Lage- und Streuungsparameter

- Beziehen sich auf die Verteilung der Ausprägungen von intervall- und ratio-skalierten Variablen
- Versuchen, diese Verteilung durch Zahlen zu beschreiben, statt sie graphisch darzustellen
- Soweit die Berechnung ohne einen Computer erfolgt, ist es sinnvoll, die Daten zu gruppieren, d.h. eine Häufigkeitstabelle zu erstellen, mit der die Berechnungen vereinfacht werden können

Alter der Kursteilnehmer: Häufigkeitstabelle

Ausprägung	Häufigkeit
20	4
21	6
22	7
23	3
24	1
25	0
26	5
27	0
28	0
29	0
30	1
31	1

Lageparameter (=Mittelwerte)

- Beziehen sich auf die zentrale Tendenz einer Verteilung
- Berechnung nur dann sinnvoll, wenn die Verteilung tatsächlich eine zentrale Tendenz aufweist, d.h. möglichst eingipflig ist
- Vielzahl von Mittelwerten. Hier behandelt werden
  – Modus (=Modalwert)
  – Median
  – Arithmetisches Mittel (=„Durchschnitt“)

Modus

- Ist der Meßwert, der in der Verteilung am häufigsten vorkommt. Abkürzung x_{Mo}
- Im Beispiel: 22 Jahre
- Wenn zwei Werte gleich häufig vorkommen
  - arithmetisches Mittel bestimmen, soweit sinnvoll und möglich
  - ansonsten auf Angabe des Modus verzichten
- Kann unmittelbar aus der Verteilung bestimmt werden
- Ist unempfindlich gegenüber Extremwerten
- Kann für alle Skalenniveaus bestimmt werden
- Niedriger Informationsgehalt

Median

- Ist der Meßwert, der die aufsteigend angeordneten Meßwerte in der Mitte teilt
- Bei einer ungeraden Anzahl von Meßwerten gibt es genau einen Wert, der diese Bedingung erfüllt
- Bei einer geraden Anzahl von Werten wird das arithmetische Mittel aus den beiden mittleren Werten gebildet
- Berücksichtigt alle Meßwerte und ist robust gegen Ausreißer
- In diesem Fall (22+22)/2=22

$$\tilde{x}$$

20	20	20	20	21	21	21	21	21	21	22	22	22	22	22	22	22	23	23	23	24	26	26	26	26	26	30	31

Arithmetisches Mittel

- Ist der bekannteste Mittelwert
- Prinzip: alle Werte werden aufsummiert, dann wird durch die Zahl der Fälle geteilt
- In unserem Fall: 644/28=23
- Empfindlich gegen Ausreißer: Bleiben die ältesten beiden Kursteilnehmer unberücksichtigt, sinkt das arithmetische Mittel auf 22,42 Jahre. Modus und Median bleiben unbeeinflußt
- Nutzt das Maximum an Information, das in den Daten enthalten ist
  - Die Summe der Abweichungen vom arithmetischen Mittel ist null
  - Die Summe der quadrierten Abweichungen vom arithmetischen Mittel ist minimal

$$\bar{x}$$

Rechnen mit dem Summenzeichen

- In der Statistik wird häufig das Summenzeichen (Sigma) benutzt
- Dieses Zeichen symbolisiert eine Aufsummierung über alle Fälle
- Unter dem Summenzeichen steht meist i=1, was den Laufindex über die Fälle und dessen Startwerte repräsentiert
- Über dem Summenzeichen steht n, was bedeutet, das bis zum letzten (n-ten) Fall aufsummiert werden soll
- Beide Zusätze werden häufig weggelassen

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

Form der Verteilung

- Aus dem Verhältnis der drei Mittelwerte zueinander kann man auf die Form der Verteilung schließen:
- Linkssteile Verteilung
  - Die häufigste Ausprägung liegt weit links, deshalb ist der Modus der kleinste der drei Werte
  - Der Median, der die Verteilung halbiert, liegt in der Mitte und ist deshalb größer als der Modus
  - Das arithmetische Mittel reagiert auf die wenigen Ausreißer am rechten Rand der Verteilung und ist deshalb nochmals größer als der Median
- Rechtssteile Verteilung: Reihenfolge kehrt sich um
- Symmetrische Verteilung: alle drei Mittelwerte fallen zusammen

Streuungsparameter

- Mittelwerte informieren über die zentrale Tendenz einer Verteilung
- Streuungsparameter informieren darüber, wie homogen eine Verteilung ist:
  - im Falle einer absolut homogenen Verteilung sind alle Werte mit dem Mittelwert identisch, die Streuung ist gleich null
  - je heterogener die Verteilung, desto stärker streuen die Meßwerte um den Mittelwert
  - im Falle einer zweigipfeligen Verteilung unterscheiden sich die die Meßwerte extrem vom Mittelwert
- Zur Beschreibung der Streuung existieren verschiedene Maße. Im Kurs behandelt werden
  - Variationsweite (auch Spannweite, range)
  - Varianz und die daraus abgeleitete Standardabweichung

Variationsweite

- Die Variationsweite (V) ist einfach die Differenz zwischen dem größten und dem kleinsten Meßwert
- In unserem Fall: 31-20= 11 Jahre
- Kann direkt aus der Verteilung abgelesen werden
- Einfache Interpretation
- Basiert auf lediglich zwei Meßwerten
  - Wertet nur einen Bruchteil der in den Daten enthaltenen Information aus
  - Extrem anfällig für Ausreißer

Varianz

- Berücksichtigt alle Meßwerte
- Entspricht der durchschnittlichen quadrierten Abweichung der Meßwerte von ihrem Mittelwert
- Abkürzung s^2
- In unserem Fall: 8,07
- Warum werden Abweichungen quadriert?
  - Damit das Vorzeichen verschwindet (Summe der einfachen Abweichungen wäre null)
  - Damit große Abweichungen vom Mittelwert stärker gewichtet werden
- Problem: Die Maßeinheit geht verloren bzw. kann nicht interpretiert werden (Quadratjahre)

Varianz II

- Berechnung der Varianz
  - Mittelwert berechnen
  - Für jeden Meßwert Differenz zum Mittelwert berechnen
  - Differenz quadrieren und aufaddieren
  - Summe durch Zahl der Fälle teilen
- s^2 = SAQ (Summe der Abweichungs-Quadrate)/n
- Die Varianz einer Stichprobe unterschätzt die Varianz der Grundgesamtheit um den Faktor (n-1)/n
  - Deshalb muß die Stichprobenvarianz mit dem Faktor n/(n-1) multipliziert werden, wenn eine Schätzung für die Grundgesamtheit gewünscht wird
  - Das führt zur Vereinfachung SAQ/n*n/(n-1) = SAQ/n-1
- Diese Korrektur wird *nur* dann vorgenommen, wenn eine Schätzung erforderlich ist

Varianz III

ID	Meßwert		arith. Mittel		Abweichung	AQ	Varianz
2	20	-	23	=	-3	9	
17	26	-	23	=	3	9	
10	24	-	23	=	1	1	
...							
						226	226/28=8,07

Standardabweichung

- Die Standardabweichung (s) wird direkt aus der Varianz berechnet, indem aus der Varianz die Quadratwurzel gezogen wird
- Wird teilweise mit S.D. (standard deviation) abgekürzt
- Vorteil: Liegt wieder in der ursprünglichen Einheit vor
- In unserem Fall: 2,84 Jahre
- Interpretation trotzdem nicht ganz einfach
  - s ist *nicht* die mittlere Abweichung vom Durchschnitt
  - s entspricht *auch nicht* dem Mittel der Beträge der einzelnen Abweichungen

z-Transformation

- Meßwerte aus unterschiedlichen Verteilungen (bspw. unterschiedliche Statistikklausuren) sollen miteinander verglichen werden
- Dazu wird zunächst von jedem Meßwert der Mittelwert seiner Verteilung abgezogen (Zentrierung), um Niveauunterschiede auszugleichen
- Die unterschiedliche Streuung (z.B. durch unterschiedliche Maßeinheiten) der Verteilungen wird kompensiert, indem die zentrierten Werte durch die Standardabweichung ihrer Verteilung geteilt werden

z-Werte

- Sind das Ergebnis der z-Transformation
- Repräsentieren die Abweichung des ursprünglichen Meßwertes vom Mittelwert seiner Verteilung
- Diese Abweichung wird in Standardabweichungen der ursprünglichen Verteilung ausgedrückt
- Damit informieren z-Werte skalenunabhängig über die Lage des ursprünglichen Meßwertes innerhalb seiner Verteilung
- z-Werte weisen ihrerseits einen Mittelwert von 0 und eine Standardabweichung von 1 auf