Klausur: **Statistik**

Jürgen Meisel

Zugelassene Hilfsmittel: Taschenrechner; Formelsammlung
Bearbeitungszeit: **60 Minuten**

1.) Mittelwerte und Streumaße

20

In einer Vorlesung auf der Universität sitzen 30 Studenten folgenden Alters:

20	**21**	**19**	**18**	**22**	**20**	**21**	**20**	**20**	**23**	**20**	**21**	**20**	**19**	**20**
21	**22**	**20**	**21**	**20**	**20**	**23**	**25**	**20**	**24**	**20**	**23**	**20**	**20**	**19**

a) Erstellen Sie eine Häufigkeitstabelle.

b) Welches durchschnittliche Alter haben die Studenten?

c) Wie groß ist die Standardabweichung?

d) Wo liegen der Median und die beiden Quartile?

e) Wie groß ist der Modalwert?

f) Da betritt noch ein weiterer „Seniorenstudent" den Raum und nimmt an der Vorlesung teil. Sein Alter ist 65. Wie lauten nun der Mittelwert und Median?

Lösung:

1

a)

Alter	18	19	20	21	22	23	24	25	
Anzahl	1	3	14	5	2	3	1	1	$n = 30$

b) $\bar{x} = 20{,}73$

c) $\sigma = 1{,}5477$

d) $\bar{x}_M = \frac{1}{2}(x_{15} + x_{16}) \leadsto \bar{x}_M = \frac{1}{2}(20+20)$

$\bar{x}_M = 20$

$q_1 = x_{[n \cdot p]+1} \rightsquigarrow q_1 = x_{[30 \cdot 0{,}25]+1}$

$q_1 = x_{7+1} = x_8$

$q_1 = 20$

$q_3 = x_{[30 \cdot 0{,}75]+1}$

$q_3 = x_{22+1} = x_{23}$

$q_3 = 21$

e) $\bar{x}_{Mod} = 20$

f) $\bar{x} = 22{,}16$ $[n = 31]$

$\bar{x}_M = x_{\frac{n+1}{2}} \rightsquigarrow \bar{x}_M = x_{16} \rightsquigarrow \bar{x}_M = 20$

2.) Mittelwerte und Streumaße

20	

Der Schatzmeister eines Tennisclubs hat Daten für die Aufstellung der Mitgliederstatistik gesammelt.

Alter (Jahre)	Absolute Häufigkeit	Relative Häufigkeit	Klassen-mitte	Klassen-breite	Abs. Häufig-keitsdichte	Relative Sum-menhäufigkeit
[0;10[	**8**	**0,04**	**5**	**10**	**0,8**	**0,04**
[10;15[	**16**	**0,08**	**12,5**	**5**	**3,2**	**0,12**
[15;20[	**22**	**0,11**	**17,5**	**5**	**4,4**	**0,23**
[20;30[	**54**	**0,27**	**25**	**10**	**5,4**	**0,50**
[30;50]	**100**	**0,50**	**40**	**20**	**5,0**	**1,00**
Summe	**200**	**1,00**	**---**	**---**	**---**	**---**

a) Zeichnen Sie das zugehörige Histogramm.

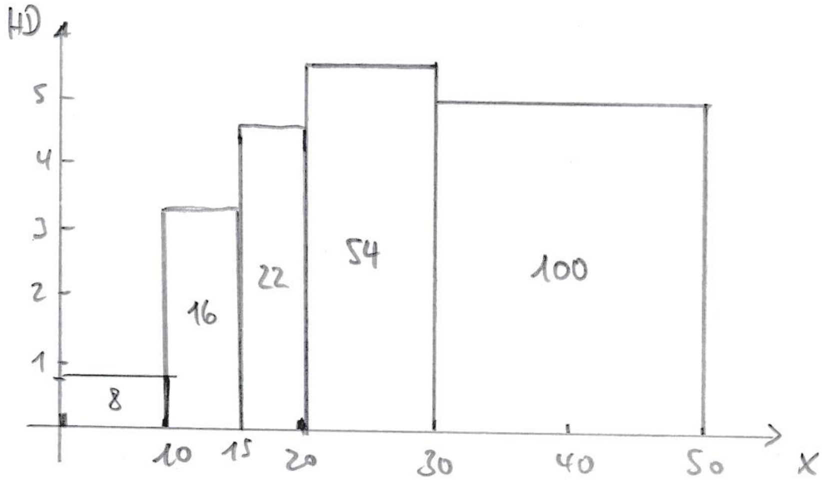
b) Bestimmen Sie den arithmetischen Mittelwert, die modale Klasse und den Modalwert.

c) Bestimmen Sie den Median, das untere Quartil und das obere Quartil.

Lösung:

(2)

a) Histogramm



b) arithmetisches Mittel : $\bar{x} = 25{,}875$ [Klassenmitten]

modale Klasse : $[20; 30[$ wegen $HD_{max} = 5{,}4$

Modalwert : 25 ~> Klassenmitte der mod. Klasse

c) Median : $\bar{x}_M = 30$ (wegen rel. Summenhäufigkeit 0,5)

Quartile 1: $q_1 = 20 + \frac{10 \cdot (0{,}25 - 0{,}23)}{0{,}27}$

$q_1 = 20 + 0{,}74 = 20{,}74$

Quartile 3: $q_3 = 30 + \frac{20 \cdot (0{,}75 - 0{,}50)}{0{,}5}$

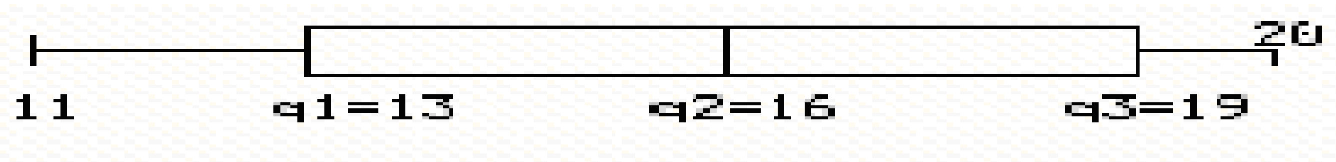
$q_3 = 30 + 10 = 40$

20	

3.) Interpretation Boxplot und Preisindizes

Teil 1:

Folgender Boxplot zeigt die Verteilung der Studiendauer(in Semestern) für eine Studienrichtung. Lesen Sie die Quartile-Werte ab und beantworten Sie folgende Fragen:



a) Was kann man über die durchschnittliche Studiendauer sagen?

b) Wie lange hat der „schnellste" Student für sein Studium gebraucht?

c) Wie viel Semester hat der „langsamste" Student benötigt?

d) Wie hoch ist der Prozentsatz der Studenten, die zwischen 11 und 19 Semestern studiert haben?

e) Geben Sie Semestergrenzen an, innerhalb derer jeweils 50 % der Studenten ihr Studium absolviert haben.

Lösung:

3

Teil 1:

a) Median : $\bar{x}_M = 16$ $(= q_2)$

b) $x_1 = 11$

c) $x_n = 20$

d) 75 %

e) $[11; 16]$

$[13; 19]$

$[16; 20]$

Teil 2:

Student Klaus Page hat seinen Verbrauch und die Preise einiger Konsumgüter genau notiert. Folgende Tabelle kam hierbei heraus:

	2007		**2015**	
	Preis	**Verbrauch**	**Preis**	**Verbrauch**
Zigaretten	4,00	10	5,00	8
Pizza	5,00	8	6,50	6
Kino	8,00	5	12,00	2

a) Berechnen Sie den Preisindex nach Laspeyres und nach Paasche.

b) Wie hoch ist die jährliche Inflationsquote während der Jahre 2007 bis 2015, wenn Sie den berechneten Preisindex nach Laspeyres zugrunde legen?

Lösung:

Teil 2:

a) $$L_P = \frac{5 \cdot 10 + 6{,}5 \cdot 8 + 12 \cdot 5}{4 \cdot 10 + 5 \cdot 8 + 8 \cdot 5} = \frac{162}{120} = 1{,}35$$

$$P_P = \frac{5 \cdot 8 + 6{,}5 \cdot 6 + 12 \cdot 2}{4 \cdot 8 + 5 \cdot 6 + 8 \cdot 2} = \frac{103}{78} = 1{,}3205$$

b) $$q = \sqrt[8]{1{,}35} \approx 1{,}03823 \Rightarrow i = 0{,}03823$$

$$p = 3{,}823\ \%$$

4.) Lineare Regression bzw. Korrelation

20	

Teil 1:

In einem bestimmten Bereich hängt der Ernteertrag eines landwirtschaftlichen Gutes von der Menge eines eingesetzten Düngemittels je Hektar ab.
Auf 6 Versuchsfeldern wird der Düngemitteleinsatz getestet.

Dabei wurden die folgenden Erträge je Hektar erzielt:

Versuchsfeld	1	2	3	4	5	6
Düngemitteleinsatz [100 kg]	6	3	8	2	7	2
Ernteertrag [t]	30	10	22	14	36	24

a) Stellen Sie den Ernteertrag in Abhängigkeit von dem Düngemitteleinsatz durch eine Funktion $y = b_0 + b_1 x$ dar.

b) Wie hoch wäre demnach der Ernteertrag bei 1 Tonne Düngemittel?

Lösung:

4

Teil 1:

a) $A = 12{,}19 \quad B = 2{,}245 \quad \rightsquigarrow \quad y = 2{,}245x + 12{,}19$

b) $f(10) = 22{,}45 + 12{,}19 \quad \rightsquigarrow \quad f(10) = 34{,}64$

Teil 2:

Ein Landwirt möchte feststellen, ob ein Zusammenhang zwischen Blütebeginn und Erntebeginn von hellen Süßkirschen besteht.
Im Jahre 2015 machte er an 5 Bäumen folgende Beobachtungen:

Baum	Blütebeginn	Erntebeginn
A	28.04.	02.07.
B	29.04.	25.06.
C	01.05.	27.06.
D	02.05.	03.07.
E	03.05.	26.06.

Berechnen Sie den geeigneten Korrelationskoeffizienten.

Lösung:

Teil 2:

Blütebeginn	Ernte	d	d^2
1	4	3	9
2	1	1	1
3	3	0	0
4	5	1	1
5	2	3	9
			20

$$\Rightarrow r_s = 1 - \frac{6 \cdot 20}{5 \cdot 24}$$

$$r_s = 0$$

5.) Lorenzkurve und Gini-Koeffizient

20	

Aus nachfolgender Tabelle soll die Konzentration der Lebensmittelbranche eines Landkreises untersucht werden:

Tagesumsatz (in T€)	Anzahl der Unternehmen (x-Achse)
50	14
120	5
150	8
250	2
1.000	1

a) Zeichnen Sie die Lorenzkurve

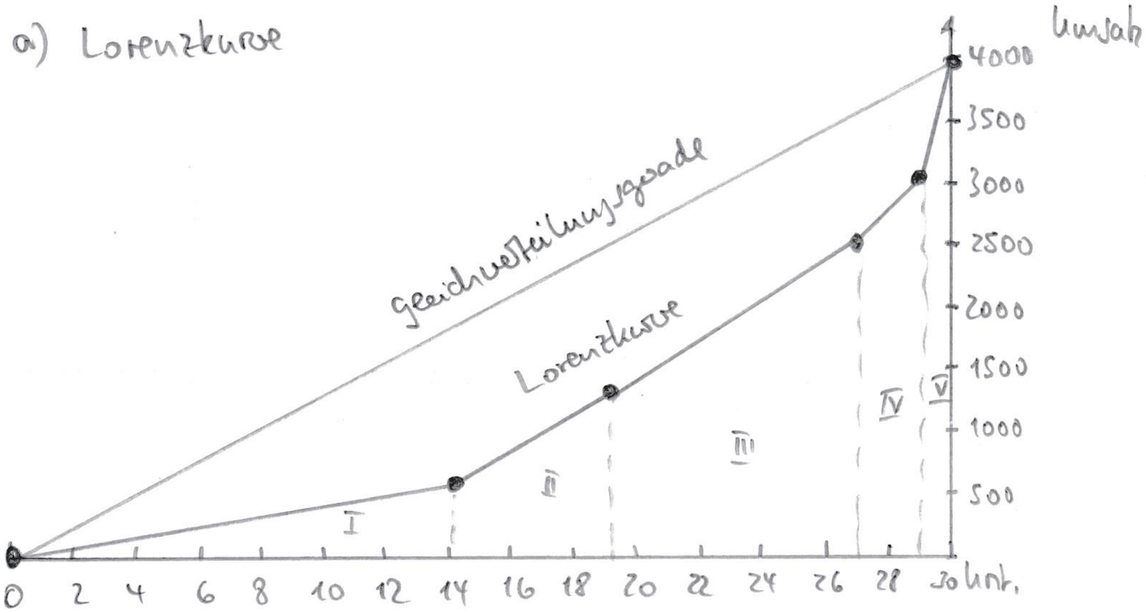
b) Bestimmen Sie den Ginikoeffizient.

Lösung:

(5)

Anzahl Unt.	Σ Unt.	Tagesumsatz	Tagesum. gesamt	Σ Umsatz
14	14	50	700	700
5	19	120	600	1300
8	27	150	1200	2500
2	29	250	500	3000
1	30	1.000	1.000	4000

a) Lorenzkurve



b) Ginikoeffizient

$$KF_{max} = \frac{1}{2} \cdot 30 \cdot 4000 = 60.000$$

I $A_1 = \frac{1}{2} \cdot 14 \cdot 700 = 4.900$

II $A_2 = 5 \cdot \frac{1}{2}(700+1300) = 5.000$

III $A_3 = 8 \cdot \frac{1}{2}(1300+2500) = 15.200$

IV $A_4 = 2 \cdot \frac{1}{2}(2500+3000) = 5.500$

V $A_5 = 1 \cdot \frac{1}{2}(3000+4000) = 3.500$

Fläche unter Lorenzkurve 34.100

$$GK = \frac{60\,000 - 34\,100}{60000}$$

$$GK = \frac{25.900}{60.000}$$

$$GK = 0{,}43167$$

6.) Binomialverteilung

20	

Eine Multiple Choice Prüfung im Rahmen des Studiengangs Statistik an der DHS BW setzt sich aus insgesamt **10 Fragen** zusammen.
Zu jeder gestellten Frage sind 5 mögliche Antwortalternativen gegeben, von denen jeweils **genau eine Antwort** richtig ist.

Eine Studentin hat es mal wieder komplett versäumt für die Prüfung zu lernen, weiß überhaupt nichts und behilft sich daher damit, bei jeder Frage eine Antwort zu raten.

a) Wie groß ist die erwartete Anzahl der richtigen Antworten?

b) Wie groß ist die Wahrscheinlichkeit, genau zwei richtige Antworten zu geben?

c) Wie groß ist die Wahrscheinlichkeit, höchstens zwei richtige Antworten zu geben?

d) Wie groß ist die Wahrscheinlichkeit, mehr als zwei richtige Antworten zu geben?

e) Wie groß ist die Wahrscheinlichkeit, die Multiple Choice Prüfung zu bestehen, wenn dabei mindestens die Hälfte der Fragen richtig beantwortet werden muss?

Lösung:

6

a) $\mu = n \cdot p \rightsquigarrow \mu = 10 \cdot 0{,}2 = 2$

b) $B(X=2) = \binom{10}{2} 0{,}2^2 \cdot 0{,}8^8 = 0{,}302 \approx 30{,}2\,\%$

c) $B(X \leq 2) = \sum_{k=0}^{2} \binom{10}{k} 0{,}2^k \cdot 0{,}8^{10-k} = 0{,}6778$

d) $B(X > 2) = 1 - B(X \leq 2)$

$= 1 - 0{,}6778 = 0{,}3222$

e) $B(X \geq 5) = \sum_{k=5}^{10} \binom{10}{k} 0{,}2^k \, 0{,}8^{10-k} = 0{,}0328$