BSc PHY – Angewandte Statistik

LE2 Zuvallsvariablen und Verteilungen

Patric Eichelberger & Aglaja Busch
aF&E Physiotherapie

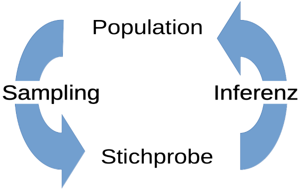
patric.eichelberger@bfh.ch | aglaja.busch@bfh.ch
Moodlekurs

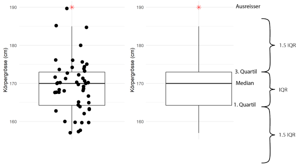
9. Mai 2025

Intro

Rückblick LE1 Deskriptive Statistik

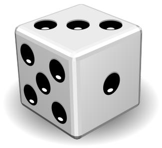
- Es muss zwischen der Grundgesamtheit und einer betrachteten Stichprobe unterschieden werden.
- Eine Stichprobe ist nicht zwingend repräsentativ was die Gültigkeit einer Statistik stark beeinflussen kann.
- Wir unterscheiden zwischen beschreibender und schliessender Statistik sowie quantitativen (rational- und intervallskaliert) und qualitativen (nominal, ordinal) Merkmalen.
- Wir benötigen graphische Werkzeuge wie Balkendiagramm, Histogramm und Boxplot um Daten visuell zu beschreiben.
- Wir haben Lage- (Mittelwert, Median, Perzentile) und Streumasse (Varianz, Spannweite, IQR) gesehen um Daten zusammenfassend mit Zahlen zu beschreiben.

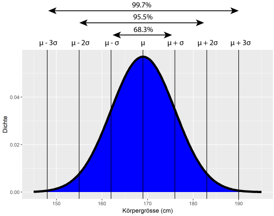




Lernziele Zufallsvariablen und Verteilungen

- Die Studierenden kennen den Begriff der Wahrscheinlichkeit im Zusammenhang mit Häufigkeiten und Verteilungen.
- Die Studierenden kennen die Bedeutung der Normalverteilung.
- Die Studierenden können Wahrscheinlichkeiten und Quantile der Normalverteilung in jamovi berechnen.
- Die Studierenden können einen QQ-Plot erstellen und anhand eines QQ-Plots beurteilen, ob eine Variable normalverteilt ist oder nicht.





Verteilung von Zufallsvariablen

Zufallsvariablen

Eine Zufallsvariable beschreibt das Ergebnis eines Zufallsexperiments.
Zufallsvariablen werden häufig mit X bezeichnet, das Ergebnis davon mit x.

Beispiele

- Einen (fairen) Würfel werfen.
- Die Körpergrösse einer zufällig ausgewählten Person.
- M&M’s ziehen.

Verteilung einer Zufallsvariablen

Die Verteilung einer Zufallsvariablen gibt an, wie wahrscheinlich welches Ergebnis ist.

Beispiel 1

Die Wahrscheinlichkeit eine bestimmte Augenzahl zu bekommen ist immer gleich (Gleichverteilung).

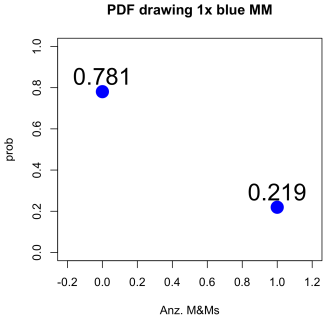
Augenzahl	1	2	3	4	5	6
P(Augenzahl)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

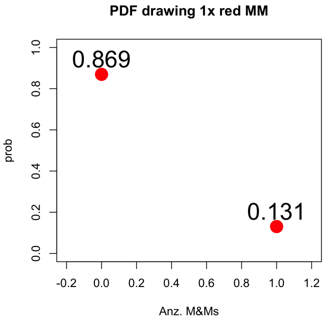
Beispiel 2

Verteilung von Farben von M&M’s sind nicht gleichverteilt

Bern University of Applied Sciences

1x M&M's ziehen

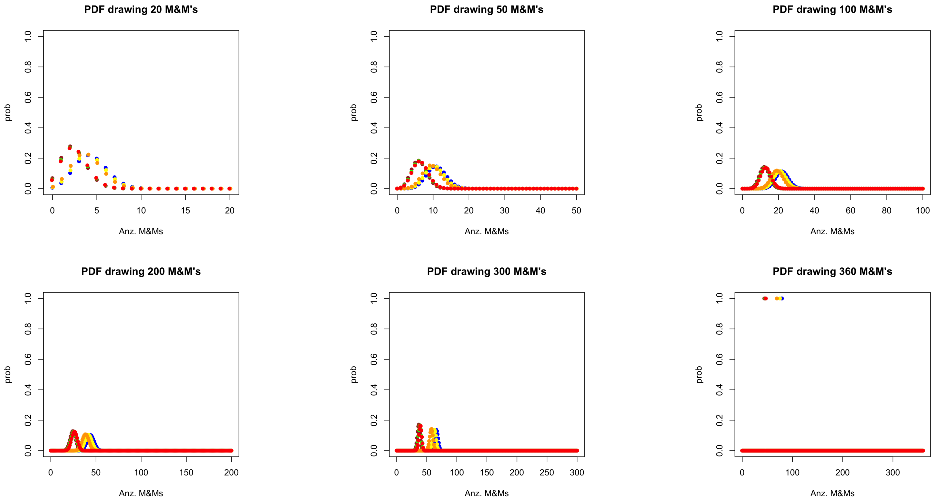




Color	N	Prop
blue	79	0.22
brown	44	0.12
green	46	0.13
orange	69	0.19
red	47	0.13
yellow	75	0.21

Die relative Häufigkeit in der Stichprobe: Wahrscheinlichkeit ein M&M's der Farbe zu ziehen.

N-mal M&M's ziehen

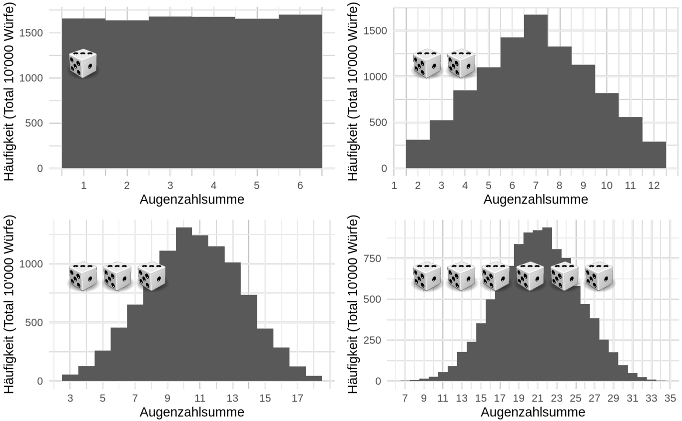


Urnenmodell ohne Zurücklegen » hypergeometrische Verteilung

Verteilung einer Zufallsvariablen (cont.)

Beispiel 3

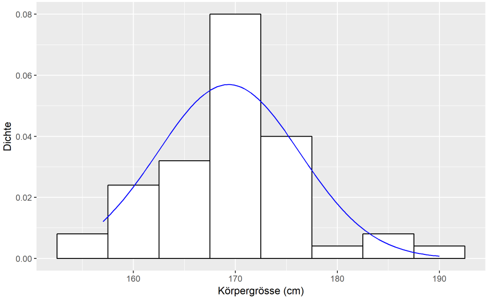
Mittlere Augenzahlsumme bei mehr als einem Würfel ist normalverteilt.



Verteilung einer Zufallsvariablen (cont.)

Beispiel 4

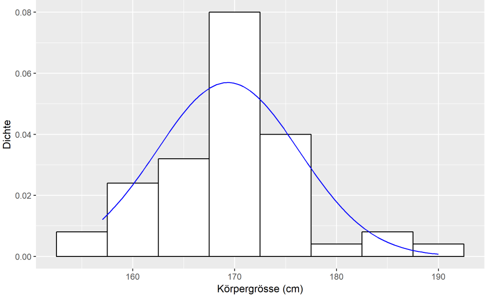
Körpergrösse folgt einer Normalverteilung

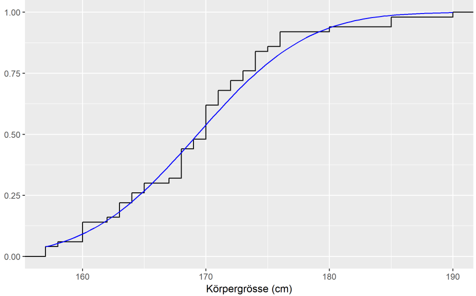


Beschrieb von Verteilungen mittels Dichte-, Verteilungs- und Quantilfunktion. Alle drei sind gleichwertig.

Dichte- und Verteilungsfunktion

Dichte- und Verteilungsfunktion

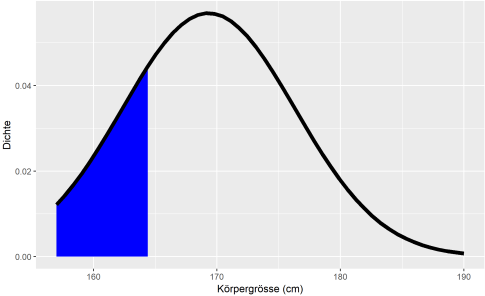


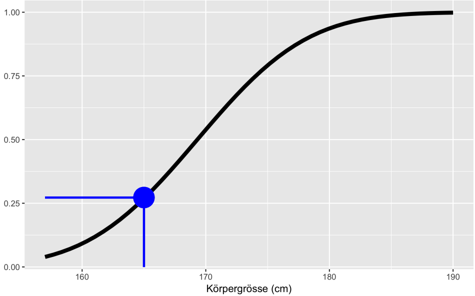


Ein Histogramm schätzt die Dichtefunktion (auch Wahrscheinlichkeitsdichtefunktion/Eng. probability density function PDF) f (blaue Kurve).

Die empirische Verteilungsfunktion F_n schätzt die Verteilungsfunktion (auch kumulative Verteilungsfunktion/Eng. cumulative distribution function CDF) F (blaue Kurve).

Dichte- und Verteilungsfunktion (cont.)

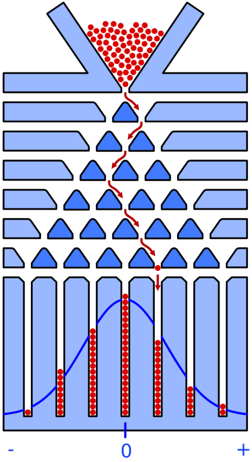




- Bedeutung der Verteilungsfunktion der Variable X:
$$F(x) = P(X \leq x)$$
- Wenn z.B. X die Körpergrösse von ist, ist $F(165)$ die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person kleiner oder gleich 165 cm gross ist.
- Die Verteilungsfunktion an der Stelle x ist die Fläche unter der Dichtefunktion bis zu Stelle x.

Normalverteilung

Zentraler Grenzwertsatz: Galtonbrett



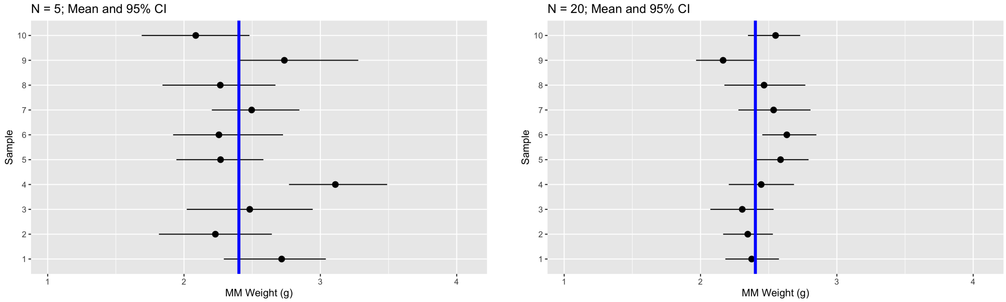
Quelle: wikipedia.org

Merke

Der Mittelwerte einer genügend grossen Anzahl Zufallsexperimente ist näherungsweise normalverteilt.

- Zufallsexperiment im Bild: Nach rechts oder links fallen über sechs Ebenen.
- Wiederholung mit unendlich vielen Kugeln führt zu einer Normalverteilung über die Fächer.
- Praktische Beispiele: (i) Zufälliger Messfehler bei wiederholten Messungen, (ii) Ziehen von zufälligen Stichproben aus einer Population

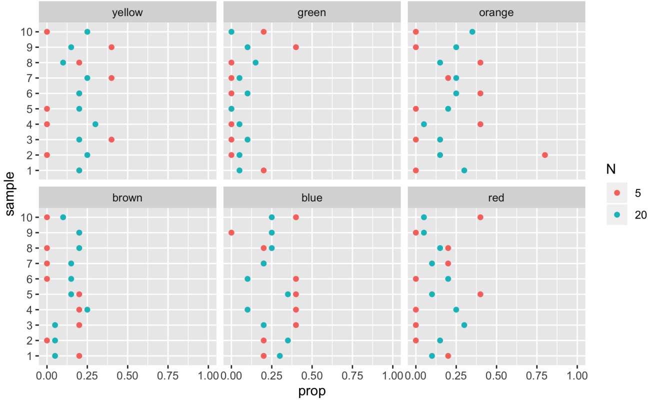
Übung M&Ms ziehen: Gewicht



Beobachtung

Die Streuung der Stichprobenmittelwerte wird kleiner wenn die Stichproben grösseren Umfang haben.

Übung M&Ms ziehen: Farben

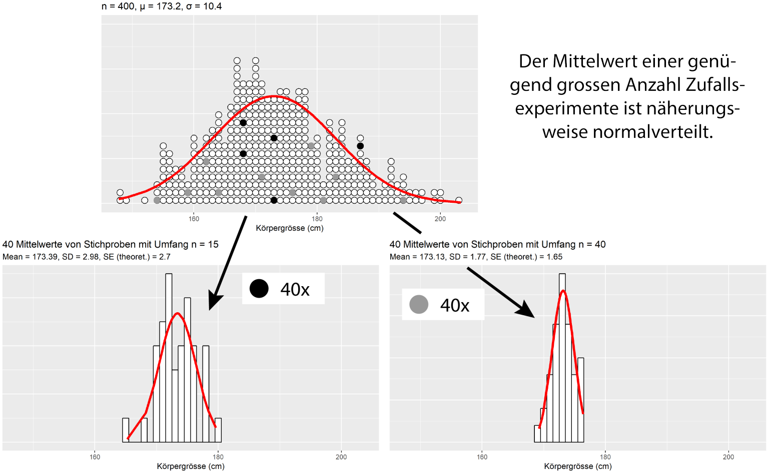


Color	N	Prop
blue	79	0.22
brown	44	0.12
green	46	0.13
orange	69	0.19
red	47	0.13
yellow	75	0.21

Beobachtung

Die Streuung der Stichprobenmittelwerte wird kleiner wenn die Stichproben grösseren Umfang haben.

Zentraler Grenzwertsatz: Stichprobenmittelwerte



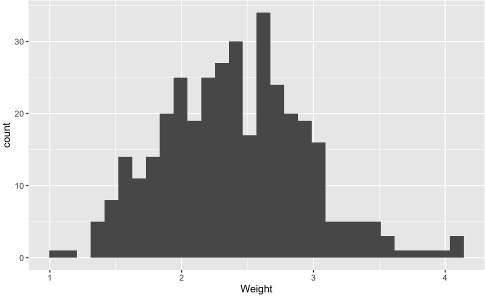
Der Mittelwert einer genügend grossen Anzahl Zufallsexperimente ist näherungsweise normalverteilt.

Stichprobenumfang → Grösser

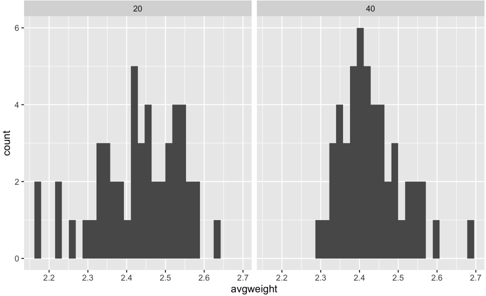
Verteilung der Mittelwerte → Normaler und schmaler

Gewicht M&M’s: 50 Stichproben

Population

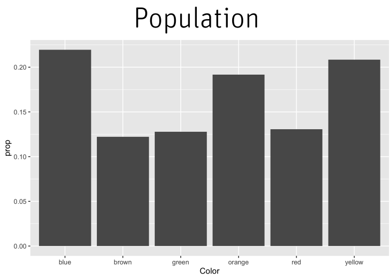


Stichproben



- Gleiches Schema wie Übung aber mit N = 20 und N =40
- Mit 50 Stichprobenmittelwerten macht ein Histogram Sinn
- Annäherung Stichprobenmittelwerte an Normalverteilung

Farben M&M's: 50 Stichproben



- Farbproportionen in Population offensichtlich nicht normalverteilt
- Stichprobenmittelwerte hingegen schon bei genügend grossen Stichproben

Normalverteilung

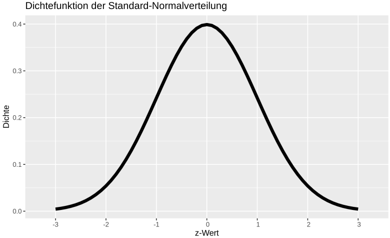
Die Normalverteilung, auch Gaussverteilung oder Gausssche Glocke genannt, spielt eine sehr wichtige Rolle in der Wahrscheinlichkeitstheorie und Statistik:

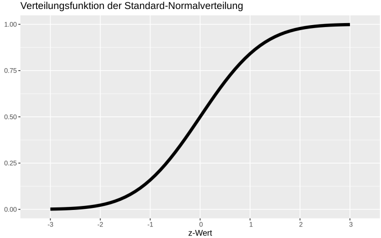
- Mittelwerte von genügend vielen unabhängigen Zufallsexperimenten sind annäherungsweise normalverteilt (nach dem sogenannten zentralen Grenzwertsatz).
- Zufällige Messfehler sind bei einigen Versuchen (z.B. physikalische Laborversuche) annäherungsweise normalverteilt.
- Einige quantitative Merkmale sind normalverteilt (z.B. Körpergrösse nach Geschlecht).
- Für normalverteilte Variablen gelten besonders einfache mathematische Sätze.

Aus diesen Gründen gibt es viele Tests und Schätzer, die die Normalverteilung als Voraussetzung haben.

Normalverteilung (cont.)

- Die Normalverteilung hat die zwei Parameter μ (Mittelwert) und σ (Standardabweichung).
- Der Spezialfall mit $\mu = 0$ und $\sigma = 1$ führt zur Standard-Normalverteilung.
- Die Normalverteilung ist eine symmetrische Verteilung.

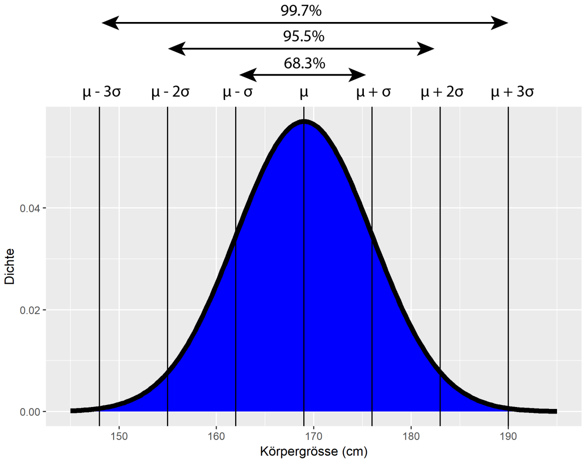




z-Transformation: Umrechnung in Standard-Normalverteilung

$$z_n = \frac{x_n - \mu}{\sigma}$$

Dichte der Normalverteilung

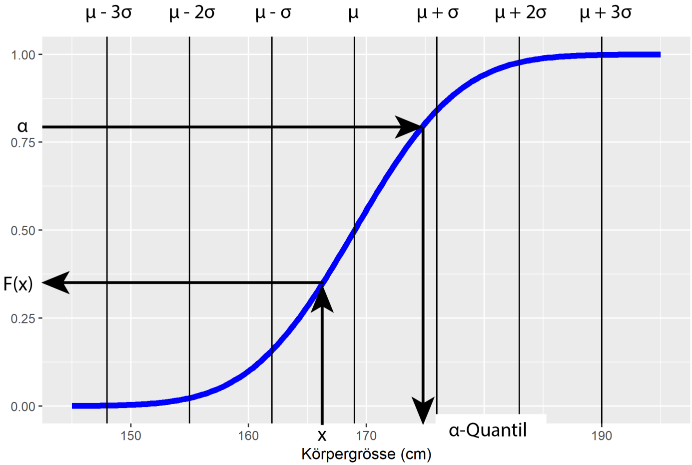


Weitere wichtige z-Werte

z	Fläche
1.96	95%
2.58	99%
3.29	99.9%

Annahme: Körpergrösse in der Population ist normalverteilt mit Mittelwert $\mu = 169$ cm und Standardabweichung $\sigma = 7$ cm.

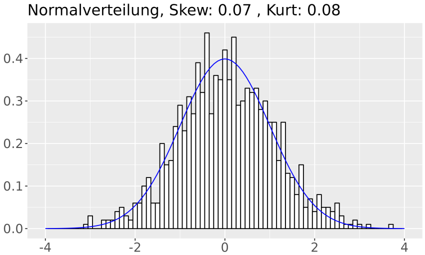
Zusammenhang Quantile und Verteilungsfunktion

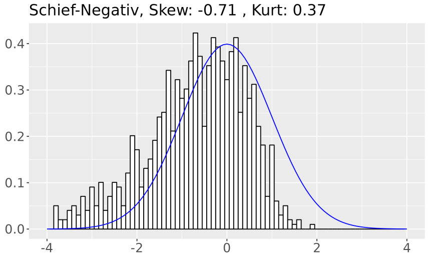


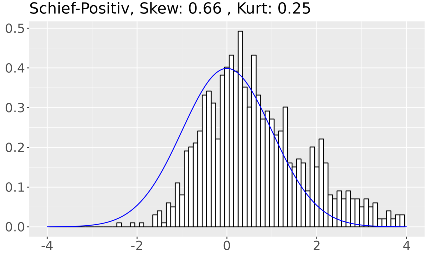
Das α-Quantil ist die Umkehrfunktion von der Verteilungsfunktion an der Stelle α.

Beurteilung von Verteilungen

Schiefe (Skewness): Asymmetrie



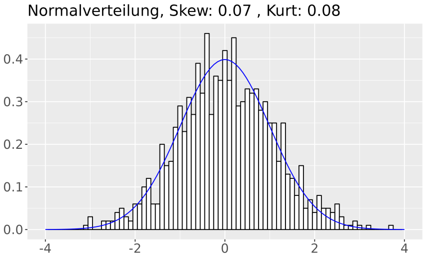


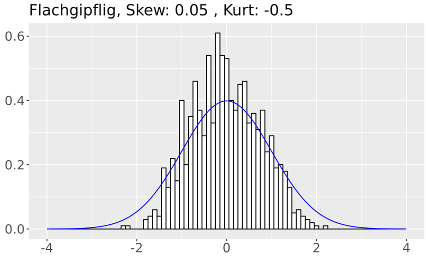


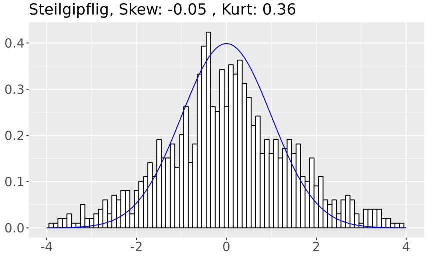
Kennzahl misst die Asymmetrie

- < 0: linksschief, rechtssteil
- > 0: rechtsschief, linkssteil

Wölbung (Kurtosis): Spitzigkeit

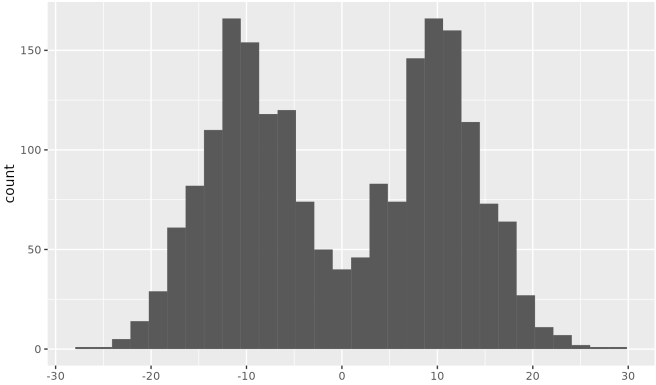






- < 0: Flacher und breiter als Normalverteilung
- > 0: Steiler und schmaler als Normalverteilung

Modus

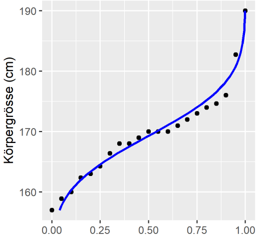


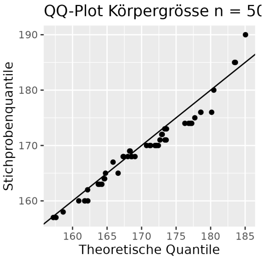
- Wert der am häufigsten in einem Datensatz vorkommt.
- ACHTUNG: es kann mehrere Moden geben (hier bimodal)

Bern University of Applied Sciences

QQ-Plot

- Zur Beurteilung, ob Daten normalverteilt sind, benutzen wir das sogenannte Quantil-Quantil-Diagramm, kurz QQ-Plot.
- Darin werden die geordneten Stichprobenwerte (schwarze Punkte) gegen die theoretischen Quantile (blaue Linie) der Standardnormalverteilung aufgetragen.
- Wenn die Daten normalverteilt sind, sollten alle Punkte auf einer Geraden liegen.
- ACHTUNG: Ein QQ-Plot kann erst ab einer Stichprobengrösse von mindestens 30 eingesetzt werden! [1]

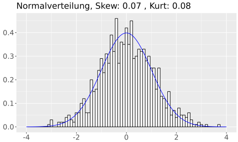


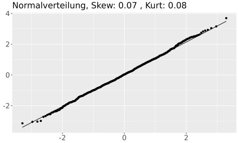


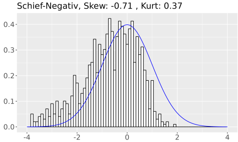
[1] Razali N. M. and Wah Y. B. (2011) J Stat Mod Ana, 2(1), 21-33

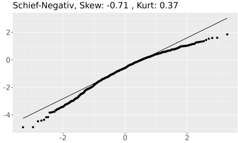
QQ-Plot: Beispiele

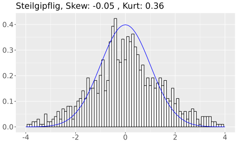
Normalverteilung	Schief negativ	Steilgipflig	Gleichverteilung

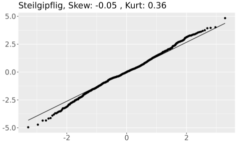


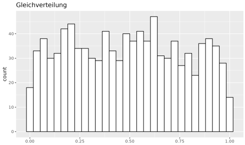


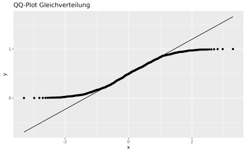










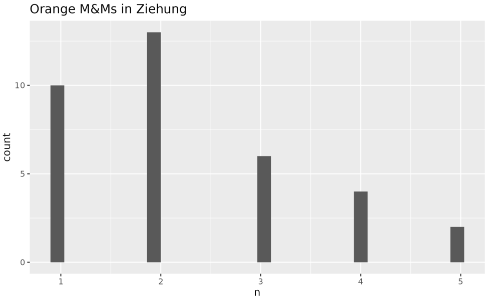


Normalverteilt	Nicht normalverteilt	Nicht normalverteilt	Nicht normalverteilt

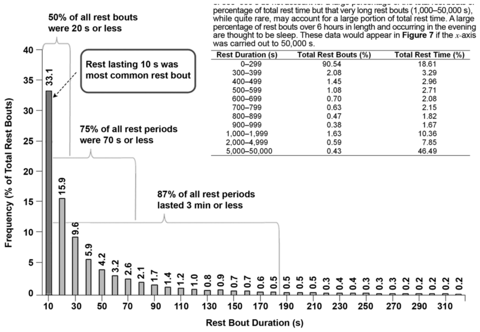
Intro
Verteilung von Zufallsvariablen
Dichte- und Verteilungsfunktion
Normalverteilung
Beurteilung von Verteilungen
Praxis mit R
Zusammenfassung

Andere wichtige Verteilungen

Poisson-Verteilung bei Zählgrössen: Wahrscheinlichkeit, dass ein Ereignis eine bestimmte Anzahl Male passiert.



Exponentialverteilung bei Zeitintervallen: Wahrscheinlichkeit der Zeitdauer bis zum Eintreten eines Ereignisses.



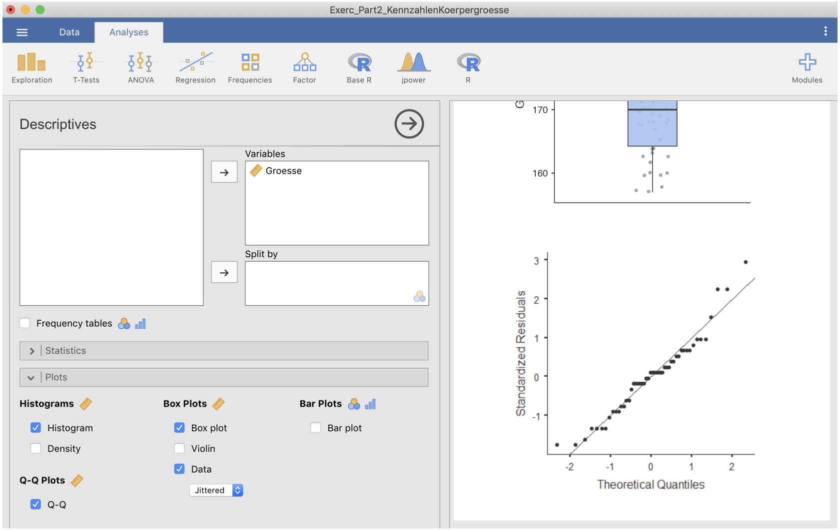
Rest Duration (s)	Total Rest Bouts (%)	Total Rest Time (%)
0–299	90.54	18.61
300–399	2.08	3.29
400–499	1.45	2.96
500–599	1.08	2.71
600–699	0.70	2.08
700–799	0.63	2.15
800–899	0.47	1.82
900–999	0.38	1.67
1,000–1,999	1.63	10.36
2,000–4,999	0.59	7.85
5,000–50,000	0.43	46.49

Orendurf et al. 2008 J Rehab Res Dev

Praxis mit R

QQ-Plot Körpergrösse

Datengrundlage: `KennzahlenKoerpergroesse_M_F.xlsx`

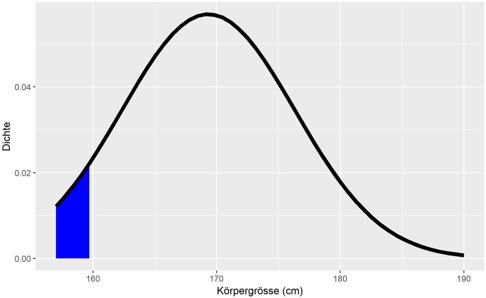


Wahrscheinlichkeit Körpergrösse

Nehmen wir an die Körpergrösse sei normalverteilt mit Mittelwert 169 cm und Standardabweichung 7 cm.

Aufgabe

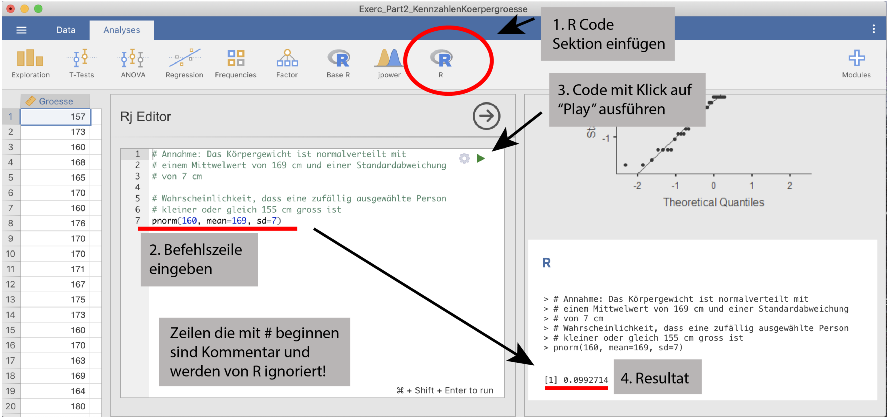
Wie gross ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person kleiner oder gleich 160 cm gross ist?



Wahrscheinlichkeit Körpergrösse in Jamovi

```
pnorm(160, mean=169, sd=7)
```

Wir erhalten 0.1, also ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person kleiner oder gleich 160 cm gross ist 10%.

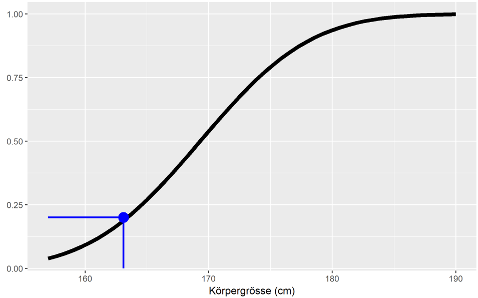


Quantile Körpergrösse

Nehmen wir an die Körpergrösse sei normalverteilt mit Mittelwert 169 cm und Standardabweichung 7 cm.

Aufgabe

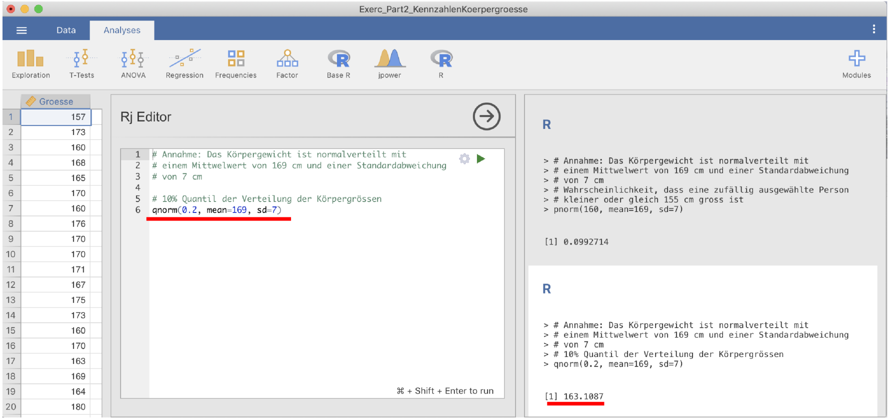
Wie gross ist das 20%-Quantil der Verteilung der Körpergrössen?



Quantile Körpergrösse in Jamovi

```
qnorm(0.2, mean=169, sd=7)
```

Wir erhalten 163.1 cm. Wenn eine Normalverteilung mit den gegebenen Parametern vorliegt sind also 20% der Personen kleiner oder gleich 163.1 cm.



Bern University of Applied Sciences

Zusammenfassung

Zusammenfassung

- Ein einfaches Zufallsexperiment ist einen (fairen) Würfel werfen, dessen Ergebnis gleichverteilt ist.
- Merkmale sind oft nicht gleichverteilt. Deren Verteilung kann mittels Histogramm geschätzt werden.
- Die Normalverteilung spielt eine zentrale Rolle da viele Merkmale oft normalverteilt sind.
  - $\mu \pm \sigma \Rightarrow$ 68.3% der Daten
  - $\mu \pm 2\sigma \Rightarrow$ 95.5% der Daten
  - $\mu \pm 3\sigma \Rightarrow$ 99.7% der Daten
- Die Normalverteilung ist eine Grundvoraussetzung für viele statistische Tests und Schätzer, welche über einen QQ-Plot beurteilt werden kann.