

Statistik 1

Seminar

Einheit 3

13.06.2025 | Janika Saretzki, MSc.

Termine

Einheit 1	02.05.25	14:45-16:15 Uhr	A + B	HS Audimax / P3
Einheit 2	15.05.25	13:05-15:30 Uhr	A + B	HS Audimax / P3

Einheit 3	12.06.25	13:05-15:30 Uhr	A	HS P5 005
Einheit 3	13.06.25	13:50-16:15 Uhr	B	HS P1 105

Einheit 4	26.06.25	13:20-15:45 Uhr	A	HS P5 005
Einheit 4	27.06.25	13:50-16:15 Uhr	B	HS P1 105

Einheit 5	17.07.25	13:20-15:45 Uhr	A	HS P5 005
Einheit 5	18.07.25	13:50-16:15 Uhr	B	HS P1 105

Taschenrechner-Hacks

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
59	61	59	63	57	61	46	63	58	56	72	54	45	57	57	72	66

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}
55	39	45	51	47	55	52	51	61	47	53	46	59	51	63	59	45

Arithmetisches Mittel: Mode → 2: STAT → Eine Gruppe: 1-VAR **oder** zwei Gruppen: A+BX → Werte für Gruppe 1 und/oder Gruppe 2 eintragen (jeden Wert jeweils mit "Enter/=" bestätigen) → AC → Shift + 1 → 4: Var → 2: \bar{x} (Mittelwert für Gruppe x/1), 5: \bar{y} Mittelwert für Gruppe y/2 → "Enter/="

$$\bar{x} = 59.18, \quad \bar{y} = 51.71$$

Standardabweichung auf Populationsebene: Mode → 2: STAT → Eine Gruppe: 1-VAR **oder** zwei Gruppen: A+BX → Werte für Gruppe 1 und/oder Gruppe 2 eintragen (jeden Wert jeweils mit "Enter/=" bestätigen) → AC → Shift + 1 → 4: Var → 4: s_x (Standardabweichung auf Populationsebene für Gruppe x/1), 7: s_y (Standardabweichung auf Populationsebene für Gruppe y/2) → "Enter/="

$$\sigma_x = 7.02, \quad \sigma_y = 6.29$$

Taschenrechner-Hacks

Median: Mode → 2: STAT → Eine Gruppe: 1-VAR → Werte für Gruppe 1/Gruppe 2 eintragen (jeden Wert jeweils mit "Enter/=" bestätigen) → AC → Shift + 1 → 5: MinMax → 4: med → "Enter/="

$$\text{Mdn}_x = 59.0, \quad \text{Mdn}_y = 51.0$$

Unteres Quartil: Mode → 2: STAT → Eine Gruppe: 1-VAR → Werte für Gruppe 1/Gruppe 2 eintragen (jeden Wert jeweils mit "Enter/=" bestätigen) → AC → Shift + 1 → 5: MinMax → 3: Q1 → "Enter/="

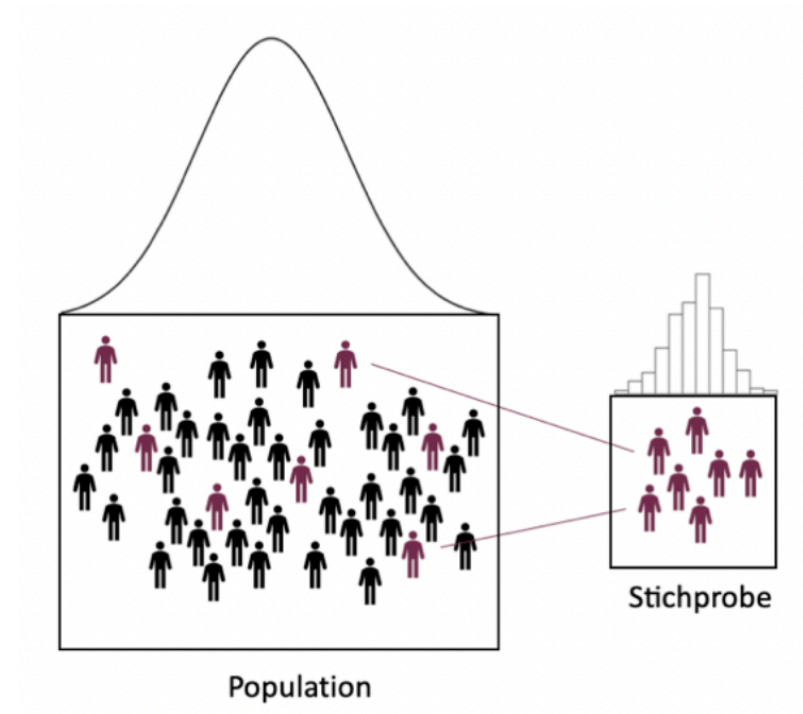
$$\text{Q1}_x = 56.5, \quad \text{Q1}_y = 46.5$$

Oberes Quartil: Mode → 2: STAT → Eine Gruppe: 1-VAR → Werte für Gruppe 1/Gruppe 2 eintragen (jeden Wert jeweils mit "Enter/=" bestätigen) → AC → Shift + 1 → 5: MinMax → 5: Q3 → "Enter/="

$$\text{Q3}_x = 63.0, \quad \text{Q3}_y = 57.0$$

Wiederholung

Logik des Schließens von Stichprobe auf Population



Wiederholung

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

Logik des Schließens von einer Stichprobe auf die Population: Inferenzstatistik (Grundlage: **Wahrscheinlichkeitsrechnung**)

Stochastik = Die Kunst des Vermutens (altgriechisch)

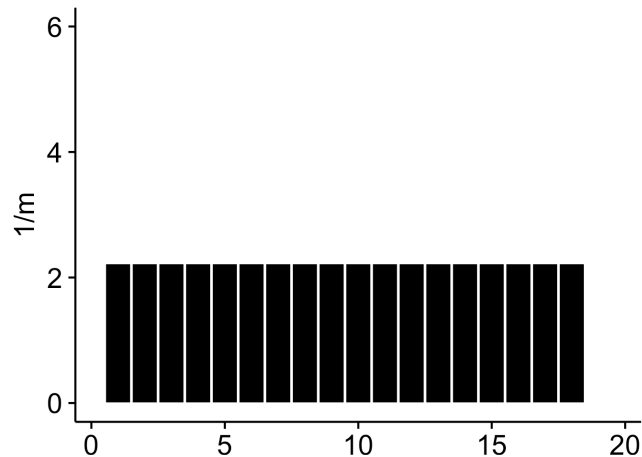
- Mathematik setzt Vorstellung von Zufall voraus (d.h. Modelle von Situationen, deren Ausgang ungewiss ist)
- Keine Einzelergebnisse vorhersagbar, aber:
→ Erkennen von Regelmäßigkeiten bei Vorgängen, deren Ergebnisse vom Zufall abhängen
- Zentraler Begriff **Zufallsexperiment**
 - Die Möglichen Ergebnisse eines Zufallsexperiments heißen **Elementarereignisse**
 - Die Menge aller möglichen Ergebnisse eines Zufallsexperiments bezeichnet man als **Ergebnisraum**
 - Der Ergebnisraum heißt **diskret**, wenn er aus abzählbar vielen Elementarereignissen besteht.
 - Der Ergebnisraum heißt **stetig**, wenn er aus überabzählbar vielen Elementarereignissen besteht.

Übungsaufgabe 1

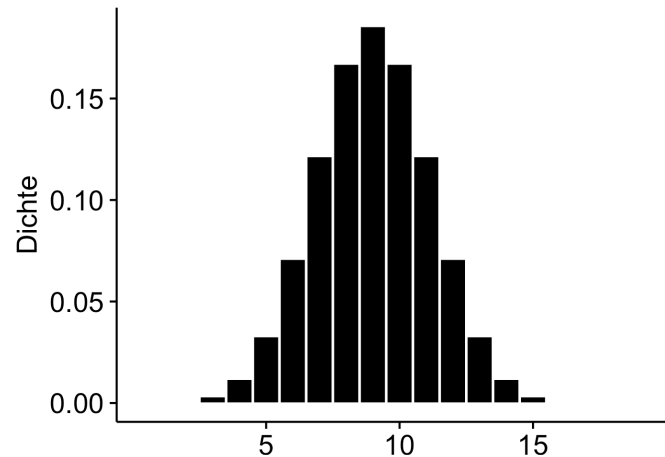
Unten sehen Sie drei Verteilungen. Bitte ordnen Sie folgende Begriffe der passenden Grafik zu:

Binomialverteilung, Gleichverteilung, Normalverteilung.

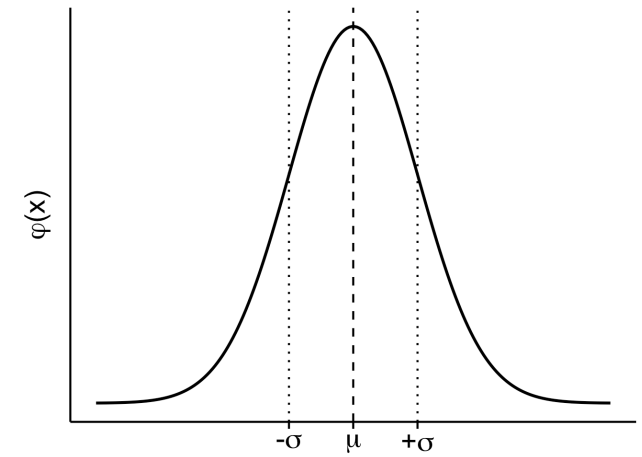
Bestimmen Sie außerdem, ob es sich um eine spezielle diskrete oder eine spezielle stetige Verteilung handelt.



$$P(X = x) = \frac{1}{m} \quad \text{für alle } x = 1, 2, \dots, m$$
$$E[X] = \frac{m+1}{2}$$
$$\sigma^2 = \frac{m^2 - 1}{12}$$



$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k = 0, 1, \dots, n$$
$$E[X] = np$$
$$\sigma^2 = np(1-p)$$



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

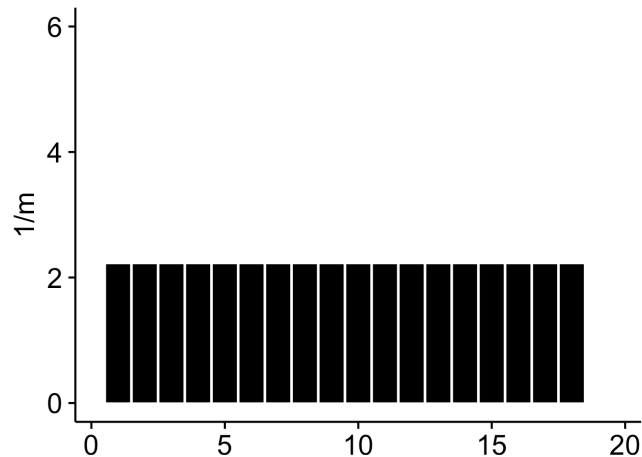
Übungsaufgabe 1 - Lösung

Unten sehen Sie drei Verteilungen. Bitte ordnen Sie folgende Begriffe der passenden Grafik zu:

Binomialverteilung, Gleichverteilung, Normalverteilung.

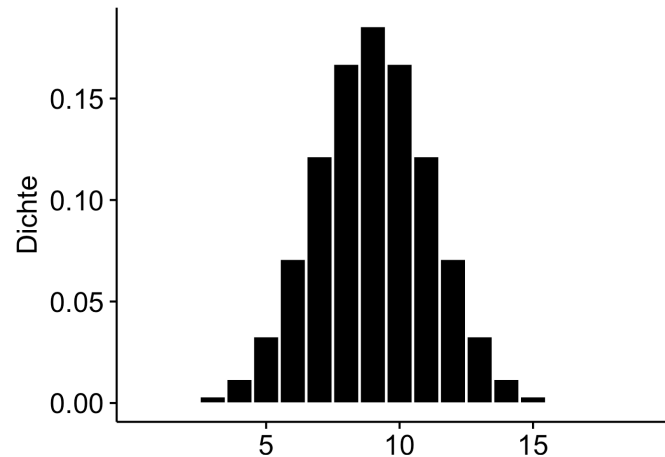
Bestimmen Sie außerdem, ob es sich um eine spezielle diskrete oder eine spezielle stetige Verteilung handelt.

Gleichverteilung (diskrete Verteilung)



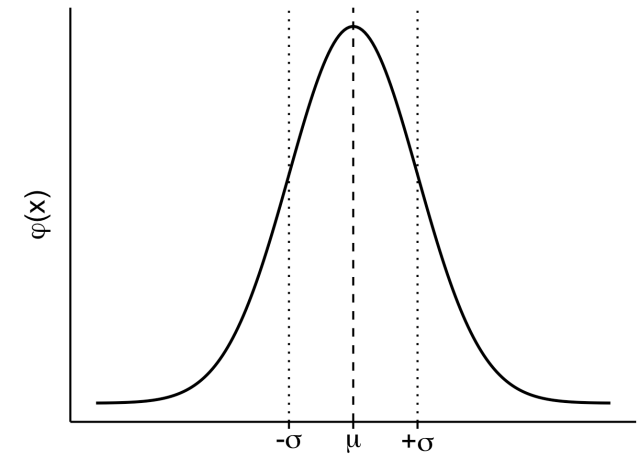
$$P(X = x) = \frac{1}{m} \quad \text{für alle } x = 1, 2, \dots, m$$
$$E[X] = \frac{m+1}{2}$$
$$\sigma^2 = \frac{m^2 - 1}{12}$$

Binomialverteilung (diskrete Verteilung)



$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k = 0, 1, \dots, n$$
$$E[X] = np$$
$$\sigma^2 = np(1-p)$$

Normalverteilung (stetige Verteilung)



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

Übungsaufgabe 1 - Lösung

Gleichverteilung (diskrete Verteilung): Die Balken sind alle gleich hoch, d.h. jeder Wert ist gleich wahrscheinlich (z.B. fairer Würfel).

Binomialverteilung (diskrete Verteilung): Beschreibt die Wahrscheinlichkeit, wie oft ein bestimmtes Ereignis in einer festen Anzahl von Versuchen vorkommen, wenn jeder Versuch nur zwei mögliche Ereignisse hat: Erfolg oder Misserfolg. Die Verteilung ist symmetrisch bei $p = 0.50$.

- "Wie oft passiert etwas, wenn ich n mal probiere?"
- Wahrscheinlichkeiten ergeben sich aus "Anzahl Erfolge in n Versuchen"
- Zählbare Ereignisse!

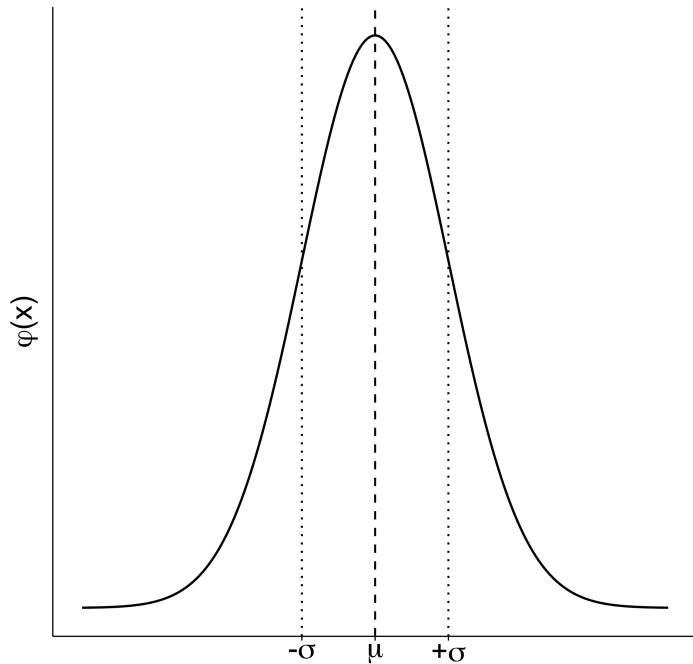
Die Binomialverteilung ist oft die Grundlage für Tests und wird dann genähert an die Normalverteilung (binomiale Annahmen "im Hintergrund")

Normalverteilung (stetige Verteilung): Die Kurve ist glockenförmig und symmetrisch um den Mittelwert

- Stetig, d.h. das Ergebnis kann jeden beliebigen Wert annehmen; glatt und glockenförmig
- Keine Balkendiagramm, sondern Dichtefunktion (durchgezogene Linie)
- "Wie ist etwas in der Natur verteilt?"
- Messbare, kontinuierliche Merkmale

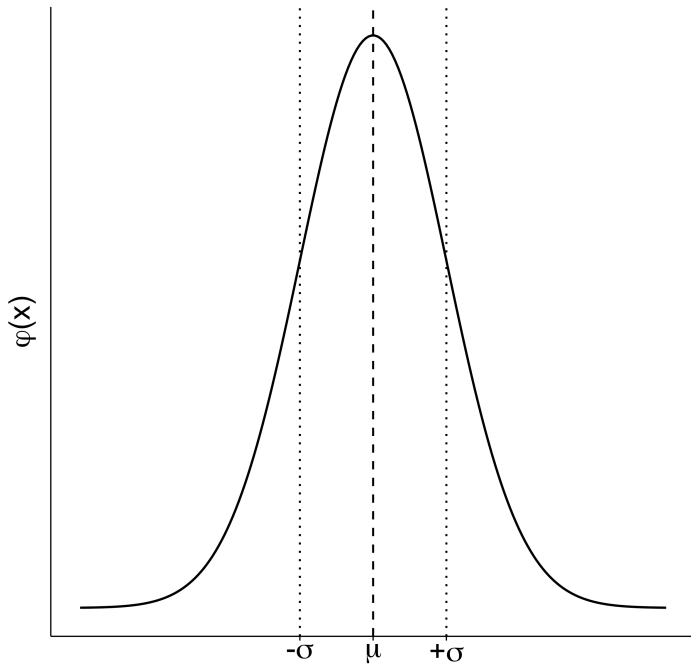
Übungsaufgabe 2

- a) Warum handelt es sich bei der Standardnormalverteilung um eine spezielle Normalverteilung?
- b) Wie kann eine Normalverteilung in eine Standardnormalverteilung überführt werden?



Übungsaufgabe 2 - Lösung

- a) Warum handelt es sich bei der Standardnormalverteilung um eine spezielle Normalverteilung?
b) Wie kann eine Normalverteilung in eine Standardnormalverteilung überführt werden?



a) Mittelwert = 0, Standardabweichung = 1 Jede Normalverteilung hat grundsätzlich irgendeinen Mittelwert und irgendeine Standardabweichung. Die Standardnormalverteilung ist genau die Normalverteilung, die auf 0 zentriert ist und eine "Standard Breite" von 1 hat → z-standardisierte Normalverteilung

b) Standardisierung (z-Transformation)

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

- Von jedem x-Wert ziehen wir den Mittelwert ab (Verschiebung von μ auf 0)
- Dann teilen wir durch die Standardabweichung (Streuung wird auf 1 skaliert)
- Jeder Wert bekommt einen z-Wert
- Nach der Transformation ist die Verteilung eine Standardnormalverteilung

Übungsaufgabe 3

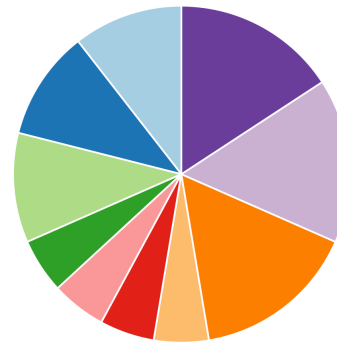
Unten sehen Sie zwei Glücksräder.

- a) In welchem Fall macht die Berechnung einer Laplace-Wahrscheinlichkeit Sinn?
- b) Wie ist die Zufallsvariable, eine bestimmte Farbe zu drehen, verteilt?
- c) Wie hoch ist die Wahrscheinlichkeit, ein grünes (hell oder dunkel) Feld zu drehen?

Glücksrad 1



Glücksrad 2



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

Übungsaufgabe 3 - Lösung

Unten sehen Sie zwei Glücksräder.

- a) In welchem Fall macht die Berechnung einer Laplace-Wahrscheinlichkeit Sinn?
- b) Wie ist die Zufallsvariable, eine bestimmte Farbe zu drehen, verteilt?
- c) Wie hoch ist die Wahrscheinlichkeit, ein grünes (hell oder dunkel) Feld zu drehen?

a) Die Laplace-Wahrscheinlichkeit setzt voraus, dass alle Ausgänge gleich wahrscheinlich sind. Das ist nur dann der Fall, wenn alle Felder gleich groß sind, denn dann hat jede Farbe die gleiche Chance, getroffen zu werden. Das trifft auf Glücksrad 1 zu.

b) Glücksrad 1: Diskrete Gleichverteilung, da jede Farbe mit der gleichen Wahrscheinlichkeit auftreten kann. Bei Glücksrad 2 liegt keine Gleichverteilung vor, die Felder unterscheiden sich in ihrer Größe; einige Farben sind wahrscheinlicher als andere.

c) Glücksrad 1 hat insgesamt 10 gleich große Felder, davon zwei grüne Felder (hell + dunkel):

$$P(\text{grün}) = \frac{2}{10} = \frac{1}{5} = 0,2$$

Die Wahrscheinlichkeit, ein grünes Feld zu drehen, liegt bei 20%.

Übungsaufgabe 4

Sie untersuchen, ob das Geben eines Zuckerwürfels dazu führt, dass eine Ratte innerhalb von fünf Minuten hyperaktiv wird. Sie interessieren sich dafür, mit welcher Wahrscheinlichkeit dieses Verhalten nach der Zuckeraufnahme auftritt.

- a) Warum handelt es sich hierbei um ein Zufallsexperiment?
- b) Welche Zufallsvariable erkennen Sie in diesem Experiment und wie ist diese verteilt?
- c) Ist der Ereignisraum diskret oder stetig?
- d) Bestimmen Sie Erwartungswert und Varianz der Zufallsvariablen, wenn die Wahrscheinlichkeit, dass die Ratte nach fünf Minuten hyperaktiv wird, 85% beträgt.
- e) Welche Verteilung würde man erhalten, würde das Experiment an mehreren Ratten durchgeführt werden?

Übungsaufgabe 4 - Lösung

Sie untersuchen, ob das Geben eines Zuckerwürfels dazu führt, dass eine Ratte innerhalb von fünf Minuten hyperaktiv wird. Sie interessieren sich dafür, mit welcher Wahrscheinlichkeit dieses Verhalten nach der Zuckeraufnahme auftritt.

a) Warum handelt es sich hierbei um ein Zufallsexperiment?

b) Welche Zufallsvariable erkennen Sie in diesem Experiment und wie ist diese verteilt?

a) 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 Versuch mit "gleichartigen" (unabhängigen) Objekten
- Ergebnis kann nicht sicher vorhergesagt werden; ist vom Zufall abhängig

b) Bernoulli-verteilte Zufallsvariable ("Ausgang des Experiments"), da zwei mögliche Ausgänge bei $n = 1$: Hyperaktiv vs. Nicht-Hyperaktiv. Eine Bernoulli-Verteilung ist die einfachste Wahrscheinlichkeitsverteilung für einen Zufallsversuch mit genau zwei möglichen Ausgängen: Erfolg (kodiert mit 1) und Misserfolg (kodiert mit 0)

→ Dichotome Zufallsvariable mit nur zwei möglichen Ausgängen

Übungsaufgabe 4 - Lösung

Sie untersuchen, ob das Geben eines Zuckerwürfels dazu führt, dass eine Ratte innerhalb von fünf Minuten hyperaktiv wird. Sie interessieren sich dafür, mit welcher Wahrscheinlichkeit dieses Verhalten nach der Zuckeraufnahme auftritt.

c) Ist der Ereignisraum diskret oder stetig?

d) Bestimmen Sie Erwartungswert und Varianz der Zufallsvariablen, wenn die Wahrscheinlichkeit, dass die Ratte nach fünf Minuten hyperaktiv wird, 85% beträgt.

c) Ereignisraum ist diskret; Es gibt nur zwei mögliche, abzählbare Ereignisse/Ausgänge: Hyperaktiv vs. Nicht-Hyperaktiv

d) Wir haben eine Zufallsvariable X , die den Wert ...

- ... $X = 1$ annimmt, wenn die Ratte hyperaktiv wird;
- ... $X = 0$ annimmt, wenn die Ratte nicht hyperaktiv wird.

Wir wissen: Die Wahrscheinlichkeit für $X = 1$ ist $p = 0.85$

$$E(X) = p = 0.85$$

Der Erwartungswert gibt an, wie oft wir im Durchschnitt bei vielen Versuchen "Erfolg" erwarten dürfen. Wenn wir das Experiment sehr oft wiederholen, werden im Durchschnitt 85% der Ratten hyperaktiv.

Übungsaufgabe 4 - Lösung

Sie untersuchen, ob das Geben eines Zuckerwürfels dazu führt, dass eine Ratte innerhalb von fünf Minuten hyperaktiv wird. Sie interessieren sich dafür, mit welcher Wahrscheinlichkeit dieses Verhalten nach der Zuckeraufnahme auftritt.

e) Welche Verteilung würde man erhalten, würde das Experiment an mehreren Ratten durchgeführt werden?

Die Varianz $\text{Var}(X)$ beschreibt, wie stark die einzelnen Untersuchungsergebnisse um den Erwartungswert streuen:

$$\text{Var}(X) = p \cdot (1 - p) = 0,85 \cdot (1 - 0,85) = 0,85 \cdot 0,15 = 0,1275$$

Die Varianz gibt an, wie unterschiedlich die Einzelergebnisse ausfallen können. Da es nur zwei mögliche Ergebnisse gibt (0 oder 1), ist die Varianz in diese Fall sehr gering. Je näher p and 0 oder 1 liegt, desto kleiner ist die Varianz. Am größten ist sie bei $p = 0.50$.

e) Wenn das Experiment an n Ratten durchgeführt wird und jede Ratte unabhängig von den anderen entweder hyperaktiv wird oder nicht, dann ist die Verteilung binomial: Zwei mögliche Ausgänge, $n > 1$

$X \sim \text{Binomial}(n, p)$: X ist binomialverteilt mit den Parameter

- n : Anzahl der unabhängigen Versuche (Ratten)
- p : Wahrscheinlichkeit für Hyperaktivität

Übungsaufgabe 4 - Lösung

Sie untersuchen, ob das Geben eines Zuckerwürfels dazu führt, dass eine Ratte innerhalb von fünf Minuten hyperaktiv wird. Sie interessieren sich dafür, mit welcher Wahrscheinlichkeit dieses Verhalten nach der Zuckeraufnahme auftritt.

Bernoulli- und Binomialverteilung hängen eng miteinander zusammen:

- Eine Bernoulli Verteilung beschreibt einen einzigen Versuch mit zwei möglichen Ausgängen: Erfolg vs. Misserfolg
- Eine Binomialverteilung beschreibt die Gesamtheit mehrerer unabhängiger Bernoulli-Versuche

Übungsaufgabe 5

In einer Schule befinden sich 750 Schülerinnen und Schüler. Bei 30% wurde eine depressive Symptomatik festgestellt. Der zuständige Schulpsychologe untersucht die achten Klassen mit insgesamt 123 Schülerinnen und Schülern.

- a) Sie interessieren sich für die Wahrscheinlichkeit, dass ein bestimmter Anteil der untersuchten Schülerinnen und Schüler eine depressive Symptomatik aufweise. Was bedeutet in diesem Zusammenhang ein "Erfolg" der Zufallsvariable, und wie ist diese verteilt?
- b) Wie groß ist die Wahrscheinlichkeit, dass genau 35 der 123 untersuchten Schülerinnen und Schüler eine depressive Symptomatik zeigt?

Übungsaufgabe 5 - Lösung

In einer Schule befinden sich 750 Schülerinnen und Schüler. Bei 30% wurde eine depressive Symptomatik festgestellt. Der zuständige Schulpsychologe untersucht die achten Klassen mit insgesamt 123 Schülerinnen und Schülern.

a) "Erfolg" ist in diesem Fall, dass eine Schülerin bzw. ein Schüler eine depressive Symptomatik aufweist (das, was wir zählen; binomialverteilt). Wenn wir aus den 123 untersuchten Schülerinnen und Schülern zufällig eine Person untersuchen, ist die Zufallsvariable

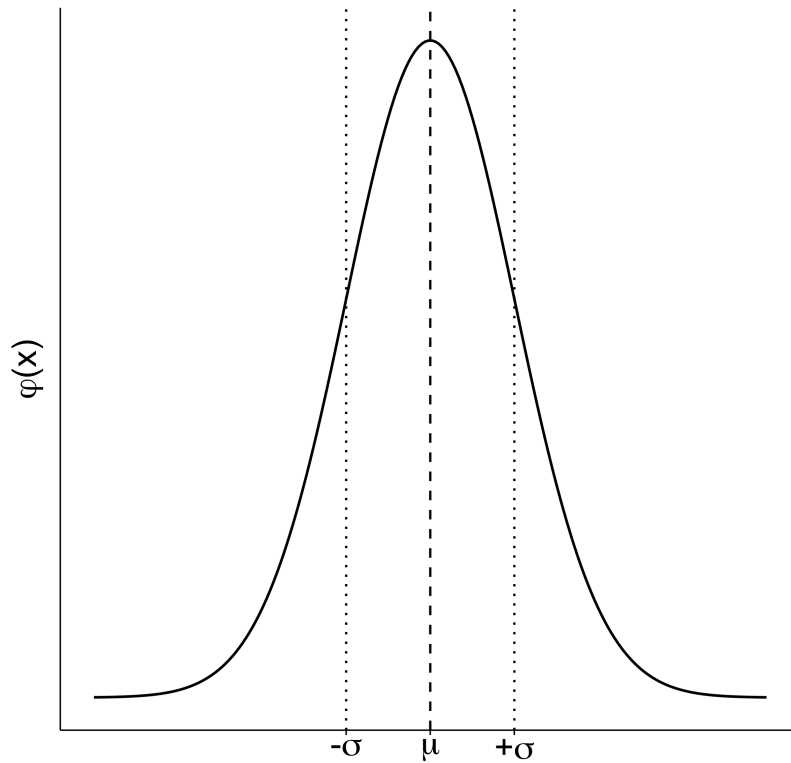
- $X = 1$, wenn die Person depressiv ist ("Erfolg")
- $X = 0$, wenn die Person nicht depressiv ist ("Misserfolg")

b) Wir berechnen die Binomialwahrscheinlichkeit für $k = 35$:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$
$$P(X = 35) = \binom{123}{35} \cdot 0,30^{35} \cdot (1 - 0,3)^{123-35} = 0,0740$$

Die Wahrscheinlichkeit, dass genau 35 der 123 Schülerinnen und Schüler eine depressive Symptomatik zeigt, liegt bei 7,4%.

Wiederholung

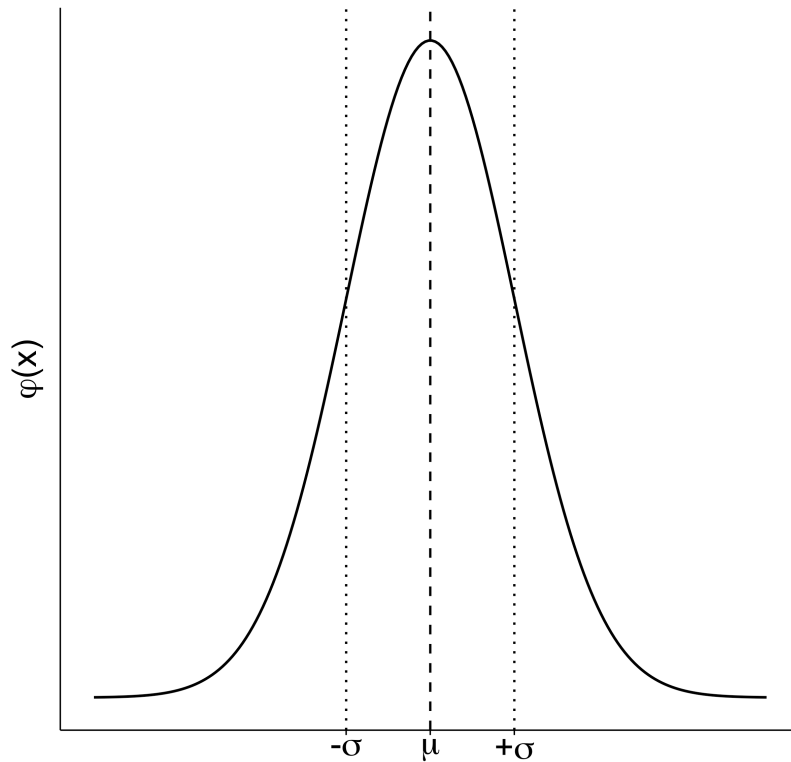


Dichtefunktion:

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

Integral über der Dichtefunktion → Bestimmung von Flächen
(Wahrscheinlichkeiten) unter der Normalverteilung

Wiederholung



Dichtefunktion:

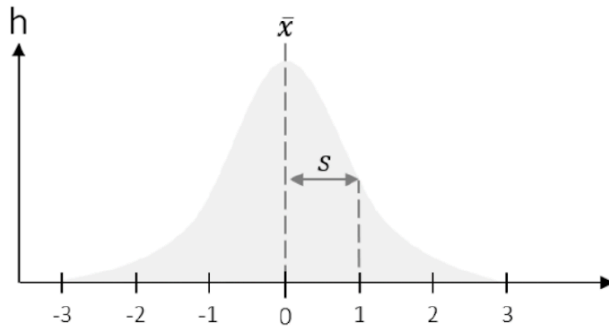
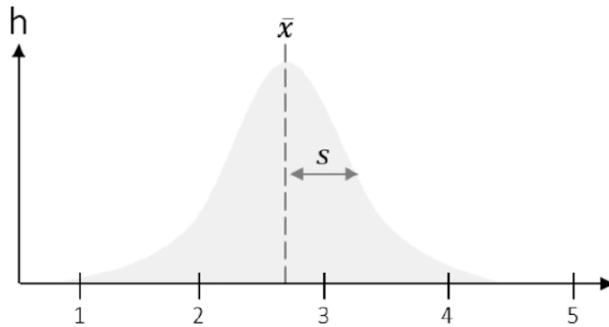
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

Integral über der Dichtefunktion → Bestimmung von Flächen
(Wahrscheinlichkeiten) unter der Normalverteilung

Einfacher: z-Standardisierung und z-Tabelle!

- Das aufwändige Berechnen von Integralen entfällt
- Sie können die Wahrscheinlichkeiten direkt ablesen!

Wiederholung



Z-Transformation:

$$Z_i = \frac{x_i - \mu}{\sigma}$$

Normalverteilung → Standardnormalverteilung

z-Wert*	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

z-Tabelle

- Wahrscheinlichkeit für jeden z-Wert kann abgelesen werden
- Zeile 1 (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
- **Voraussetzung ist, dass die Größe selbst normalverteilt ist**
- Um die Standardnormalverteilungstabelle 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

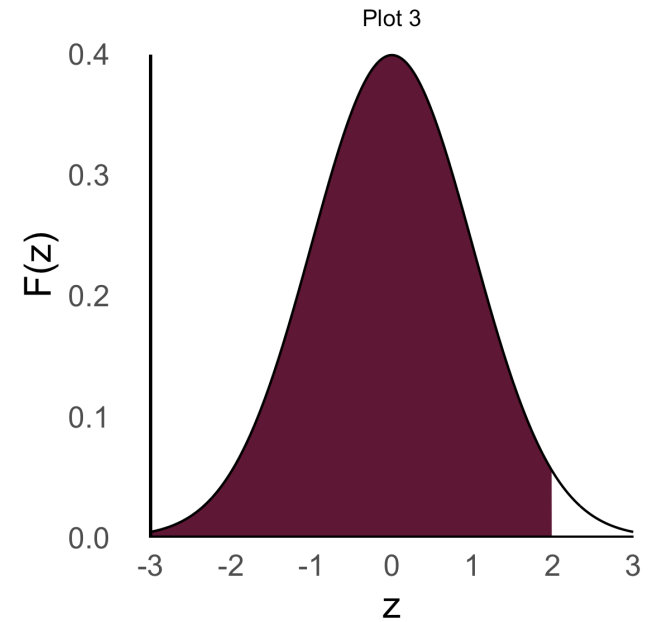
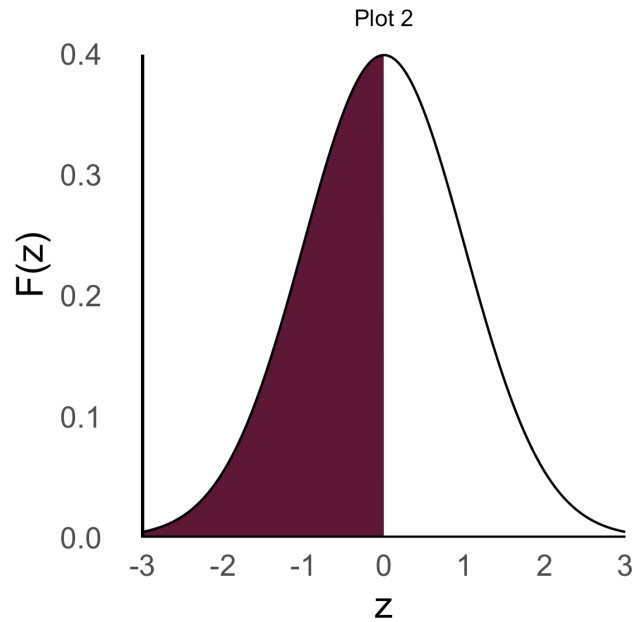
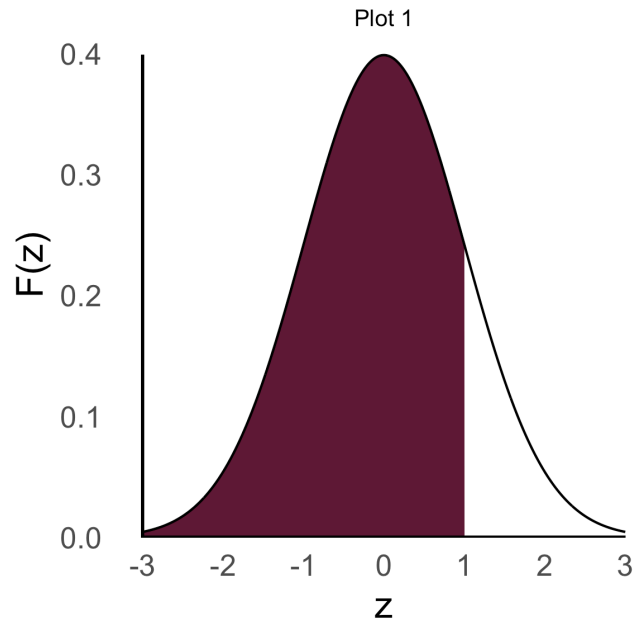
Vorteil der Standardisierung:

- Messwerte von Personen verschiedener Populationen sind oft nicht direkt vergleichbar
- Dennoch möchte man ausdrücken können, wie gut die beiden Leistungen innerhalb der Bezugsgruppe sind

z-Wert*	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767

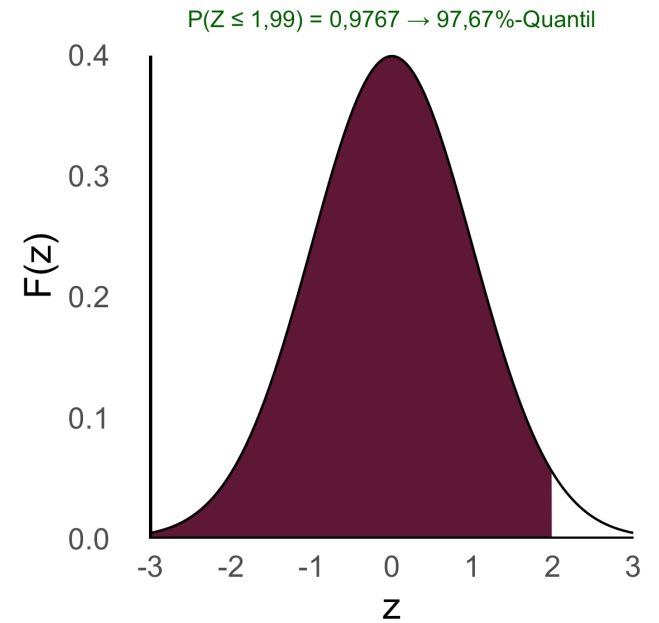
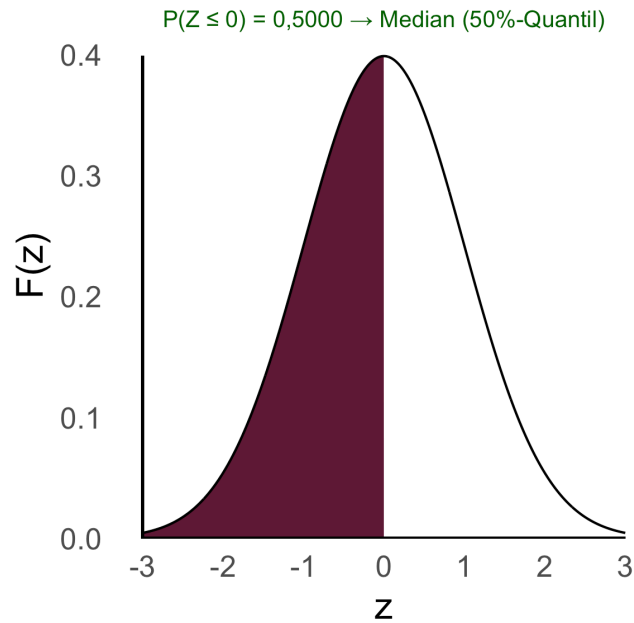
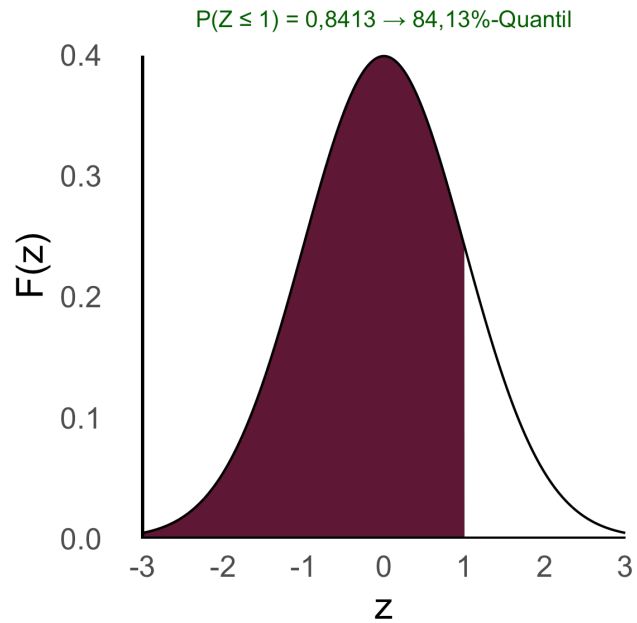
z-Tabelle

Ordnen Sie die markierten Werte in der Tabelle der letzten Folie der richtigen Visualisierung zu.



z-Tabelle - Lösung

Ordnen Sie die markierten Werte in der Tabelle der letzten Folie der richtigen Visualisierung zu.



Übungsaufgabe 6

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

Männer: $\mu = 1,80 \text{ m}$, $\sigma = 0,10 \text{ m}$ Frauen: $\mu = 1,70 \text{ m}$, $\sigma = 0,10 \text{ m}$

- a) Wie groß ist der Anteil der Männer, deren Körpergröße höchstens 1.95m beträgt?
- b) Wie groß ist der Anteil der Männer, deren Körpergröße mehr als 1.95m beträgt?
- c) Für welche Körpergröße liegt der Anteil der Männer, die mindestens so groß sind, knapp unter 5%?
- d) Wie groß müsste eine Frau sein, um - bezogen auf ihre Verteilung - denselben z-Score (Abstand zum Mittelwert in Standardabweichungen) zu erreichen wie ein Mann mit 1.95m Körpergröße?

Übungsaufgabe 6 - Lösung

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

Männer: $\mu = 1,80 \text{ m}$, $\sigma = 0,10 \text{ m}$ Frauen: $\mu = 1,70 \text{ m}$, $\sigma = 0,10 \text{ m}$

a) Wie groß ist der Anteil der Männer, deren Körpergröße höchstens 1.95m beträgt?

Höchstens = kleiner oder gleich, d.h. wir suchen nach $P(X \leq 1.95)$. Dafür müssen wir, in einem ersten Schritt, das gesuchte $x = 1,95$ in einen z-Wert umwandeln, wenn nur $P(Z \leq z)$ können wir in der Tabelle ablesen.

1. Standardisierung (Z-Transformation): Die Normalverteilung wird mit der z-Transformation auf die Standardnormalverteilung (Mittelwert 0, Standardabweichung 1) gebracht.

$$Z_i = \frac{x_i - \mu}{\sigma} = \frac{1,95 - 1,80}{0,1} = 1,5$$

Das bedeutet: 1,95m liegt 1,5 Standardabweichungen über dem Mittelwert.

Nachschlagen in der z-Tabelle: Nun, im zweiten Schritt, lesen wir aus der z-Tabelle den folgenden Wert ab: $P(Z \leq 1,5) = 0,9332$. Das bedeutet: 93,32% der Männer sind höchstens 1,95m groß.

→ Es besteht eine 93,32%ige Wahrscheinlichkeit, dass ein zufällig ausgewählter Mann höchstens 1,95m groß ist.

Übungsaufgabe 6 - Lösung

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

$$\text{Männer: } \mu = 1,80 \text{ m, } \sigma = 0,10 \text{ m} \quad \text{Frauen: } \mu = 1,70 \text{ m, } \sigma = 0,10 \text{ m}$$

b) Wie groß ist der Anteil der Männer, deren Körpergröße mehr als 1.95m beträgt?

Gesucht ist $P(X > 1,95) = P(Z > 1,5)$, wir können die Gegenwahrscheinlichkeit nutzen (siehe Ergebnis a))

$$1 - P(X \leq 1,95) = 1 - P(Z \leq 1,5) = 1 - 0,9332 = 0,0668$$

→ 6,68% der Männer sind größer als 1,95m

→ Die Wahrscheinlichkeit, dass ein zufällig ausgewählter Mann über 1,95m groß ist, beträgt 6,68%.

Die Standardnormalverteilung ist symmetrisch, aber: Die meisten Tabellen zeigen nur die linke Fläche unter der Kurve, also $P(Z \leq z)$. Wenn Sie wissen möchten, wie viel rechts von einem bestimmten z.Wert liegt, müssen Sie den Rest berechnen:

$$P(Z > z) = \text{Gesamtfläche (1)} - \text{Fläche links von } z$$

Übungsaufgabe 6 - Lösung

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

$$\text{Männer: } \mu = 1,80 \text{ m, } \sigma = 0,10 \text{ m} \quad \text{Frauen: } \mu = 1,70 \text{ m, } \sigma = 0,10 \text{ m}$$

c) Für welche Körpergröße liegt der Anteil der Männer, die mindestens so groß sind, knapp unter 5%?

Anders formuliert: Wie groß muss ein Mann mindestens sein, um zu den größten 5% zu gehören?

Gesucht ist der x-Wert (Körpergröße), für den gilt: $P(X \geq x) = 0,05$ bzw. $P(X < x) = 0,95$

Wir suchen also den 95. Perzentilwert der Verteilung.

Ein Perzentilwert gibt an, bis zu welchem Wert ein bestimmter Prozentsatz der Daten liegt. Der 95. Perzentilwert ist der Wert, unter dem 95% aller beobachteten Werte liegen. Oder anders gesagt: 5% der Werte sind größer als dieser Wert.

Übungsaufgabe 6 - Lösung

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

$$\text{Männer: } \mu = 1,80 \text{ m, } \sigma = 0,10 \text{ m} \quad \text{Frauen: } \mu = 1,70 \text{ m, } \sigma = 0,10 \text{ m}$$

c) Für welche Körpergröße liegt der Anteil der Männer, die mindestens so groß sind, knapp unter 5%?

Schritt 1: z-Wert zum 95%-Quantil bestimmen

In der z-Tabelle suchen wir den Wert, bei dem $P(Z \leq z) = 0,9505$. Dieser liegt bei $z = 1,65$. Das bedeutet, ein Wert, der 1,65 Standardabweichungen über dem Mittelwert liegt, hat eine Wahrscheinlichkeit von ca. 95%, unterschritten zu werden.

Schritt 2: Rückrechnung von z auf x (Originaleinheit Meter)

$$Z_i = \frac{x_i - \mu}{\sigma}$$

$$Z_i \cdot \sigma = x_i - \mu$$

$$x_i = Z_i \cdot \sigma + \mu = 1,65 \cdot 0,10 + 1,80 = 1,965$$

Ein Mann muss mind. 1,97m groß sein, um zu den größten 5% zu gehören. Oder anders: Nur 5% der Männer sind größer als 1,97m.

Übungsaufgabe 6 - Lösung

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

Männer: $\mu = 1,80 \text{ m}$, $\sigma = 0,10 \text{ m}$ Frauen: $\mu = 1,70 \text{ m}$, $\sigma = 0,10 \text{ m}$

c) Für welche Körpergröße liegt der Anteil der Männer, die mindestens so groß sind, knapp unter 5%?

Schritt 3: Zusatz: Untere 5%

So wie 5% der Männer größer als ein bestimmter Wert sind (oberes 5%-Perzentil), sind auch 5% der Männer kleiner als ein entsprechender Wert auf der linken Seite der Verteilung, d.h. auch bei $z = -1,65$ liegt die Fläche unter der Kurve bei ca 4,95%

$$Z_i = \frac{x_i - \mu}{\sigma}$$

$$Z_i \cdot \sigma = x_i - \mu$$

$$x_i = Z_i \cdot \sigma + \mu = -1,65 \cdot 0,10 + 1,80 = 1,635$$

Nur 5% der Männer sind kleiner als 1,64m.

Übungsaufgabe 6 - Lösung

Für die normalverteilte Zufallsvariable „Körpergröße“ sind die Parameter für die männliche und weibliche Bevölkerung eines Landes wie folgt gegeben:

$$\text{Männer: } \mu = 1,80 \text{ m, } \sigma = 0,10 \text{ m} \quad \text{Frauen: } \mu = 1,70 \text{ m, } \sigma = 0,10 \text{ m}$$

d) Wie groß müsste eine Frau sein, um - bezogen auf ihre Verteilung - denselben z-Score (Abstand zum Mittelwert in Standardabweichungen) zu erreichen wie ein Mann mit 1.95m Körpergröße?

$$Z_i = \frac{x_i - \mu}{\sigma}$$

$$1,5 = \frac{x_i - 1,70}{0,1}$$

$$x_i - 1,70 = 1,5 * 0,1$$

$$x_i = 1,5 * 0,1 + 1,70 = 1,85$$

Eine Frau müsste 1,85m groß sein, um denselben z-Wert (1,5) zu erreichen wie ein Mann mit 1,95m.

Möglichkeit der Visualisierung

