



von
nd
al-
tigter
Rolle.
Schaus
die,
nerscheinungen
Merkmale

Statistik

Deskriptive Statistik

Lügen mit Statistik



Kontraste, ARD, Do 16.10.2025, 21:45 Uhr





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

Statistik

Beschreibung von Häufigkeitsverteilungen

Beschreibung von Häufigkeitsverteilungen



- Beschreibende Statistik
- Ziel: Übersichtliche Darstellung von empirischen Daten
 - Tabellen ✓
 - Diagramme / Grafiken ✓
 - Kennzahlen (Parameter) ⇐
 - Lageparameter
 - Streuungsparameter
 - Gestaltparameter

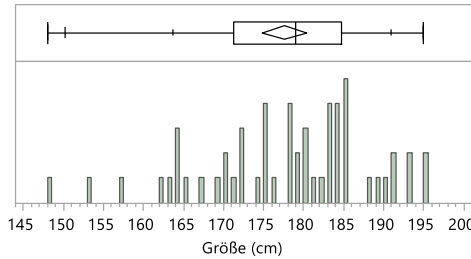
	Aggregation der Daten	Übersichtlichkeit	Informationsgehalt
Tabelle	Niedrig	Niedrig	Hoch
Diagramm	Mittel	Mittel	Mittel
Parameter	Hoch	Hoch	Niedrig



Lageparameter – Modus, Median, Mittelwert

- Beschreibung zentrale Tendenz (Zentrum) einer (Häufigkeits-)Verteilung
 - Wo befindet sich ein Großteil der Merkmalswerte einer Verteilung?
- **Modus, Modalwert** x_{Mod}
 - Häufigster Wert
 - Anwendung auch bei nominal skalierten Daten
- **Median** x_{Med}
 - „Mittlerer“ Wert: Gleichviele Werte ober- und unterhalb
 - Median = 50%-Quantile
 - Urliste aufsteigend sortieren (nach Rang geordnet): $x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}$
 - Mindestens ordinal skaliert
 - $$x_{Med} = \begin{cases} x_{[\frac{n+1}{2}]} & \text{falls } n \text{ ungerade} \\ \frac{x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}}{2} & \text{falls } n \text{ gerade} \end{cases}$$
- **Arithmetischer Mittelwert** (Durchschnittswert) $\bar{x}$
 - $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
 - Sinnvoll nur bei kardinalskalierten Werten
 - Meist nur sinnvoll bei symmetrischer Verteilung (Schiefe ≈ 0)
 - Nicht robust gegen Ausreißer

Lageparameter – Modus, Median, Mittelwert

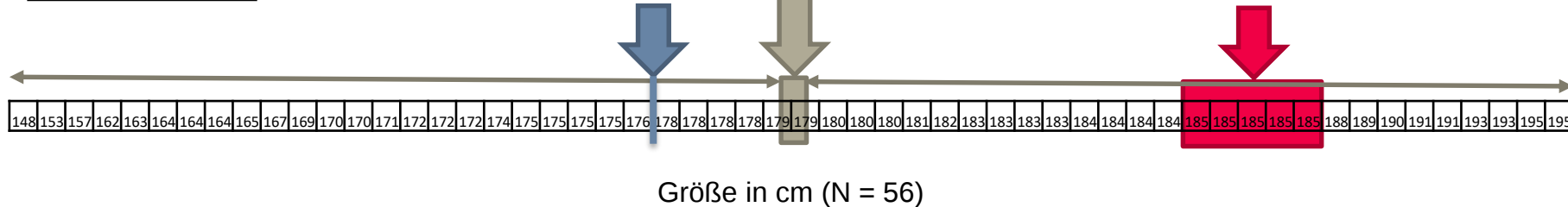


Modus	185
Median	179
Mittelwert	177,57
N	56

$$\text{Median} = \frac{x_{[\frac{56}{2}]} + x_{[\frac{56}{2}+1]}}{2} = \frac{x_{[28]} + x_{[29]}}{2} = \frac{179 + 179}{2} = 179$$

$$\text{Mittelwert} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 177,57$$

$$\text{Modus} = \max h(x_j) = h(185) = 5$$



Streuungsparameter

- Beschreiben Streubreite einer (Häufigkeits-)Verteilung um Lageparameter
- Voraussetzung: Kardinalskalierte Daten
- **Spannweite** (*range*)
 - $SP = \max(x_i) - \min(x_i)$
 - Nicht robust gegen Ausreißer

Streuungsparameter um arithmetischen Mittelwert



- Mittlere absolute Abweichung (*mean absolute deviation*)
 - $MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$
- Summe der Abweichungsquadrate (*sum of squared differences*)
 - $SQ_x = \sum_{i=1}^n (x_i - \bar{x})^2$
- **Empirische Varianz (variance)**
 - $\tilde{s}_x^2 = \frac{1}{n} SQ_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
- **Empirische Standardabweichung (standard deviation)**
 - $\tilde{s} = \sqrt{\tilde{s}_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
- Empirischer Variationskoeffizient (*coefficient of variation*)
 - $v = \frac{\tilde{s}}{\bar{x}}$

Streuungsparameter um Stichprobenmittelwert

Stichprobenvarianz und Stichprobenstandardabweichung



- Stichprobenvarianz

- $$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Anzahl Freiheitsgrade
 $n - 1$

- Stichprobenstandardabweichung

- $$\bar{s} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Größe	Mittelwert	Größe-Mittelwert
148	156,6	-8,6
153	156,6	-3,6
157	156,6	0,4
162	156,6	5,4
163	156,6	6,4
Summe		0

$$(-8,6) + (-3,6) + 0,4 + 5,4 + 6,4 = 0$$

Frei!

Fest!

Empirische Varianz vs. Stichprobenvarianz

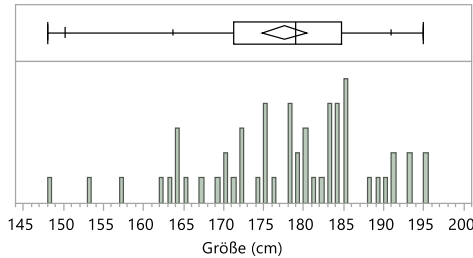
- **Empirische Varianz**

- $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
- „Varianz der Daten“
- **Streuung der Werte in beobachteter Stichprobe**

- **Stichprobenvarianz**

- $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- Verwendung bei Daten **einer Stichprobe** aus größerer Grundgesamtheit
- **Schätzen der Streuung der Grundgesamtheit**
- *Unverzerrter Schätzer der wahren Varianz der Grundgesamtheit*

Median und Quartile

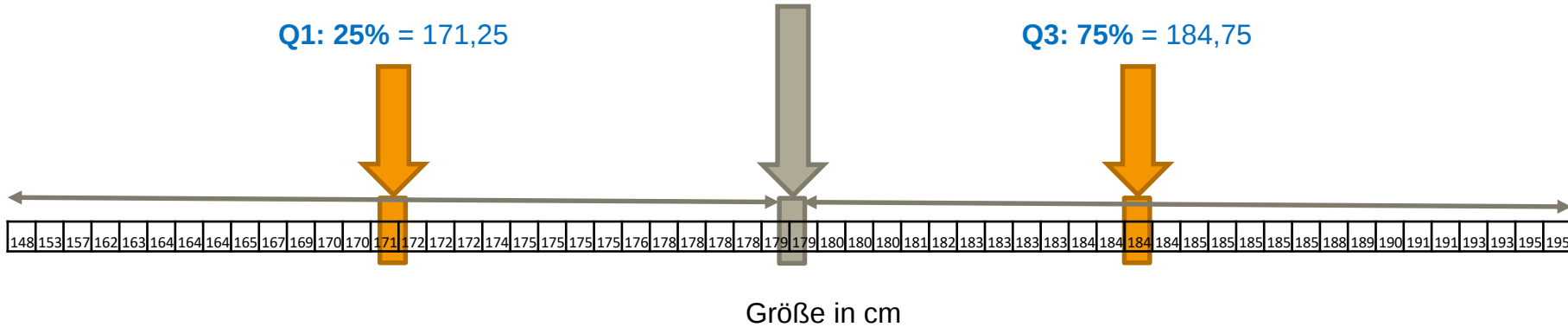


Median (Q2): 50%

$$\frac{x_{\lceil \frac{56}{2} \rceil} + x_{\lceil \frac{56}{2} + 1 \rceil}}{2} = \frac{x_{[28]} + x_{[29]}}{2} = \frac{179 + 179}{2} = 179$$

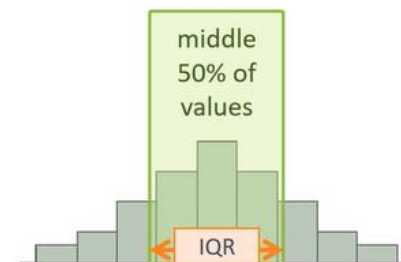
Q1: 25% = 171,25

Q3: 75% = 184,75



Streuungsparameter um Median

- **p-Quantil**
 - Teilt Werte in Anteil $< p$ und Anteil $> p$
 - **Median** $x_{Med}, Q_{0,5}$: $p = 0,5$
 - Oberes **Quartil** ($p = 0,75$), unteres **Quartil** ($p = 0,25$)
 - Terzile ($1/3, 2/3$), Quintile ($0,2; 0,4; 0,6; 0,8$), Dezile ($0,1$ Schritte)
 - **Perzentile** ($0,01$ Schritte, z. B. 5% oder 95% Perzentil)
- **Interquartilsabstand** (*interquartile range*)
 - enthält 50% aller Werte
 - $IQR = Q_{0,75} - Q_{0,25}$
- Mittlere absolute Abweichung vom Median (*mean absolute deviation from the median*)
 - $MD = \frac{1}{n} \sum_{i=1}^n |x_i - x_{Med}|$



Wie berechnet man Quantile (Median, Quartil, Perzentil, etc.)

- Quantile unterteilen sortierte Datenreihe in gleich große Teile
 - z. B. Terzile 3, Quartile 4, Quintile 5, Dezile 10, Perzentile 100 Teile
- Quantil
 - Datenwert, bei dem der Anteil der Datenreihe der kleiner als der Wert ist, kleiner oder gleich dem Quantil p und der Anteil der größer als der Wert ist, höchstens $1 - p$ ist
- Allgemeine Formel

$$x_p = \begin{cases} \frac{1}{2}(x_{n \cdot p} + x_{n \cdot p + 1}) & \text{wenn } n \cdot p \text{ ganzzahlig} \\ x_{\lfloor n \cdot p + 1 \rfloor} & \text{wenn } n \cdot p \text{ nicht ganzzahlig} \end{cases}$$

n Länge der Liste
 p Anteil
 $\lfloor \rfloor$ Abrunden

- Formel in JMP ([Quelle](#))

$$x_p = \begin{cases} x_{\lfloor p \cdot (n+1) \rfloor} & \text{wenn } n \cdot p \text{ ganzzahlig} \\ (1 - (p \cdot n - \lfloor p \cdot n \rfloor)) \cdot x_{\lfloor p \cdot n \rfloor} + (p \cdot n - \lfloor p \cdot n \rfloor) \cdot x_{\lfloor p \cdot (n+1) \rfloor} & \text{wenn } n \cdot p \text{ nicht ganzzahlig} \end{cases}$$

Mittelwert vs. Median

- Warum ist der Mittelwert anfällig für Ausreißer (extreme Werte)?

Statistik	Mathematik	Programmieren	Datenbanken
4	1	1	1

- Median: $x_{Med} = \frac{x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}}{2} = \frac{x_{[2]} + x_{[3]}}{2} = \frac{1+1}{2} = 1$
- Mittelwert: $\bar{x} = \frac{1}{n} * \sum_i^n x_i = \frac{x_1 + x_2 + x_3 + x_4}{n} = \frac{4 + 1 + 1 + 1}{4} = \frac{4}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1 + \frac{3}{4} = 1,75$

⇒ Einfluss von Statistik auf Mittelwert 4x stärker als der anderer Fächer

Mittelwert vs. Median

- Warum ist der Mittelwert anfällig für Ausreißer (extreme Werte)?

Statistik	Mathematik	Programmieren	Datenbanken	G. d. W.-Inf.	BWL
4	1	1	1	1	1

- Median: $x_{Med} = \frac{x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}}{2} = \frac{x_{[3]} + x_{[4]}}{2} = \frac{1+1}{2} = 1$

- Mittelwert: $\bar{x} = \frac{1}{n} * \sum_i^n x_i = \frac{4 + 1 + 1 + 1 + 1 + 1 + 1}{6} = \frac{4}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} + \frac{5}{6} = 1,5$

⇒ Mindestens eine weitere 1 nötig, um Einfluss von Statistik auf Mittelwert auszugleichen

⇒ Mindestens zwei weitere 1en nötig, um Einfluss von Statistik auf Mittelwert zu übertreffen

⇒ **Strengen Sie sich in Statistik an!**



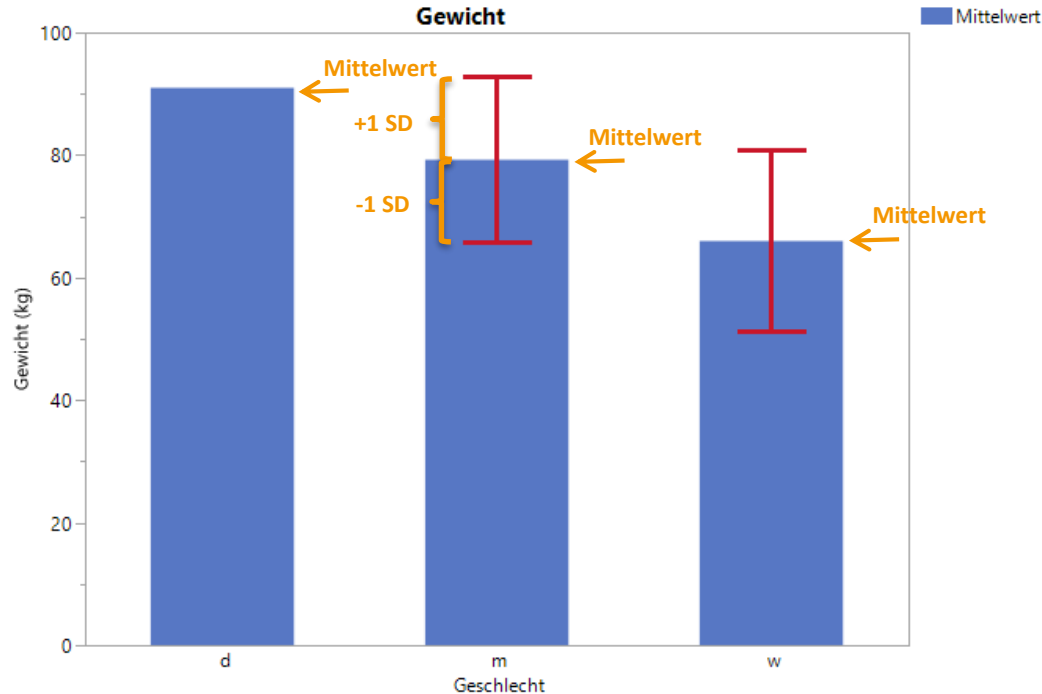
von
nd
al-
tigter
Rolle.
Schaus
die
verschieden
Merkmale

Statistik

Graphische Darstellung von Lage- und Streuungsparametern

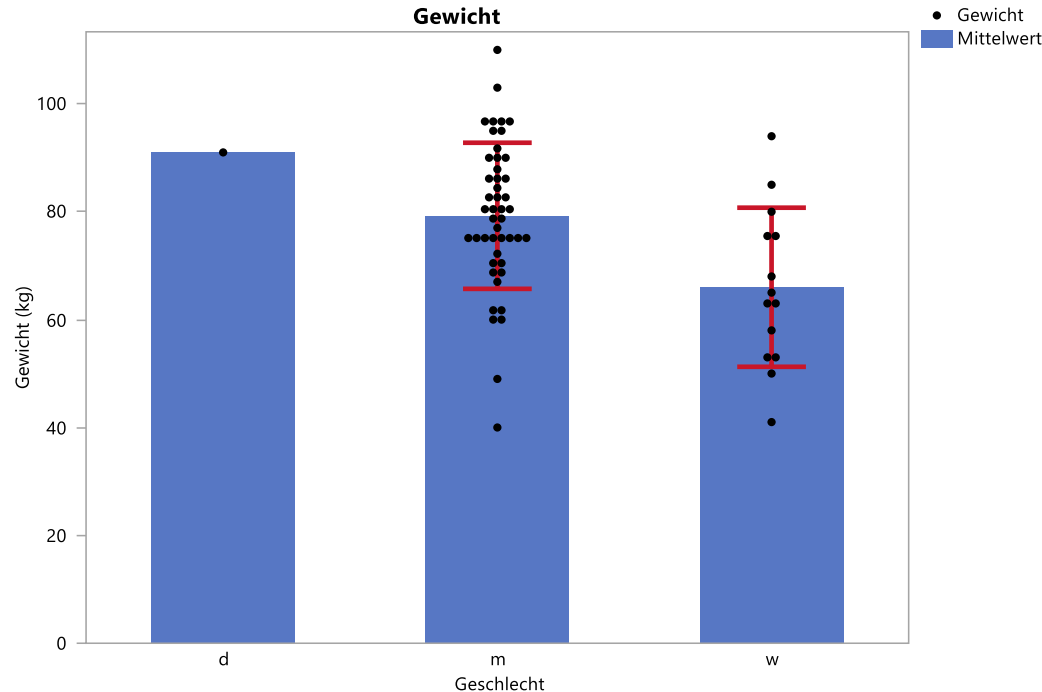
Graphische Darstellung einer Verteilung: Balkendiagramm

Mittelwert und Standardabweichung



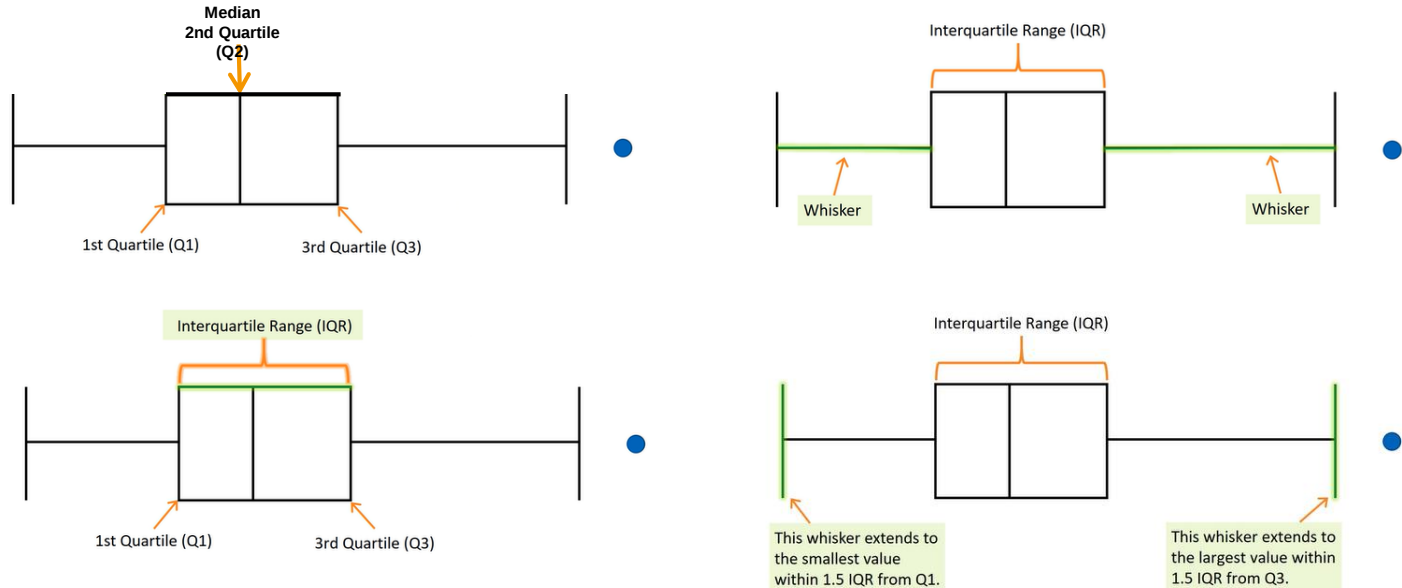
Graphische Darstellung einer Verteilung: Balkendiagramm

Mittelwert und Standardabweichung



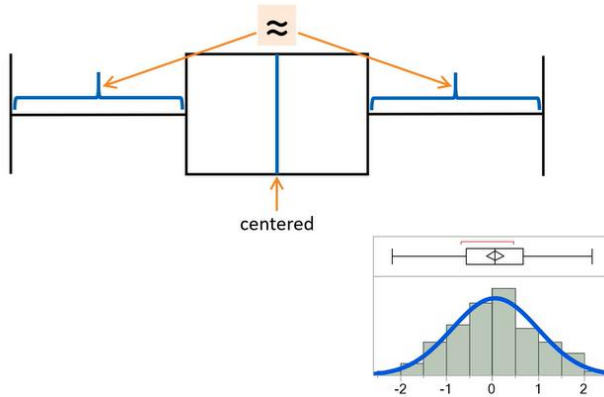
Graphische Darstellung einer Verteilung: Box-Plots

Median und Quartile

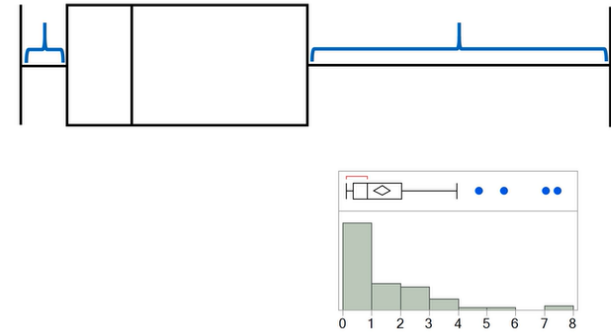


Graphische Darstellung einer Verteilung: Box-Plots

Median und Quartile



Verteilung ist symmetrisch



Verteilung ist asymmetrisch

Statistik

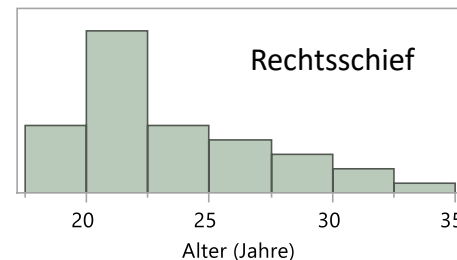
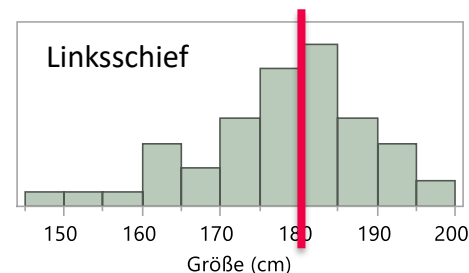
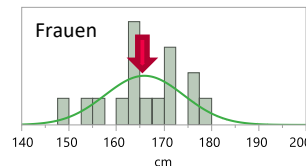
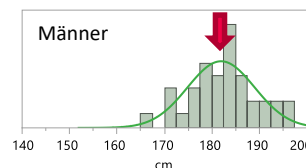
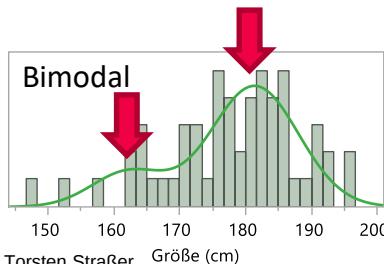
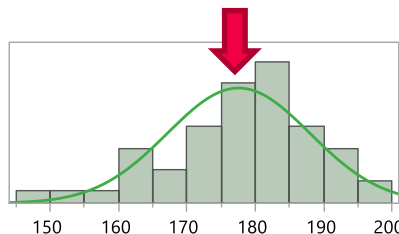
Gestaltparameter: *The Shape of Data*

Deskriptive Statistik

Die Form der Häufigkeitsverteilung (*The Shape of Data*)



- Eigenschaften von Verteilungen
 - Symmetrie
 - Symmetrisch
 - Asymmetrisch
 - Schiefe (*skewness*)
 - Stärke der Asymmetrie
 - Linksschief, negative Schiefe
 - Rechtsschief, positive Schiefe
 - Unimodal / multimodal

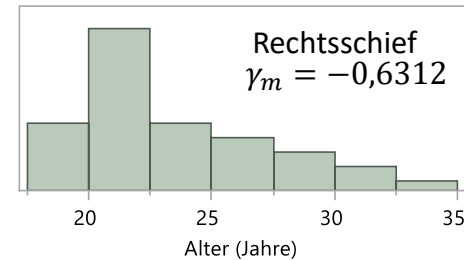
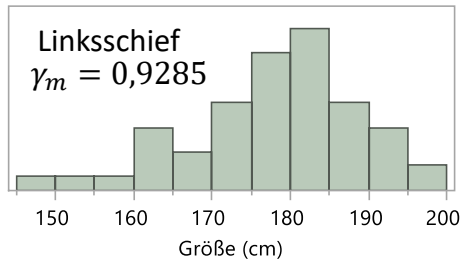


Gestaltparameter

Schiefe und Kurtosis



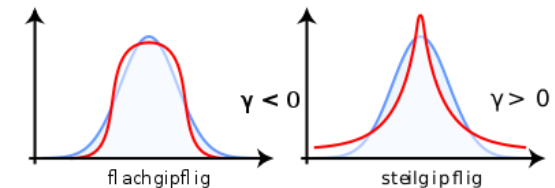
- Schiefe (*skewness*)
 - Maßzahl für Symmetrie einer (Häufigkeits-)Verteilung



- Kurtosis (Wölbung, *kurtosis*)
 - Maßzahl für Steilheit bzw. „Spitzigkeit“ einer (Häufigkeits-)Verteilung

$$w = 0,2595$$

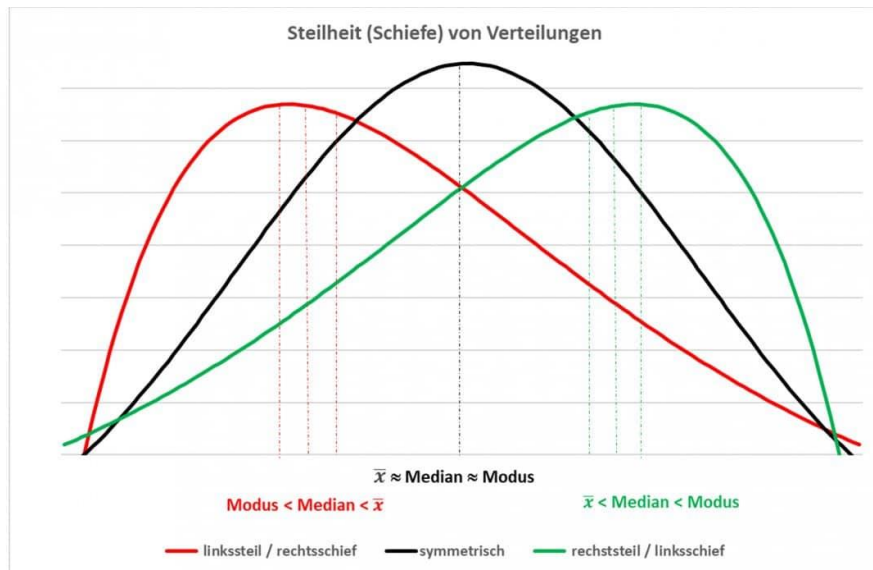
$$w = 0,1746$$



Gestaltparameter

Schiefe und Symmetrie

- Schiefe und Symmetrie



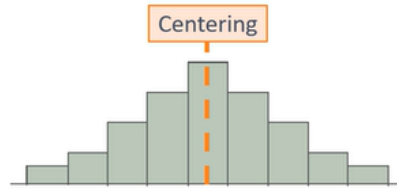
<https://www.statistik-nachhilfe.de/ratgeber/statistik/deskriptive-statistik/masszahlen/parameter-der-form/rechtssteile-verteilung-linkssteile-verteilung>

$$v = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Beschreibung von Häufigkeitsverteilungen

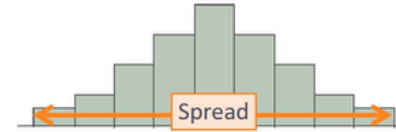
Übersicht

Lageparameter



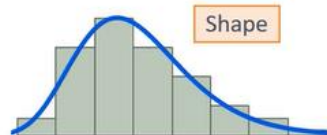
- Modus
- Median
- Mittelwert

Streuungsparameter

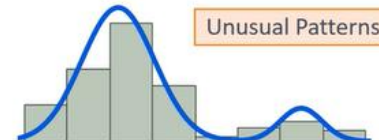


- Spannweite
- Quantilen (Quartile, Perzentile), IQR
- Varianz / Standardabweichung

Gestaltparameter



- Schiefe
- Kurtosis (Wölbung)



- Unimodal / Multimodal



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

Statistik

Standardnormalverteilung

Mittelwert und Standardabweichung

Standardnormalverteilung



- Gauß-Kurve, Glockenkurve, Gaußsche Normalverteilung
- Nach Carl Friedrich Gauß (1777 – 1855),
Mathematiker, Statistiker, Physiker, Astronom
- $f(x) = e^{-x^2} \Rightarrow \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$
- 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: $\mu = 0, \sigma = 1 \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}}$



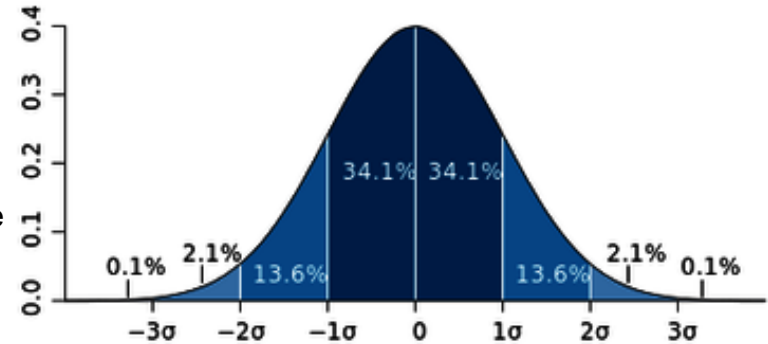
Wikipedia, Tigran Mittr am, Public domain

Mittelwert und Standardabweichung

Standardnormalverteilung



- Es gilt näherungsweise
 - $\pm 1 * \sigma$ enthält 68,27% aller Werte
 - $\pm 2 * \sigma$ enthält 95,73% aller Werte
 - $\pm 3 * \sigma$ enthält 99,73% aller Werte
 - $\pm 5 * \sigma$ enthält 99,9999426697% aller Werte
- Umgekehrt
 - 50% aller Werte sind zwischen $\pm 0,675 * \sigma$
 - 90% aller Werte sind zwischen $\pm 1,645 * \sigma$
 - 95% aller Werte sind zwischen $\pm 1,96 * \sigma$
 - 99% aller Werte sind zwischen $\pm 2,576 * \sigma$



Wikipedia, Intervalle der Normalverteilung, M. W. Toews, CC-BY 2.5

- Wendepunkt der Gaußkurve liegt bei einer Standardabweichung σ (SD)



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

Statistik

Sigma-Regeln der Normalverteilung

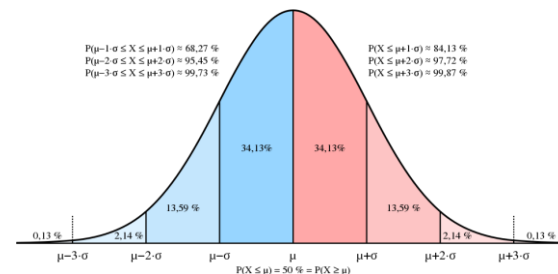
k-σ-Bereiche, Sigma-Regeln

- > 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

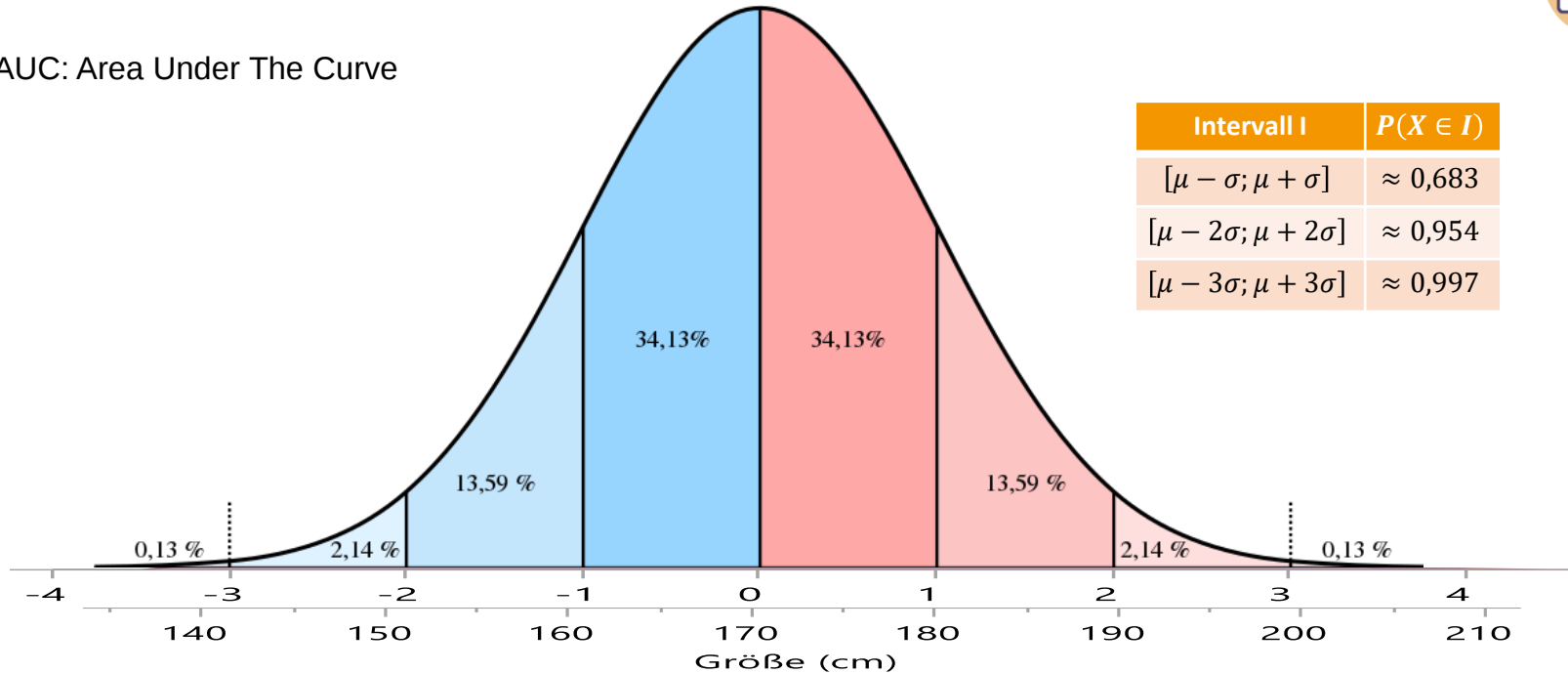
Intervall I	$P(X \in I)$	Intervall I	$P(X \in I)$
$[\mu - \sigma; \mu + \sigma]$	$\approx 0,683$	$[\mu - 1,64\sigma; \mu + 1,64\sigma]$	$\approx 0,90$
$[\mu - 2\sigma; \mu + 2\sigma]$	$\approx 0,954$	$[\mu - 1,96\sigma; \mu + 1,96\sigma]$	$\approx 0,95$
$[\mu - 3\sigma; \mu + 3\sigma]$	$\approx 0,997$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



Normalverteilung



AUC: Area Under The Curve





von
nd
al-
tigter
Rolle.
Schaus
die,
erscheinungen
Merkmale

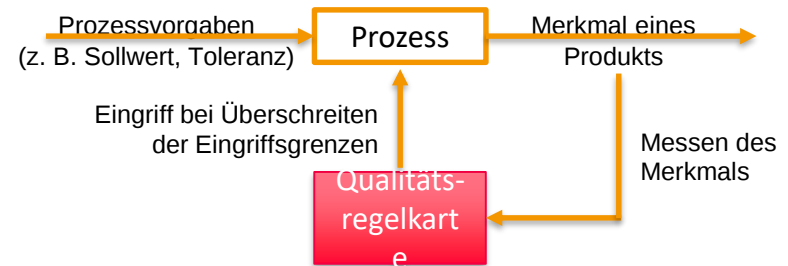
Statistik

Statistische Qualitätskontrolle

Statistische Qualitätskontrolle

Qualitätsregelkarten

- Werkzeug zur statistischen Prozesslenkung (*statistical process control*, SPC)
- Bewertung von (zeitlicher) Qualitätskonstanz (Prozessstabilität)
 - Optimierung von Produktions- und Serviceprozessen
- (Periodische) Auswertung von Prüfdaten (Stichproben)
- Einfache, graphische Bewertung von Prozessveränderungen
 - Änderung der Lage eines Prozessmerkmals
 - Änderung der Streuung eines Prozessmerkmals
- Darstellung Prozessmerkmals
 - Stichprobenmittelwert
 - Stichprobenstreuung
 - Spannweite
 - Stichprobenstandardabweichung
 - Warn- und Eingriffsgrenzen



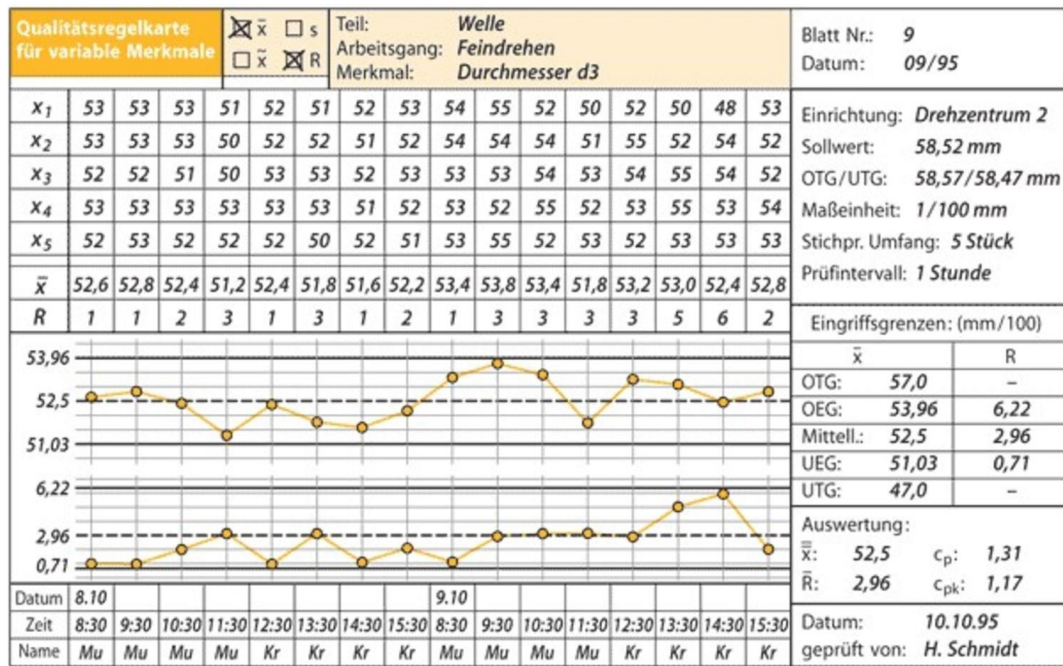
Statistische Qualitätskontrolle

Qualitätsregelkarten

- XbarR-Karte / **XbarS-Karte**
 - Mittelwert $\bar{x}$ „x-quer“, englisch „*x-bar*“
 - Spannweite „R“, englisch „*range*“
 - **Standardabweichung** „S“
- Darstellung Stichprobendaten als
 - Mittelwert $\bar{x}$
 - Spannweite / **Standardabweichung**
- Grenzen (aus Vorwissen)
 - Obere/untere Warngrenze (OWG/UWG)
 - Obere/untere Eingriffsgrenze (OEG/UEG)

Statistische Qualitätskontrolle

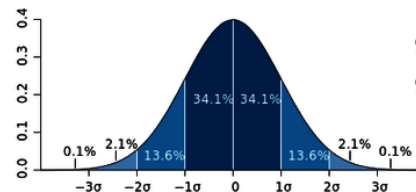
Beispiel XbarR-Karte


 Quelle: <https://www.qz-online.de/a/grundlagenartikel/die-qualitaetsregelkarte-316052>

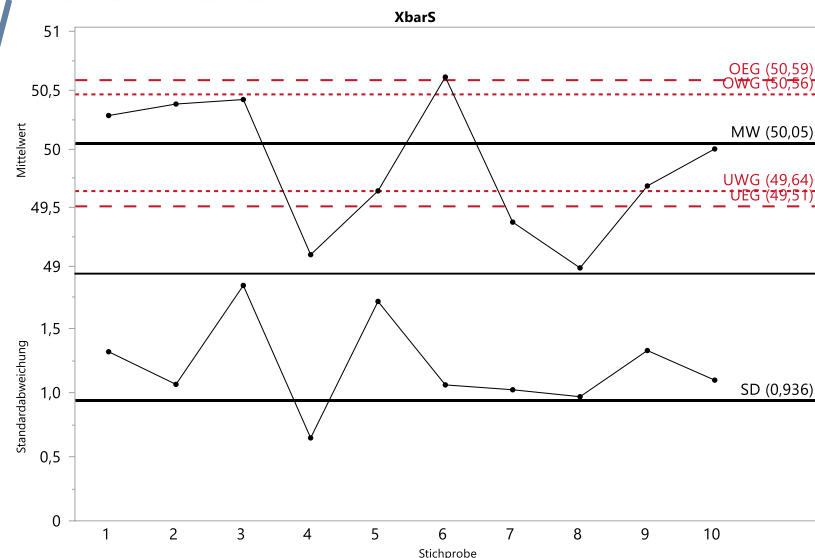
Statistische Qualitätskontrolle

Aufbau XbarS-Karte

- Mittelwert μ (aus Vorwissen)
- Standardabweichung σ
- Eingriffsgrenzen (aus Vorwissen)
 - $OWG = \mu + 1,96\sigma * \frac{1}{\sqrt{n}}$ Standardfehler
 - $UWG = \mu - 1,96\sigma * \frac{1}{\sqrt{n}}$
 - $OEG = \mu + 2,576\sigma * \frac{1}{\sqrt{n}}$
 - $UEG = \mu - 2,576\sigma * \frac{1}{\sqrt{n}}$
- Laufend aktualisierte Stichprobendaten
 - i.d.R. $n = 8$ bis 20
 - Stichprobenmittelwert $\bar{x}$
 - Stichprobenstandardabweichung s



95% aller Werte zwischen $\pm 1,96 * \sigma$
 99% aller Werte zwischen $\pm 2,576 * \sigma$

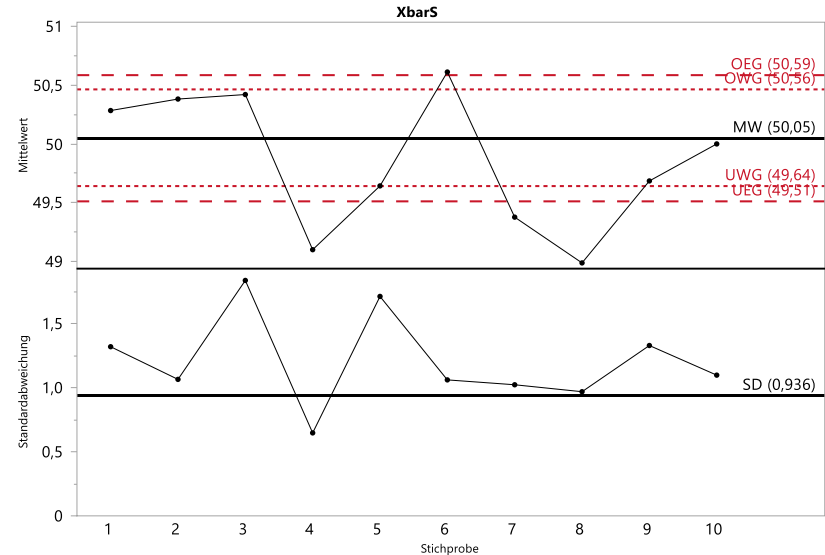


Statistische Qualitätskontrolle

Entscheidungsregeln XbarS-Karte

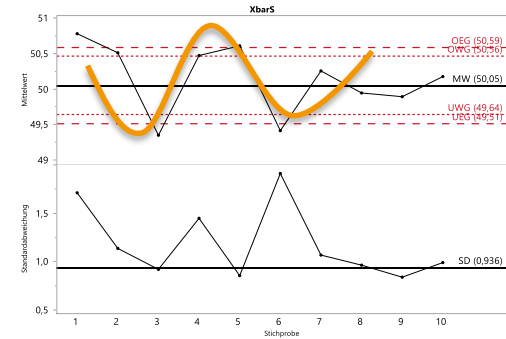
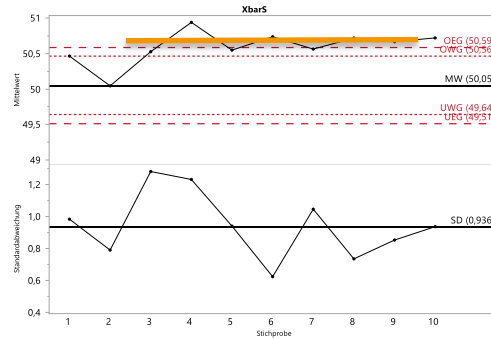
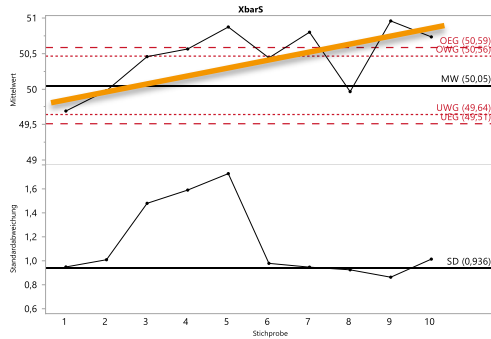


- Datenpunkt zwischen Warngrenzen
 - $UWG < \bar{x} < OWG$
 - Vermutlich keine Störung
- Datenpunkt zwischen Warngrenzen und Eingriffsgrenzen
 - $UEG < \bar{x} < UWG$ bzw. $OWG < \bar{x} < OEG$
 - Verdacht auf Störung
 - Sofort weitere Stichprobe(n)
- Datenpunkt außerhalb Eingriffsgrenzen
 - $\bar{x} < UEG$ bzw. $\bar{x} > OEG$
 - Vermutlich Störung
 - Sofortiger Eingriff erforderlich



Statistische Qualitätskontrolle

Auswertung XbarS-Karte

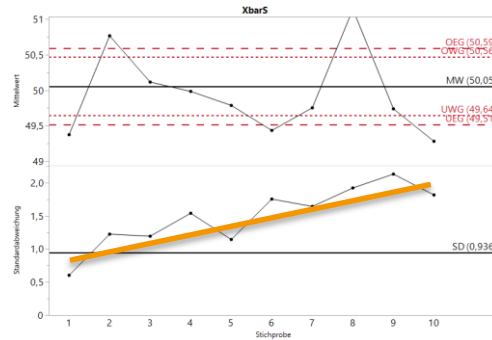


- Trend im Mittelwert
 - > 7 Punkte lineare Steigung
 - Zunehmender Werkzeugverschleiß
- Durchlauf, *Run*
 - > 7 Punkte über/unter Mittelwert
 - Werkzeugschneide beschädigt

- Muster, *Pattern*
 - Temperaturschwankungen, Schichtwechsel

Statistische Qualitätskontrolle

Auswertung XbarS-Karte



- Trend in Standardabweichung
 - Streuung in Stichproben
 - Werkzeugverschleiß

Statistik

Six Sigma

Statistische Qualitätskontrolle

Six Sigma als Qualitätsziel

6σ

- Managementsystem zur Prozessverbesserung, statistisches Qualitätsziel und Methode für Qualitätsmanagement
- Beschreibung, Messung, Analyse, Verbesserung und Überwachung von Geschäftsvorgängen mit statistischen Werkzeugen
- Prozessoptimierung DMAIC
 - *Define – Measure – Analyse – Improve – Control*
- Entwicklung 1987 von Motorola, USA
 - Bekannt durch erfolgreiche Einführung bei General Electric 1996 durch Jack Welch
- Einsatz auch in Softwareentwicklung



Wikipedia, C64zottel, CC BY-SA 4.0

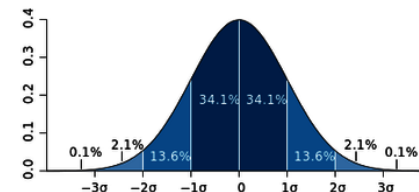
Statistische Qualitätskontrolle

Six Sigma als statistisches Qualitätsziel

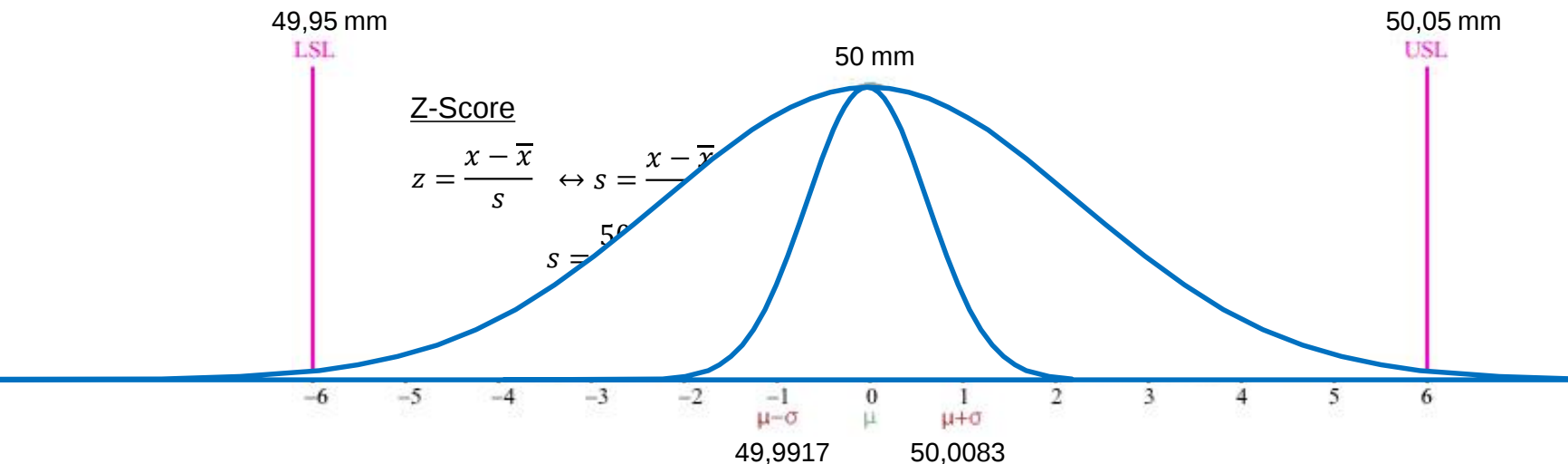
- Qualitätsmerkmale besitzen (unerwünschte) Streuung
- Idealwert wird in Beziehung zum Toleranzbereich gesetzt
 - Je größer die Standardabweichung, desto wahrscheinlicher Überschreitung der Toleranzgrenzen
- Ansatz 6σ
 - Stichprobenmittelwert mindestens 6 Standardabweichungen (Sigma) von Toleranzgrenzen entfernt
 - Praktisch „Nullfehlerproduktion“

Statistische Qualitätskontrolle

Six Sigma als statistisches Qualitätsziel



- Beispiel Schraubenfertigung
 - Zielwert: Länge 50 mm
 - Toleranzgrenzen: $\pm 0,05$ mm
- 68,27 % der Schrauben zwischen 49,9917 und 50,0083 mm
- Fehlerfreie Schrauben (Abweichung $< \pm 0,05$ mm)
 - 99,99966 %





von
nd
al-
tigter
Rolle.
Schaus
die,
senerscheinung
Merkmale au

Statistik

Take Home

Take Home

- Deskriptive Statistik
 - Häufigkeitsverteilungen
 - Beschreibung von Häufigkeitsverteilungen
 - Gestalt
 - Lageparameter
 - Streuungsparameter
 - Darstellung
 - Box-Plot
 - Balkendiagramm
- Standardnormalverteilung
 - k-Sigma-Regeln
- Statistische Qualitätskontrolle
 - XbarS-Karte
- Six Sigma