gibt den Abstand zwischen μ und den Wendepunkten der Dichtefunktion an
- Wendepunkte an den Stellen $\mu \pm \sigma$
- Wenn σ groß ist, ist die Verteilung breit und niedrig, wenn σ klein ist, ist die Verteilung schmal und hoch
- Fläche unter $\varphi(x)$ zwischen $-\infty$ und $+\infty$ ist gleich 1
- Die Fläche $\mu \pm \sigma$ umfasst ca. 68% aller Fälle
- Die Fläche $\mu \pm 2\sigma$ umfasst ca. 95% aller Fälle
- Es existieren unendlich viele NV durch beliebige Auswahl von μ und σ

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle stetige Verteilungen

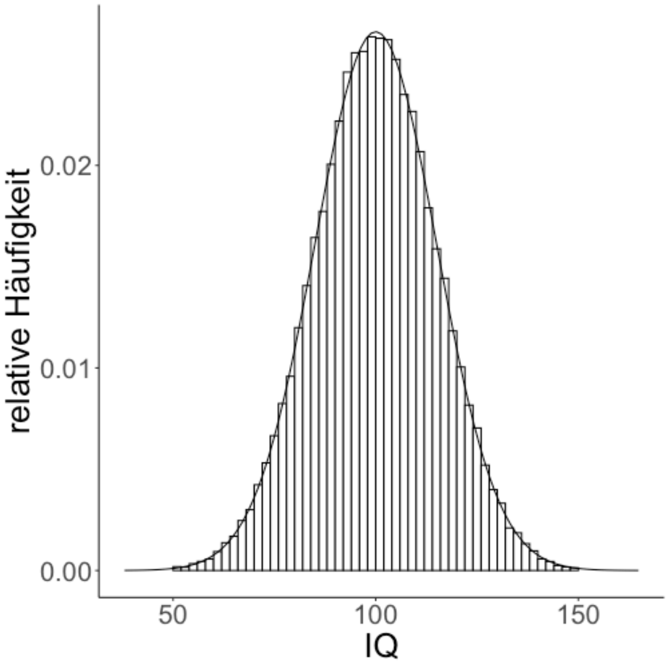
Normalverteilung (NV)



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Beispiel IQ:

- Wir interessieren uns für den Mittelwert $\bar{x}_{IQ}$ und die empirische Varianz s^2_{empIQ} des IQs von Personen in Europa
- Wir setzen voraus, dass das Histogramm der Variable IQ in der Population der Personen in Europa durch die Wahrscheinlichkeitsdichtefunktion einer Normalverteilung approximiert werden kann, d.h. dass das Histogramm die „Form“ der Dichte einer Normalverteilung hat.
- Dies ist eine **Annahme**, von der wir nicht wissen, ob sie zutrifft. Wir werden jedoch Methoden kennenlernen, um die Plausibilität dieser Annahme zu überprüfen.



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Beispiel IQ:

- Außerdem setzen wir wieder voraus, dass alle Personen die **gleiche Wahrscheinlichkeit** haben, gezogen zu werden.
- Sei nun X eine Zufallsvariable, die für den IQ der zufällig gezogenen Person steht.
- Man kann dann beweisen, dass diese Zufallsvariable X einer Normalverteilung folgt und der Parameter μ dieser Normalverteilung dem Mittelwert des IQs in der Population entspricht:

$$\mu = \bar{x}_{IQ}$$

- der Parameter σ^2 dieser Normalverteilung der empirischen Varianz des IQs in der Population entspricht:

$$\sigma^2 = s^2_{empIQ}$$

- Der Beweis hierfür funktioniert ähnlich wie bei der Bernoulli-Verteilung, ist aber deutlich aufwendiger.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Beispiel IQ:

Zusammengefasst: Unter der Voraussetzung, dass

- das Histogramm des IQs in der Population der Personen in Deutschland durch die Wahrscheinlichkeitsdichtefunktion einer Normalverteilung approximiert werden kann,
- jede Person in der Population die gleiche Wahrscheinlichkeit hat, gezogen zu werden,
- X eine Zufallsvariable ist, die für den IQ der gezogenen Person steht

folgt X einer Normalverteilung und

- der Wert des Parameters $\boldsymbol{\mu}$ dieser Normalverteilung ist identisch mit dem Mittelwert $\bar{x}_{IQ}$ des IQs in der Population,
- der Wert des Parameters $\boldsymbol{\sigma}^2$ dieser Normalverteilung ist identisch mit der empirischen Varianz s^2_{empIQ} des IQs in der Population.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle stetige Verteilungen

Standardnormalverteilung $N(0, 1)$

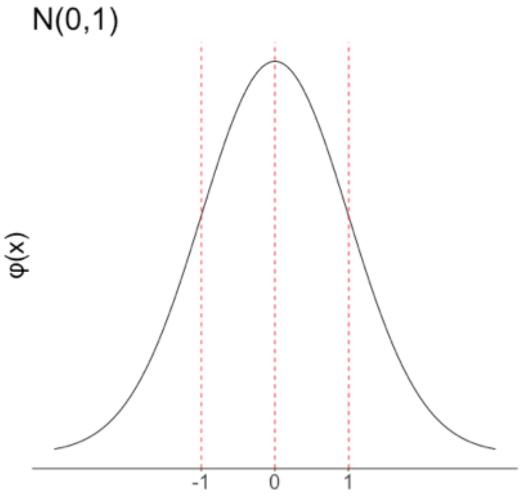
- Spezielle NV für $\mu = 0$ und $\sigma = 1$ (Gauß’sche Glockenkurve)
- Verteilung der $N(0, 1)$ ist tabelliert
- Fläche zwischen $\mu = 0$ und einem beliebigen Wert z ist ablesbar
- Quantile der NV; 1-Fläche rechts von einem Wert z, und links von −z
- Ist X $N(\mu, \sigma^2)$ verteilt dann führt die Transformation $\frac{X-\mu}{\sigma}$ auf eine $N(0, 1)$ Verteilung
- Vorteil, da Quantile in Tabellen ablesbar (es müssen nicht jedes mal Integrale für Dichtefunktion berechnet werden)

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Wahrscheinlichkeitsrechnung als Grundlage der Inferenzstatistik

Spezielle stetige Verteilungen

Standardnormalverteilung $N(0, 1)$



Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Nutzen von Wahrscheinlichkeitsverteilungen zur Quantifizierung des Stichprobenfehlers:

Z.B. Standardnormalverteilung (N~0,1):

- Quantile der Standardnormalverteilung sind tabelliert

Z-Tabelle

- Wahrscheinlichkeit für jeden z-Wert kann abgelesen werden
- Zusammensetzen des z-Werts aus Zeile (bis 1 Stelle nach dem Komma) und Spalte (2. Stelle nach dem Komma)
- Anhand der Tabelle kann abgelesen werden, wie wahrscheinlich die Werte einer Verteilung sind (angenommen die Variable ist normalverteilt)

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

Bedeutung der p-Werte

- Die Felder in der Tabelle geben Ihnen die Wahrscheinlichkeit P an, dass genau der ausgewählte z-Wert oder ein kleinerer z-Wert auftritt.
- Die Wahrscheinlichkeit, die Sie in den Feldern der z Tabelle finden, entspricht der Fläche unter der Verteilung.
- Diese Fläche ist das Integral der Dichtefunktion von $-\infty$ bis z.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

Beispiel: Orientierung in der z-Tabelle

Aufgabe: Sie suchem den z-Wert 0.35.

Schritt 1: Schauen Sie die 1. Spalte an

In der ersten Spalte (senkrecht) finden Sie die ersten zwei Ziffern 0.3 des z-Werts.

Schritt 2: Schauen Sie die 2. Spalte an

Die dritte Ziffer 0.05 (die zweite Nachkommastelle), findet sich in der 3. Spalte.

Schritt 3: Wahrscheinlichkeit finden

Das Feld, in dem sich nun die Zeile mit 0.3 und die Spalte mit 0.05 kreuzen, ist die gesuchte Wahrscheinlichkeit für 0.35 oder einen kleineren z-Wert also $P(X \leq 0.35) = 0.63683$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

Negative Werte in der z-Werte Tabelle:

Wie Sie wahrscheinlich gesehen haben, fängt die z-Tabelle bei 0 an. Was machen Sie also, wenn Ihr gegebener z-Wert negativ ist?

Dafür gibt es einen Trick: Die Standardnormalverteilung ist achsensymmetrisch (die Funktion spiegelt sich also an der y-Achse). Das heißt, sie verläuft links und rechts von der y-Achse genau spiegelverkehrt.

Es gilt: $\Phi(-x) = 1 - \Phi(x)$

Wenn Sie einen negativen z Wert haben, suchen Sie also zunächst den dazugehörigen positiven z-Wert. Dann rechnen Sie 1 minus den positiven z-Wert.

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

- Um die Standardnormalverteilung Tabelle nutzen zu können, brauchen Sie entweder einen gegebenen z-Wert oder eine gegebene Wahrscheinlichkeit.
- Die Berechnung eines z-Werts kann für jeden Wert einer normalverteilten Variable erfolgen
- Dieser Prozess nennt sich **z-Transformation** oder kurz **Standardisierung**
- Dafür braucht man nichts weiter als den Mittelwert und die Standardabweichung der Verteilung

$$z_i = \frac{x_i - \bar{x}}{s}$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

Vorteil der Standardisierung:

- Messwerte von Personen verschiedener Populationen sind oft nicht direkt **vergleichbar**, z.B. die Leistung eines Mädchens in Kugelstoßen mit jener eines Jungen
- Dennoch möchte man ausdrücken können, wie gut die beiden Leistungen innerhalb der Bezugsgruppe sind
- Der Standardmesswert z_i: bezieht den beobachteten Messwert x_i der i-ten Person auf den Mittelwert $\bar{x}$ der Gruppe und drückt die Abweichung in Standardeinheiten s aus

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

Beispiel: Standardisierung Kugelstoßen $(N = 5)$; Vergleich Frauen (w) und Männer (m)

ID	1	2	3	4	5
Meter_m	8	9	12	9	9
Meter_w	9	7	3	5	5

Lösungsweg (für 3. Mann):

$$\bar{x} = \frac{8+9+12+9+9}{5} = \frac{47}{5} = 9.4$$

$$s = \sqrt{\frac{(8-9.4)^2+(9-9.4)^2+(12-9.4)^2+(9-9.4)^2+(9-9.4)^2}{5-1}} = \sqrt{\frac{9.2}{4}} = 1.52$$

$$z_3 = \frac{12-9.4}{1.52} = 1.71$$

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Standardnormalverteilung (N~0,1):

Beispiel: Standardisierung Kugelstoßen $(N = 5)$; Vergleich Frauen (w) und Männer (m)

ID	1	2	3	4	5
Meter_m	8	9	12	9	9
Meter_w	9	7	3	5	5

Nach der Standardisierung jeden Werts anhand Mittelwert und Standardabweichung der Referenzgruppe:

ID	1	2	3	4	5
z_m	-0.92	-0.26	1.71	-0.26	-0.26
z_w	1.40	0.53	-1.23	-0.35	-0.35

Interpretation: Während z.B. die 2. Frau absolut weniger weit gestoßen hat (7m) als der 2. Mann (9m), liegt sie relativ zum Mittel der Gruppen vor ihm (0.53 > -0.26).

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Nutzen von Wahrscheinlichkeitsverteilungen zur Quantifizierung des Stichprobenfehlers:

Z.B. Standardnormalverteilung (N~0,1):

Beispiel 1:

- gegeben sei eine normalverteilte Variable X mit Mittelwert von 11 und Varianz von 5.53
- Wie hoch ist die Wahrscheinlichkeit das X nicht mehr als 14.5 Punkte aufweist?
- Zunächst berechnen wir den z-Wert für $X = 14.5$ (siehe Standardisierung)

$$z = \frac{14.5 - 11}{\sqrt{5.53}} = 1.49$$

- In der z-Tabelle schlagen wir nach, wie wahrscheinlich ein z-Wert von höchstens 1.49 ist

$$P(Z \leq 1.49) = 0.9319$$

- Mit einer 93%-igen Wahrscheinlichkeit ist ein zufällig aus der Verteilung gezogener Wert nicht größer als 14.5

Stichprobe, Grundgesamtheit - Wahrscheinlichkeitstheorie und Verteilungen

Nutzen von Wahrscheinlichkeitsverteilungen zur Quantifizierung des Stichprobenfehlers:

Z.B. Standardnormalverteilung (N~0,1):

Beispiel 2:

- gegeben sei eine normalverteilte Variable X mit Mittelwert von 11 und Varianz von 5.53
- Wie hoch ist die Wahrscheinlichkeit, dass X mehr als 14.5 Punkte aufweist?
- Zunächst berechnen wir den z-Wert für $X = 14.5$ (siehe Standardisierung)

$$z = \frac{14.5 - 11}{\sqrt{5.53}} = 1.49$$

- In der z-Tabelle schlagen wir nach, wie wahrscheinlich ein z-Wert von größer als 1.49 ist

$$P(Z > 1.49) = 1 - P(Z \leq 1.49) = 1 - 0.9319 = 0.0681$$

- Mit einer 6.8%-igen Wahrscheinlichkeit ist ein zufällig aus der Verteilung gezogener Wert größer als 14.5.

Take-aways

- **Inferenzstatistik** ist ein wahrscheinlichkeitsbasierter Schluss von Zufallsstichprobe auf Population
- Variablen in der Population sind nicht vollständig beobachtbar und daher **Zufallsvariablen** (diskret vs. stetig)
- **Wahrscheinlichkeitsfunktion** definiert welche Werte wir beim zufälligen Ziehen mit welcher Wahrscheinlichkeit erwarten
- Der **Erwartungswert** ist das Zentrum der Verteilung und der wahrscheinlichste Wert
- Unter der **Gleichverteilung** ist jedes Ereignis gleich wahrscheinlich
- **Binomialverteilung** lässt uns Wahrscheinlichkeit für ein diskretes Ereignis mit 2 Ausgängen berechnen
- **Normalverteilung** ist stetige Verteilung, die extremen Ereignissen geringere und durchschnittlichen Ereignissen höhere Wahrscheinlichkeit zuweist