Anmerkungen zur Vorlesung Statistik vom 18.10.2016

Inhalt:

- Streuungsparameter: SP, IQ, MQA, Standardabweichung
- Konzentration: Lorenzkurve und Gini-Koeffizient
- weitere Konzentrationsmaße
- Klausur SS2016: Aufgabe 1 (wird am 25.10. gelöst)

Aufgabensammlung Statistik:

04.10.2016
Hausaufgabe
A1 - A7

11.10.2016
Hausaufgabe
A8 - A13
A10: ohne
Boxplot

18.10.2016
Hausaufgabe
Boxplot in A10
A14 - A20

F 53

Warum quadratische Abweichung:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}x_i}_{=\bar{x}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}\bar{x}}_{=n\cdot\bar{x}} = 0 \quad //$$

s^2 ist def. als:

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

weil:

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

$$= \frac{1}{n}\sum x_i^2 - \frac{1}{n}\cdot 2\cdot\bar{x}\sum_{i=1}^{n}x_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2$$

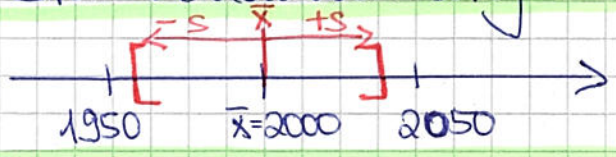
$$= \frac{1}{n}\sum x_i^2 - 2\cdot\bar{x}\cdot\bar{x} + \frac{1}{n}\cdot n\cdot\bar{x}^2$$

$$= \frac{1}{n}\sum x_i^2 - \bar{x}^2 \quad /\square$$

F54

Zur Standardabweichung s:

Fall a)



Maß für die Streuung um den Mittelwert

Vorteil Variationskoeff. & Maßstabsunabhängigkeit

→ Vergleichbarkeit versch. Merkmale

Taschenrechnerfunktionen!

(Ich: CASIO fx-85MS)

gerne auch Nachfolgemodell (≈15€)

(Vorteil: Datenliste)

- Beginnen Sie die Dateneingabe immer mit der Tastenfolge [SHIFT] [CLR] [1] (Scl) [=], um den statistischen Speicher zu löschen.
- Geben Sie die Daten unter Verwendung der folgenden Tastenfolge ein.
  <x-Datenwert> [DT] [M+] (≙ Dateneingabe)
- Die Eingabedaten werden verwendet, um die Werte für n, Σx, Σx^2, $\bar{x}$, σ_n und σ_{n-1} zu berechnen, die Sie dann unter Verwendung der folgenden Tastenbetätigungen aufrufen können.

Um diesen Wert aufzurufen:	Führen Sie diese Tastenoperation aus:
Σx^2	[SHIFT] [S-SUM] [1]
Σx	[SHIFT] [S-SUM] [2]
n	[SHIFT] [S-SUM] [3]
$\bar{x}$	[SHIFT] [S-VAR] [1]
σ_n	[SHIFT] [S-VAR] [2]
σ_{n-1}	[SHIFT] [S-VAR] [3]

→ $\uparrow^2 = s^2$

- **Beispiel:** Zu berechnen sind σ_{n-1}, σ_n, $\bar{x}$, n, Σx und Σx^2 für die folgenden Daten: 55, 54, 51, 55, 53, 53, 54, 52

  In dem SD-Modus:

  [SHIFT] [CLR] [1] (Scl) [=] (Stat clear)

  55 [DT] n= SD 1.

  Mit jedem Drücken der [DT]-Taste für die Registrierung Ihrer Eingabe, wird die Anzahl der bis zu diesem Zeitpunkt eingegebenen Daten am Display angezeigt (n-Wert).

  54 [DT] 51 [DT] 55 [DT]
  53 [DT] [DT] 54 [DT] 52 [DT]

Stichproben-Standardabweichung (σ_{n-1}) = *1,407885953* [SHIFT] [S-VAR] [3] [=]

Gesamtheits-Standardabweichung (σ_n) = *1,316956719* [SHIFT] [S-VAR] [2] [=]

Arithmetischer Mittelwert ($\bar{x}$) = *53,375* [SHIFT] [S-VAR] [1] [=]

Anzahl der Daten (n) = *8* [SHIFT] [S-SUM] [3] [=]

Summe der Werte (Σx) = *427* [SHIFT] [S-SUM] [2] [=]

Quadratsumme der Werte (Σx^2) = *22805* [SHIFT] [S-SUM] [1] [=]

Vorsichtsmaßnahmen bei der Dateneingabe

- Mit [DT] [DT] wird der gleiche Datenwert zwei Mal eingegeben.

-S22-

■ Koordinatenumwandlung (Pol (x, y), Rec (r, θ))

- Die Rechenergebnisse werden automatisch den Variablen E und F zugeordnet.
- **Beispiel 1:** Die Polarkoordinaten (r=2, θ=60°) sind in die kartesischen Koordinaten (x, y) (Deg) umzuwandeln.

  x = *1* [SHIFT] [Rec] 2 [,] 60 [)] [=]

  y = *1,732050808* [RCL] [F]

- Die Tastenfolge [RCL] [E] oder [RCL] [F] zeigt den Wert für x bzw. y an.
- **Beispiel 2:** Die kartesischen Koordinaten (1, $\sqrt{3}$) sind in die Polarkoordinaten (r, θ) (Rad) umzuwandeln.

  r = *2* [Pol] 1 [,] [√] 3 [)] [=]

  θ = *1,047197551* [RCL] [F]

- Die Tastenfolge [RCL] [E] oder [RCL] [F] zeigt den Wert für r bzw. θ an.

■ Berechnungen mit technischer Schreibweise

- **Beispiel 1:** Umzuwandeln sind 56.088 Meter in Kilometer.

  → *56,088* ×10³ (km) 56088 [=] [ENG]

- **Beispiel 2:** Umzuwandeln sind 0,08125 Gramm in Milligramm.

  → *81,25* ×10⁻³ (mg) 0.08125 [=] [ENG]

Statistische Rechnungen SD REG

Standardabweichung SD

Verwenden Sie die [MODE]-Taste, um den SD-Modus aufzurufen, wenn Sie statistische Rechnungen mit der Standardabweichung ausführen möchten.

SD ..... [MODE] [2]

- In dem SD-Modus und in dem REG-Modus arbeitet die [M+]-Taste als [DT]-Taste.

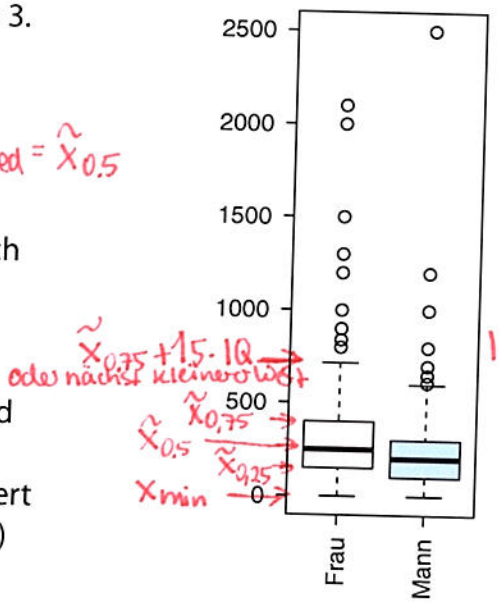
zur Dateneingabe

-S21-

Lage und Streuung als Grafik: Boxplot

- Graphische Darstellung von Lage und Streuung
- Box: Oberer/Unterer Rand: 3. bzw. 1. Quartil ($\tilde{x}_{0,75}$ bzw. $\tilde{x}_{0,25}$),
- Linie in Mitte: Median $x_{Med} = \tilde{x}_{0.5}$
- Whiskers: Länge: Max./Min Wert, aber beschränkt durch das 1,5-fache des Quartilsabstands (falls größter/kleinster Wert größeren/kleineren Abstand von Box: Länge Whiskers durch größten/kleinsten Wert innerhalb dieser Schranken)
- Ausreißer: Alle Objekte außerhalb der Whisker-Grenzen

```
boxplot(AusgSchuhe ~ Geschlecht,
        col=c("mistyrose", "lightblue"),
        data=MyData, main="", las=2)
```



Ausgaben für Schuhe

j	1	2	3	4	5	6	7	8
a_j	70	150	160	200	260	340	400	550
$h(a_j)$	1	2	1	2	1	1	1	1

$n = 10$

1) x_{MOD} = nicht eindeutig

$x_{MED} = \tilde{x}_{0.5} = \frac{1}{2}(x_{[5]} + x_{[6]}) = \frac{1}{2} \cdot (200 + 200) = 200$

$10 \cdot 0.5 = 5 \in \mathbb{N}_0$ $\quad \| \; n \cdot p \quad \| \; n \cdot p + 1$

$\bar{x} = \frac{1}{n} \cdot \sum_j a_j \cdot h(a_j) = \frac{1}{10} \cdot (70 \cdot 1 + 150 \cdot 2 + \dots + 550 \cdot 1)$

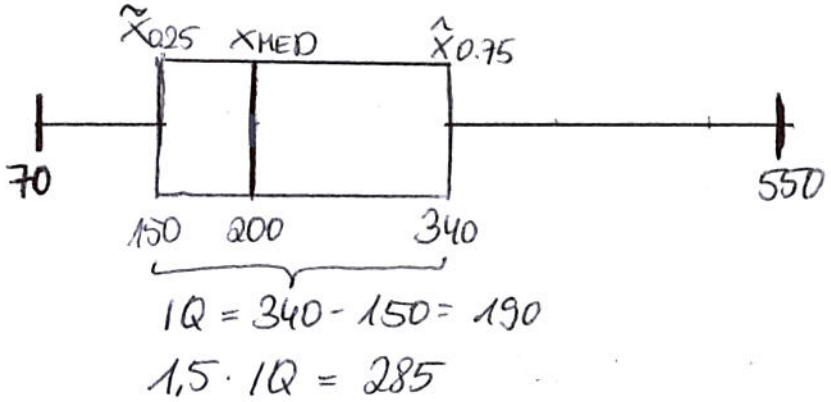
$= 2480/10 = 248$

2) $\tilde{x}_{0.25} = x_{\lceil 2{,}5 \rceil} = x_{[3]} = 150$

$10 \cdot 0{,}25 = 2{,}5 \notin \mathbb{N}_0$

$\hat{x}_{0.75} = x_{\lceil 7{,}5 \rceil} = x_{[8]} = 340$

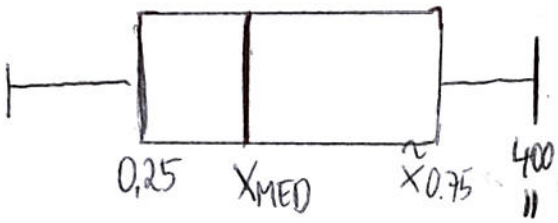
3)



$IQ = 340 - 150 = 190$

$1{,}5 \cdot IQ = 285$

Exkurs: Falls $x_{[10]} = 1000$



$400 = \min(\max_j a_j ; a_j)$ mit $a_j \leq \tilde{x}_{0.75} + 1{,}5 \cdot IQ$

F63:

$$G = \frac{\Delta}{\blacktriangle} = \frac{A}{\frac{1}{2}} = 2 \cdot A$$

G_{max} : maximales Ginikoeffizient, d.h. maximales Maß der Konzentration, falls ein MM-Träge allein Anteil hat, d.h. $p_n = 1$ und $p_i = 0 \quad \forall \; i = 1, \dots, n-1$

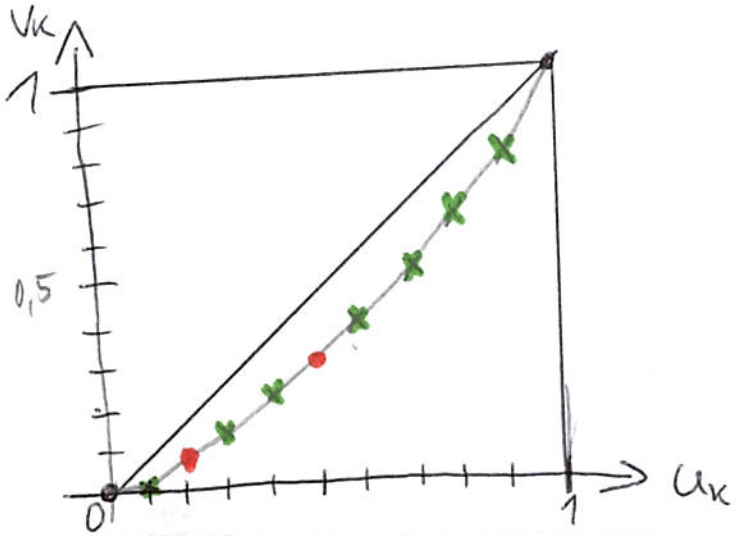
$$\Rightarrow G_{max} \overset{\text{Formel für } G}{=} \dots = \frac{2 \cdot n \cdot 1 - (n+1)}{n} = \frac{2n - n - 1}{n} = \frac{n-1}{n}$$

Erwünscht zur Vergleichbarkeit ist stets $[0;1]$

$$\Rightarrow \text{Norm. Gini-Koeff. } G^* = \frac{n}{n-1} \cdot G = \frac{G}{G_{max}}$$

i/k	1	2	3	4	5	6	7	8	9	10
$x_{[i]/k}$	70	150	150	160	200	200	260	340	400	550
$p_{i/k}$	70/2480	150/2480	150/..	160/..	200/..	200/..	260/..	340/..	400/..	550/..
$V_{i/k}$	70/2480	220/..	370/..	530/..	730/..	930/..	1190/..	1530/..	1930/..	1
	0,028		0,149	0,214		0,375	0,480	0,617	0,778	
$U_{i/k}$	1/10	2/10	3/10	4/10	5/10	6/10	7/10	8/10	9/10	10/10

Lorenzkurve



für Häufigkeitstabellen:

j	1	2	3	4	5	6	7	8
a_j	70	150	160	200	260	340	400	550
$h(a_j)$	1	2	1	2	1	1	1	1
$H(a_j)=k$	1	3	4	6	7	8	9	10
V_k	70/2480	$\frac{70\cdot1+150\cdot2}{2480}=\frac{370}{2480}$						1
$U_k=\frac{k}{n}=F(a_j)$	1/10	3/10	4/10	...				1

$$V_k = \frac{\sum_{j=1}^{k} a_j \cdot h(a_j)}{\sum_{j} a_j \cdot h(a_j)}$$

(alles sind Knickstellen!)

$$G \overset{F63}{=} \frac{2 \cdot \sum_{i=1}^{n} i \cdot p_i - (n+1)}{n} = \frac{2 \cdot 6\frac{121}{124} - 11}{10} \simeq 0{,}295$$

$$G_{max} = \frac{n-1}{n} = \frac{9}{10}$$

$$G^* = \frac{G}{G_{max}} = G \cdot \frac{10}{9} = 0{,}328$$

Kontrolle: (↗)!

F 67 : weitere Konzentrationsmaße ... mit Bsp. von F.64

... Berechnungen für unsere Beispieldaten:

$$CR_2 = \frac{550 + 400}{2480} \simeq 0{,}3831$$

$$H = \sum_{i=1}^{n} \cdot p_i^2 \simeq 0{,}1303$$

Klausur Statistik vom 20. Juli 2016

Aufgabe 1

15 Punkte

Für ein metrisches Merkmal X wurden 30 Beobachtungen erfasst. Für X ergibt sich die empirische Verteilungsfunktion F mit

$$F(x) = \begin{cases} 0 & \text{für } x < 4 \\ 0.2 & \text{für } 4 \leqq x < 8 \\ 0.4 & \text{für } 8 \leqq x < 12 \\ 0.4 & \text{für } 12 \leqq x < 15 \\ 0.7 & \text{für } 15 \leqq x < 22 \\ 0.9 & \text{für } 22 \leqq x < 24 \\ 1 & \text{für } x \geq 24 \end{cases}$$

a) Erstellen Sie eine Tabelle der absoluten Häufigkeiten.

b) Berechnen Sie mit Hilfe der angegebenen empirischen Verteilungsfunktion

(1) den Modus des Merkmals X.

(2) die relative Häufigkeit der Ausprägung 21.

(3) die absolute Häufigkeit der Ausprägung 15.

Für die Teilaufaben c) bis e) sei ein weiteres metrisches Merkmals Y mit ebenfalls $n = 30$ Beobachtungen gegeben. Für Y sind die Ausprägungen a_i und die relativen Häufigkeiten $f(a_j)$ in der folgenden Tabelle aufgeführt:

a_j	3	6	16	22	25
$f(a_j)$	0.1	0.3	0.2	0.2	0.2

c) Bestimmen Sie den Median von Y.

d) Bestimmen Sie die kumulierte relative Häufigkeit für die Ausprägung 17.

e) Berechnen Sie den Anteil der Beobachtungen von Y, an denen eine Ausprägung von mindestens 12, aber weniger als 23 vorliegt?

R Nehmen Sie für die Teilaufgaben f) bis h) an, dass eine Urliste $x_1, \ldots, x_n$ in einem R-Vektor `data` gespeichert. Geben Sie das (die) R-Kommando(s) an, mit dem (denen) Sie

f) einen horizontal dargestellten Boxplot der Daten erstellen.

g) ein Histogramm mit den Klassengrenzen 75, 80, 95 und 105 erstellen.

h) eine Tabelle der kumulierten absoluten Häufigkeiten erstellen.

Für Teilaufgabe i) ist folgende Tabelle zu den Daten der Urliste $x_1, \ldots, x_7$ gegeben.

k	1	2	3	4	5	6	7
x_k	2	2	4	8	8	10	10
p_k	$2/44$	$2/44$	$4/44$	$8/44$	$8/44$	$10/44$	$10/44$
v_k	$2/44$	$4/44$	$8/44$	$16/44$	$24/44$	$34/44$	1
u_k	$1/7$	$2/7$	$3/7$	$4/7$	$5/7$	$6/7$	1

i) Bestimmen Sie die Knickstellen der zugehörigen Lorenzkurve.
(*Hinweis*: Die Lorenzkurve muss nicht gezeichnet werden)