



von
nd
al-
tigter
Rolle.
Schaus
die,
erscheinung
Merkmalen

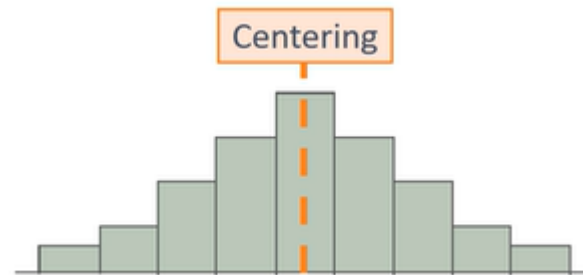
Statistik

Häufigkeitsverteilungen

Beschreibung von Häufigkeitsverteilungen

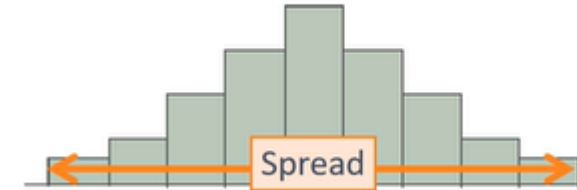


Lageparameter



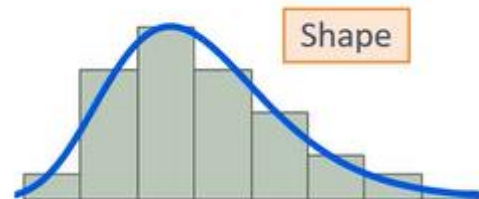
- Modus
- Median
- Mittelwert

Streuungsparameter

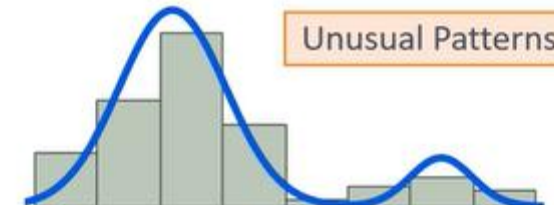


- Spannweite
- Quantilen (Quartile, Perzentile), IQR
- Varianz / Standardabweichung

Gestaltparameter



- Schiefe
- Kurtosis (Wölbung)



- Unimodal / Multimodal

Lageparameter und Streuungsparameter



- Lageparameter beschreiben zentrale Tendenz (Zentrum) einer (Häufigkeits-)Verteilung
 - **Modus, Modalwert** x_{Mod}
 - Häufigster Wert
 - **Median** x_{Med}
 - „Mittlerer“ Wert: Gleichviele Werte ober- und unterhalb
 - Median = 50%-Quantile
 - **Arithmetischer Mittelwert (mean)**
 - $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
 - Meist nur sinnvoll bei symmetrischer Verteilung (Schiefe ≈ 0)
 - Nicht robust gegen Ausreißer
- Beschreiben Streuung der Werte einer (Häufigkeits-)Verteilung um Lageparameter
 - p-Quantil (**Quartil, Perzentile**, etc.)
 - Teilt Werte in Anteil $< p$ und Anteil $> p$
 - **Interquartilsabstand** (*interquartile range*)
 - enthält 50% aller Werte, $IQR = Q_{0,75} - Q_{0,25}$
 - **Varianz (variance)**
 - $\tilde{s}_x^2 = \frac{1}{n} SQ_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
 - **Standardabweichung (standard deviation)**
 - $\tilde{s} = \sqrt{\tilde{s}_x^2}$

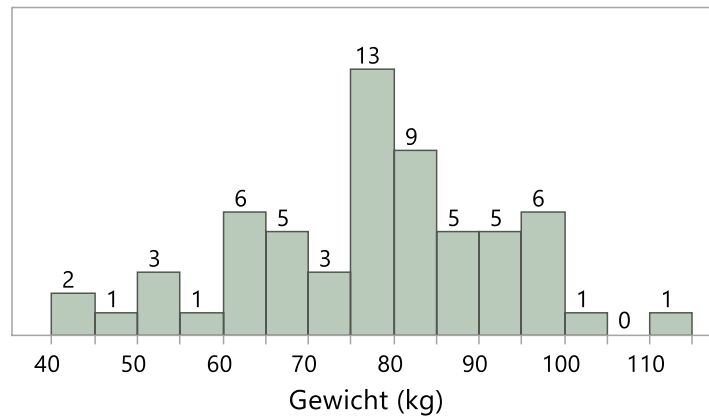
Häufigkeitsverteilungen

Absolute und relative Häufigkeit



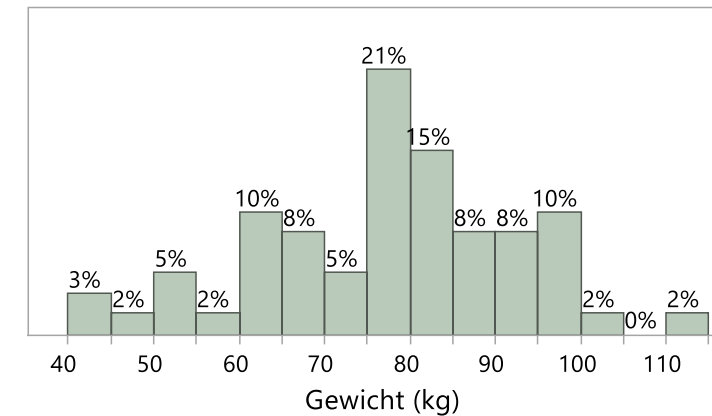
- Absolute Häufigkeit

$$h(x_i) = h_i \quad 1 < i \leq n$$



- Relative Häufigkeit

$$f(x_i) = \frac{h(x_i)}{n} = \frac{h_i}{n} \quad 1 < i \leq n$$



n : Anzahl Klassen (Merkmalsausprägungen)
 x_i : i^{te} -Klasse
 h_i : Anzahl Merkmalsträger in Klasse i

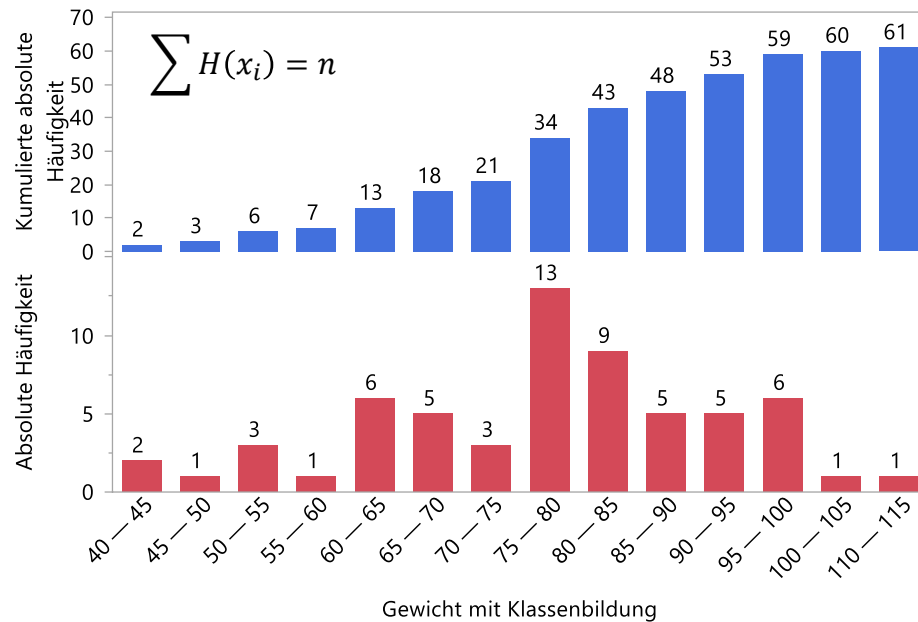
Häufigkeitsverteilungen

Kumulierte absolute und relative Häufigkeit



- Kumulierte absolute Häufigkeit

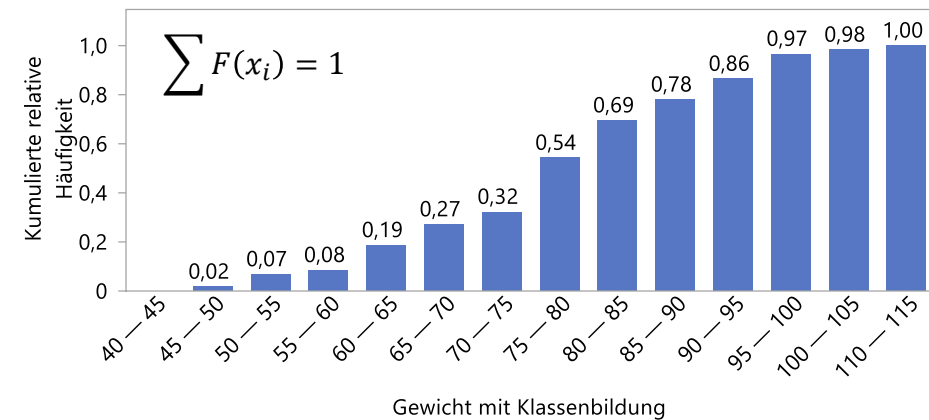
$$H(x_i) = \sum_{k=1}^i h(x_k) \quad 1 < i \leq n$$



n :
 x_i :
 h_i :

- Kumulierte relative Häufigkeit

$$F(x_i) = \sum_{k=1}^i f(x_k) = \sum_{k=1}^i \frac{h(x_k)}{n} \quad 1 < i \leq n$$



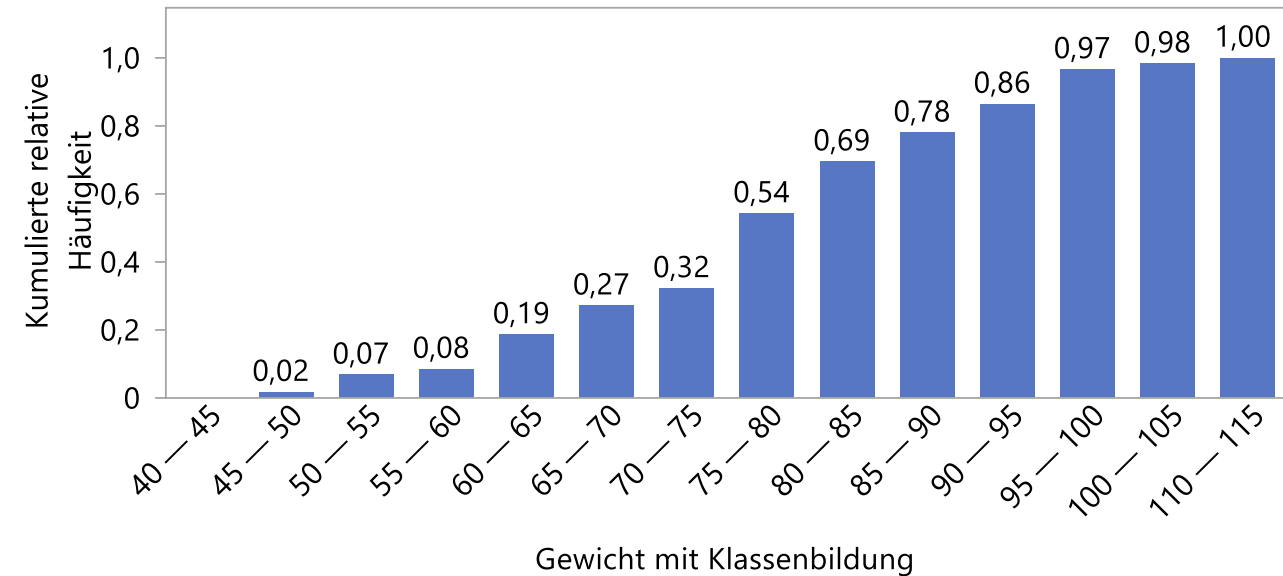
Anzahl Klassen (Merkmalsausprägungen)
 i^{te} -Klasse
 Anzahl Merkmalsträger in Klasse i

Häufigkeitsverteilungen

Kumulierte absolute und relative Häufigkeit



- Anteil < 80 kg?
 $F(75 - 80) = 0,54 \triangleq 54\%$
- Anteil ≥ 80 kg?
 $1 - F(75 - 80)$
 $= 1 - 0,54 = 0,46 \triangleq 46\%$
- Anteil ≥ 50 und < 100 kg?
 $F(75 - 80) - F(50 - 55)$
 $= 0,54 - 0,07 = 0,47 \triangleq 47\%$

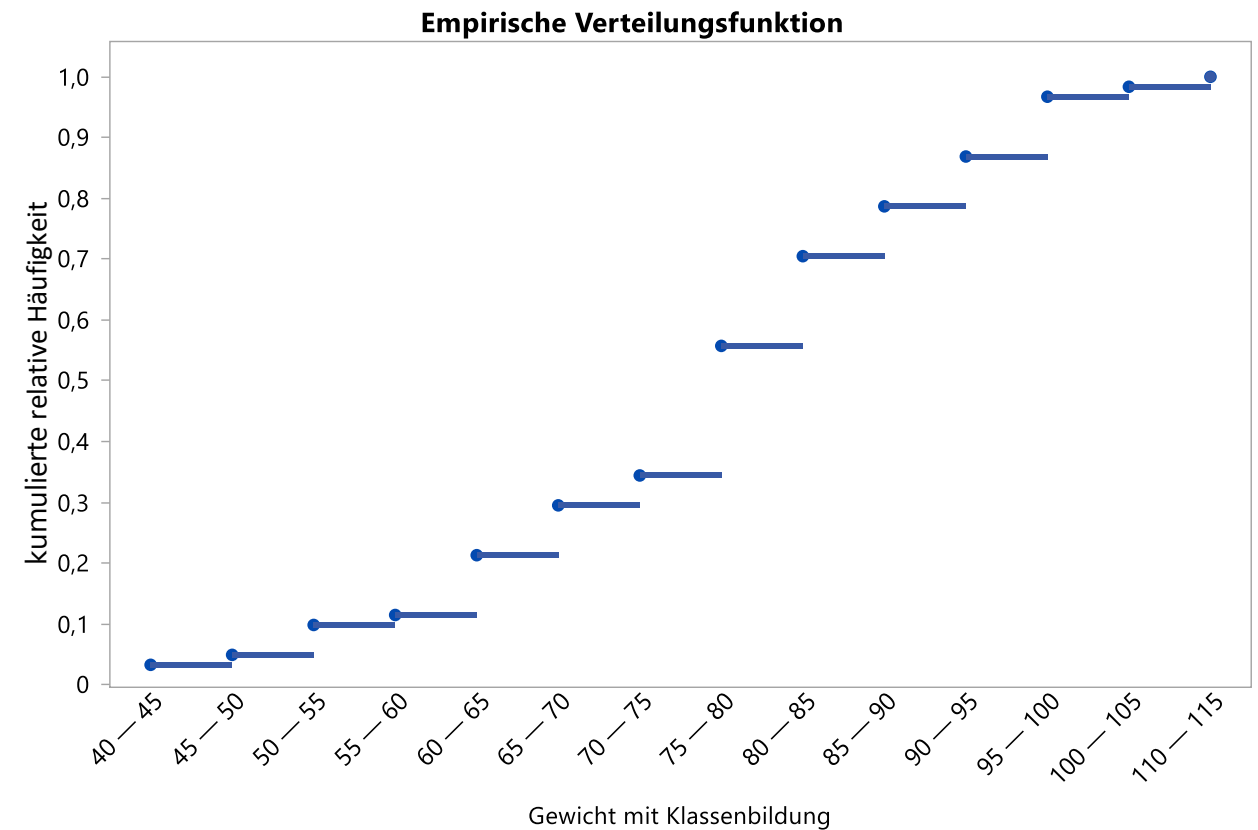


Häufigkeitsverteilungen

Empirische Verteilungsfunktion / Verteilungsfunktion der Stichprobe



- Summenhäufigkeitsfunktion
- $$F_n(x) = \frac{\text{Anzahl Werte} \leq x}{n} = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}}$$
- Ordnet jedem X Anteil der Stichprobenwerte $\leq X$ zu
- Auch bei Ordinalskalierung



The background of the slide is a close-up photograph of a document. The word 'Statistik' is prominently displayed in the center, written in a bold, black, sans-serif font. It is highlighted with a thick, orange, brush-stroke-like underline. Surrounding this central text are various other words from the document, which are out of focus. These include 'von', 'nd', 'al-', 'tigter', 'Rolle.', 'Schaus', 'die,', 'Scheinung', and 'Merkmalen'. The overall color palette is dominated by the orange of the highlight and the white background of the slide header and footer.

Statistik

Anteile von Häufigkeitsverteilungen

Häufigkeit und Anteil



- > Größe von 46 Studierenden
- > Anteil Studierender größer als mindestens 1,80 m?
 - $P(\text{Groesse} \geq 180 \text{ cm}) = ?$
- > Möglichkeiten
 - Sortieren, Scrollen, Selektieren
 - Datenfilter
- > $P(\text{Groesse} \geq 180 \text{ cm}) = \frac{26}{46} \approx 0,57$

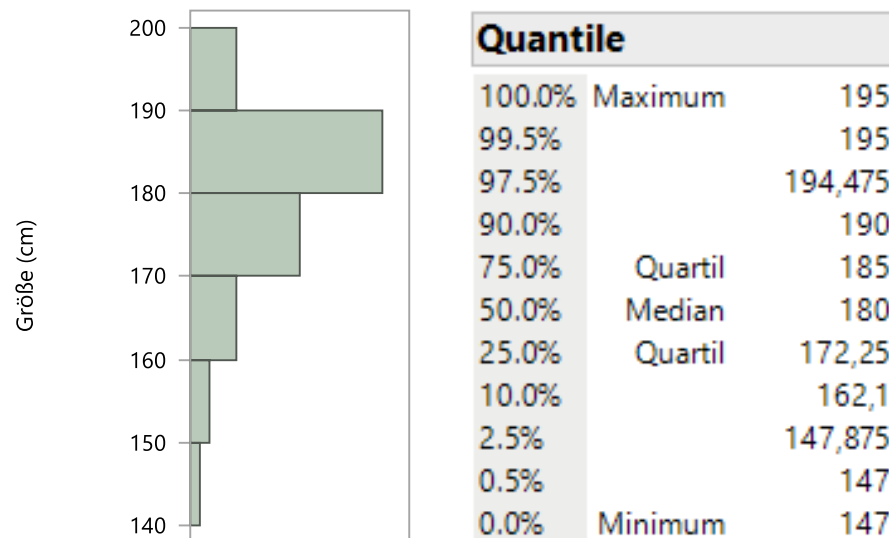
Anteil einer Häufigkeitsverteilung

Quantile



- > Für welche Körpergröße muss die Sicherheitseinrichtungen in einem Auto (Sitz, Airbag, Gurt, etc.) planen, so dass die meisten Menschen bei einem Unfall geschützt werden?
 - Z. B. 95% der Menschen

- > Stichprobe (n = 46)



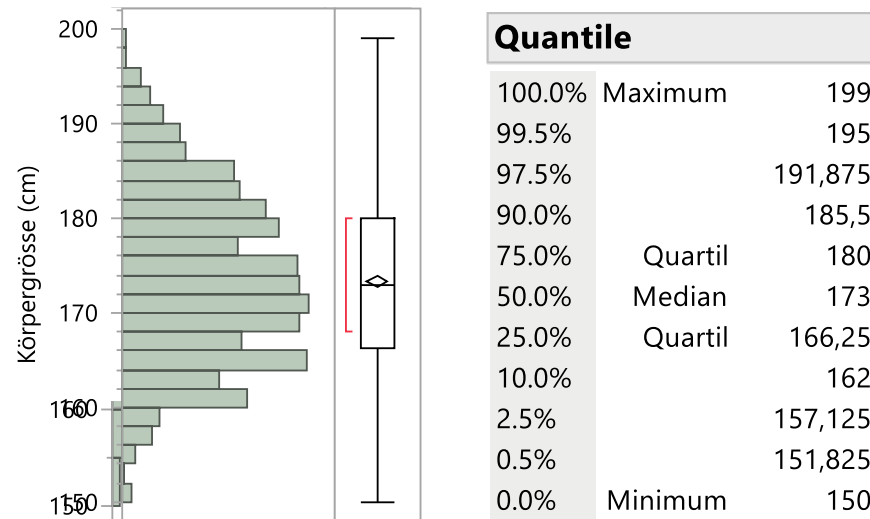
- > 95% $\equiv$ 2,5% bis 97,5% Quantilen
 $\Rightarrow$ Körpergröße von 148 cm bis 194 cm

Anteil einer Häufigkeitsverteilung

Quantile



- > Für welche Körpergröße muss die Sicherheitseinrichtungen in einem Auto (Sitz, Airbag, Gurt, etc.) planen, so dass die meisten Menschen bei einem Unfall geschützt werden?
 - Z. B. 95% der Menschen
- > Stichprobe (n = 1.164, Prof. Etschberger)



> 95% $\equiv$ 2,5% bis 97,5% Quantilen
 $\Rightarrow$ Körpergröße von 157,125 cm bis 191,875 cm

Anteil einer Häufigkeitsverteilung

Quantile

- > Für welche Körpergröße muss die Sicherheitseinrichtungen in einem Auto (Sitz, Airbag, Gurt, etc.) planen, so dass die meisten Menschen bei einem Unfall geschützt werden?
 - Z. B. 95% der Menschen

- > Noch größere Stichprobe!?
 - NaKo Studie



Statistik

NaKo Studie

NAKO Gesundheitsstudie (Nationale Kohorte)

- > Ziel: Erforschung der Volkskrankheiten
- > Start 2014, Laufzeit 20 – 30 Jahre
- > 200.000 freiwillige Menschen zwischen 20 – 69 Jahren
 - Entsprechend der Deklaration von Helsinki
- > 18 Zentren in Deutschland
- > Lebensumstände, Krankheitsgeschichte, medizinische Untersuchungen
 - Herz-Kreislauf
 - Krebs
 - Diabetes mellitus
 - Neurologische und psychiatrische Erkrankungen
 - Infektionskrankheiten
 - Krankheiten des Bewegungsapparates
- > Untersuchungsprogramm Level 1
 - Messung von Körpergröße und Körpergewicht, Taillenumfang und Körperzusammensetzung
 - ...



Deklaration von Helsinki

- > 1964 während der 18. Versammlung des Weltärztebundes verabschiedet
 - Ethische Standards für medizinische Forschung am Menschen
 - Letzte Überarbeitung 2013
 - Weltweit anerkannt

- > Prinzipien
 - Aufklärung der Studienteilnehmer/-innen
 - Erfordernis einer Einwilligungserklärung
 - Freiwillige Teilnahme
 - Schutz nicht-einwilligungsfähiger Patienten/-innen
 - Nutzen-Risiko-Abwägung
 - Vorrang des Wohlergehens der Versuchsperson vor Interessen der Wissenschaft
 - Verpflichtung der Genehmigung eines Forschungsvorhabens durch eine unabhängige Ethikkommission
 - Nichtveröffentlichung von Forschungsergebnissen aus unethischen Versuchen
 - Datenschutz

The background of the slide is a close-up photograph of a document. The word 'Statistik' is prominently displayed in the center, written in a bold, black, sans-serif font. It is highlighted with a thick, orange, brush-stroke-like underline. Surrounding this central text are various other words from the document, which are out of focus. These include 'von', 'nd', 'al-', 'tigter', 'Rolle.', 'Schaus', 'die,', 'Scheinung', and 'Merkmalen'. The overall color palette is dominated by the orange of the highlight and the white background of the slide header and footer.

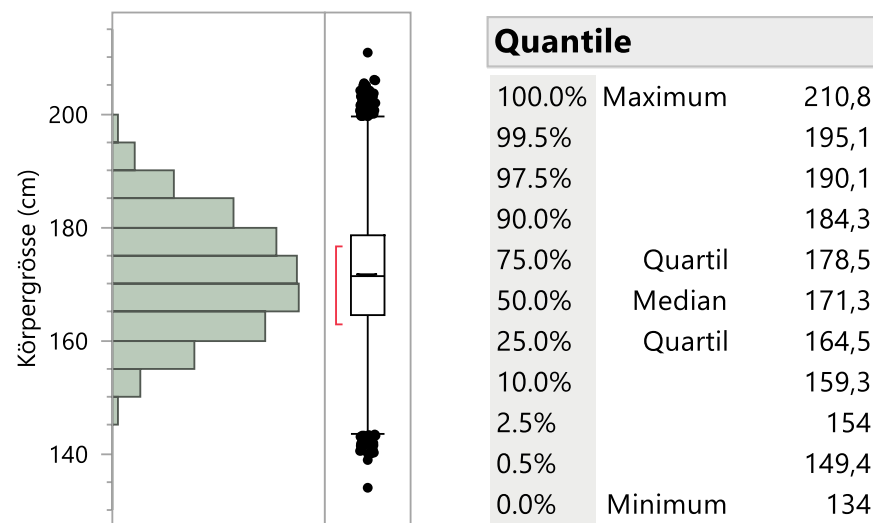
Statistik

Anteile von Häufigkeitsverteilungen

Anteil einer Häufigkeitsverteilung

Quantilen

- > Für welche Körpergröße muss die Sicherheitseinrichtungen in einem Auto (Sitz, Airbag, Gurt, etc.) planen, so dass die meisten Menschen bei einem Unfall geschützt werden?
 - Z. B. 95% der Menschen
- > Stichprobe (n = 89.725, NaKo)



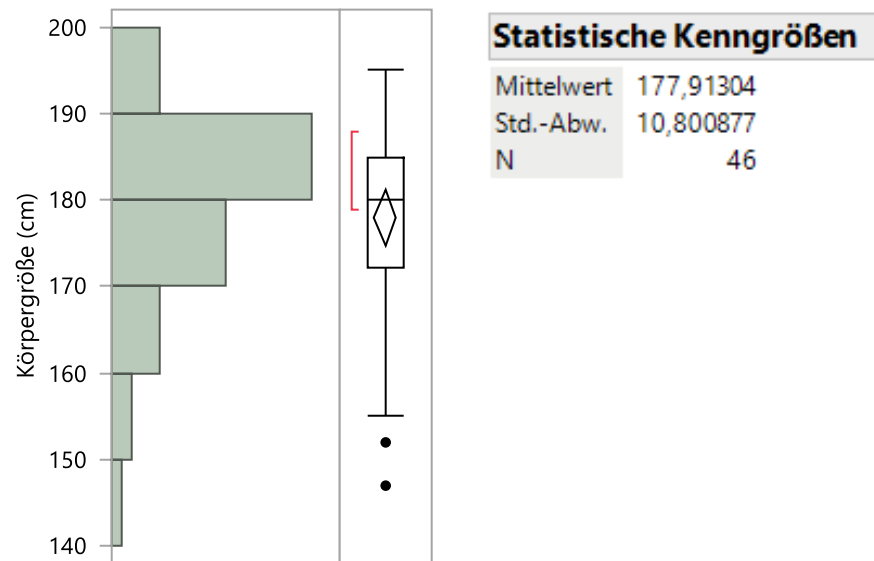
> 95% $\equiv$ 2,5% bis 97,5% Quantilen
 $\Rightarrow$ Körpergröße von 154 cm bis 190,1 cm

Anteil einer Häufigkeitsverteilung

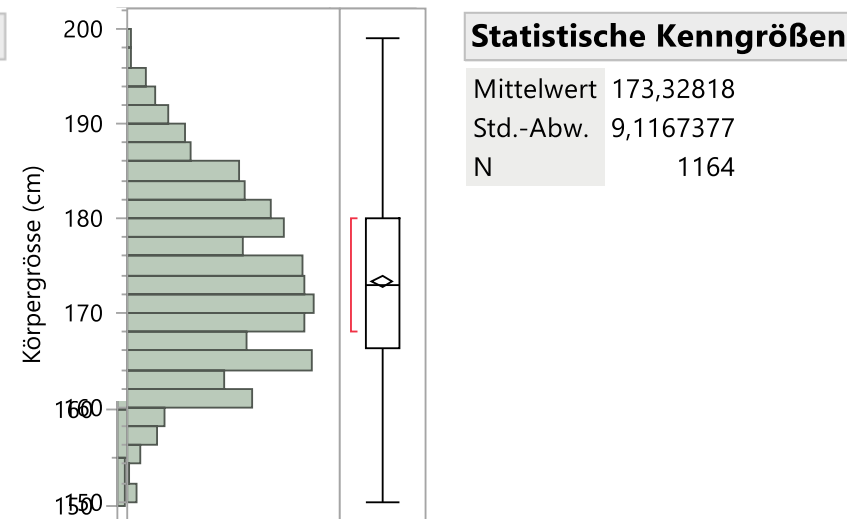
Geht das auch anders?

- > Für welche Körpergröße muss die Sicherheitseinrichtungen in einem Auto (Sitz, Airbag, Gurt, etc.) planen, so dass die meisten Menschen bei einem Unfall geschützt werden?
 - Z. B. 95% der Menschen

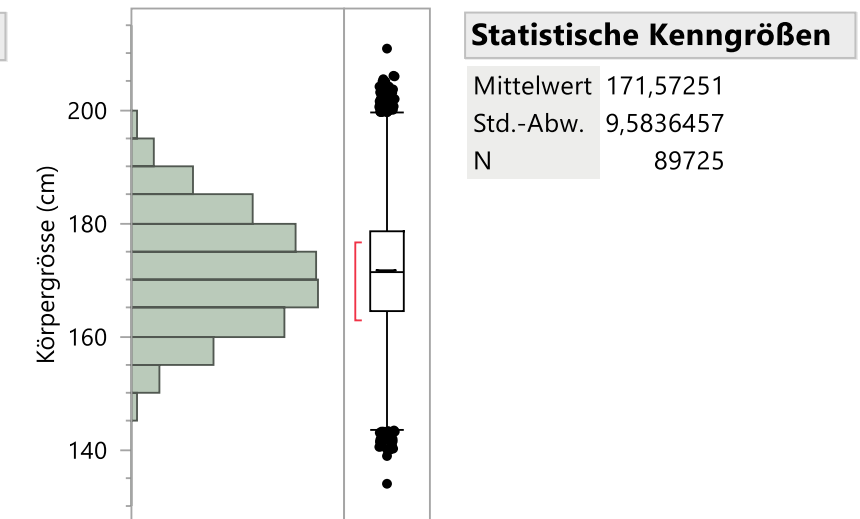
Stichprobe (n = 46)



Stichprobe (n = 1.164, Etschberger)



Stichprobe (n = 89.725, NaKo)



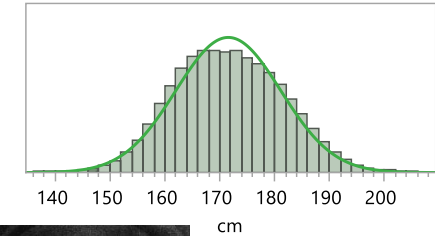
Statistik

Normalverteilung

Normalverteilung

Geschichte

- > *The Doctrine of Chances*, Abraham de Moivre, 1733
 - Vorform der Normalverteilung aus Grenzwertsatz für Binomialverteilungen
- > Pierre-Simon Laplace, 1782 (oder Siméon Denis Poisson)
 - Berechnung des Integrals für Normierung der Normalverteilung zur Wahrscheinlichkeitsdichte $\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}$
- > *Theorie der Bewegung der in Kegelschnitten sich um die Sonne bewegenden Himmelskörper*, Gauß, 1809
 - Definition Normalverteilung $f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- > Pierre-Simon Laplace, 1810
 - Zentraler Grenzwertsatz, Grundlage theoretische Bedeutung der Normalverteilung
- > Adolphe Quetelet, 1844
 - Normalverteilung des Brustumfangs von Soldaten
 - Normalverteilung für angewandte Statistik

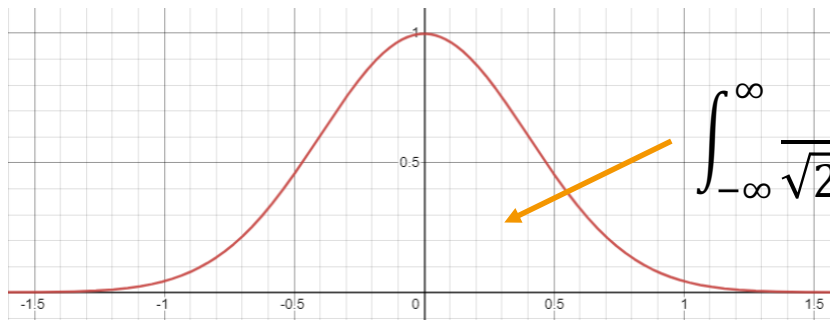


Normalverteilung



- Gauß-Kurve, Glockenkurve, Gaußsche Normalverteilung
- $f(x) = e^{-x^2} \Rightarrow \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$
- Nach Carl Friedrich Gauß (1777 – 1855), Mathematiker, Statistiker, Physiker, Astronom

- Normalverteilung $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

- Standardnormalverteilung: $\mu = 0, \sigma = 1 \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}}$



Wikipedia, Tigran Mitran, Public domain

Normalverteilung

Was hat die Größe von Menschen mit π zu tun?

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

> $\pi = \frac{\text{Kreisumfang}}{\text{Kreisdurchmesser}}$

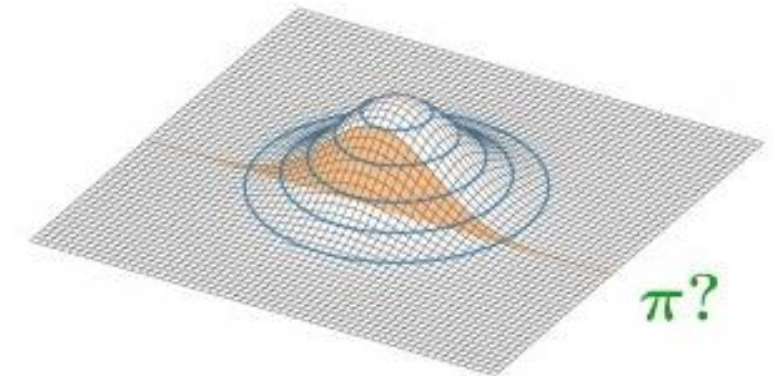
> Was hat ein Kreis mit der Größe von Menschen zu tun?

> Glockenkurve

- $f(t) = e^{-\frac{1}{2}t^2}$
- Fläche unter Kurve: $F(t) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}$
- Fläche unter Kurve $\stackrel{!}{=} 1$: Korrektur mit Faktor $\frac{1}{\sqrt{2\pi}}$

Warum π ?

1. Substitution: $t^2 = (x^2 + y^2)$
 $\Rightarrow$ 2-dimensional



2. Polarkoordinaten

$$\Rightarrow r = \sqrt{x^2 + y^2}, \varphi = \arctan\left(\frac{y}{x}\right)$$

<https://www.youtube.com/watch?v=jkxoOIAIcYA>

Statistik

Sigma-Regeln der Normalverteilung

k-σ-Bereiche, Sigma-Regeln



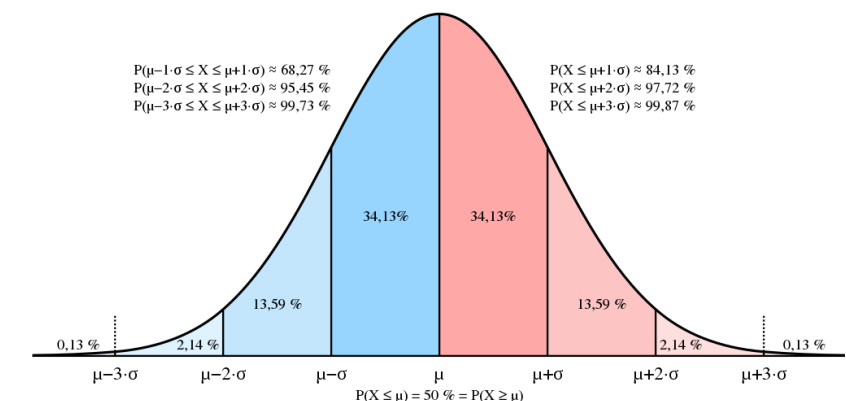
> Ermittlung eines Intervalls für gegebenen Anteil

- Normalverteilte Zufallsgrößen $X \sim N(\mu, \sigma^2)$
 - Symmetrisch um Erwartungswert μ
- Intervall i.d.R. durch Vielfachens der Standardabweichung σ zum Erwartungswert μ
 - Z-Transformation (siehe spätere Folie)

$$P(|X - \mu| \leq k * \sigma) = P(\mu - k * \sigma \leq X \leq \mu + k * \sigma)$$

> Sigma-Regeln

Intervall I	$P(X \in I)$	Intervall I	$P(X \in I)$
$[\mu - \sigma; \mu + \sigma]$	$\approx 0,683$	$[\mu - 1,64\sigma; \mu + 1,64\sigma]$	$\approx 0,90$
$[\mu - 2\sigma; \mu + 2\sigma]$	$\approx 0,954$	$[\mu - 1,96\sigma; \mu + 1,96\sigma]$	$\approx 0,95$
$[\mu - 3\sigma; \mu + 3\sigma]$	$\approx 0,997$	$[\mu - 2,58\sigma; \mu + 2,58\sigma]$	$\approx 0,99$



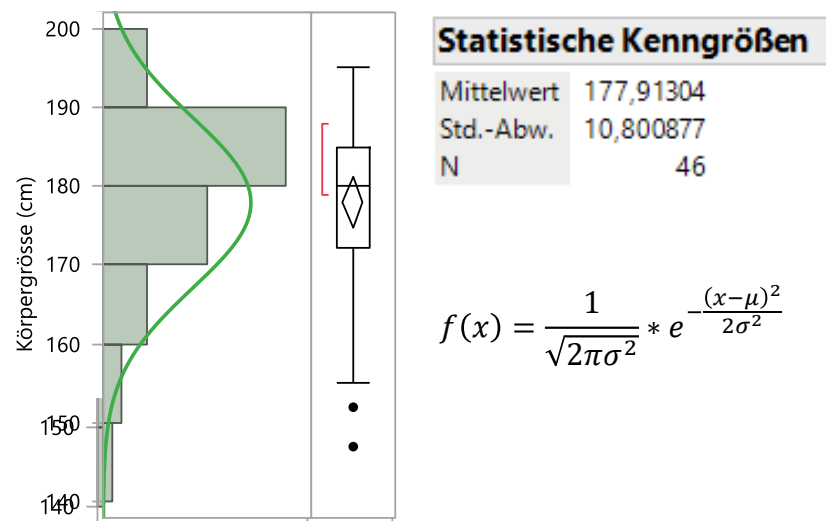
Anteil einer Häufigkeitsverteilung

Geht das auch anders? Ja!

- > Für welche Körpergröße muss die Sicherheitseinrichtungen in einem Auto (Sitz, Airbag, Gurt, etc.) planen, so dass die meisten Menschen bei einem Unfall geschützt werden?
 - Z. B. 95% der Menschen

Intervall I	$P(X \in I)$
$[\mu - 1,64\sigma; \mu + 1,64\sigma]$	$\approx 0,90$
$[\mu - 1,96\sigma; \mu + 1,96\sigma]$	$\approx 0,95$
$[\mu - 2,58\sigma; \mu + 2,58\sigma]$	$\approx 0,99$

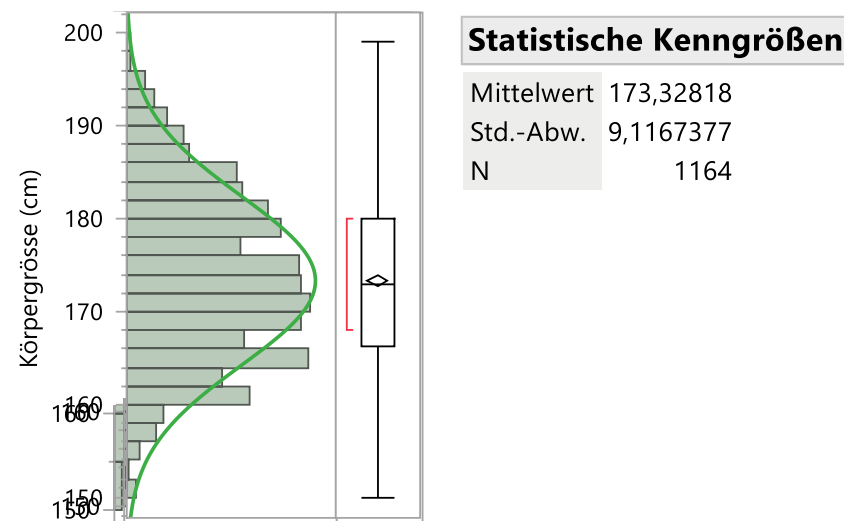
Stichprobe (n = 46)



$$[\mu \pm 1,96\sigma] = [177,91 \pm 1,96 * 10,80]$$

$$= [156,74; 199,08]$$

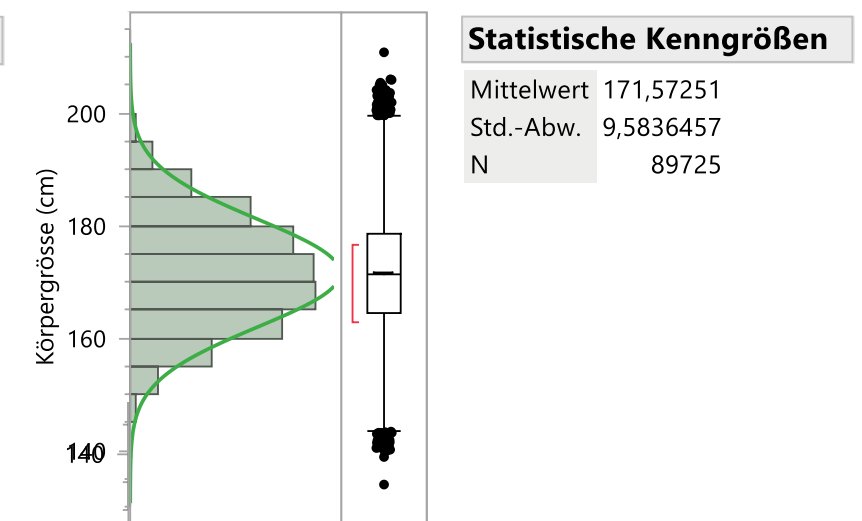
Stichprobe (n = 1.164)



$$= [173,33 \pm 1,96 * 9,12]$$

$$= [155,45; 191,21]$$

Stichprobe (n = 89.725)



$$= [171,56 \pm 1,96 * 9,58]$$

$$= [152,78; 190,34]$$

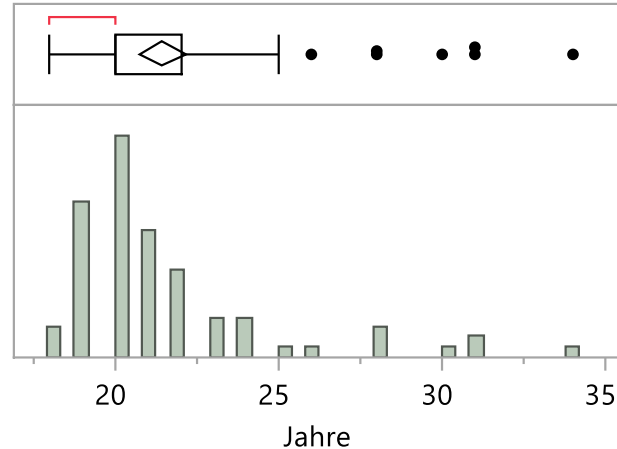
Anteil einer Häufigkeitsverteilung

Geht das auch anders? Nicht immer!

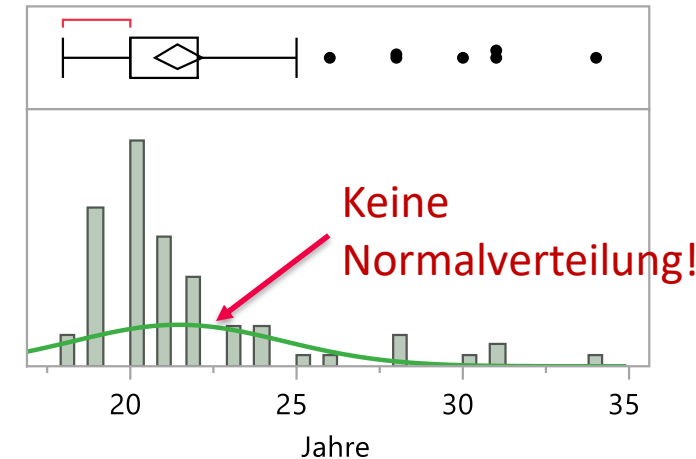


- > Welches Alter haben die meisten Studierenden in der Vorlesung?
 - Z. B. 95% der Studierenden

Stichprobe (n = 81)



Quantile		
100.0%	Maximum	34
99.5%		34
97.5%		31
90.0%		25,8
75.0%	Quartil	22
50.0%	Median	20
25.0%	Quartil	20
10.0%		19
2.5%		18
0.5%		18
0.0%	Minimum	18



Statistische Kenngrößen	
Mittelwert	21,45679
Std.-Abw.	3,1545577
N	81

2,5% bis 97,5% Quantile
 ⇒ 95% im Alter von 18 cm bis 31 Jahre

$$[\mu \pm 1,96\sigma] = [21,46 \pm 1,96 * 3,15]$$

$$= [15,27; 27,63]$$

Statistik

Wie „normal“ ist eine Verteilung?

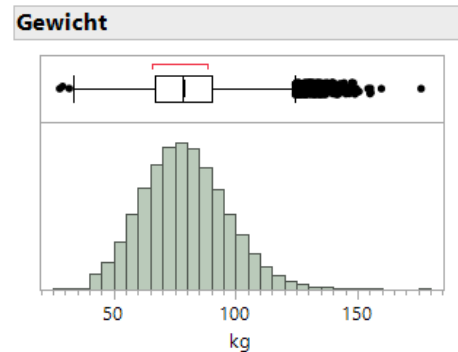
Quantil-Quantil-Diagramm (QQ-Plot)

Visueller Test auf Normalverteilung

- > Exploratives, grafisches Werkzeug
 - Vergleich von Verteilungen durch Auftrag der Quantilen zweier statistischer Variablen in Diagramm
 - **Vergleich empirischer Häufigkeitsverteilung mit theoretischer Verteilung (z. B. Normalverteilung)**
 - Visuelle Prüfung von Verteilungsvoraussetzungen
- > Vorgehen
 - Bestimmung empirische Quantile x_i
 - Bestimmung theoretischer Quantile y_i aus Normalverteilung
 - $p_i = F_{\text{empirisch}}(x_i)$
 - $y_i = F_{\text{theoretisch}}^{-1}(p_i)$
 - Darstellung als Wertepaare (x/y) im Diagramm
 - Wenn Merkmalswerte der theoretischen Verteilung folgen, stimmen empirische und theoretischen Quantile annähernd überein, die Punkte liegen auf der Winkelhalbierenden

Quantil-Quantil-Diagramm (Q-Q-Plot)

Visueller Test auf Normalverteilung

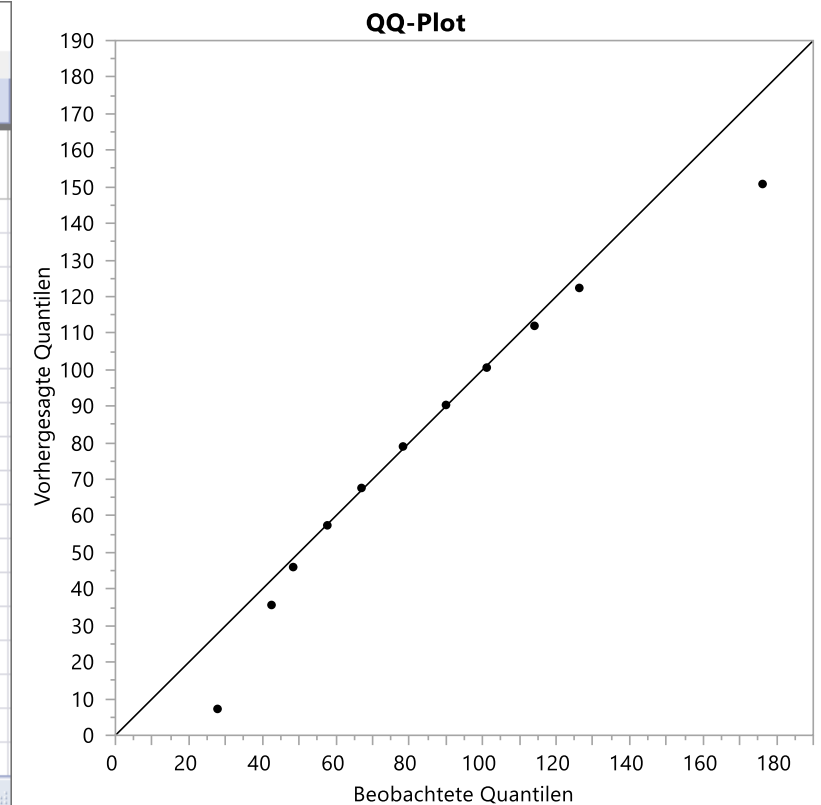


Quantile		
100.0%	Maximum	176,2
99.5%		126,3
97.5%		114,1
90.0%		101,1
75.0%	Quartil	90
50.0%	Median	78,3
25.0%	Quartil	67
10.0%		57,7
2.5%		48,4
0.5%		42,5
0.0%	Minimum	27,8

Statistische Kenngrößen	
Mittelwert	79,004467
Std.-Abw.	16,825147
Std.-Fehler Mittelwert	0,0561697
95% KI oben Mittelwert	79,114559
95% KI unten Mittelwert	78,894375
N	89725
Anzahl fehlender Werte	0

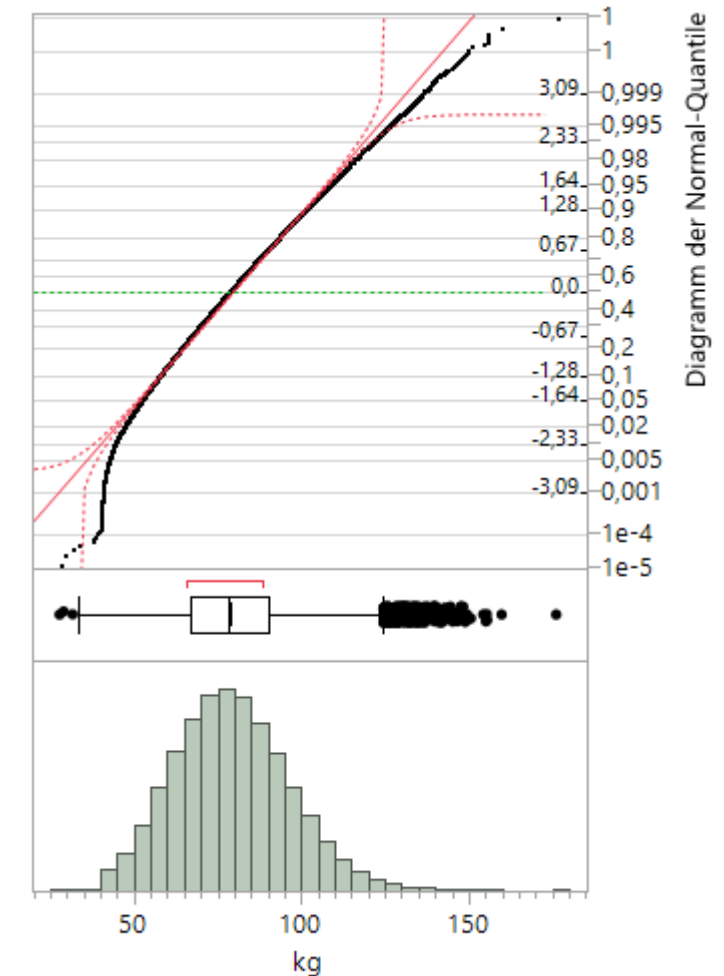
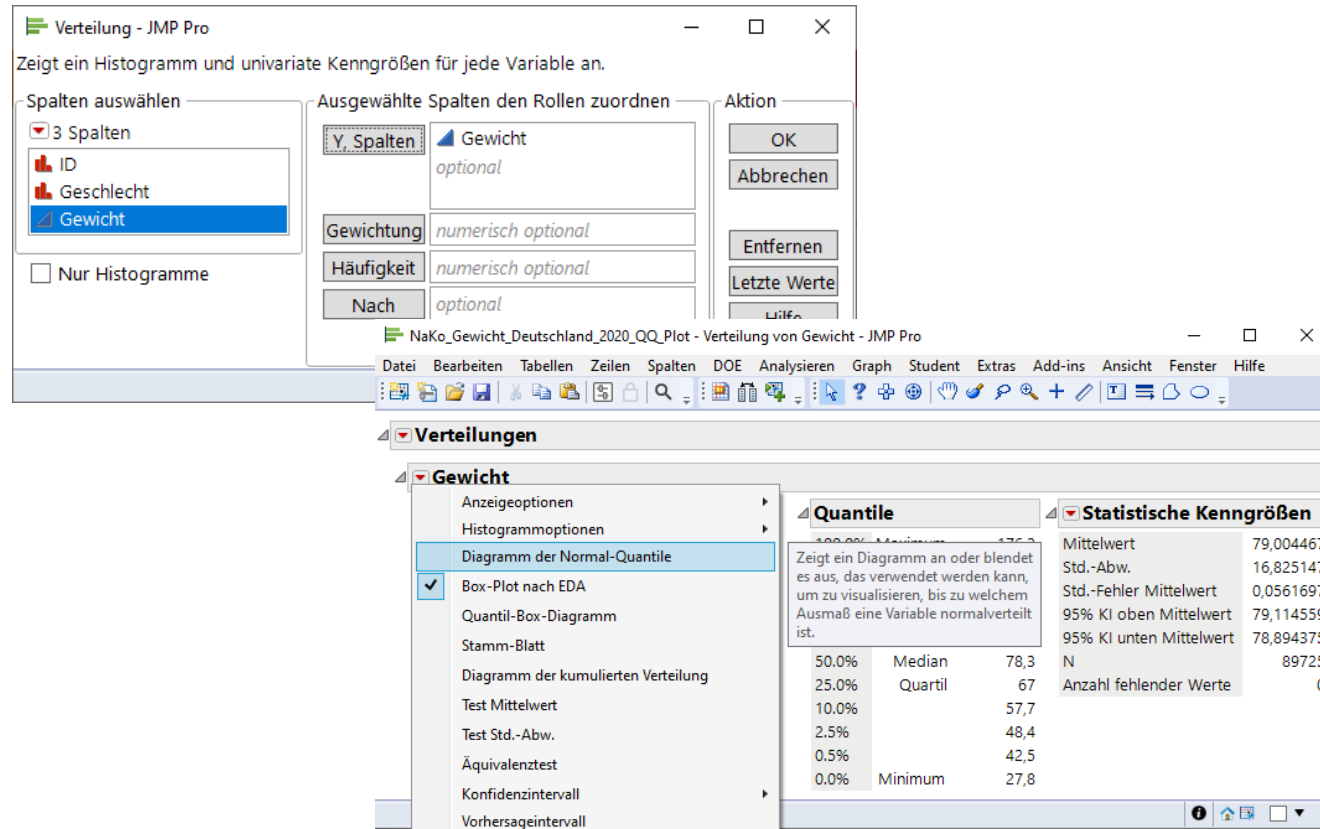
QQ_Plot_Gewicht - JMP Pro

Quantile	p	Beobachtete Quantilen	Vorhergesagte Quantilen
1	100.0%	176,2	150,77811206
2	99.5%	126,3	122,35120718
3	97.5%	114,1	111,98619386
4	90.0%	101,1	100,56851285
5	75.0%	90	90,351662496
6	50.0%	78,3	79
7	25.0%	67	67,648337504
8	10.0%	57,7	57,431487152
9	2.5%	48,4	46,01380614
10	0.5%	42,5	35,648792821
11	0.0%	27,8	7,2218879383



Quantil-Quantil-Diagramm (Q-Q-Plot)

Visueller Test auf Normalverteilung



Statistik

Z-Transformation

Lage von Werten in verschiedenen Verteilungen



- > Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen

The screenshot shows the JMP software window titled 'Pub_Quiz - JMP'. The main data table has columns: Name, Quiz 1, and Quiz 2. The 'Name' column lists 21 participants (Torsten, Stephanie, Gabriele, Renate, Jens, Sabine, Uwe, Claudia, Bernd, Birgit, Daniel, Stefan, Klaus, Jörg, Andrea, Thomas, Ralf, Karin, Martina, Sven, Susanne). The 'Quiz 1' and 'Quiz 2' columns show their respective scores. The left sidebar shows the 'Spalten (3/0)' section with 'Name', 'Quiz 1', and 'Quiz 2' listed. The 'Zeilen' section shows 'Alle Zeilen' as 50, and other row counts as 0. The status bar at the bottom indicates 'Auswertungen fertig'.

Name	Quiz 1	Quiz 2
1 Torsten	88	86
2 Stephanie	82	89
3 Gabriele	84	92
4 Renate	82	92
5 Jens	77	86
6 Sabine	87	90
7 Uwe	95	88
8 Claudia	72	92
9 Bernd	75	86
10 Birgit	86	86
11 Daniel	74	91
12 Stefan	100	88
13 Klaus	81	86
14 Jörg	85	86
15 Andrea	96	89
16 Thomas	80	93
17 Ralf	73	87
18 Karin	76	85
19 Martina	67	91
20 Sven	75	88
21 Susanne	87	88

	Quiz 1	Quiz 2
Torsten	88	86
Stephanie	82	89

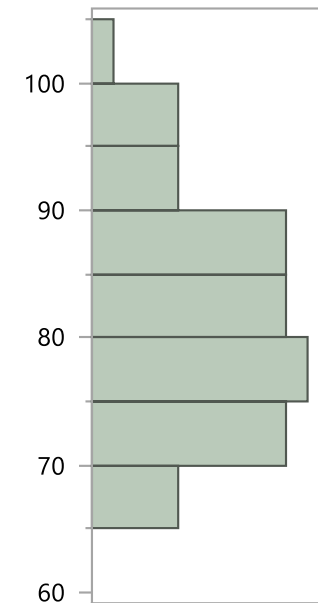
Lage von Werten in verschiedenen Verteilungen

- > Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen

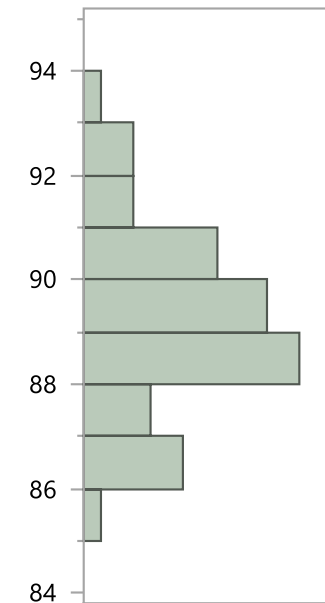
	Name	Quiz 1	Quiz 2
1	Torsten	88	86
2	Stephanie	82	89
3	Gabriele	84	92
4	Renate	82	92
5	Jens	77	86
6	Sabine	87	90
7	Uwe	95	88
8	Claudia	72	92
9	Bernd	75	86
10	Birgit	86	86
11	Daniel	74	91
12	Stefan	100	88
13	Klaus	81	86
14	Jörg	85	86
15	Andrea	96	89
16	Thomas	80	93
17	Ralf	73	87
18	Karin	76	85
19	Martina	67	91
20	Sven	75	88
21	Susanne	87	88

Verteilungen

Quiz 1



Quiz 2



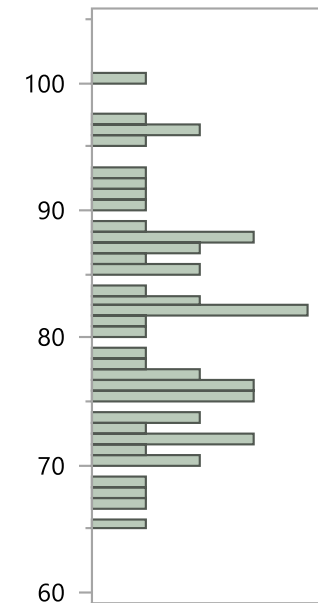
Lage von Werten in verschiedenen Verteilungen

- > Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen

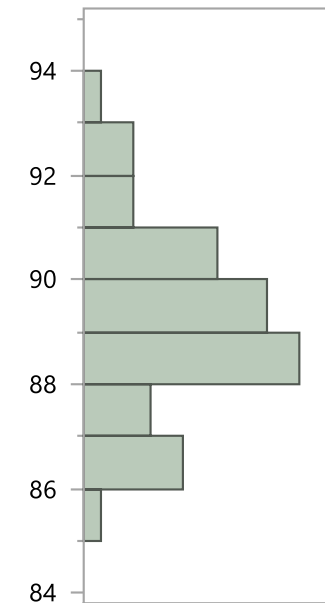
	Name	Quiz 1	Quiz 2
1	Torsten	88	86
2	Stephanie	82	89
3	Gabriele	84	92
4	Renate	82	92
5	Jens	77	86
6	Sabine	87	90
7	Uwe	95	88
8	Claudia	72	92
9	Bernd	75	86
10	Birgit	86	86
11	Daniel	74	91
12	Stefan	100	88
13	Klaus	81	86
14	Jörg	85	86
15	Andrea	96	89
16	Thomas	80	93
17	Ralf	73	87
18	Karin	76	85
19	Martina	67	91
20	Sven	75	88
21	Susanne	87	88

Verteilungen

Quiz 1



Quiz 2



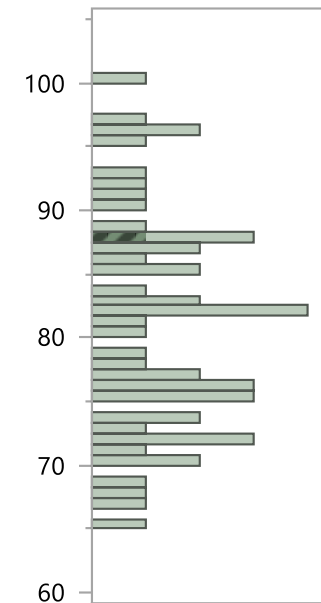
Lage von Werten in verschiedenen Verteilungen

- > Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen

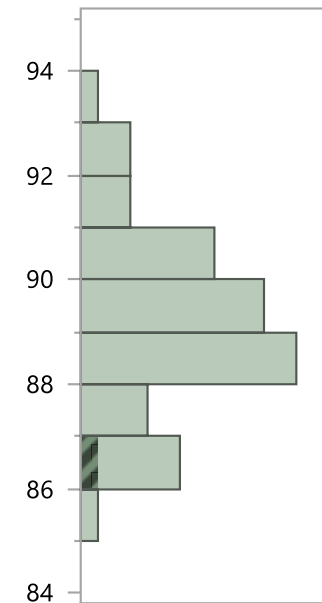
	Name	Quiz 1	Quiz 2
1	Torsten	88	86
2	Stephanie	82	89
3	Gabriele	84	92
4	Renate	82	92
5	Jens	77	86
6	Sabine	87	90
7	Uwe	95	88
8	Claudia	72	92
9	Bernd	75	86
10	Birgit	86	86
11	Daniel	74	91
12	Stefan	100	88
13	Klaus	81	86
14	Jörg	85	86
15	Andrea	96	89
16	Thomas	80	93
17	Ralf	73	87

Verteilungen

Quiz 1



Quiz 2



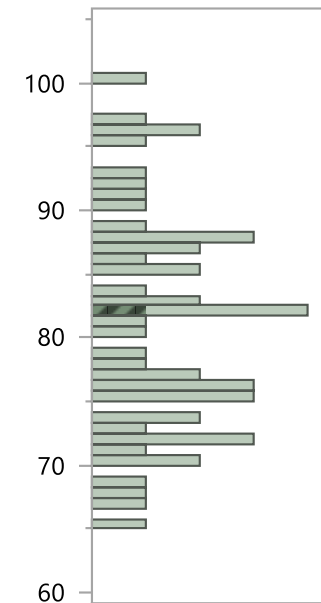
Lage von Werten in verschiedenen Verteilungen

- > Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen

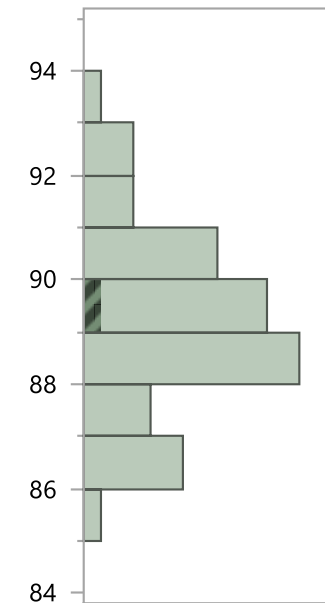
	Name	Quiz 1	Quiz 2
1	Torsten	88	86
2	Stephanie	82	89
3	Gabriele	84	92
4	Renate	82	92
5	Jens	77	86
6	Sabine	87	90
7	Uwe	95	88
8	Claudia	72	92
9	Bernd	75	86
10	Birgit	86	86
11	Daniel	74	91
12	Stefan	100	88
13	Klaus	81	86
14	Jörg	85	86
15	Andrea	96	89
16	Thomas	80	93
17	Ralf	73	87

Verteilungen

Quiz 1

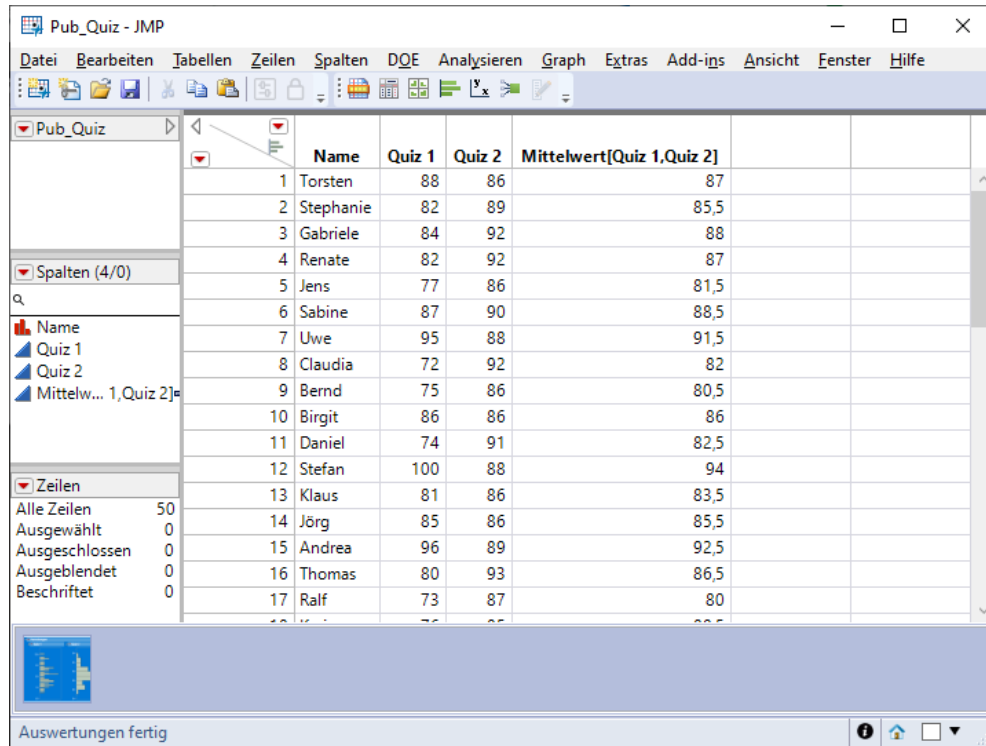


Quiz 2



Lage von Werten in verschiedenen Verteilungen

- Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen



Name	Quiz 1	Quiz 2	Mittelwert[Quiz 1, Quiz 2]
1 Torsten	88	86	87
2 Stephanie	82	89	85,5
3 Gabriele	84	92	88
4 Renate	82	92	87
5 Jens	77	86	81,5
6 Sabine	87	90	88,5
7 Uwe	95	88	91,5
8 Claudia	72	92	82
9 Bernd	75	86	80,5
10 Birgit	86	86	86
11 Daniel	74	91	82,5
12 Stefan	100	88	94
13 Klaus	81	86	83,5
14 Jörg	85	86	85,5
15 Andrea	96	89	92,5
16 Thomas	80	93	86,5
17 Ralf	73	87	80

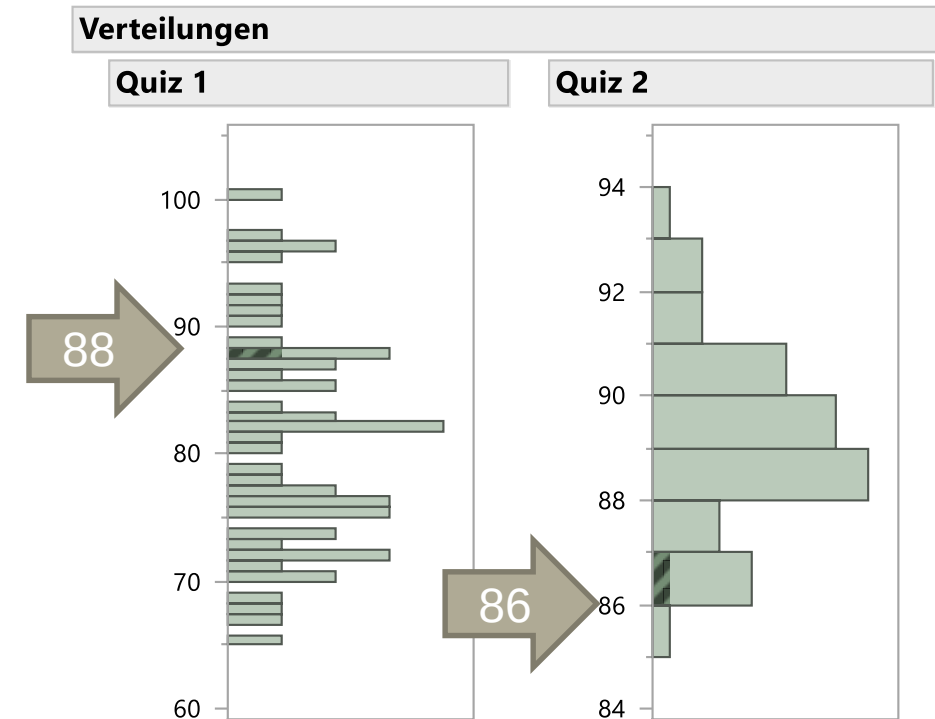
	Quiz 1	Quiz 2	Durchschnitt
Torsten 	88	86	87
Stephanie	82	89	85,5

- Durchschnitt unabhängig von Verteilung!
- Berücksichtigt nicht relative Position in Verteilung
- Werte von Quiz 1 haben andere Bedeutung als Werte von Quiz 2

Lage von Werten in verschiedenen Verteilungen

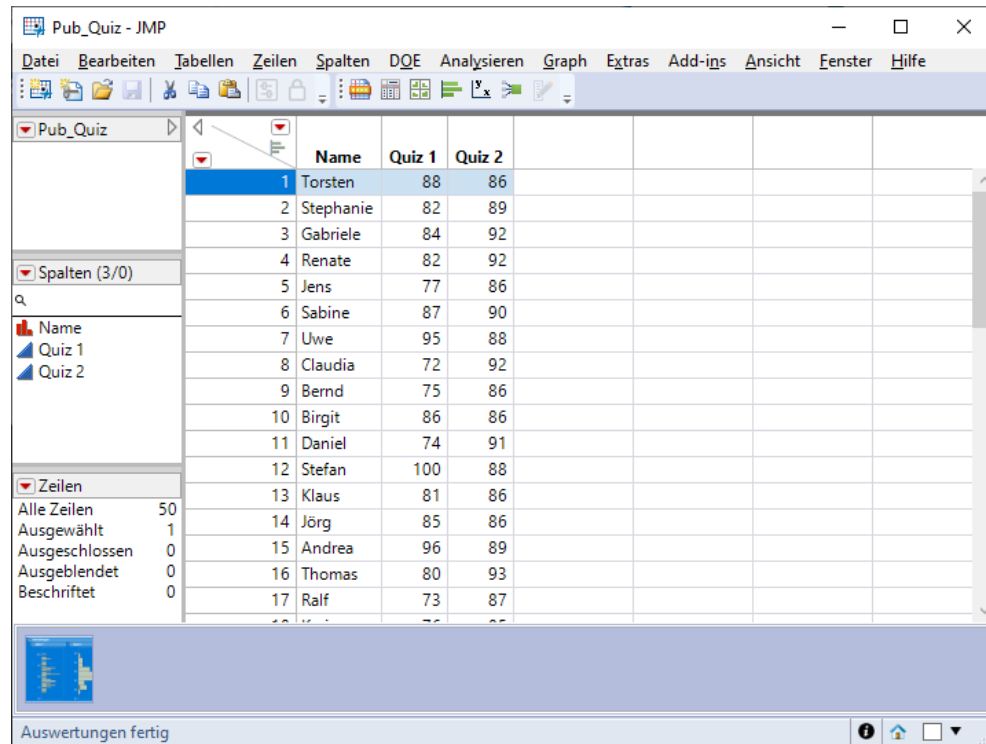
- Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen

	Name	Quiz 1	Quiz 2
1	Torsten	88	86
2	Stephanie	82	89
3	Gabriele	84	92
4	Renate	82	92
5	Jens	77	86
6	Sabine	87	90
7	Uwe	95	88
8	Claudia	72	92
9	Bernd	75	86
10	Birgit	86	86
11	Daniel	74	91
12	Stefan	100	88
13	Klaus	81	86
14	Jörg	85	86
15	Andrea	96	89
16	Thomas	80	93
17	Ralf	73	87



Lage von Werten in verschiedenen Verteilungen

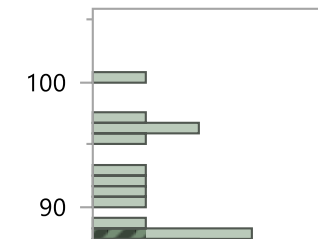
- Beispiel: Erreichte Punkte in Pub-Quiz
 - Ergebnisse von 50 Teilnehmern und Teilnehmerinnen



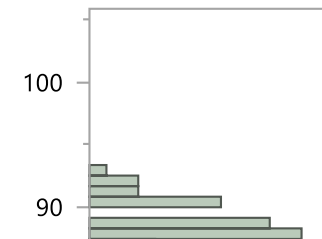
	Name	Quiz 1	Quiz 2
1	Torsten	88	86
2	Stephanie	82	89
3	Gabriele	84	92
4	Renate	82	92
5	Jens	77	86
6	Sabine	87	90
7	Uwe	95	88
8	Claudia	72	92
9	Bernd	75	86
10	Birgit	86	86
11	Daniel	74	91
12	Stefan	100	88
13	Klaus	81	86
14	Jörg	85	86
15	Andrea	96	89
16	Thomas	80	93
17	Ralf	73	87
18	...	...	...

Verteilungen

Quiz 1



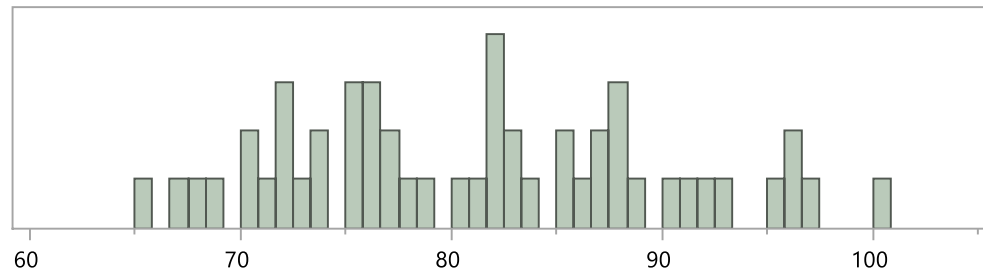
Quiz 2



- Streuung sehr unterschiedlich
- Änderung um 1 Einheit (1 Punkt) in Quiz 1 unwesentlich
- Ziel: Vergleichbarkeit herstellen
- **Standardisierung**

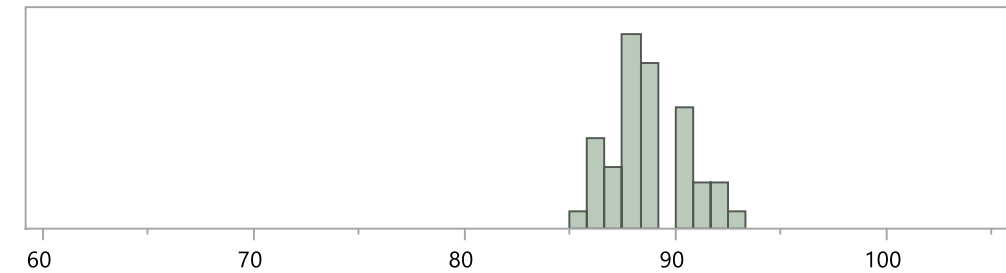
Standardisierung (z-Transformation $\Rightarrow$ z-Score)

Quiz 1



$$\bar{x}_1 = 81,06$$

Quiz 2

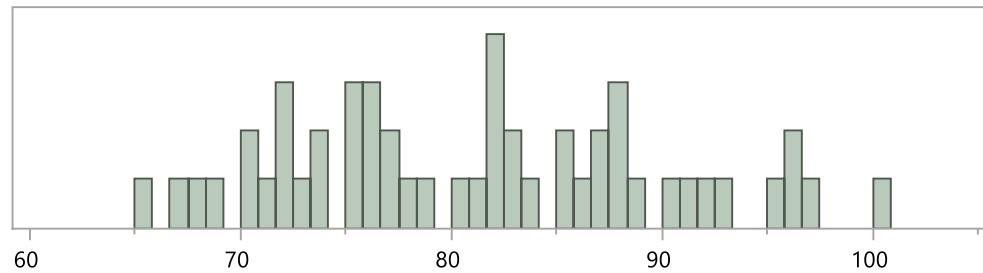


$$\bar{x}_2 = 88,68$$

	Quiz 1	Quiz 2
Torsten	$88 - \bar{x}_1$	$86 - \bar{x}_2$
Stephanie	$82 - \bar{x}_1$	$89 - \bar{x}_2$

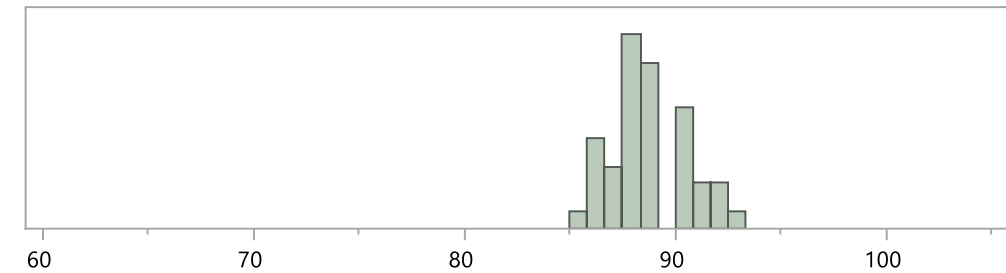
Standardisierung (z-Transformation $\Rightarrow$ z-Score)

Quiz 1



$$\bar{x}_1 = 81,06$$

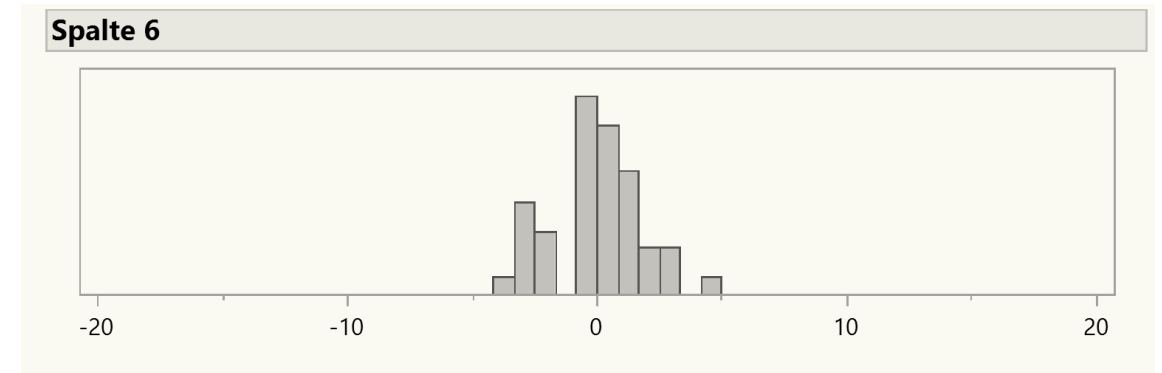
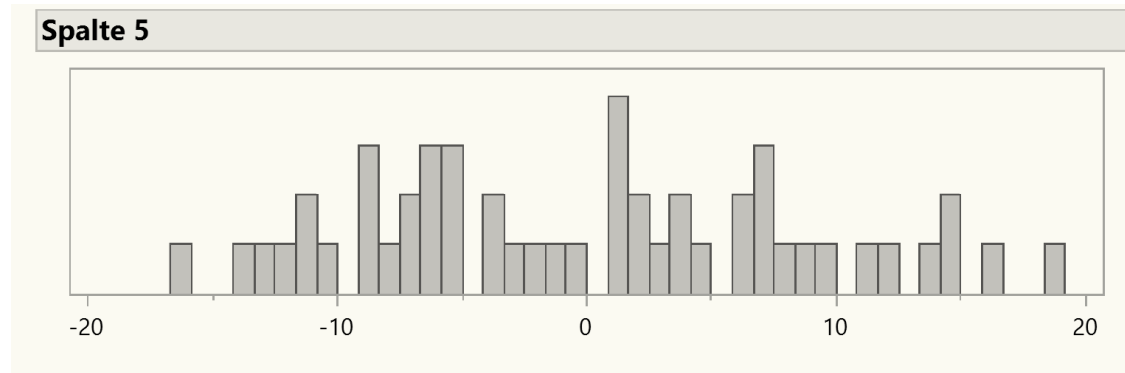
Quiz 2



$$\bar{x}_2 = 88,68$$

	Quiz 1	Quiz 2
Torsten	88 – 81,06	86 – 88,68
Stephanie	82 – 81,06	89 – 88,86

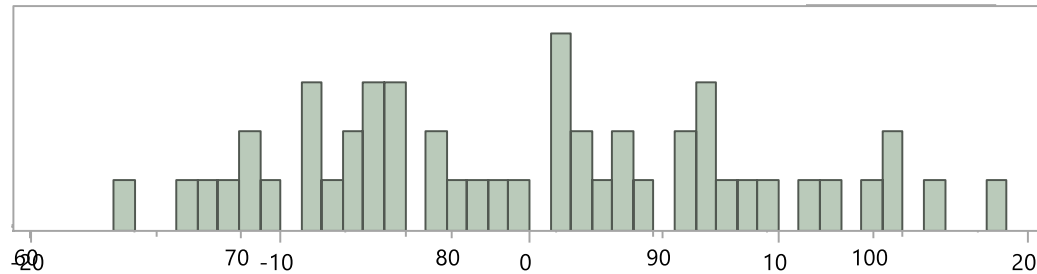
Standardisierung (z-Transformation $\Rightarrow$ z-Score)



	Quiz 1	Quiz 2
Torsten	6,94	-2,86
Stephanie	0,94	0,14

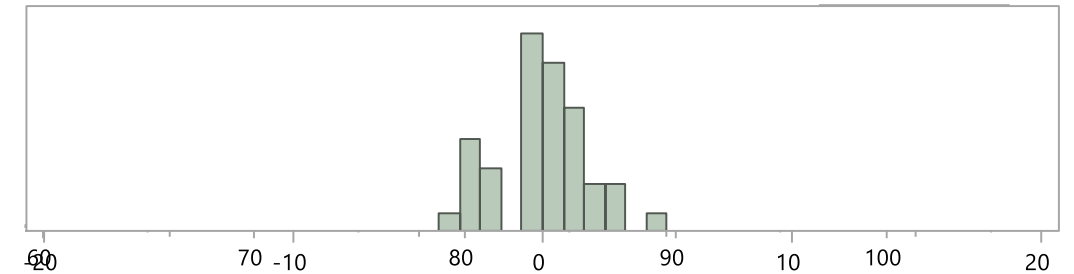
Standardisierung (z-Transformation $\Rightarrow$ z-Score)

Spalte 5



$$s_1 = 8,82$$

Spalte 6



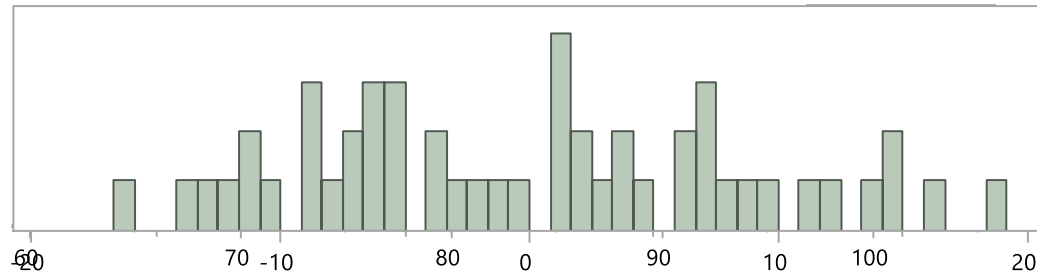
$$s_2 = 1,79$$

	Quiz 1	Quiz 2
Torsten	$6,94 \div s_1$	$-2,86 \div s_2$
Stephanie	$0,94 \div s_1$	$0,14 \div s_2$

- Standardabweichung
 - Für Verteilung: Durchschnittliche Differenz (Abstand) zum Mittelwert

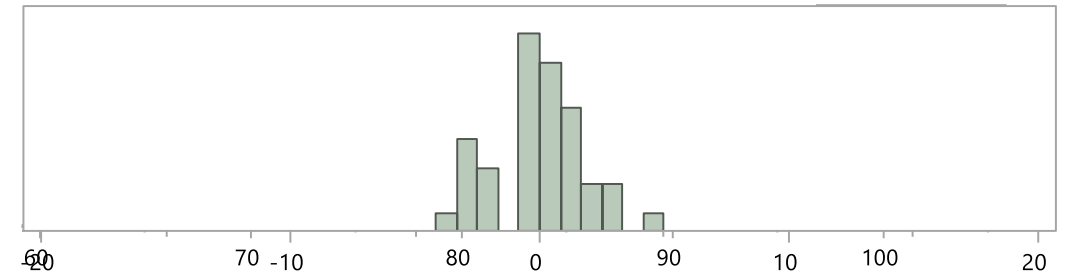
Standardisierung (z-Transformation $\Rightarrow$ z-Score)

Spalte 5



$$s_1 = 8,82$$

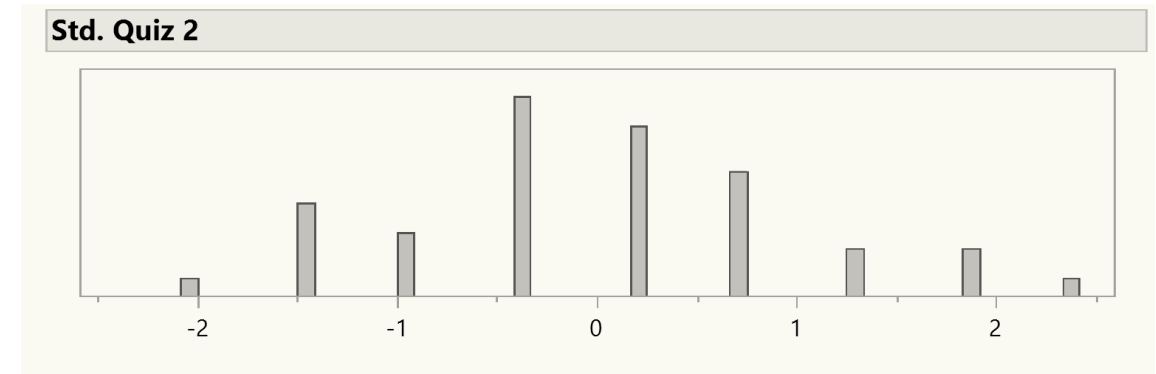
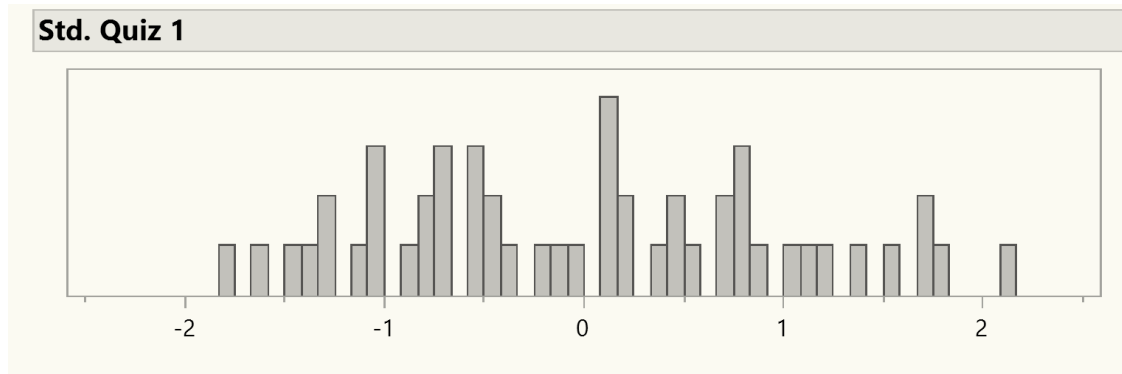
Spalte 6



$$s_2 = 1,79$$

	Quiz 1	Quiz 2
Torsten	$6,94 \div 8,82$	$-2,86 \div 1,79$
Stephanie	$0,94 \div 8,82$	$0,14 \div 1,79$

Standardisierung (z-Transformation $\Rightarrow$ z-Score)



	Quiz 1	Quiz 2
Torsten	0,79	-1,50
Stephanie	0,11	0,18

- Gleiche Einheit für beide Verteilungen: Z-Score
- Vergleichbarkeit
- Anzahl Einheiten (Standardabweichungen) über/unter Mittelwert

Standardisierung (z-Transformation $\Rightarrow$ z-Score)

- **Z-Score**

- $z_i = \frac{deviation_i}{s} = \frac{X_i - \bar{x}}{s}$
 - $\bar{x}$: Mittelwert der Verteilung
 - s : Standardabweichung der Verteilung
- Vorzeichen
 - Über (+) oder unter (-) Mittelwert der Population
- Wert
 - Anzahl der Standardabweichungen, um die ein bestimmter Wert vom zugehörigen Mittelwert abweicht

Umrechnung zwischen Rohdaten und z-Score

- Rohdaten zu Z-Score (Standardisieren, Z-Transformation)

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

- Z-Score zu Rohdaten

$$x_i = \bar{x} + z_i * s$$

Standardisierung (z-Transformation $\Rightarrow$ z-Score)

- Beispiel: Standardisierte Punkte in Pub-Quiz

	Name	Quiz 1	Quiz 2	Mittelwert[Quiz 1,Quiz 2]	Std. Quiz 1	Std. Quiz 2	Mittelwert[Std. Quiz 1,Std. Quiz 2]
1	Torsten	88	86	87	0,79	-1,50	-0,36
2	Stephanie	82	89	85,5	0,11	0,18	0,14
3	Gabriele	84	92	88	0,33	1,86	1,09
4	Renate	82	92	87	0,11	1,86	0,98
5	Jens	77	86	81,5	-0,46	-1,50	-0,98
6	Sabine	87	90	88,5	0,67	0,74	0,71
7	Uwe	95	88	91,5	1,58	-0,38	0,60
8	Claudia	72	92	82	-1,03	1,86	0,41
9	Bernd	75	86	80,5	-0,69	-1,50	-1,09
10	Birgit	86	86	86	0,56	-1,50	-0,47
11	Daniel	74	91	82,5	-0,80	1,30	0,25
12	Stefan	100	88	94	2,15	-0,38	0,88
13	Klaus	81	86	83,5	-0,01	-1,50	-0,75
14	Jörg	85	86	85,5	0,45	-1,50	-0,53
15	Andrea	96	89	92,5	1,69	0,18	0,94
16	Thomas	80	93	86,5	-0,12	2,41	1,15
17	Ralf	73	87	80	-0,91	-0,94	-0,93
18	Karin	76	85	80,5	-0,57	-2,06	-1,32
19	Martina	67	91	79	-1,59	1,30	-0,15
20	Sven	75	88	81,5	-0,69	-0,38	-0,53
21	Susanne	87	88	87,5	0,67	-0,38	0,15

	Quiz 1	Quiz 2	Durchschnitt
Torsten	0,79	-1,50	
Stephanie	0,11	0,18	

- Torsten:
 - 0,36 SD schlechter als der Durchschnitt aller Mitspieler/-innen
- Stephanie:
 - 0,14 SD besser als der Durchschnitt aller Mitspieler/-innen

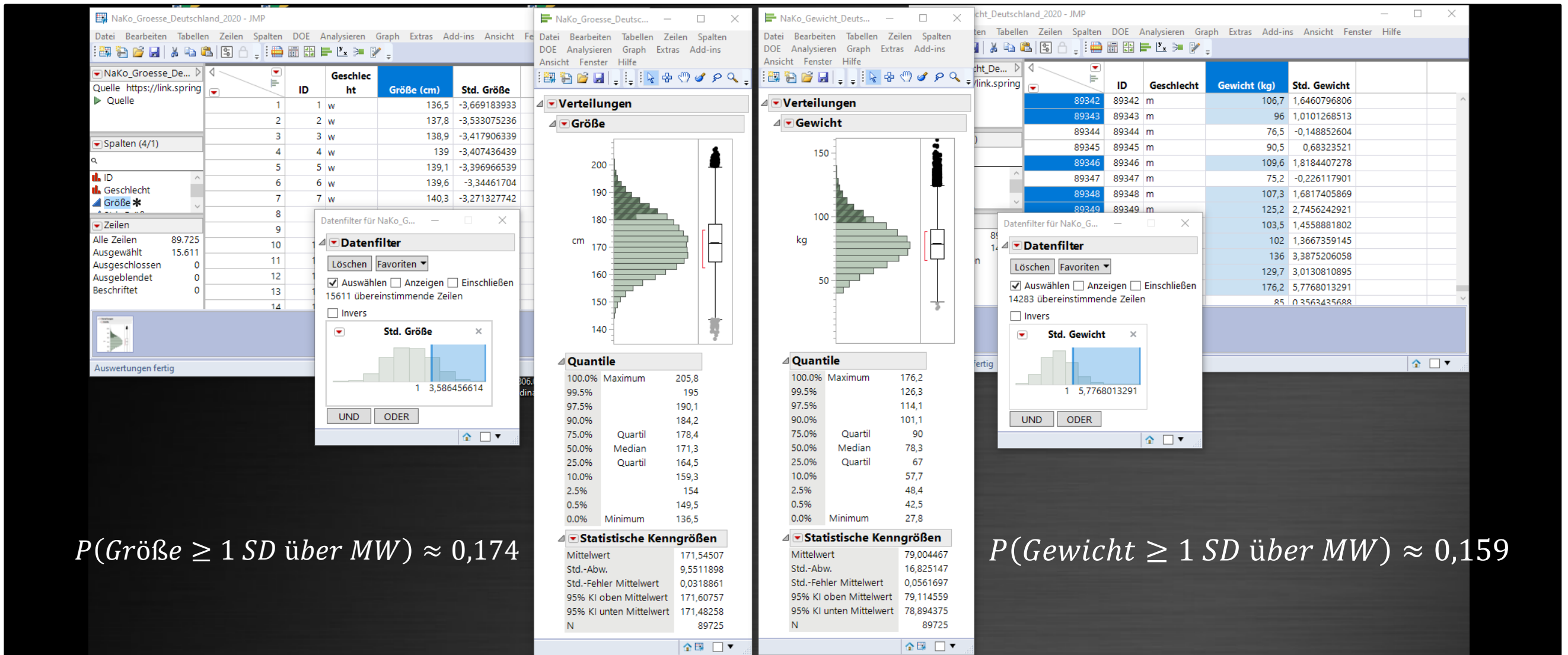
Verwendung der z-Standardisierung / z-Transformation



- Beschreibung eines Wertes (scores) in einer Verteilung
 - mit einziger Zahl
- Zentrierung und Neuskalierung einer gesamten Verteilung (Relativierung)
 - Mittelwert: **Immer 0**
 - Standardabweichung: **Immer 1**
 - ➡ **Standardnormalverteilung** (falls ursprüngliche Verteilung normalverteilt)
 - Form der Verteilung: **Unverändert**
- Vergleichbarkeit von „ungleichen“ (unterschiedliche Größenordnung / unterschiedliche Streuung) Verteilungen
- Beispiel: Sprachtest (TOEFL vs. TOEIC)

Anteile von standardnormalverteilten Häufigkeitsverteilungen

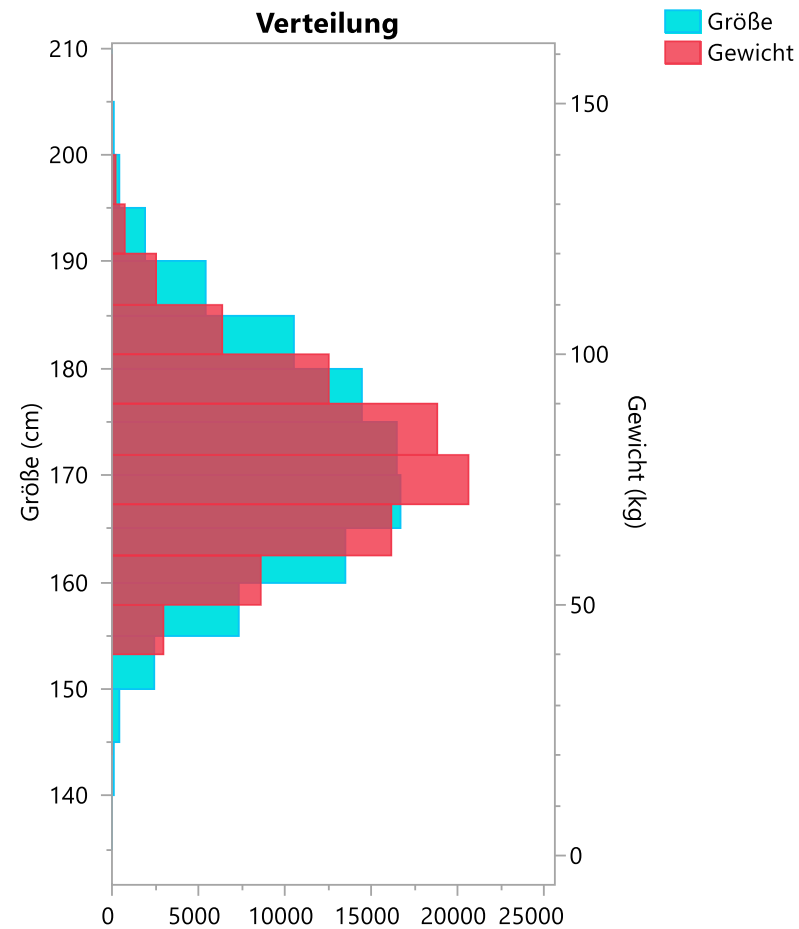
Z-Transformation



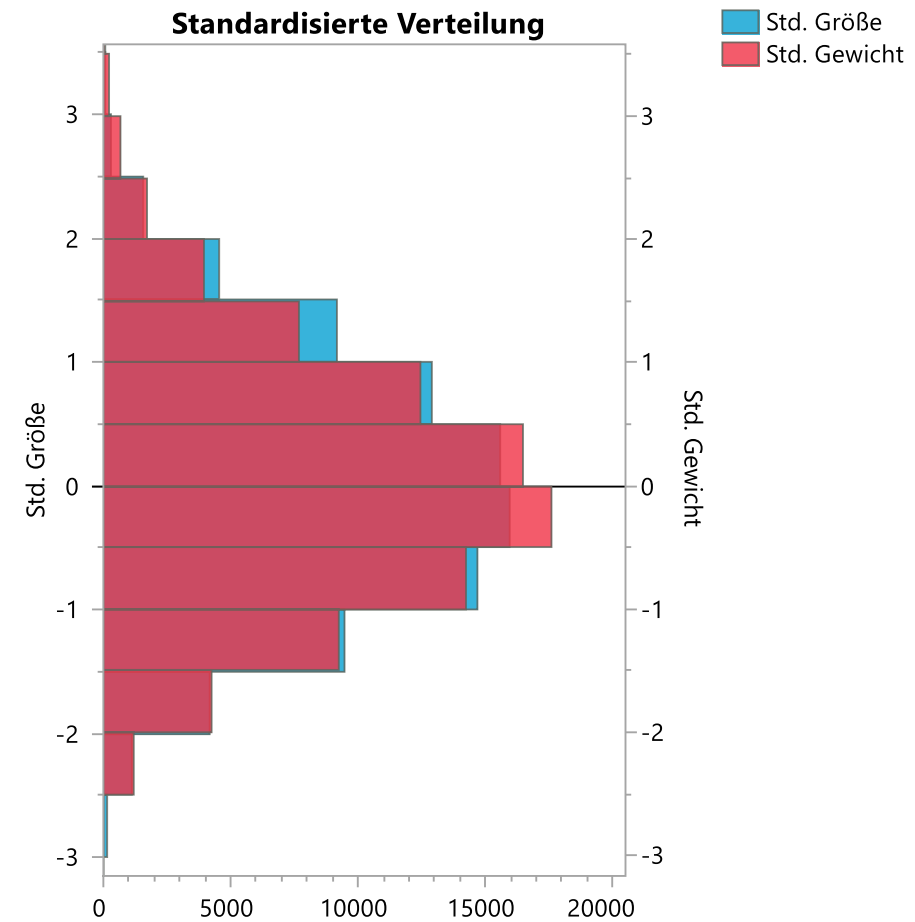
$$P(\text{Größe} \geq 1 \text{ SD über MW}) \approx 0,174$$

$$P(\text{Gewicht} \geq 1 \text{ SD über MW}) \approx 0,159$$

Normalverteilung / Standardnormalverteilung



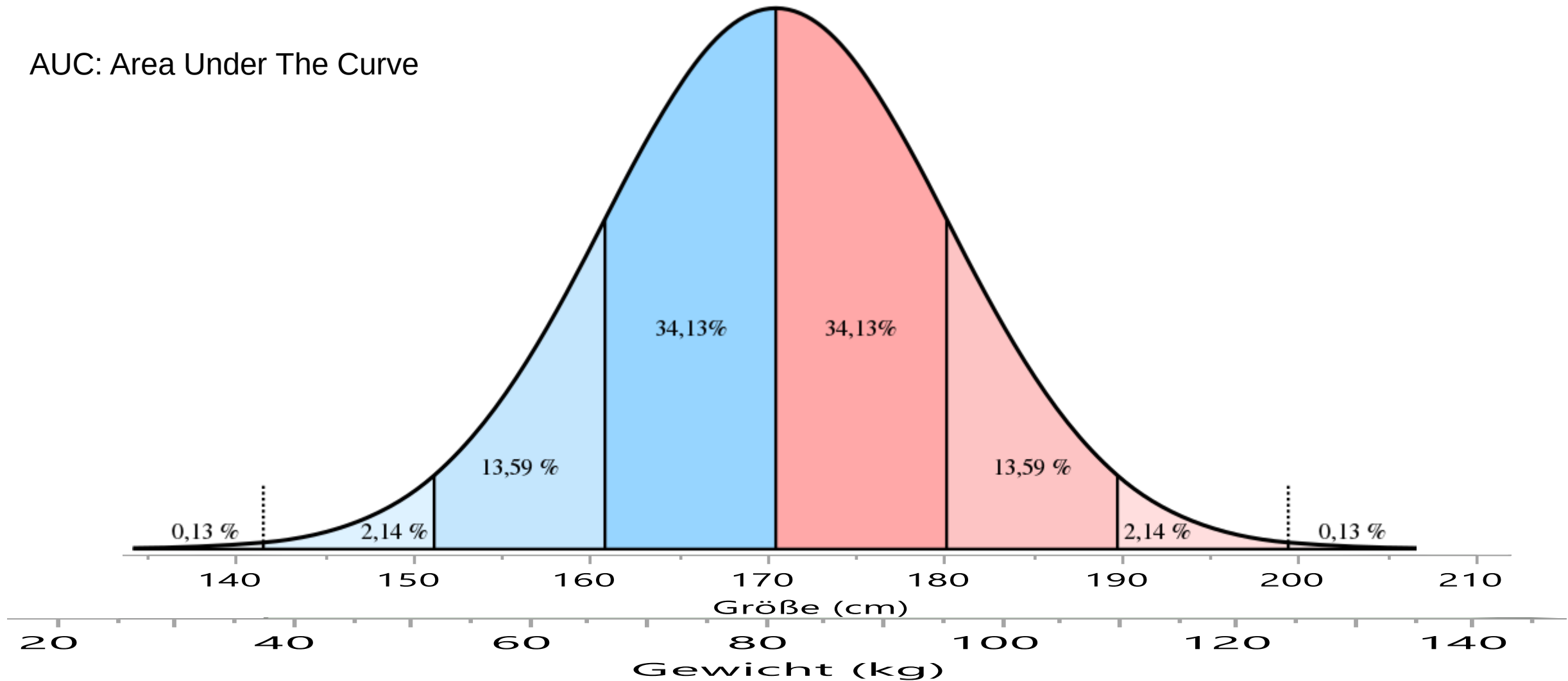
$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



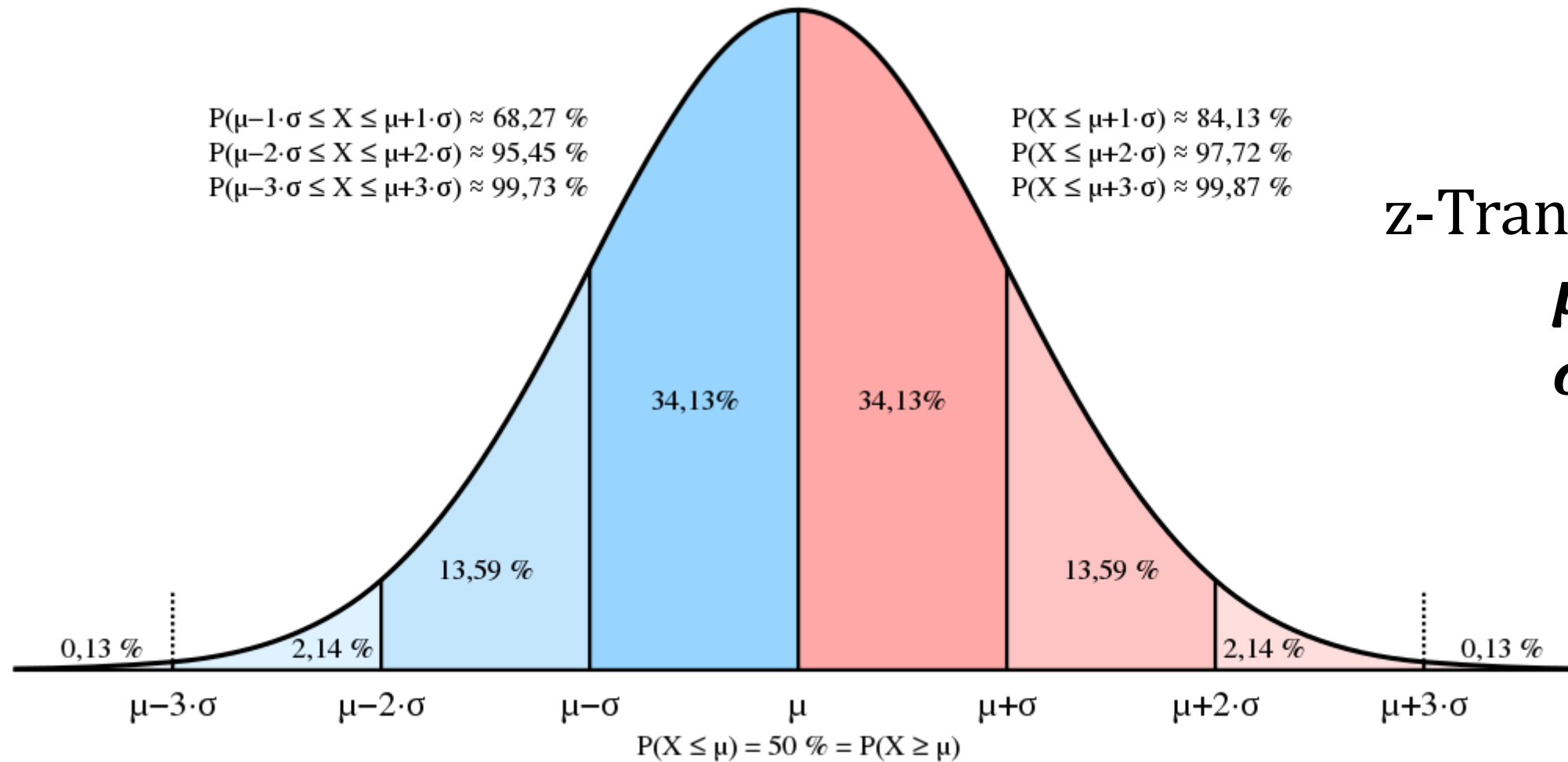
$$f(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{1}{2}x^2}$$

Normalverteilung

AUC: Area Under The Curve



Standardnormalverteilung (Z-Verteilung)



Statistik

Take Home

Take Home

- (Kumulierte) absolute und relative Häufigkeit
 - Empirische Verteilungsfunktion
- Anteil und Häufigkeit
- Normalverteilung
 - Q-Q-Plot
- z-Transformation
- Standardnormalverteilung