Klausur: **Statistik**

Jürgen Meisel

Zugelassene Hilfsmittel: nicht progr. Taschenrechner
Bearbeitungszeit: **120 Minuten**

Statistik

Die Firma Gebäudetec International in Ludwigstadt hat mehrere Industrie-Immobiliengrundstücke im Angebot.
Es ergibt sich folgende Verteilung:

Grundstück	1	2	3	4	5	6	7	8	9	10
qm	300	250	800	500	300	200	250	500	700	200
Preis pro qm 2010	100	80	120	110	250	300	800	150	140	90
Preis pro qm 2009	95	90	110	100	265	300	450	140	110	120

1.) **Mittelwerte und Streumaße: Grundstückspreise**

a) Bestimmen Sie den durchschnittlichen qm-Preis.

b) Berechnen Sie die Varianz und die Standardabweichung.

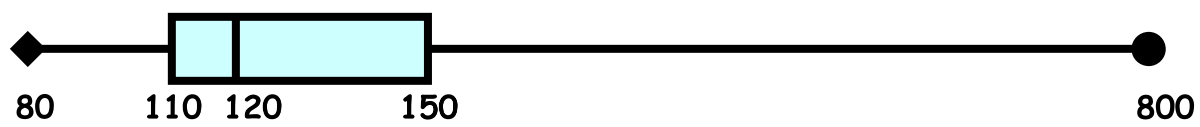
c) Bestimmen Sie nun noch den Median.

d) Erstellen/zeichnen Sie jetzt einen kompletten Boxplot mit einer Berechnung der Quartile 1 und 3.

Lösung:

Sinnvoll: Vorher ordnen der Werte nach den qm-Preisen!

Mittelwert:	**181,75**	**Quartile 1**	**110,00**	**[a(1000) + a(1001)]/2**
		Quartile 3	**150,00**	**[a(3000) + a(3001)]/2**
Varianz	**28.344,44**			
Stdabw.	**168,35806**			
Median	**120,00**	**[a(2000) + a(2001)]/2**		



e) Welcher Mittelwert würde in obigem Fall geeignet die Situation realitätsnah darzustellen. Begründen Sie kurz Ihre Entscheidung.

Lösung: *Median ist der geeignete und realitätsnähere MW, da er den Ausreißer für Grundstück 7 mit 800,00 € pro qm nicht berücksichtigt.*

f) Bestimmen Sie das klassierte arithmetische Mittel nach folgender Klasseneinteilung ***Preis pro qm 2010***.

[0; 100[[100; 200[[200; 500[[500; 1000]

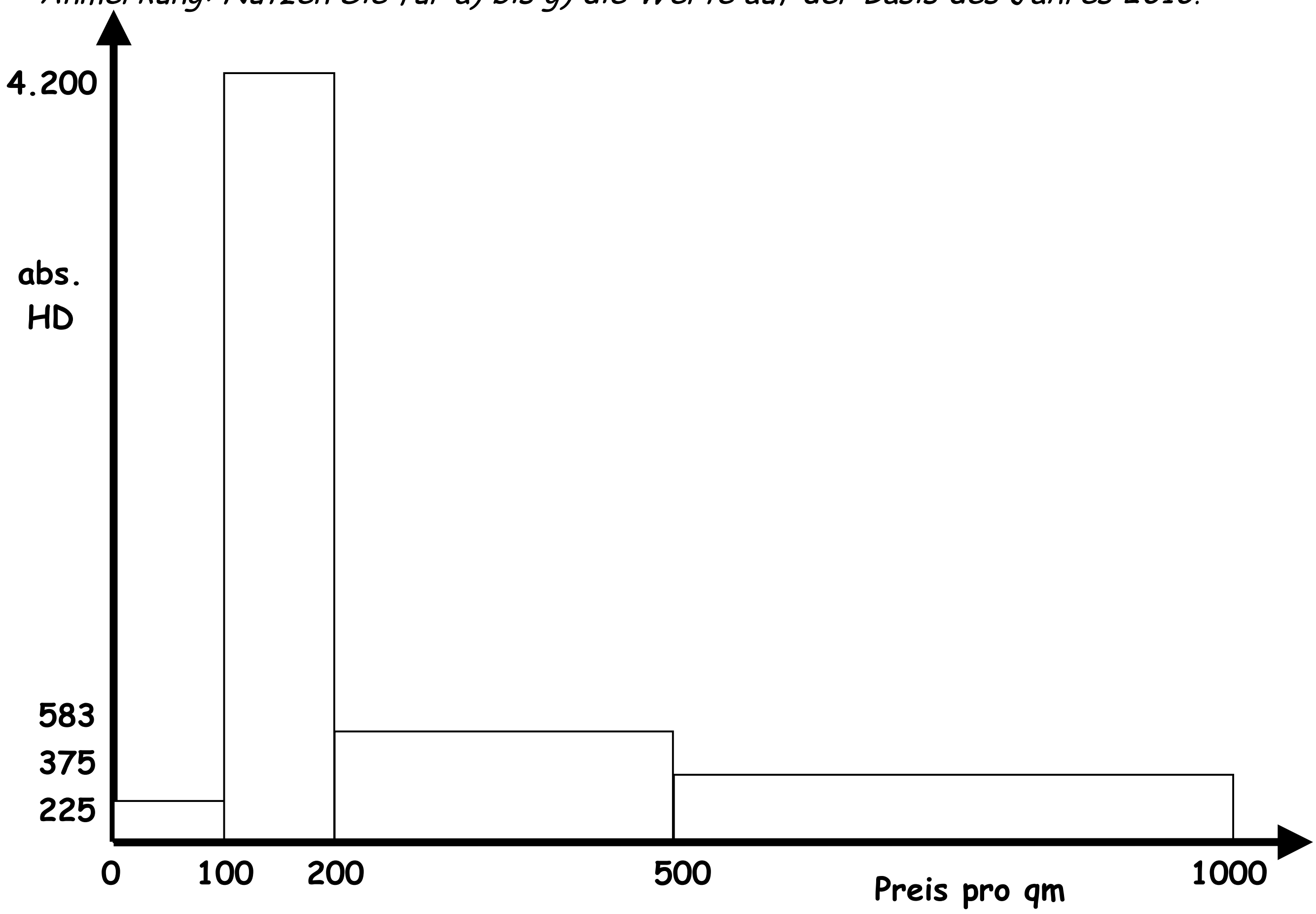
Lösung:

klassierter MW					**Summe**
qm	**450**	**2.800**	**500**	**250**	**4.000**
Klassenmitte	**50**	**150**	**350**	**750**	
Klassenbreite	**100**	**100**	**300**	**500**	
HD (Höhe)	**225,00**	**4.200,00**	**583,33**	**375,00**	

Produkt	**22.500,00**	**420.000,00**	**175.000,00**	**187.500,00**	**805.000,00**
Mittelwert:	**201,25**				

g) Zeichnen Sie das zu f) gehörende Histogramm.

Anmerkung: Nutzen Sie für a) bis g) die Werte auf der Basis des Jahres 2010!



2.) **Preisindizes**

Die qm-Preise unterliegen gewissen Schwankungen

a) Bestimmen Sie den Preisindex nach Laspeyres zur Basis 2009.

b) Wie wäre der Preisindex nach Paasche zur Basis 2009?
Begründen Sie ohne neue Berechnung Ihren Wert.

c) Bestimmen Sie den Preisindex nach Laspeyres zur Basis 2009
ohne das Grundstück 7.
Macht dies Sinn? Erläutern Sie kurz Ihre Ansicht.

Lösung:

Laspeyres:	**1,187908**	
Paasche:	**1,187908**	**(keine Veränderung der Menge)**
Laspeyres (ohne 7)	**1,055055**	

=> Sinnvolle Alternative, da dadurch eine realitätsnahere Preissteigerung ermittelt werden kann.

3.) **Regressionsgerade und Korrelation**

Die Tabelle gibt Daten aus seriösen Quellen über die Anzahl der Storchenpaare und die Einwohneranzahl in den Jahren 1930 bis 1936 in Oldenburg wieder.

Jahr	1930	1931	1932	1933	1934	1935	1936	
Anzahl der Storchenpaare X	132	142	166	188	240	250	252	
Anzahl der Einwohner Y	55.400	55.400	65.000	67.700	69.800	72.300	76.000	
$x_i \cdot y_i$								
$(x_i)^2$								
$(y_i)^2$								

a) Ergänzen Sie die Tabelle und bestimmen Sie die Gleichung der linearen Regressionsgeraden $y = b_0 + b_1 x$.

b) Bestimmen Sie den Korrelationskoeffizienten r zwischen X und Y **und** beurteilen Sie Ihr Ergebnis (kritisch!).

Lösung:

Korrelation und Regression

	x-Werte	y-Werte		x-Werte
01	132	55400	11	
02	142	55400	12	
03	166	65000	13	
04	188	67700	14	
05	240	69800	15	
06	250	72300	16	
07	252	76000	17	
08			18	
09			19	
10			20	

Korrelation und Regression berechnen

Mittelwerte: x-Werte = 195.7143 y-Werte = 65942.8571

Standardabweichung: x-Werte = 51.5807 y-Werte = 7986.6257

Gleichungen der Regressionsgeraden: y = b1*x + b0 bzw. x = a1*y + a0

b1 = 146.2396 **b0 =** 37321.6745 **a1 =** 0.0061 **a0 =** -206.5218

Korrelationskoeffizient r = 0.944472 Neustart Grafik

Quelle: http://www.mathematik.ch/anwendungenmath/Korrelation_Regression/

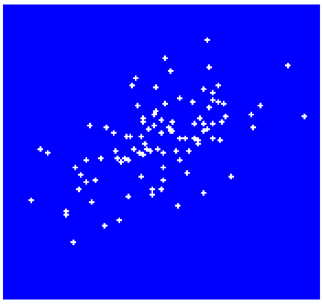
Excel-Lösung:

3	Jahr	1930	1931	1932	1933	1934	1935	1936	Summe	MW
4	Anzahl der Storchenpaare X	132	142	166	188	240	250	252	1.370	195,71
5	Anzahl der Einwohner Y	55.400	55.400	65.000	67.700	69.800	72.300	76.000	461.600	65.942,86
6	x * y	7.312.800	7.866.800	10.790.000	12.727.600	16.752.000	18.075.000	19.152.000	92.676.200	13.239.457,14
7	x * x	17.424	20.164	27.556	35.344	57.600	62.500	63.504	284.092	40.584,57
8	y * y	3.069.160.000	3.069.160.000	4.225.000.000	4.583.290.000	4.872.040.000	5.227.290.000	5.776.000.000	30.821.940.000	4.403.134.285,71
9										
10	b1:	146,2396								
11	b0:	37.321,6745								

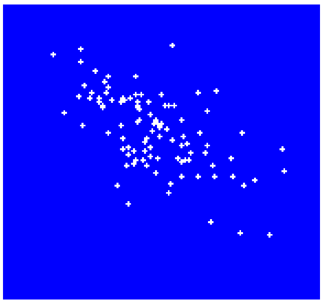
$$y = 146,24x + 37.321,67$$

c) Betrachten Sie die folgenden Punktdiagramme und ordnen Sie jeweils die Steigung der Regressionsgerade und den Korrelationskoeffizienten einander zu.

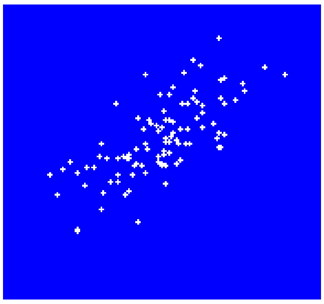
Plot A



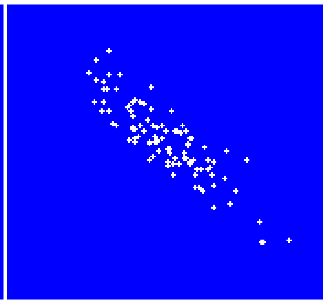
Plot B



Plot C



Plot D



Korrelationskoeffizienten:

	Plot A	Plot B	Plot C	Plot D
r = - 0,87				XXX
r = - 0,59		XXX		
r = 0,54	XXX			
r = 0,70			XXX	

Stochastik

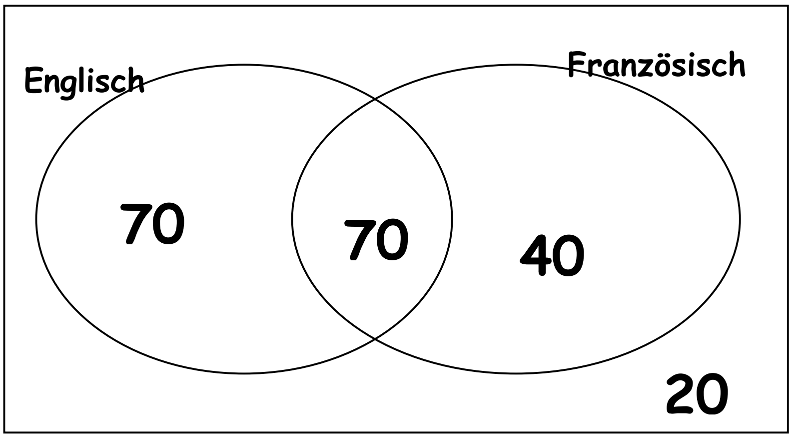
1.) **Relative Häufigkeiten**

Von 200 Mitarbeitern der Firma Gebäudetec International sprechen 140 Mitarbeiter Englisch, 110 Mitarbeiter Französisch aber 20 Mitarbeiter keine der beiden Sprachen.

Wie viel Prozent der Mitarbeiter sprechen ...

a) ... beide Sprachen? b) ... Französisch aber kein Englisch?
c) ... nur eine der beiden Sprachen?

Lösung:



Ω = 200

$$h_a = \frac{70}{200} = 0,35$$

$$h_b = \frac{40}{200} = 0,20$$

$$h_c = \frac{110}{200} = 0,55$$

2.) **Erwartungswert** - *Jedes Los gewinnt!*

Bei einer Jahresabschlussfeier muss jeder der 100 Teilnehmer ein Los kaufen. Der 1. Preis hat einen Wert von 150,00 €, der 2. von 50,00 € und der 3. von 25,00 €. Jeder, der keinen dieser Gewinne bekommt, erhält einen Trostpreis in Höhe von 1,00 €.

a) Jedes Los wird für 7,50 € verkauft. Der Erlös geht ans Friedensdorf. Wie groß ist der Erlös?

Lösung:

$X = x_i$	150,00	50,00	25,00	1,00
$P(X = x_i)$	0,01	0,01	0,01	0,97
$x_i * P(X = x_i)$	1,50	0,50	0,25	0,97

=> Erwartungswert: 3,22 €

Der Erlös für das Friedensdorf beträgt bei 100 Teilnehmer demnach:

Erlös = 100 * (7,50 – 3,22) = 428,00 €

b) Wie teuer müsste ein Los sein, damit Einnahmen und Ausgaben überein stimmen?

Lösung: **Der Preis müsste in Höhe des EW liegen, also bei 3,22 €.**

3.) **Laplace-Wahrscheinlichkeit und Axiome Kolmogorow**

Die Auswahl einiger europäischer Städte lautet wie folgt:

Ω = {Athen, Frankfurt, Hamburg, Innsbruck, London, München, Nürnberg, Paris, Rom, Salzburg, Wien, Würzburg}

Wie groß ist die Wahrscheinlichkeit bei einmaligem Ziehen ...

a) ... eine Stadt am Mittelmeer zu ziehen?

b) ... eine deutsche oder österreichische Stadt zu ziehen?

c) ... eine deutsche Stadt nördlich der Donau zu ziehen?

d) ... eine europäische Hauptstadt zu ziehen?

e) ... eine Stadt zu ziehen, die nicht in Bayern liegt, wenn man aber weiß, dass man eine deutsche Stadt gezogen hat.

Lösung:

$$h_a = \frac{2}{12} = \frac{1}{6} \qquad h_b = \frac{8}{12} = \frac{2}{3} \qquad h_c = \frac{4}{12} = \frac{1}{3}$$

$$h_d = \frac{5}{12} \qquad h_e = \frac{2}{12} : \frac{5}{12} = \frac{2}{5}$$

4.) **Binomialverteilung & Bayes**

Von einer großen Ladung Apfelsinen sind 20% verdorben. Es werden 5 Stück entnommen. Wie groß ist die Wahrscheinlichkeit für folgende Ereignisse:

a) Eine Apfelsine ist verdorben.

Lösung: $$B_{5;0,2}(X=1) = \binom{5}{1} 0,2^1 \cdot 0,8^4 = 0,4096$$

b) Alle Apfelsinen sind in Ordnung.

Lösung: $$B_{5;0,2}(X=0) = \binom{5}{0} 0,2^0 \cdot 0,8^5 = 0,32768$$

c) Mindestens zwei Apfelsinen sind verdorben.

Lösung:

$$B_{5;0,2}(X \geq 2) = 1 - B_{5;0,2}(X \leq 1) = 1 - 0,32768 - 0,4096 = 0,26272$$

d) Wie viele Apfelsinen muss man mindestens entnehmen, um mit einer Wahrscheinlichkeit von mindestens 99 % mindestens eine verdorbene zu erhalten?

Lösung:

$$B_{n;0,2}(X \geq 1) \geq 0,99 \rightarrow 1 - B_{n;0,2}(X = 0) \geq 0,99$$

$$\rightarrow 1 - \binom{n}{0} 0,2^0 \cdot 0,8^n \geq 0,99$$

$$\rightarrow 0,8^n \leq 0,01$$

$$\rightarrow n \geq \frac{\ln 0,01}{\ln 0,8} \approx 20,63$$

$$\rightarrow n \geq 21$$

Die Lieferung kommt von drei verschiedenen Plantagen, die eine unterschiedliche Qualitätsgüte aufweisen:

Die Apfelsinen von Plantage A haben eine Verderblichkeitsquote von 25 %, die von Plantage B haben eine Quote von 15 % und die von Plantage C besitzen eine Quote von 12 %.

Die Aufteilung der Gesamtlieferung auf die jeweiligen Herkunftsplantagen beträgt 2 : 5 : 3.

e) Wie groß ist die Wahrscheinlichkeit für eine verdorbene Orange?

Lösung:

$$P(\text{"verdorben"}) = \frac{2}{10}\cdot 0,25+\frac{5}{10}\cdot 0,15+\frac{3}{10}\cdot 0,12 = 0,161$$

f) Der Lieferung wird eine verdorbene Orange entnommen.
Mit welcher Wahrscheinlichkeit stammt sie von Plantage A?

Lösung: => Satz von Bayes

$$P(\text{"verdorben von A"}) = \frac{\frac{2}{10}\cdot 0,25}{\frac{2}{10}\cdot 0,25+\frac{5}{10}\cdot 0,15+\frac{3}{10}\cdot 0,12} = \frac{0,05}{0,161} = 0,3106$$

5.) **Normalverteilung und ...**

In Hamburg führt die Polizei an einer bestimmten Stelle in regelmäßigen Abständen in der Nacht von Sonnabend auf Sonntag zwischen 1 Uhr und 4 Uhr Verkehrskontrollen durch.

Dabei muss der Fahrer "in die Röhre pusten", um festzustellen, ob der Alkoholgehalt im Blut im gesetzlich erlaubten Rahmen liegt oder nicht. Aus mehrjähriger Erfahrung weiß die Polizei, dass ungefähr 14 % aller männlichen und 6,5 % aller weiblichen Fahrer um diese Zeit an dieser Stelle die "Promillegrenze" überschreiten. Wir nennen diese Personen hier kurz "Alkoholsünder".

Am letzten Wochenende wurden 1.000 Verkehrsteilnehmer überprüft.

Darunter befanden sich 400 Frauen.

a) Ermitteln Sie den durchschnittlichen Wert p für einen Alkoholsünder.

Lösung: $$P(\text{"}Alkohol\text{"}) = \frac{600}{1000} \cdot 0{,}14 + \frac{400}{1000} \cdot 0{,}065 = 0{,}11$$

Gehen Sie nun unabhängig von Ihrer Lösung in a) von p = 0,1 aus.

=> Berechnen Sie die gesuchten Werte <u>ohne Stetigkeitskorrektur</u>!

b) Mit wie vielen Fahrverboten kann die Polizei bei der Überprüfung von 1.000 Verkehrsteilnehmern rechnen?

Lösung: Erwartungswert: $\mu = 1000 \cdot 0{,}1 = 100$

c) Berechnen Sie die Wahrscheinlichkeit $P(X \leq \mu)$.

Lösung: Vor.: $\sigma^2 = 1000 \cdot 0{,}1 \cdot 0{,}9 = 90 > 9$

$$P(X \leq 100) = \Phi(0) = 0{,}5$$

$$Umrechnung: \quad x = \frac{100-100}{\sqrt{90}} = 0$$

d) Mit welcher Wahrscheinlichkeit liegt die Anzahl der Alkoholsünder zwischen 75 und 125?

Lösung: Symmetrisches Intervall

$$P(75 \leq X \leq 125) \quad = \quad 2 \cdot \Phi(2,635) - 1 \quad = \quad 2 \cdot 0,996 - 1 \quad = \quad 0,992$$

$$Umrechnung: \quad x \quad = \quad \frac{125-100}{\sqrt{90}} \quad = \quad 2,635$$

e) Mit welcher Wahrscheinlichkeit liegt die Anzahl der Alkoholsünder über 120?

Lösung:

$$P(X > 120) \quad = \quad 1 - P(X \leq 120) \quad = \quad 1 - \Phi(2,108) \quad = \quad 1 - 0,983 \quad = \quad 0,017$$

$$Umrechnung: \quad x \quad = \quad \frac{120-100}{\sqrt{90}} \quad = \quad 2,108$$

f) Ermitteln Sie die Grenzen, damit die Wahrscheinlichkeit bei mind. 95 % im Intervall um den Erwartungswert liegt?

Lösung:

$$Bekannt: \quad 2\Phi(x) - 1 \quad \geq \quad 0,95 \quad \rightarrow \quad \Phi(x) \quad \geq \quad 0,975 \quad \xrightarrow{Tabelle} \quad x \quad \geq \quad 1,96$$

$$Umrechnung:$$

$$1,96 = \frac{k_2 - 100}{\sqrt{90}} \quad \rightarrow \quad k_2 \quad = \quad 118,59 \qquad und \qquad -1,96 \quad = \quad \frac{k_1 - 100}{\sqrt{90}} \quad \rightarrow \quad k_1 = 81,40$$

$$Intervall: \quad I \quad = \quad [81;119]$$

Quelle zum Nachrechnen:

http://www.arndt-bruenner.de/mathe/scripts/normalverteilung1.htm