

Grundlagen sportwissenschaftlicher Forschung

Deskriptive Statistik

Dr. Jan-Peter Brückner

jpbrueckner@email.uni-kiel.de



R.216
Tel. 880 4717

Deskriptive Statistik - Ziele

- Beschreiben der Daten
- Zusammenfassen der Daten
- Überblick über die Daten
- Datenanalyse
- Fehlerprüfung/Plausibilitätsprüfung

Beispiel

*Eine Stichprobe von $N=10$ Schüler absolviert einen Fitness-Test.
Die Leistung X wird mit 0 (schlechte Fitness) bis 10 Punkten (sehr gute Fitness) bewertet*

Rohdaten (Beispiel 1):

j	1	2	3	4	5	6	7	8	9	10
x_j	3	9	2	10	9	9	4	3	3	8

sortierte Rohdaten:

i	1	2	3	4	5	6	7	8	9	10
x_i	2	3	3	3	4	8	9	9	9	10

Deskriptive Statistik - Kennwerte

- Häufigkeiten und Häufigkeitsverteilung
- Maße der zentralen Tendenz
- Streuungsmaße
- Zusammenhangsmaße
- Unterschiedsmaße

Häufigkeit

Absolute Häufigkeit

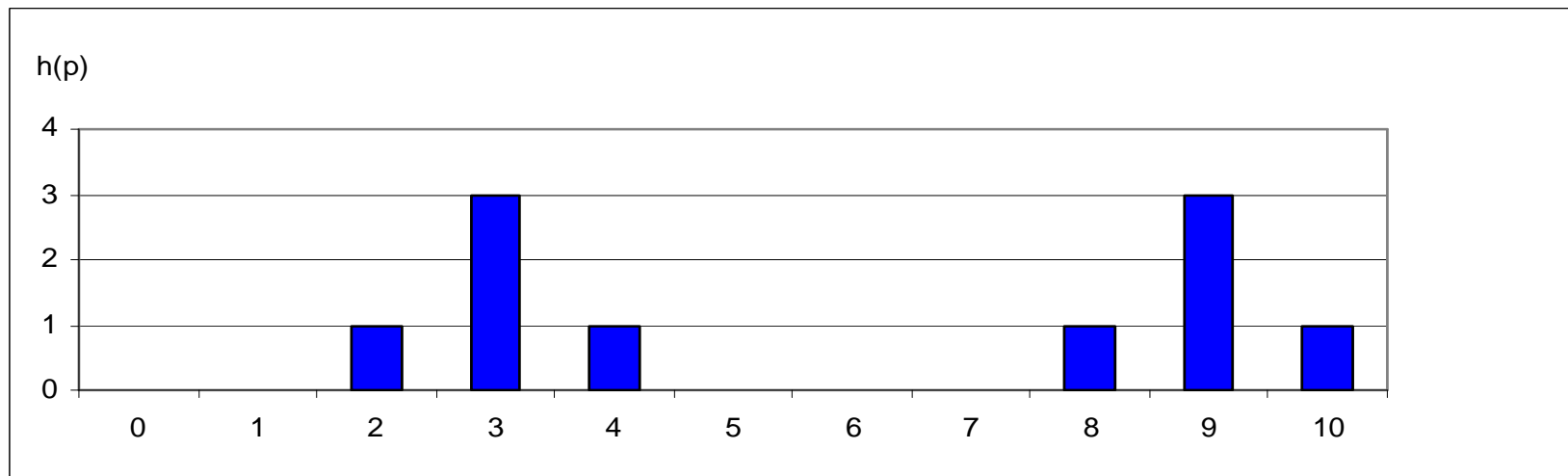
$h(a)$ Häufigkeit für $x_i = a$ ($i = 1, 2, 3, \dots, N$)

Relative Häufigkeit

$$f(a) = \frac{h(a)}{N}$$

Häufigkeitsverteilung

p	0	1	2	3	4	5	6	7	8	9	10
$h(p)$			1	3	1				1	3	1
$f(p)$	0	0	0,1	0,3	0,1	0	0	0	0,1	0,3	0,1



p : Punkte im Fitness-Test

Maße der zentralen Tendenz

Modalwert, Median, Arithmetisches Mittel: Wann welcher Wert?

Arithmetisches Mittel:

- *bei Intervallskalierung*
- *symmetrische Verteilung*

bei Verallgemeinerungen von der Stichprobe auf die Grundgesamtheit • bei Verallgemeinerungen von der Stichprobe auf die Grundgesamtheit

Median:

- *bei Ordinalskalierung*
- *bei Extremwerten*
- *bei Abweichungen von der Normalverteilung*

Modalwert (Modus):

- *bei Ordinalskalierung*
- *bei mehrgipfligen Verteilungen*
- *grober Überblick*

Maße der zentralen Tendenz

- *Arithmetisches Mittel, Mittelwert*
- *Median*
- *Modalwert (Modus)*

Summenzeichen

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$$

Arithmetisches Mittel - Mittelwert

$$M_x = \frac{1}{N} \sum_{i=1}^N x_i$$

- Auch: AM oder \bar{x}
- erfordert Intervallskalenniveau
- keine Aussage über die Verteilung/Abweichung der Einzelwerte

Beispiel:

i	1	2	3	4	5	6	7	8	9	10
x_i	2	3	3	3	4	8	9	9	9	10

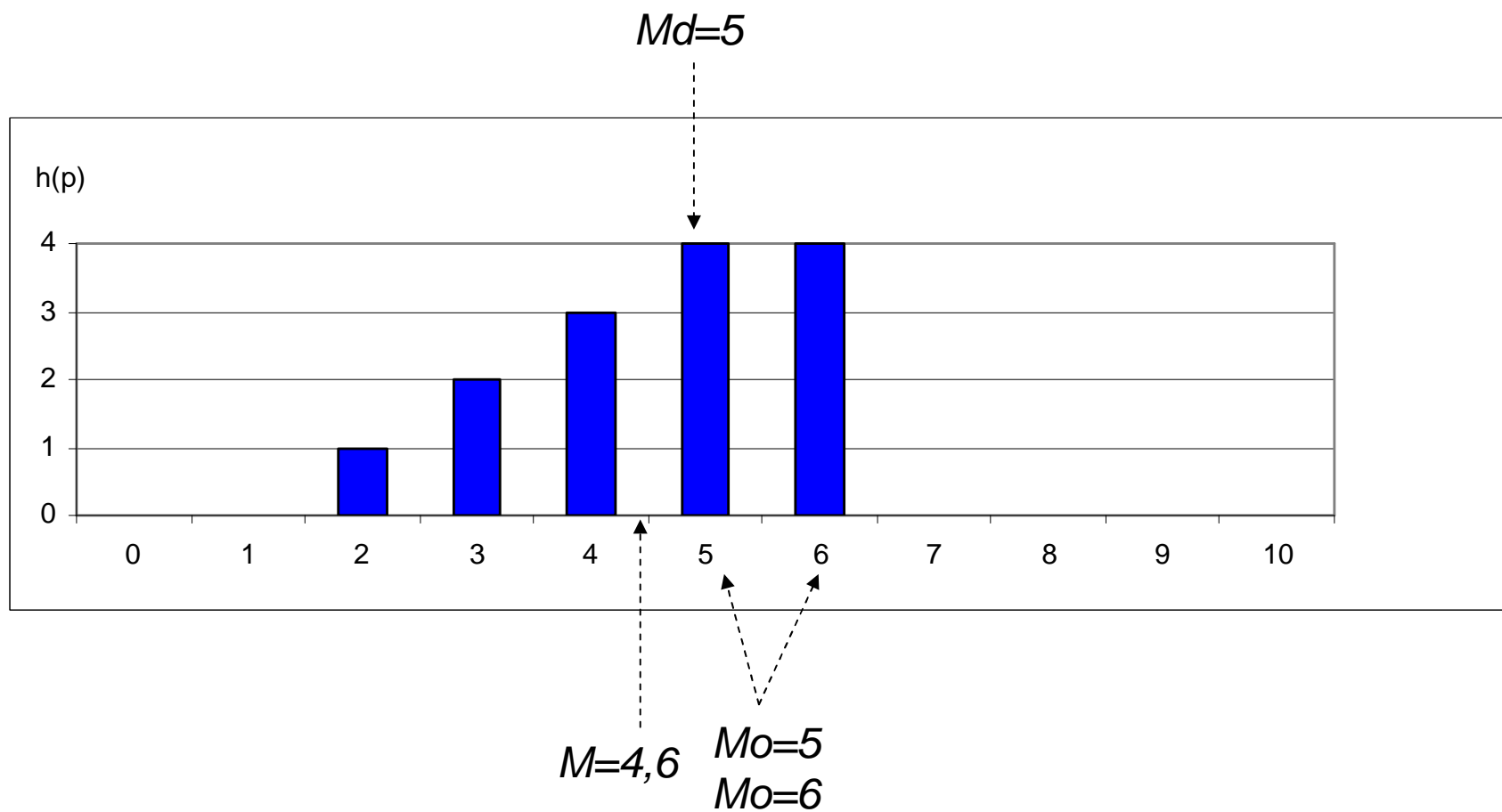
$$M_x = \frac{1}{10} * 60 = 6$$

Median - Modus

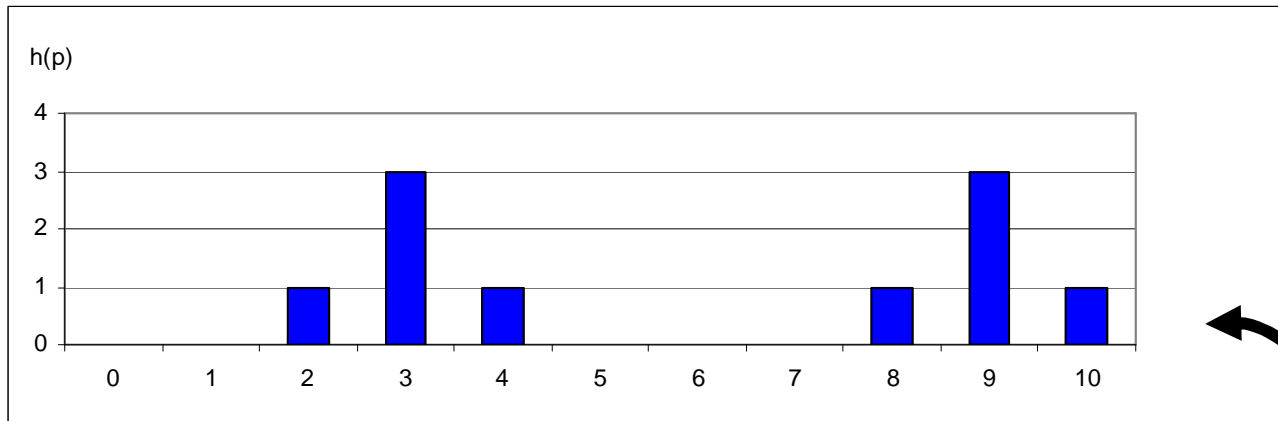
- **Median Md:**
50% der Werte liegen über,
50% der Werte unter dem Median.
- **Modus Mo (auch: Mod):**
Häufigster Wert der Verteilung

Wo liegen Median und Modus im Verhältnis zum Mittelwert???

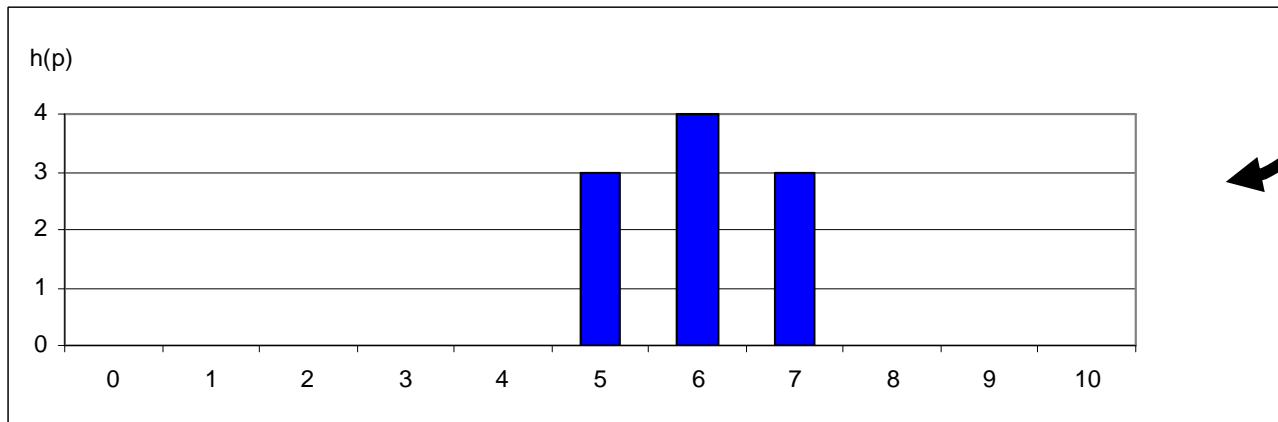
→ Symmetrie der Verteilung (Schiefe)



Beispiel 1:



Beispiel 2:



$$M = 6$$

Streuungsmaße

- Range
- Perzentilen
- Varianz
- Standardabweichung
- Variationskoeffizient

Spannweite

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

- erfordert Intervallskalenniveau
- Anfällig gegenüber Ausreißern
- Insbesondere angeben bei Personendaten (Alter,...)

Beispiel 1:

$$x_{\max} = 10$$

$$x_{\min} = 2$$

$$\textit{Range} = 10 - 2 = 8$$

Beispiel 2:

$$x_{\max} = 7$$

$$x_{\min} = 5$$

$$\textit{Range} = 7 - 5 = 2$$

Varianz

$$SD_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - M_x)^2$$

*Summe der Abweichungs-
quadrate*

- Auch: SD^2 , s^2 , $Var(X)$
- SD^2 ist immer positiv
- Extremwerte haben einen besonders großen Einfluss

Standardabweichung

$$SD_x = \sqrt{SD_x^2}$$

$$SD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - M_x)^2}$$

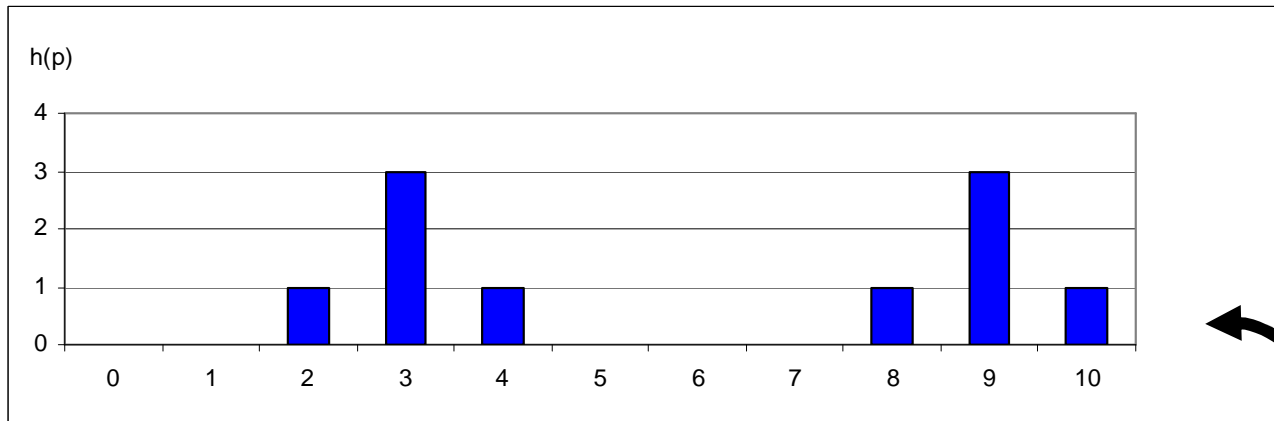
- Auch: s
- Extremwerte haben einen besonders großen Einfluss
- *Achtung:* In Statistiksoftware wird häufig durch **N-1** geteilt!
Diese Berechnung ist nur sinnvoll, wenn die Stichprobe zur Schätzung der Varianz/Standardabweichung auf Populations-ebene genutzt wird!

Beispiel 1:

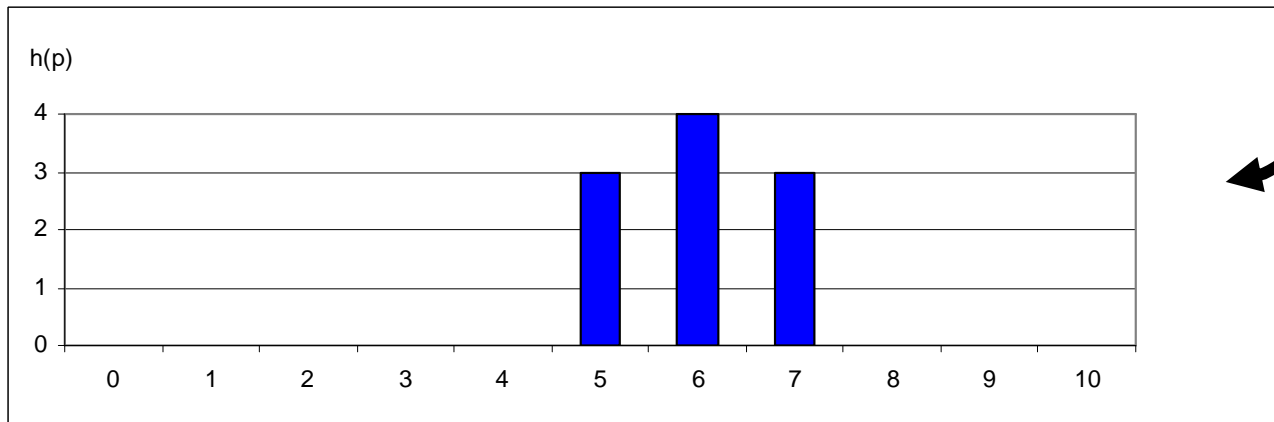
i	1	2	3	4	5	6	7	8	9	10	Summe	
x_i	2	3	3	3	4	8	9	9	9	10	60	-----> $M = 6$
$x_i - M$	-4	-3	-3	-3	-2	2	3	3	3	4	0	
$(x_i - M)^2$	16	9	9	9	4	4	9	9	9	16	94	--> $\text{Var}(X) = 9,4$

$$SD_X = \sqrt{\text{Var}(X)} = \sqrt{9,4} \approx 3,1$$

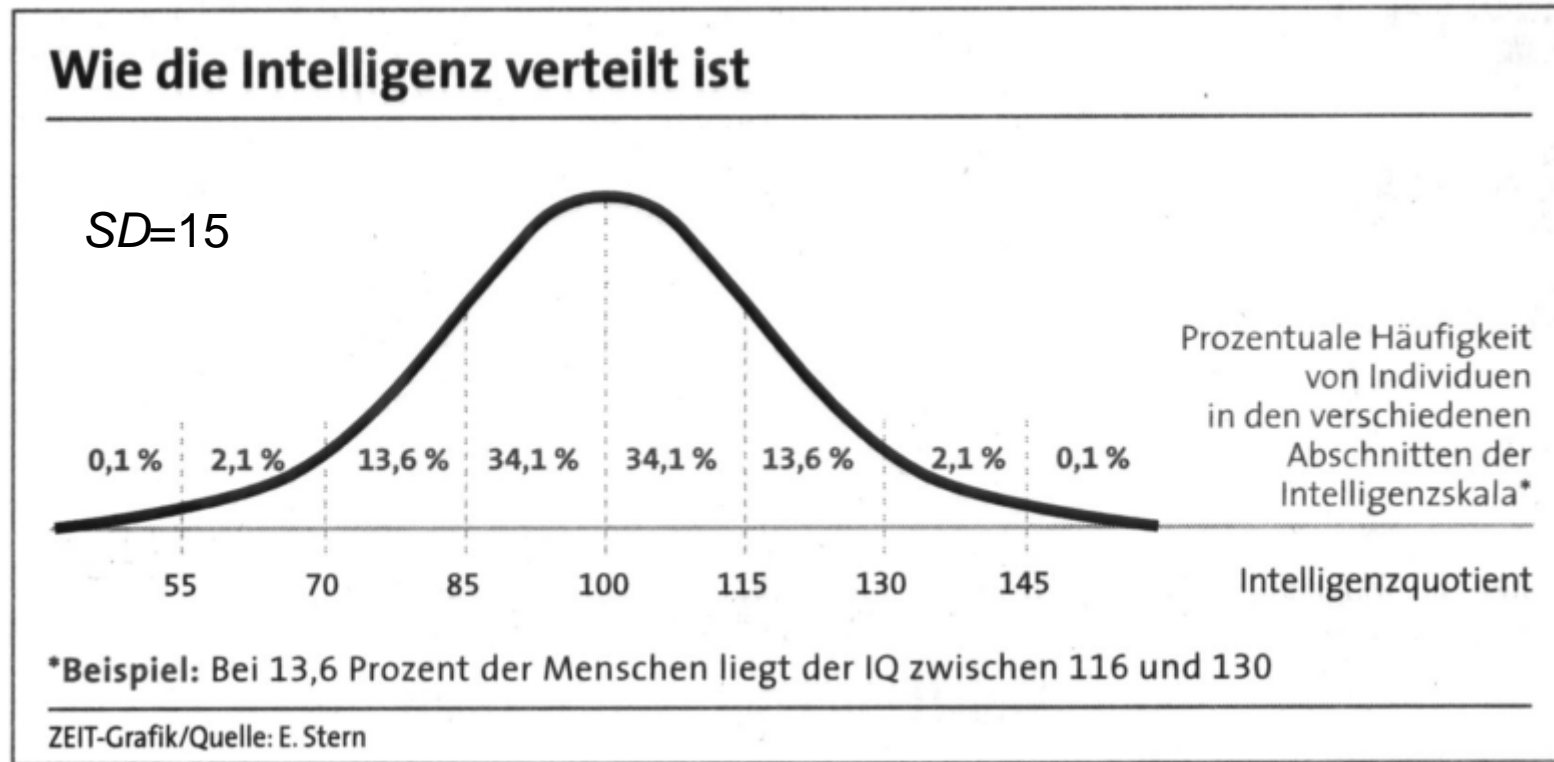
Beispiel 1:

 $SD=3,1$

Beispiel 2:

 $M = 6$ $SD=0,8$

Normalverteilung



$2*SD$

$4*SD$

Variationskoeffizient

Ist die Standardabweichung hoch oder niedrig?
→ Standardisierung am Mittelwert:

$$V(X) = \frac{SD_X}{M_X}$$

→ Vergleichen der Streuungen von unterschiedlich skalierten Variablen

z-Standardisierung

Wie stark weicht ein Wert x_i im Vergleich zu anderen Werten vom Mittelwert ab?

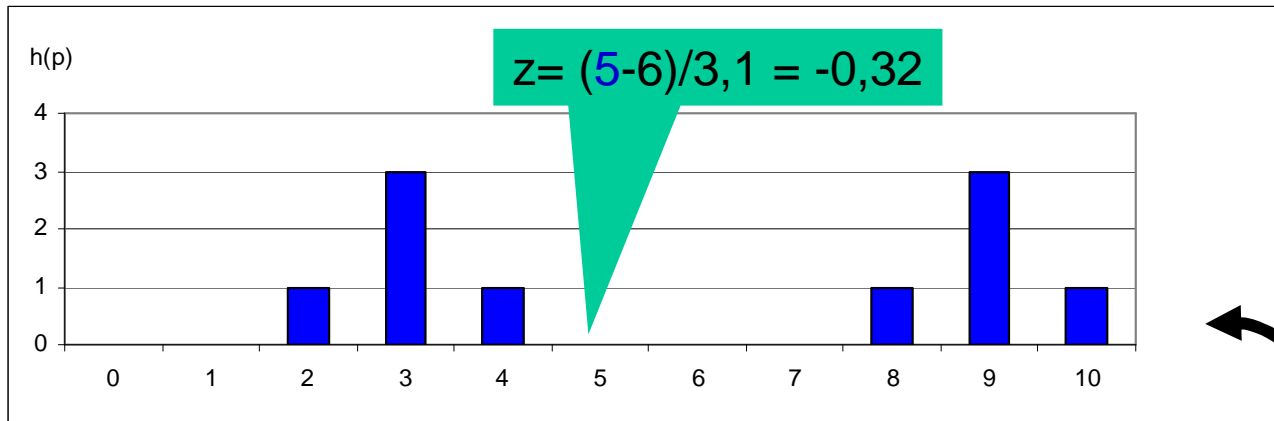
→ Standardisierung an Mittelwert und Standardabweichung:

$$z_i = \frac{x_i - M_X}{SD_X}$$

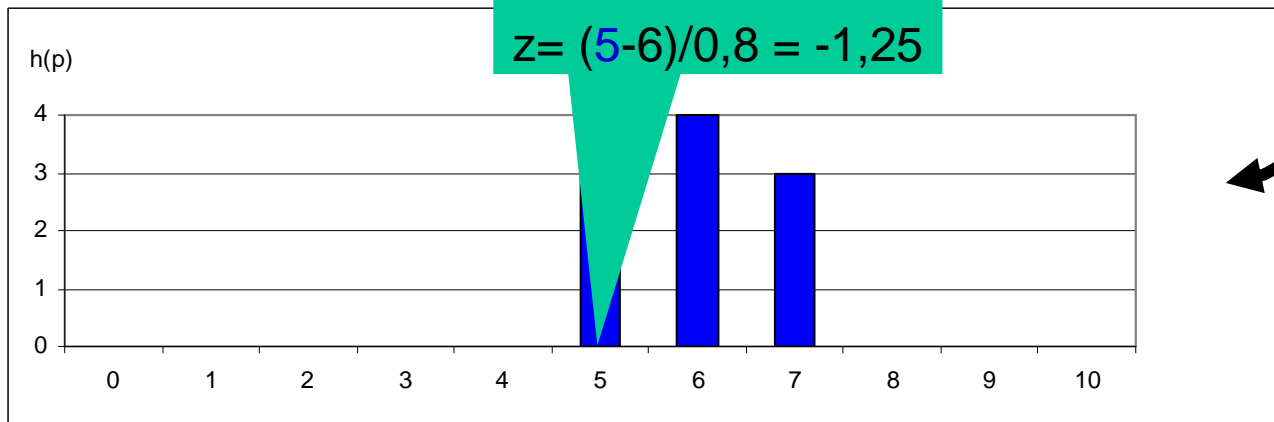
→ Vergleich verschiedener Werte x_i und x_j (Person i mit Person j)

→ Vergleich der Merkmalsausprägungen x_i und y_i einer Person i auf verschiedenen Variablen X und Y

Beispiel 1:

 $SD=3,1$ $M = 6$

Beispiel 2:

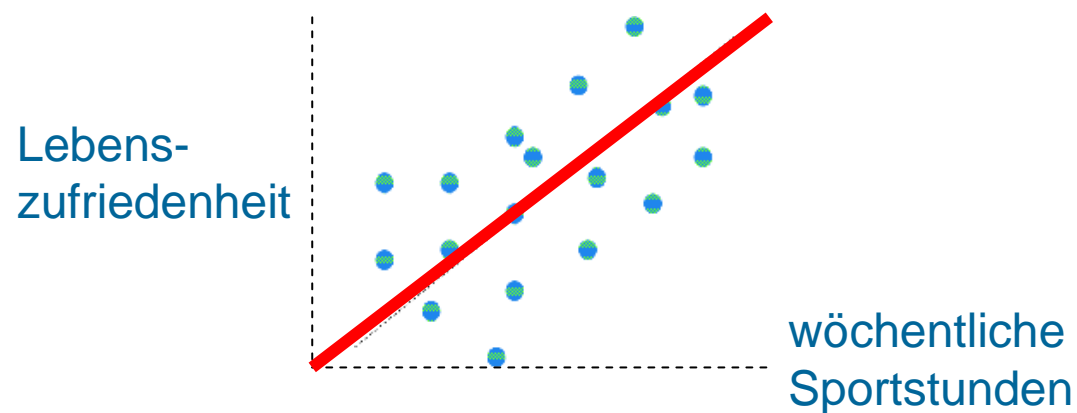
 $SD=0,8$

Kovarianz

„Gemeinsame Varianz“ zweier Variablen X und Y:

In wiefern ist X hoch ausgeprägt, wenn Y hoch (niedrig) ausgeprägt ist?

Beispiel aus der Einheit „Untersuchungsplanung“:

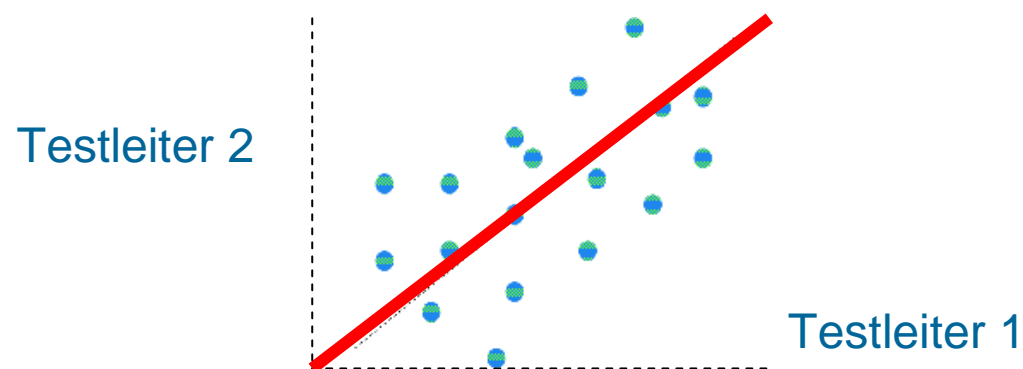


Kovarianz

„Gemeinsame Varianz“ zweier Variablen X und Y:

In wiefern ist X hoch ausgeprägt, wenn Y hoch (niedrig) ausgeprägt ist?

Beispiel zur Objektivität eines Tests:

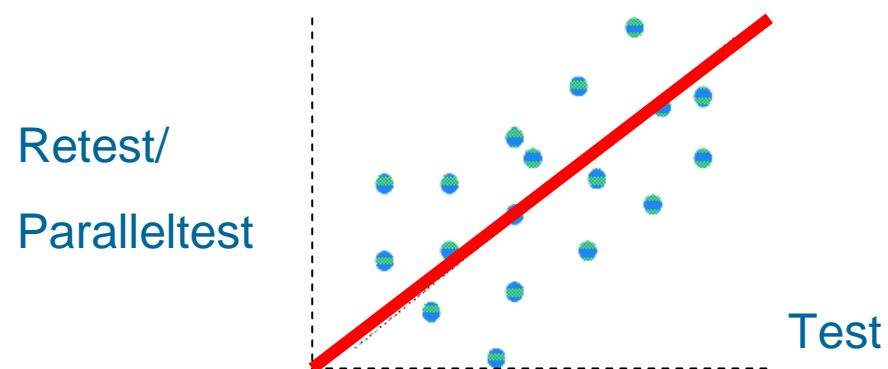


Kovarianz

„Gemeinsame Varianz“ zweier Variablen X und Y:

In wiefern ist X hoch ausgeprägt, wenn Y hoch (niedrig) ausgeprägt ist?

Beispiel zur Reliabilität eines Tests:

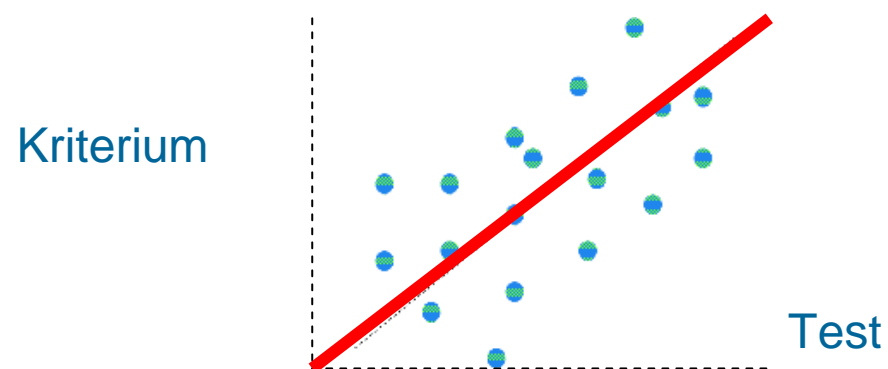


Kovarianz

„Gemeinsame Varianz“ zweier Variablen X und Y:

In wiefern ist X hoch ausgeprägt, wenn Y hoch (niedrig) ausgeprägt ist?

Beispiel zur Validität eines Tests:



Kovarianz

„Gemeinsame Varianz“ zweier Variablen X und Y:

$$COV(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - M_x)(y_i - M_y)$$

- $COV(X, Y)$ ist von $Var(X)$ und $Var(Y)$ abhängig
- Interpretation nur unter Berücksichtigung von $Var(X)$ und $Var(Y)$ möglich

Korrelationskoeffizient

Standardisierung der Kovarianz an der Varianz von X und Y:

$$r_{X,Y} = \frac{COV(X,Y)}{\sqrt{VAR(X)} * \sqrt{VAR(Y)}} = \frac{COV(X,Y)}{SD_X * SD_Y}$$

- Auch: r oder $Kor(X, Y)$
- Wertebereich: $-1 \leq r \leq 1$
- $r > 0$: positiver Zusammenhang
- $r = 0$: kein Zusammenhang
- $r < 0$: negativer Zusammenhang
- Voraussetzung: Intervallskalenniveau

Korrelationskoeffizient

$$r_{X,Y} = KOR(X,Y) = \frac{COV(X,Y)}{\sqrt{VAR(X)} * \sqrt{VAR(Y)}}$$

$$r_{X,Y} = KOR(X,Y) = \frac{COV(X,Y)}{SD_X * SD_Y}$$

Korrelationskoeffizient

$$r_{X,Y} = \frac{COV(X,Y)}{\sqrt{VAR(X)} * \sqrt{VAR(Y)}}$$

$$\rightarrow r_{X,Y} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - M_x)^2} * \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - M_y)^2}}$$

$$\rightarrow r_{X,Y} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^N (x_i - M_x)^2} * \sqrt{\sum_{i=1}^N (y_i - M_y)^2}}$$

Berechnung des Korrelationskoeffizienten

i	X	Y	$x_i - M$	$y_i - M$	$(x_i - M)^2$	$(y_i - M)^2$	$(x_i - M) \cdot (y_i - M)$
	Lebens- zufrieden- heit	wöch. Sport- stunden					
1	1	3	-2,00	-1,17	4,00	1,36	2,33
2	2	3	-1,00	-1,17	1,00	1,36	1,17
3	2	4	-1,00	-0,17	1,00	0,03	0,17
4	4	5	1,00	0,83	1,00	0,69	0,83
5	4	4	1,00	-0,17	1,00	0,03	-0,17
6	5	6	2,00	1,83	4,00	3,36	3,67
Summe	18	25			12,00	6,83	8,00
M	3,0	4,2			2,00	1,14	1,33
Wurzel					1,41	1,07	
					SD	SD	Cov

$$r = \frac{1,33}{1,41 \cdot 1,07} = 0,88$$

Testgütekriterien

Hauptgütekriterien	Nebengütekriterien
<p>Objektivität (anwenderunabhängig)</p> <ul style="list-style-type: none">• Durchführung• Auswertung• Interpretation	<p>Normierung</p> <p>Vergleichbarkeit</p> <p>Nicht-Verfälschbarkeit</p>
<p>Reliabilität (zuverlässig)</p> <ul style="list-style-type: none">• Paralleltest• Retest• Innere Konsistenz (Testhalbierung)	<p>Ökonomie</p> <p>Nützlichkeit</p> <p>Zumutbarkeit</p>
<p>Validität (gültig)</p> <ul style="list-style-type: none">• inhaltlich / Experten• Konstrukt• kriterienbezogen	<p>Fairness</p>

Beurteilung von Testkennwerten

	niedrig	mittel	hoch
Schwierigkeit	>.80	.80-.20	<.20
Trennschärfe (korrigiert)	<.30	.30-.50	>.50
Objektivität (Auswerter)	<.60	.60-.90	>.90
Reliabilität	<.80	.80-.90	>.90
Validität (unkorrigiert)	<.40	.40-.60	>.60
Eichstichprobe	$N < 150$	$150 < N < 300$	$N > 300$

AUFGABE

N=5 Schüler werden zu Ihren *Sportnoten* und zur *Einstellung zum Sport* (0: sehr negative Einstellung; 5: sehr positive Einstellung) befragt:

i	1	2	3	4	5
Sportnote	1	2	2	3	4
Einstellung	4	5	3	2	2

Berechne jeweils M, SD, Var, Range und Md sowie Kov und r für den Zusammenhang der beiden Variablen! (Einschl. Rechenweg!)

Abgabe bis Di, 05.06., 10:15 h

im Fach von J.-P. Brückner

mit Angabe des Namens,

und des Tutors und Tutoriums (rechts oben)