Zusammenfassung der letzten Vorlesung

Ziele und Aufgaben der deskriptiven Statistik

↓

Formen der statistischen Beschreibung

↓

1. Tabellen	2. Graphiken	3. Kenngrößen
↓	↓	↓
Häufigkeitstabellen	Kreisdiagramm	Maße der zentralen Tendenz
	Histogramm	Streuungsmaße
	Polygon	Formmaße

Maße der zentralen Tendenz:
Beispiel für Modus = Median = arithm. Mittel

- Häufigkeitsverteilung mit gleicher zentraler Tendenz, aber unterschiedlicher Streuung:

- Messung 1: 100 100 110 150 150 150 190 200 200

  Messung 2: 145 146 147 150 150 150 153 154 155

- Modus = Median = arithmetisches Mittel

- Unterschied:

  Die beiden Messwertreihen unterscheiden sich in der Variabilität der Messwerte!

Streuungsmaße: Varianz

- Definition:

  Die Varianz ist die Summe aller quadrierten Abweichungen vom Mittelwert, dividiert durch die Anzahl der Freiheitsgrade (= Anzahl aller Messwerte (N) minus 1).

- Information:

  Sie gibt die durchschnittliche quadrierte Abweichung der Einzelwerte vom Mittelwert an.

- Keine Information:

  Über die Einheit der Daten

- Anwendung:

  Intervallskalierten Daten

Formel: Varianz

- Formel:

$$s^2 = \frac{1}{N-1} \sum (x_i - \overline{x})^2$$

- Wobei:

s^2 = Varianz

N = Anzahl

x_i = Platzhalter für die einzelnen Daten

$\overline{x}$ = arithmetisches Mittel

Streuungsmaße: Standardabweichung

- Definition:

  Die Standardabweichung ist die Wurzel der Varianz.

- Information:

  Sie gibt die durchschnittliche Abweichung der Einzelwerte vom Mittelwert an.

- Keine Information:

  Durch die Quadrierung der Abweichungen bei der Varianz erhalten stärker abweichende Daten mehr gewicht.

- Anwendung:

  Intervallskalierte Daten

Formel: Standardabweichung

- Formel:

$$s = \sqrt{\frac{1}{N-1}\sum\left(x_i - \overline{x}\right)^2}$$

Streuungsmaße: Range

- Definition:

  Unter der Spannweite (Range) versteht man die Different zwischen dem größten und dem kleinsten Messwert.

- Formel:

  Range = $x_{max} - x_{min}$

- Information:

  Bandbreite der Daten

- Anwendung:

  Intervallskalierte Daten

Beispiel: Range

- Messwertreihe:
  10, 45, 46, 47, 52, 57, 58, 59, 59, 100

- Range = 100 - 10 = 90

Streuungsmaße: Quartilmaße

- Prozentränge:
  Prozentränge (cum f%) sind kumulierte Prozentwerte

- Quartilmaße:
  Dezildifferenz = Prozentrang 90 – Prozentrang 10
  Quartilabstand = Prozentrang 75 – Prozentrang 25
  Mittlerer Quartilabstand = Quartilabstand / 2

- Information:
  Anteilige Verteilung der Anzahl von Messwerten pro Merkmalsausprägung.

- Anwendung:
  Ordinal- und intervallskalierten Daten

Standardabweichung: Fehler und Fallen

- Bei der Berechnung von Varianz und Standardabweichung wird man leicht zu Fehlern verleitet. Zu beachten ist:

- Die Summe der Quadrate ist ungleich dem Quadrat der Summe.

$$\sum x_i^2 \neq \left(\sum x_i\right)^2$$

$$F\ddot{u}r\ n = 2\ z.B.\ gilt\ f\ddot{u}r\ alle\ x_1, x_2 \neq 0:$$

$$x_1^2 + x_2^2 \neq (x_1 + x_2)^2$$

$$x_1^2 + x_2^2 \neq x_1^2 + 2x_1 x_2 + x_2^2$$

- Daher:

1. Berechne erst die Quadrate aller Abweichungen
2. Und bilde dann erst die Summe

Übungen zu den Kennwerten und Streuungsmaßen

- Messwertreihe:
  5, 7, 12, 14, 17, 21, 26, 29, 31

- Berechne das arithmetische Mittel (Lösung: $\overline{x} = 18$)

- Berechne die Varianz (Lösung: $s^2 = 88.25$)

- Berechne die Standardabweichung (Lösung: $s = 9.39$)

Zusatz Streuungsmaße: Variabilitätskoeffizient

- Definition:

  Der Variabilitätskoeffizient (VK) bringt in relativer Form zum Ausdruck, wie viel Prozent des arithmetischen Mittels die Standardabweichung ausmacht.

- Formel:

  $VK = (s/\bar{x}) * 100$

- Information:

  Mehrere VKs verdeutlichen die Variabilität der Messwerte von Versuchsreihen.

- Anwendung:

  Verhältnisskalenniveau

Beispiel: VK

- Versuch:

  Bei N=100 Vpn wird die Anzahl der Treffer in einem Experiment am Reaktionsgerät ermittelt. Es soll überprüft werden, wie groß die Streuung der Trefferzahlen beim 1., 5. und 10. Versuch ist.

Versuchsreihe	Mittelwert (Treffer)	Standardabweichung	VK %
1. Versuch	13.85	4.75	34.3
5. Versuch	22.60	4.65	20.6
10. Versuch	24.50	3.90	15.9

- VKs zeigen, dass die Variabilität der individuellen Leistungen im Laufe des Trainings abnimmt. Die Streuung um den Mittelwert wird geringer.

Einschub: Normalverteilung

- Definition:

  Verteilungsfunktion für stetige Zufallsvariablen, d.h. eine Variable X kann unendlich viele Ausprägungen haben.

- Eigenschaften:

  - Glockenförmiger Verlauf
  - Symmetrischer Verlauf
  - Mo, Md, und Mittel fallen zusammen
  - Die Verteilung nähert sich asymptotisch der X-Achse
  - Zwischen den Wendepunkten befinden sich ca. 2/3 der Werte

Normalverteilung

- Annahme:

  Empirische Daten, wie Körpergrößen, Körpergewicht etc. bei großer Stichproben folgen häufig dieser Verteilung.

- Folgen:

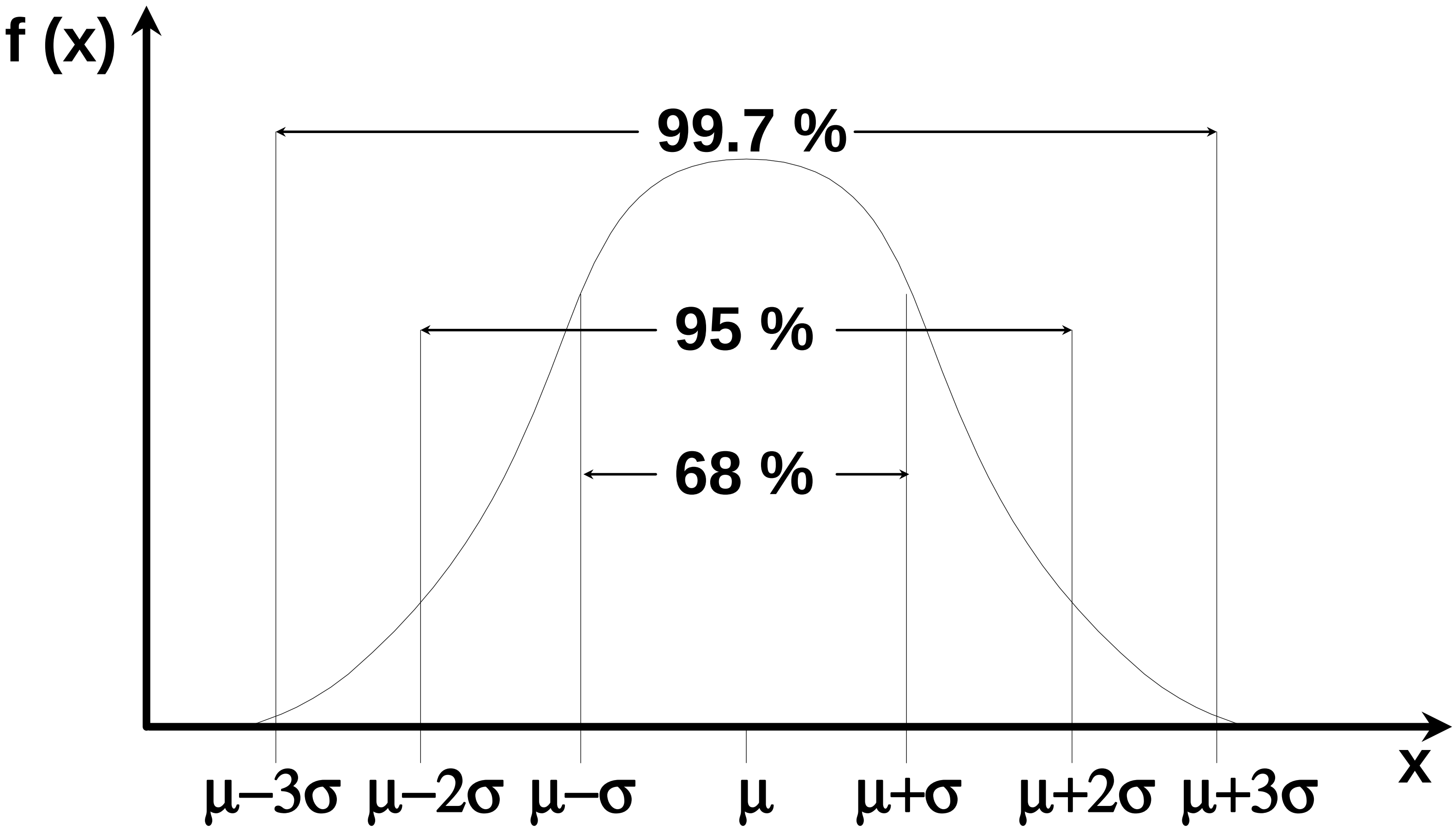
  Wegen der stetigen Zufallsvariablen können für einzelne Realisationen von X auch keine Wahrscheinlichkeiten des Auftretens mehr angegeben werden.

  Es lassen sich aber Wahrscheinlichkeiten dafür angeben, dass der Wert x in ein bestimmtes Intervall zwischen a und b fällt.

- Anwendungsbereich:

  In Abhängigkeit von dem Vorliegen einer Normalverteilung können verschiedene statistische Prüfverfahren verwendet werden.

Streuungsbereiche der Normalverteilung



Erläuterung der Graphik

- μ (mü) kennzeichnet den Mittelwert
- □ö(sigma) kennzeichnet die Standardabweichung.
- Der Wendepunkt ist $x = \mu +$ □□
- Die Gesamtfläche unter der Graphik ist 1, denn:

  Die gesamte Fläche zwischen der Kurve und der x-Achse entspricht gerade der Wahrscheinlichkeit, dass x einen Wert zwischen – unendlich und + unendlich annimmt. Dies entspricht dem sicheren Ereignis.
- Information:

  Man kann nun entscheiden, mit welcher Wahrscheinlichkeit ein Variablenwert in ein bestimmtes Intervall fällt.

Formmaße

- Neben der zentralen Tendenz und der Streuungsmaße lassen sich Häufigkeitsverteilungen auch nach ihrer Form charakterisieren.

- Funktion der Formmaße:

  Sie dienen der Überprüfung der Anpassung einer Verteilung an die Normalverteilung.

- Arten der Formmaße:

  1. Schiefe

  2. Exzess

Formmaße: Schiefe

- Definition:

  Die Schiefe bezieht sich auf die Symmetrie einer Verteilung.

- Schiefe = 0: es liegt eine Normalverteilung vor.

  Schiefe > 0: es liegt eine linksgipflige Verteilung vor

  Schiefe < 0: es liegt eine rechtsgipflige Verteilung vor.

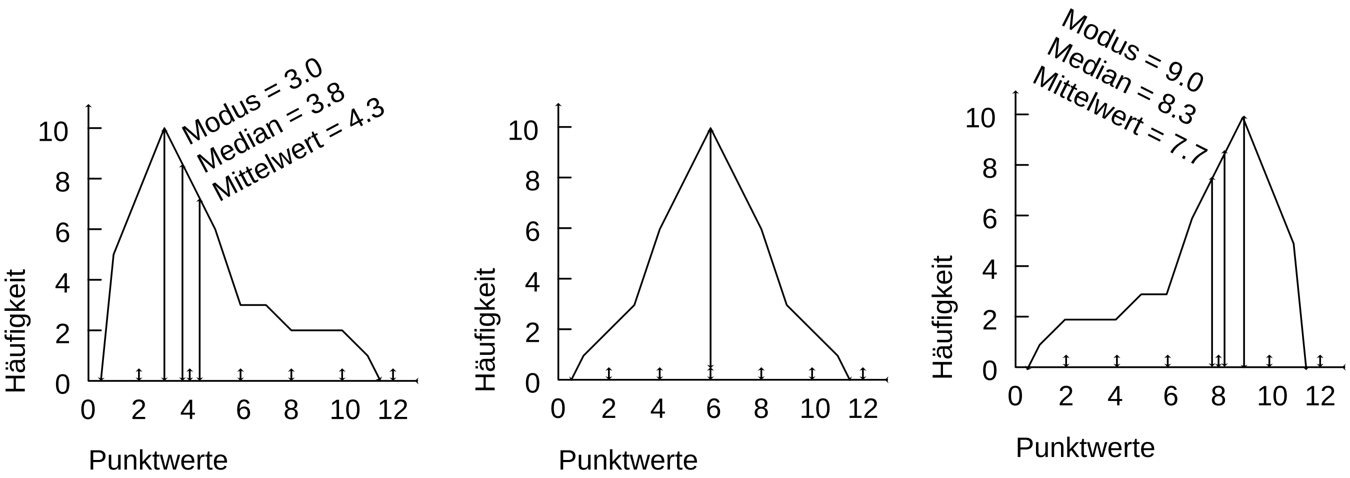
- Die Schiefe einer Kurve lässt sich auch durch die Maße der zentralen Tendenz zueinander charakterisieren:

  Bei symmetrischen Verteilungen sind Mo, Md und aM identisch.

  Bei rechtsschiefen Verteilungen gilt: Mo $\leq$ Md < aM

  Bei linksschiefen Verteilungen gilt: Mo $\geq$ Md > aM

Graphische Lösungen von Schiefen



rechtsschief,
linkssteil,
S>0

symmetrisch,
normal,
S=0

linksschief,
rechtssteil,
S<0

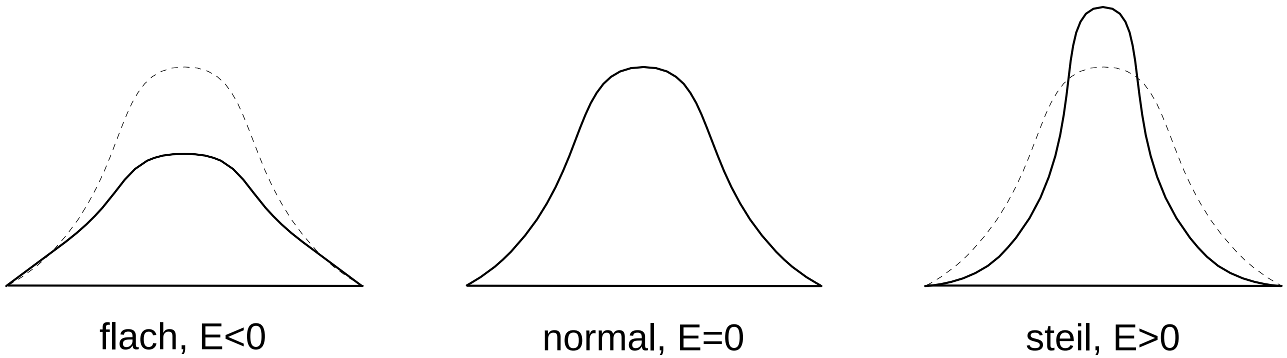
Formmaße: Exzeß

- Definition:
  Der Exzeß bezieht sich auf die Steilheit einer Verteilung.

- Exzeß = 0: es liegt eine Normalverteilung vor.
  Exzeß > 0: es liegt eine steilere Kurve als die Normalverteilung vor
  Exzeß < 0: es liegt eine flachere Kurve als die Normalverteilung vor.

- Steilere Kurven: relativ homogene Gruppe bzgl. des beobachteten Merkmals
  Flachere Kurven: relativ heterogene Gruppe bzgl. des beobachteten Merkmals

Graphische Lösungen von Exzessen



Vergleich bivariater Verteilungen

- Definition:

  Mit bivariaten Verteilungen werden Messwertreihen zweier Variablen bezeichnet, die miteinander in Beziehung gesetzt werden, z.B. die Sprunghöhe und die Laufzeit von einer Gruppe Studenten.

- Problem:

  Vergleichbarkeit verschiedener Messgrößen

- Lösung:

  1. Prozentränge
  2. z-Wert und Z-Wert

z-Wert und Z-Wert

- Vorteil:

  Die z-Werte haben gegenüber den Prozenträngen der Vorteil, dass sie weiter statisch verarbeitet werden können.

- Ziel:

  Vergleich mehrerer Verteilungen über die Standardnormierung.

- Nutzen:

  z-normierte Werte sind für verschiedene Stichproben vergleichbar und können auch, da sie im selben Messbereich variieren, addiert werden.

- Kennzeichen:

  z-normierte Verteilungen haben einen Mittelwert von 0 und eine Standardabweichung von 1. Die z-Werte variieren von -3 bis 3.

z-Wert

- Formel:

$$z_i = \frac{x_i - \overline{x}}{s_x}$$

- Beispiel:

- (N = 813)
- Jump and Reach-Test: aM_x = 34 cm; s = 11cm
- Handgrip-Test: aM_y = 43 kg; s = 13kg
- Versuchsperson x hat in beiden Fitnesstests den Wert 45

à z_x = 45 – 34 / 11 = 1.00; z_y = 45- 43 / 13 = 0.15

à Die relative Leistung im Jump and Reach-Test ist besser zu bewerten als die des Handgrip-Tests.

Z-Wert

- Definition:

  Z-normierte Verteilungen haben einen Mittelwert von 100 und eine Standardabweichung von 10.

- Vorteil:

  Die Standardwerte bewegen sich nur im positiven Zahlenbereich.

- Formel:

$$Z_i = 100 + 10 \frac{x_i - \overline{x}}{s_x} = 100 + 10\, z_i$$

- Beispiel:

  $Z_x = 100 + 10 \text{ x } 1.00 = 110$; $Z_y = 100 + 10 \text{ x } 0.15 = 101{,}5$