Forschungsmethoden II - Schliessende Statistik
Klausuraufgaben mit Lösungen

Robert Morys

30.07.2006

Ohne Haftung für Richtigkeit.

Bei Nervosität, Korrekturvorschlägen oder Fragen
einfach eine E-Mail an rob.mo@gmx.de

Die Aufgaben stammen von Bodo´s pdf-file:

`http://www-users.rwth-aachen.de/Bodo.von.der.Heiden/uni/Forschungsmethoden2.pdf`

Danke nochmal dafür.

1 Theorie

Aufgabe 1.1)
Im folgenden sind die fünf Skalenniveaus mit jeweils zwei Beispielen aufgeführt, von denen eines falsch ist. Streichen Sie das falsche Beispiel bitte durch

Skala	Beispiele	
Nominal	Geschlecht	Ausweisnummer
Ordinal	Härtegrad	Haarfarbe
Intervall	Zeugnisnote	Grad Celsius
Verhältnis	Länge	Konfektionsgröße
Absolut	Lautstärke	Häufigkeit

Lösung:

Skala	Beispiele	
Nominal	Geschlecht	
Ordinal	Härtegrad	
Intervall		Grad Celsius
Verhältnis	Länge	
Absolut		Häufigkeit

Aufgabe 1.2)
Den Personalunterlagen eines Bewerbers für einen international tätigen Konzern sind die folgenden Informationen zu entnehmen. Welches Skalenniveau haben die Daten?

	Skalenniveau				
Daten	No	Ord	Int	Rat	Abs
Nationalität					
Führerscheinklasse (ohne = 0)					
Schulabschluss (Haupt.,Real.,Gymnas.,Studium)					
Auslandsaufenthalt in Jahren					

Lösung:

	Skalenniveau				
Daten	No	Ord	Int	Rat	Abs
Nationalität	X				
Führerscheinklasse (ohne = 0)		X			
Schulabschluss (Haupt.,Real.,Gymnas.,Studium)		X			
Auslandsaufenthalt in Jahren					X

Aufgabe 1.3)
Wie viele Wendepunkte besitzt die Normalverteilung ?

LÖSUNG: Einen bei s und einen bei -s $\Rightarrow$ 2

Aufgabe 1.4)
Die Korrelation lässt sich als ein Verhältnis von Varianzen interpretieren. Bitte beschreiben Sie dieses Verhältnis.

LÖSUNG: Die Korrelation ist das Verhältnis vom Mittleren Abweichungsprodukt s_{xy}(Kovarianz) zum Geometrischen Mittel der Varianzen $s_x s_y$(Gesamtvarianz).

Aufgabe 1.5)
Welche Bedeutung haben die additiven und die multiplikativen Konstanten in der linearen Regressionsgleichung ?

LÖSUNG: Die additive und multiplikative Konstante stehen in linearer Abhängigkeit und werden benötigt um von einer Variablen x auf eine Variable y zu schließen. Z.B.
$Gewicht = a * Groesse - b$

Aufgabe 1.6)
Welche Beziehung besteht zwischen Binominalverteilung und Normalverteilung ?

LÖSUNG: Die Normalverteilung ist eine Verallgemeinerung der Binominalverteilung, bei welcher $p = \frac{1}{2}$ ist und $n \to \infty$ läuft.

Aufgabe 1.7)
Was bedeuten alpha-Fehler und beta-Fehler?

LÖSUNG: Als alpha-Fehler bezeichnet man den Fehler, den man begeht, wenn man die Nullhypothese verwirft, obwohl sie zutrifft. Als beta-Fehler bezeichnet man den Fehler, den man begeht, wenn man die Nullhypothese beibehält, obwohl sie falsch ist.

Aufgabe 1.8)
Wie ändert sich die Größe des Konfidenzintervalls des Mittelwerts

a) bei einer Vergrößerung des Stichprobenumfangs ?

b) bei einer Vergrößerung der Streuung ?

Lösung:
a) Das Intervall wird kleiner

b) Das Intervall wird größer

Aufgabe 1.9)
Wodurch unterscheidet sich die Standardnormalverteilung von der Normalverteilung ?

Lösung: Die Standardnormalverteilung ist ein Spezialfall der Normalverteilung, bei welchem der Mittelwert $\mu = 0$ und die Varianz $\sigma = 1$ ist.

Aufgabe 1.10)
Aus den Zeugnisunterlagen von Abiturienten ist zu entnehmen, welche Note sie in einem bestimmten Fach erreicht haben und wie viele Jahre sie darin Unterricht erhielten. Welches Skalenniveau haben die Variablen? Erläutern Sie ihr Urteil.

Lösung:

Variable	Skalenniveau	Begründung
Note	Ordinalskala	Ordnung, jedoch keinen Nullpunkt
Jahre	Absolutskala	gleiche Abstände und Nullpunkt vorhanden

Aufgabe 1.11)
Am Arbeitsamt Ceburg weist die Statistik verschiedene Lehrberufe aus. Sie wollen deren Verteilung als Kreisdiagramm darstellen. Bestimmen Sie die Gradzahlen der entsprechenden Sektoren.

Lehrberuf	Anzahl
Handel	716
Handwerk	504
Verwaltung	160
Banken/Versicherungen	129
Touristik	205
Medizin. Einrichtungen	28

Lösung:

Lehrberuf = x	Anzahl = n	p $=x/n*100$	Gradzahl = $3.6*p$
Handel	716	35.8	128.88°
Handwerk	504	25.2	90.72°
Verwaltung	160	8	28.8°
Banken/Versicherungen	129	6.45	23.22°
Touristik	205	10.25	36.9°
Medizin. Einrichtungen	286	14.3	51.48°
$\sum$	2000	100	360°

Aufgabe 1.12)
Im Kreis Faburg wurden von einer Fast-Food-Zentrale im Mai verschiedene Gerichte ausgeliefert: Sie wollen ihre Verteilung als Kreisdiagramm darstellen. Bestimmen Sie die Gradzahl der entsprechenden Sektoren.

Gericht	Stück		Grad
Fischstäbchen	12600		
Pizza Margaritha	3600		
Pizza Con Tutto	5400		
Lasagne	4500		
Hähnchen	720		
Schnitzel	9180		

Lösung:

Gericht	Stück	$p = x/36000*100$	$Grad = 3.6*p$
Fischstäbchen	12600	35	126
Pizza Margaritha	3600	10	36
Pizza Con Tutto	5400	15	54
Lasagne	4500	12.5	45
Hähnchen	720	2	7.2
Schnitzel	9180	25.5	91.8
$\sum$	36000	100	360

Aufgabe 1.13)
Wie hoch korrelieren zwei Variablen, wenn ihre Regressionsgerade in einem Winkel von
a) 0°
b) 90°
c) 180°
zueinander stehen?

Winkelgrad =	0°	90°	180°
Korrelation =			

Lösung:

Winkelgrad =	0°	90°	180°
Korrelation =	1	0	-1

2 Mittelwerte / Zentrale Tendenz

Formeln:

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

Gewogenes Arithmetisches Mittel: $\bar{x} = \frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i}$

Geometrisches Mittel: $\bar{x}_{geo} = \sqrt[n]{\prod_{i=1}^{n} x_i}$

Median: $ZW_u = x_{(n+1)/2}$; $ZW_g = (x_{n/2} + x_{n/2+1})/2$

Aufgabe 2.1)
Bestimmen Sie für die nebenstehenden Werte die Kenngrößen der zentralen Tendenz:

3	7	9	6	4	5	10	7	11	8

a) Modus
b) Median
c) arithmetisches Mittel
d) Handelt es sich wohl um eine symmetrische oder um eine schiefe

LÖSUNG: Liste sortieren ⇒

3	4	5	6	7	7	8	9	10	11

a) Modus = meist vorkommender Wert = 7

b) $ZW_g = (x_5 + x_6)/2 = \frac{7+7}{2} = 7$

c) $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{70}{10}$ =7

d) Da Modus = Median = arithmetisches Mittel = 7 gilt, handelt es sich um eine symmetrische Verteilung.

Aufgabe 2.2)
In fünf Schulen wurde für 12-jährige ein Intelligenztest erprobt. Man erhielt pro Schule das arithmetische Mittel $\bar{x}$ und die Anzahl n_i der getesteten Schüler. Wie lautet der Durchschnittswert über alle Schulen?

	1	2	3	4	5
$\bar{x}$	9.3	6.7	7.8	9.6	8.5
n_i	20	60	30	30	20

LÖSUNG:

	1	2	3	4	5	$\sum$
$\bar{x}_i$	9.3	6.7	7.8	9.6	8.5	
n_i	20	60	30	30	20	160
$\bar{x}_i * n_i$	186	402	234	288	170	1280

$\bar{x} = \frac{\sum_{i=1}^{5} n_i * \bar{x}_i}{\sum_{i=1}^{5} n_i} = \frac{1280}{160}$ =8

Der Durchschnittswert aller Schulen lautet 8.

Aufgabe 2.3)
Innerhalb der letzten 15 Jahre fiel am 13.12 in der Zeit von 0 bis 24 Uhr in Oberammergau die folgende Regenmenge in Millimetern:

2	6	8	5	3	4	9	10	4	5	5	7	11	13	7

a) Vereinfachen Sie die Urliste durch Zusammenfassen der Ausgangswerte auf 6 neue Werteklassen.
b) Erstellen Sie die Häufigkeitsverteilung.
c) Erstellen Sie sie kumulierte Häufigkeitsverteilung

Lösung:
a) Werte von 2-13 $\Rightarrow$ 12Werte / 6 Klassen = 2 Werte pro Klasse

$\rightarrow$

2-3	4-5	6-7	8-9	10-11	12-13

b)

2-3	4-5	6-7	8-9	10-11	12-13
2	5	3	2	2	1

c)

2-3	4-5	6-7	8-9	10-11	12-13
2	5	3	2	2	1
$2/15 = 0,1\bar{3}$	$0,1\bar{3} + 5/15 = 0,4\bar{6}$	$0,\bar{6}$	$0,8$	$0,9\bar{3}$	1

Aufgabe 2.4)
In einer kleinen Dorfsiedlung haben die Haushalte folgendes Monatseinkommen in Euro:

1200 1400 1200 1300 2000 2500
1300 1500 1700 1900 2000 2500
1400 1200 1800 1900 2100 2600

Nun zieht noch ein Millionär hinzu mit einem Monatseinkommen von 50000 Euro. Berechnen Sie das geeignete Maß der zentralen Tendenz.

Lösung: 1. Sortierte Liste erstellen:

1200	1200	1200	1300	1300	1400	1400	1500	1700	1800
1900	1900	2000	2000	2100	2500	2500	2600	50000	

2. Modus = meist vorkommender Wert = 1200

3. $ZW = x_{(19+1)/2} = x_{10} = 1800$

4. $\bar{x} = \frac{\sum_{i=1}^{19} x_i}{19} = \frac{81500}{19} \approx 4289.47$

$\Rightarrow$ Das durchschnittliche Einkommen der Siedlung beträgt 4289.47 Euro mit Tendenz nach unten.

Aufgabe 2.5)
Bei einem optischen Experiment werden einer Versuchsperson 5 gleichartige Lichtquellen mit unterschiedlicher Helligkeit präsentiert. Die Wattzahlen lauten:

25	50	75	100	125

Jede Vp soll über einen Dimmer eine sechste Leuchte so einstellen, dass sie dem Mittelwert der 5 anderen entspricht. Auf wie viel Watt wird sie voraussichtlich eingependelt?

Lösung: $\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{375}{5} = 75$

$\Rightarrow$ Die Leuchte wird voraussichtlich auf 75 Watt eingependelt.

Aufgabe 2.6)
Bestimmen Sie für die folgenden Wert die Kenngrößen der zentralen Tendenz:

10	8	7	7	9	8

a) Modus
b) Median
c) arithmetisches Mittel
d) Handelt es sich um eine symmetrische oder um eine schiefe Verteilung

Lösung: Liste sortieren $\Rightarrow$

7	7	8	8	9	10

a) Modus = meist vorkommender Wert = 7

b) $ZW_g = x_3 + x_4/2 = 8$

c) $\bar{x} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{49}{6} = 8,1\bar{6}$

d) Da Modus$\neq$ Median $\neq$arithmetisches Mittel gilt handelt es sich um eine schiefe Verteilung.

3 Streuungsmaße

Formeln:

Spannweite: $SW = x_{max} - x_{min}$

Durchschnittliche absolute Abweichung: $AD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

Varianz: $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

Varianz klassierter Merkmale: $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 * n_i}{n-1}$

Aufgabe 3.1)
In der Vorlesung wurden im Sommersemester folgende Teilnehmerzahlen festgestellt. Bestimmen Sie

90	75	80	90	100	115	80	89	91

a) die Spannweite
b) die durchschnittliche absolute Abweichung
c) die Standardabweichung
d) die Varianz

Lösung: a) $SW = 115 - 75 = 40$

b) 1. Schritt: $\overline{x}$ bestimmen: $\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{810}{9} = 90$

2. Schritt: $|x_i - \overline{x}|$ bestimmen:

x_i	90	75	80	90	100	115	80	89	91	$\sum$
$\lvert x_i - x \rvert$	0	15	10	0	10	25	10	1	1	72

3. Schritt: AD bestimmen: $AD = \frac{\sum_{i=1}^{9} |x_i - \overline{x}|}{9} = \frac{72}{9} = 8$

d)

x_i	90	75	80	90	100	115	80	89	91	$\sum$
$\lvert x_i - x \rvert$	0	15	10	0	10	25	10	1	1	72
$(x_i - \overline{x})^2$	0	225	100	0	100	625	100	1	1	1152

$s^2 = \frac{\sum_{i=1}^{9} (x_i - \overline{x})^2}{8} = \frac{1152}{8} = 144$

c) $Standardabweichung = s = \sqrt{s^2} = \sqrt{144} = 12$

Aufgabe 3.2)
Ein Experimentator braucht für die Zeitmessung in einem Wahrnehmungsexperiment eine elektronische Uhr, die im Bereich von 200 Millisekunden möglichst genau geht. Er findet in der Requisietenkammer 2 Geräte. Er nimmt bei den Physikern eine kleine Vergleichsprobe mit einer Atomuhr und erhält folgende Werte:

	1	2	3	4	5	6	7	8	9	10	11
Uhr1	198	200	203	200	201	198	202	200	199	198	201
Uhr2	202	201	200	201	201	203	200	202	201	201	199

Bestimmen Sie bitte Mittelwert und Streuung jeder Uhr. Ist eine von beiden besser geeignet?

Lösung: 1. Schritt: Mittelwerte bestimmen:

$$\overline{x}_{Uhr1} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{2200}{11} = 200$$
$$\overline{x}_{Uhr2} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{2211}{11} = 201$$

2. Schritt: $(x_i - \overline{x}_{Uhr})^2$ bestimmen:

	1	2	3	4	5	6	7	8	9	10	11
Uhr1	198	200	203	200	201	198	202	200	199	198	201
Uhr2	202	201	200	201	201	203	200	202	201	201	199
$(x_i - \overline{x}_{Uhr1})^2$	4	0	9	0	1	4	4	0	1	4	1
$(x_i - \overline{x}_{Uhr2})^2$	1	0	1	0	0	4	1	1	0	0	4

3. Schritt: s bestimmen:

$$s^2_{Uhr1} = \frac{\sum_{i=1}^{11}(x_i-\overline{x}_{Uhr1})^2}{10} = \frac{28}{10} = 2.8 \Rightarrow s = \sqrt{2.8} \approx 1.67$$

$$s^2_{Uhr2} = \frac{\sum_{i=1}^{11}(x_i-\overline{x}_{Uhr2})^2}{10} = \frac{12}{10} = 1.2 \Rightarrow s = \sqrt{1.2} \approx 1.10$$

Ergebnis: Da Uhr2 eine geringere Streuung hat, ist die besser geeignet.

Aufgabe 3.3)
Gegeben sind die folgenden Werte:

10	8	7	7	7	9	8

a) Bestimmen Sie die Spannweite
b) Bestimmen Sie die durchschnittliche absolute Abweichung
c) Bestimmen Sie die Standardabweichung

Lösung: a) $SW = 10 - 7 = 3$

b) 1. Schritt: $\overline{x}$ bestimmen:

$\overline{x} = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{56}{7} = 8$

2.Schritt: $|x_i - \overline{x}|$ bestimmen:

x_i	10	8	7	7	7	9	8
$\lvert x_i - \overline{x}\rvert$	2	0	1	1	1	1	0

3. Schritt: AD bestimmen: $AD = \frac{\sum_{i=1}^{7} |x_i - \overline{x}|}{7} = \frac{6}{7} \approx 0.86$

c)

x_i	10	8	7	7	7	9	8
$(x_i - \overline{x})^2$	4	0	1	1	1	1	0

$s^2 = \frac{\sum_{i=1}^{7} (x_i - \overline{x})^2}{6} = \frac{8}{6} = 1.\overline{3} \Rightarrow s \approx 1.15$

Aufgabe 3.4)
Eine Diskothek hatte von Montag bis Sonntag die folgende Gästezahl:

90	80	80	90	100	110	80

a) Bestimmen Sie die Spannweite
b) Bestimmen Sie die durchschnittliche absolute Abweichung
c) Bestimmen Sie die Standardabweichung
d) Bestimmen Sie die Varianz.

Lösung: a) $SW = 110 - 80 = 30$

b) 1. Schritt: $\overline{x}$ bestimmen: $\bar{x} = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{630}{7} = 90$

2. Schritt: $|x_i - \overline{x}|$ bestimmen:

x_i	90	80	80	90	100	110	80	$\sum$
$\lvert x_i - \bar{x}\rvert$	0	10	10	0	10	20	10	60

3. Schritt: AD bestimmen: $AD = \frac{\sum_{i=1}^{7} |x_i - \overline{x}|}{7} = \frac{60}{7} \approx 8.57$

d)

x_i	90	80	80	90	100	110	80	$\sum$
$\lvert x_i - \bar{x}\rvert$	0	10	10	0	10	20	10	60
$(x_i - \overline{x})^2$	0	100	100	0	100	400	100	800

$s^2 = \frac{\sum_{i=1}^{7}(x_i - \overline{x})^2}{7} = \frac{800}{7} \approx 114.29$

c) $Standardabweichung = s = \sqrt{s^2} = \sqrt{114.29} \approx 10.69$

Aufgabe 3.5)
Gegeben sind die folgenden Wert:

11	9	8	8	8	10	9

a) Bestimmen Sie den Median.
b) Bestimmen Sie den Modus.
c) Bestimmen Sie das arithmetische Mittel.
d) Bestimmen Sie die Spannweite.
e) Bestimmen Sie die durchschnittliche absolute Abweichung.
f) Bestimmen Sie die Standardabweichung.

Lösung: Liste sortieren $\Rightarrow$

8	8	8	9	9	10	11

a) $ZW_u = x_{(7+1)/2} = x_4 = 9$

b) Modus = am häufigsten vorkommender Wert = 8

c) $\overline{x} = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{63}{7} = 9$

d) $SW = 11 - 8 = 3$

e)

x_i	8	8	8	9	9	10	11
$\lvert x_i - \overline{x}\rvert$	1	1	1	0	0	1	2

$AD = \frac{\sum_{i=1}^{7}\lvert x_i - \overline{x}\rvert}{7} = \frac{6}{7} \approx 0.86$

f)

x_i	8	8	8	9	9	10	11
$(x_i - \overline{x})^2$	1	1	1	0	0	1	4

$s^2 = \frac{\sum_{i=1}^{7}(x_i - \overline{x})^2}{6} = \frac{8}{6} = 1.\overline{3} \Rightarrow s \approx 1.15$

4 Korrelation

Formeln:

PRODUKT-MOMENT-KORRELATION: $r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}}$$

$$r = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} = \sqrt{a_{\hat{y}x} * a_{\hat{x}y}}$$

Aufgabe 4.1)
Die Kurzversion eines Intelligenztests soll geprüft werden. Dazu wurde sie 7 Versuchspersonen vorgelegt, die folgende Leistungen wurde erzielt:

Vp	1	2	3	4	5	6	7
$x_i(lang)$	50	49	52	51	49	48	51
$y_i(kurz)$	11	10	12	10	8	9	10

Berechnen Sie den Korrelationskoeffizienten. Ist die Kurzversion eine gute Möglichkeit, eine schnelle Abschätzung der Langversion zu bekommen.

LÖSUNG:
$\overline{x} = 50$
$\overline{y} = 10$

Vp	1	2	3	4	5	6	7	$\sum$
$x_i(lang)$	50	49	52	51	49	48	51	350
$y_i(kurz)$	11	10	12	10	8	9	10	70
$x_i - \overline{x}$	0	1	2	1	1	2	1	
$y_i - \overline{y}$	1	0	2	0	2	1	0	
$(x_i - \overline{x})^2$	0	1	4	1	1	4	1	12
$(y_i - \overline{y})^2$	1	0	4	0	4	1	0	10
$(x_i - \overline{x}) * (y_i - \overline{y})$	0	0	4	0	2	2	0	8

$$r = \frac{\sum_{i=1}^{7}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{7}(x_i - \overline{x})^2 \sum_{i=1}^{7}(y_i - \overline{y})}} = \frac{8}{\sqrt{12*10}} = \frac{8}{\sqrt{120}} \approx 0.73$$

Da ein positiver Zusammenhang zwischen x und y besteht, ist die Kurzversion gut geeignet.

Aufgabe 4.2)
Die Kurzversion eines Intelligenztests soll geprüft werden. Dazu wurde sie 7 Versuchspersonen vorgelegt, die die folgende Leistungen erzielten:

Vp	1	2	3	4	5	6	7
$x_i(lang)$	60	59	62	61	59	58	61
$y_i(kurz)$	12	11	13	11	9	10	11

Berechnen Sie den Korrelationskoeffizienten. Ist die Kurzversion eine gute Möglichkeit, eine schnelle Abschätzung der Langversion zu bekommen ?

Lösung:
$\overline{x} = 60$
$\overline{y} = 11$

Vp	1	2	3	4	5	6	7	$\sum$
$x_i(lang)$	60	59	62	61	59	58	61	420
$y_i(kurz)$	12	11	13	11	9	10	11	77
$x_i - \overline{x}$	0	1	2	1	1	2	1	
$y_i - \overline{y}$	1	0	2	0	2	1	0	
$(x_i - \overline{x})^2$	0	1	4	1	1	4	1	12
$(y_i - \overline{y})^2$	1	0	4	0	4	1	0	10
$(x_i - \overline{x}) * (y_i - \overline{y})$	0	0	4	0	2	2	0	8

$$r = \frac{\sum_{i=1}^{7}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{7}(x_i-\overline{x})^2 \sum_{i=1}^{7}(y_i-\overline{y})}} = \frac{8}{\sqrt{12*10}} = \frac{8}{\sqrt{120}} \approx 0.73$$

Da ein positiver Zusammenhang zwischen x und y besteht, ist die Kurzversion gut geeignet.

Aufgabe 4.3)
Zwei Variablen X und Y korrelieren zu $\omega_{xy} = 0.76$ miteinander. Die Steigung der Regressionsgeraden für die Voraussage von Y aufgrund von X ist $a_{\hat{y}x} = 0.304$. Wie lautet die Steigung $a_{\hat{x}y}$ der Regressionsgeraden für die Voraussetzung von X aufgrund von Y?

Lösung: $r = \sqrt{a_{\hat{y}x} * a_{\hat{x}y}} \Rightarrow 0.76 = \sqrt{0.304 * a_{\hat{x}y}} \Rightarrow 0.5776 = 0.304 * a_{\hat{x}y} \Rightarrow a_{\hat{x}y} = 1.9$

Die Steigung der Regressionsgeraden lautet 1.9.

5 Regression

Formeln:

Multiplikative Konstante: $a_{\hat{y}x} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$

$a_{\hat{y}x} = \frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}$

Additive Konstante: $b_{\hat{y}x} = \frac{\sum_{i=1}^{n}y_i}{n} - a\frac{\sum_{i=1}^{n}x_i}{n}$

$b_{\hat{y}x} = \frac{n\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_iy_i}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}$

Aufgabe 5.1)
In einer studentischen Lerngruppe entstand die Hypothese, dass zwischen dem Umfang der Prüfungsvorbereitung und dem Prüfungsergebnis ein Zusammenhang besteht. Man wollte für die nächste Prüfung ein gut anstreben. Deshalb wurden die Noten der letzten Prüfung mit dem individuellen Prüfungsaufwand in Tagen in einer Tabelle zusammengestellt. Wie viele Tage sind für ein gut zu investieren? (Aus rechenökonomischen Gründen wurde auf Drittel-Noten verzichtet).

Student		1	2	3	4	5	6	7	8	9
Note	x_i	2.5	1.0	3.0	4.0	4.0	3.0	2.0	3.5	4.0
Tage	y_i	11	21	4	11	2	9	11	9	1

Lösung:

Student		1	2	3	4	5	6	7	8	9	$\sum$
Note	x_i	2.5	1.0	3.0	4.0	4.0	3.0	2.0	3.5	4.0	27
Tage	y_i	11	21	4	11	2	9	11	9	12	90
	x_iy_i	27.5	21	12	44	8	27	22	31.5	48	241
	x_i^2	6.25	1	9	16	16	9	4	12.25	16	89.5

$a_{\hat{y}x} = \frac{9\sum_{i=1}^{9}x_iy_i-\sum_{i=1}^{9}x_i\sum_{i=1}^{9}y_i}{9\sum_{i=1}^{9}x_i^2-(\sum_{i=1}^{9}x_i)^2} = \frac{9*241-27*90}{9*89.5-27^2} = \frac{2169-2430}{805.5-729} = \frac{-261}{76.5} \approx -3.41$

$\Rightarrow \hat{y} = -3.41 * x + b$

Mit $\bar{x} = \frac{27}{9} = 3$ und $\bar{y} = \frac{90}{9} = 10$ folgt:

$b = \bar{y} - a\bar{x} \Rightarrow b = 10 + 3.41 * 3 = 20.23$

$\Rightarrow \hat{y} = -3.41x + 20.23$

Um also ein gut (2.0) zu erreichen muss man also $\hat{y} = -3.41 * 2.0 + 20.23 = 13.41$ Tage investieren.

Aufgabe 5.2)
Zwischen Körpergröße und Gewicht scheint ein Zusammenhang zu bestehen. In der folgenden Liste sind die entsprechenden Daten von 10 Personen aufgeführt. Bestimmen Sie bitte die Regressionsgerade, um von der Größe auf das Gewicht zu schließen.

Vp	1	2	3	4	5	6	7	8	9	10
Größe	182	192	175	181	173	181	181	180	183	176
Gewicht	78	91	71	68	75	73	75	70	65	78

Lösung:
$\bar{x} = \frac{1804}{10} = 180.4$

$\bar{y} = \frac{744}{10} = 74.4$

	1	2	3	4	5	6	7	8	9	10	$\sum$
x_i	182	192	175	181	173	181	181	180	183	176	1804
y_i	78	91	71	68	75	73	75	70	65	78	744
$(x_i - \bar{x})$	1.6	11.6	-5.4	0.6	-7.4	0.6	0.6	-0.4	2.6	-4.4	
$(x_i - \bar{x})^2$	2.56	134.56	29.16	0.36	54.76	0.36	0.36	0.16	6.76	19.36	248.4
$(y_i - \bar{y})$	3.6	16.6	-3.4	-6.4	0.6	-1.4	0.6	-4.4	-9.4	3.6	
$(x_i - \bar{x})(y_i - \bar{y})$	5.76	192.56	18.36	-3.84	-4.44	-0.84	0.36	1.76	-24.44	-15.84	169.4

$$a_{\hat{y}x} = \frac{\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10}(x_i - \bar{x})^2} = \frac{169.4}{248.4} \approx 0.68$$

$b = \bar{y} - a\bar{x} = 74.4 - 0.68 * 180.4 = -48.27$

$\Rightarrow \hat{y} = 0.68x - 48.27$

Aufgabe 5.3)
Zwischen Körpergröße und Gewicht scheint ein Zusammenhang zu bestehen. In der folgenden Liste sind die entsprechenden Daten von 9 Personen aufgeführt. Bestimmen Sie bitte die Regressionsgerade, um von der Größe auf das Gewicht zu schließen.

Vp		1	2	3	4	5	6	7	8	9
Größe	x_i	182	191	175	182	179	180	184	173	183
Gewicht	y_i	75	89	71	68	75	73	76	71	68

a=	b=	Gewicht= $*$Größe

Lösung:
$\bar{x} = \frac{1629}{9} = 181$

$\bar{y} = \frac{666}{9} = 74$

	1	2	3	4	5	6	7	8	9	$\sum$
x_i	182	191	175	182	179	180	184	173	183	1629
y_i	75	89	71	68	75	73	76	71	68	666
$(x_i - \bar{x})$	1	10	-6	1	-2	-1	2	-8	2	
$(x_i - \bar{x})^2$	1	100	36	1	4	1	4	64	4	215
$(y_i - \bar{y})$	1	15	-3	-6	1	-1	2	-3	-6	
$(x_i - \bar{x})(y_i - \bar{y})$	1	150	18	-6	-2	1	4	24	-12	178

$a_{\hat{y}x} = \frac{\sum_{i=1}^{9}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{9}(x_i-\bar{x})^2} = \frac{178}{215} \approx 0.83$

$b = \bar{y} - a\bar{x} = 74 - 0.83 * 181 = -76.23$

$\Rightarrow Gewicht = 0.83 * Groesse - 76.23$

6 Kombinatorik

Formeln:

Permutation ohne Wiederholung: $P_n = n!$

Permutation mit Wiederholung: $P_n = \frac{n!}{\prod_{i=1}^{c} k_i!}$

Variation ohne Wiederholung: $V_n = \frac{n!}{(n-k)!}$

Variation mit Wiederholung: $V_n = n^k$

Kombination ohne Wiederholung: $C_n = \binom{n}{k}$

Kombination mit Wiederholung: $C_n = \binom{n+k-1}{k}$

Aufgabe 6.1)
Wie viele verschiedene Wörter lassen sich aus den Buchstaben der Stadt "toronto" bilden (das vorgegebene Wort mitgezählt)?

LÖSUNG: $k_1 = t$; $k_2 = o$; $k_3 = r$; $k_4 = n$; $n = 7$

$\Rightarrow P_n = \frac{7!}{2!*3!*1!*1!} = \frac{5040}{12} = 420$

Aus "toronto" lassen sich 420 verschiedene Wörter bilden.

Aufgabe 6.2)
Wie viele verschiedene Wörter lassen sich aus den Buchstaben des Begriffs "mississippi" bilden (das vorgegebene Wort mitgezählt)?

LÖSUNG: $k_1 = m$; $k_2 = i$; $k_3 = s$; $k_4 = p$; $n = 11$

$\Rightarrow P_n = \frac{11!}{1!*4!*4!*2!} = \frac{39916800}{1152} = 34650$

Aus "mississippi" lassen sich 34650 verschiedene Wörter bilden.

Aufgabe 6.3)
Wie viele verschiedene Wörter lassen sich aus den Buchstaben des Begriffs "barbara" bilden (das vorgegebene Wort mitgezählt)?

LÖSUNG: $k_1 = b$; $k_2 = a$; $k_3 = r$; $n = 7$

$\Rightarrow P_n = \frac{7!}{2!*3!*2!} = \frac{5040}{24} = 210$

Aus "barbara" lassen sich 210 verschiedene Wörter bilden.

Aufgabe 6.4)
Eine Bank hat 1.998.000 Kunden mit EC-Karte. Jede EC-Karte hat eine vierstellige Geheimnummer. Allerdings sind keine Karten mit vier gleichen Ziffern ausgegeben worden. Wie viele Kunden haben jeweils die gleiche Geheimnummer?

LÖSUNG: Für jede Stelle der Geheimnummer gibt es es n=10 Ziffern $\Rightarrow V_n = 10^4 = 10000$

Keine Geheimnummer mit vier gleichen Ziffern $\Rightarrow V_n = 10000 - 10 = 9990$

$\Rightarrow \frac{1998000\, Kunden}{9990\, Nummern} = 200\, Kunden/Nummer$

Daraus folgt, dass jeweils 200 Kunden die gleiche Geheimnummer haben.

Aufgabe 6.5)
Welche Augensumme kommt beim Werfen mit zwei Würfeln am häufigsten vor. Wie hoch ist dafür die Wahrscheinlichkeit?

Lösung: Tabelle zu allen möglichen Augensummen:

x	1	2	3	4	5	6
1	2	3	4	5	6	**7**
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Meist vorkommender Wert ist 7 (6mal).

$p(X) = \frac{|G|}{|M|} = \frac{6}{36} = \frac{1}{6}$

Aufgabe 6.6)
Wie groß ist die Wahrscheinlichkeit, mit zwei Würfeln die Augensumme 4 zu erhalten?

Lösung: Tabelle zu allen möglichen Augensummen:

x	1	2	3	4	5	6
1	2	3	4	5	6	**7**
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Der Wert 4 kommt drei mal vor $\Rightarrow p(X) = \frac{|G|}{|M|} = \frac{3}{36} = \frac{1}{12}$

Aufgabe 6.7)
Ein Pokerspiel enthält 52 Karten. Jeder Mitspieler erhält 5 Karten. Wie viele verschiedene Spiele kann der erste Spieler bekommen?

Lösung: $V_n = \frac{52!}{(52-5)!} = \frac{52!}{47!} = 311875200$

Der erste Spieler kann 311875200 verschiedene Spiele bekommen.

Aufgabe 6.8)
Ein Skatspiel enthält 32 Karten. Jeder der drei Spieler erhält 10 Karten. Wie viele verschiedene Spiele kann der erste Spieler bekommen?

Lösung: $V_n = \frac{32!}{(32-10)!} = \frac{32!}{22!} \approx 2.34 * 10^{14}$

Der erste Spieler kann $2.34 * 10^{14}$ verschiedene Spiele bekommen.

Aufgabe 6.9)
Wie groß ist die Wahrscheinlichkeit nicht die Augensumme 4 zu erhalten?

Lösung: Tabelle zu allen möglichen Augensummen:

x	1	2	3	4	5	6
1	2	3	4	5	6	**7**
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$p(X) = \frac{3}{36} = \frac{1}{12}$

$q(X) = 1 - p(X) = 1 - \frac{1}{12} = \frac{11}{12}$

7 Diskrete Verteilungen

Formeln:

1. BINOMIALVERTEILUNG: $f(x|n;p) = \binom{n}{x} p^x q^{n-x}$

MITTELWERT: $\mu = n * p$

STREUUNG: $\sigma^2 = n * p * q$

2. NORMALVERTEILUNG: $f(x|n;p) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})}$

MITTELWERT⇒ARITHMETISCHES MITTEL

STREUUNG ⇒VARIANZ

STICHPROBENFEHLER: $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$

3. HYPERGEOMETRISCHE VERTEILUNG: $f(x|N;n;R) = \binom{R}{x}\binom{N-R}{n-x}/\binom{N}{n}$

Aufgabe 7.1)
Ein Student nimmt an einer Klausur teil. Sie umfasst 9 Aufgaben, von denen jede zwei Antwortalternativen bietet. Die richtige ist anzukreuzen. Wie groß ist die Wahrscheinlichkeit, dass der Student per Zufall 6 "richtige" Antworten gibt?

LÖSUNG: Binomialverteilung mit $n = 9$ Aufgaben, $x = 6$ Richtige, $p = \frac{1}{2}$ (da jeweils ja oder nein ankreuzbar)

$\Rightarrow f(6|9; \frac{1}{2}) = \binom{9}{6} (\frac{1}{2})^6 (1 - \frac{1}{2})^{9-6} = 84 * \frac{1}{64} * \frac{1}{8} \approx 0.16$

Aufgabe 7.2)
10 Architekten/Innen beteiligen sich an der Ausschreibung für den Bau eines neuen Museums. Drei von ihnen sind Frauen. Die drei besten Vorschläge werden den Bürgern des Ortes zur Wahl vorgelegt. Wie groß ist die Wahrscheinlichkeit, dass eine weibliche Architektin unter den besten drei Entwürfen ist?

LÖSUNG: Hypergeometrische Verteilung mit $n = 3$ (unter den ersten drei), $x = 1$(eine Architektin), $N = 10$ (insg. 10 Architekten), $R = 3$ (drei Frauen insg.)

$\Rightarrow f(1|10;3;3) = \binom{3}{1} \binom{10-3}{3-1} / \binom{10}{3} = \frac{3*21}{120} = \frac{63}{120} = 0.525$

Aufgabe 7.3)
Man wirft eine Münze 196 mal und erhält eine Verteilung.

a) Um was für eine Verteilung handelt es sich?
b) Bestimmen Sie ihren Mittelwert.
c) Bestimmen Sie ihre Standardabweichung.

Lösung: a) Es handelt sich um eine Binomialverteilung.

b)$\mu = 196 * \frac{1}{2} = 98$

c) $\sigma = \sqrt{196 * \frac{1}{2} * \frac{1}{2}} = 7$

Aufgabe 7.4)
Die Wahrscheinlichkeit für den erfolgreichen Start einer Satelliten-Trägerrakete sei p = 0.85. Wie groß ist die Wahrscheinlichkeit, bei 10 Starts mindestens 8 Erfolge verbuchen zu können?

Lösung: Binomialverteilung mit $n = 10$Starts, $x = 8$ Erfolge, $p = 0.85$

$\Rightarrow f(8|10; 0.85) = \binom{10}{8}(0.85)^8(1-0.85)^{10-8} \approx 45 * 0.2725 * 0.0225 \approx 0.276$

8 Stetige Verteilungen

Formeln:

z-Transformation: $z = \frac{x-\overline{x}}{s}$; $x = \overline{x} + s * z$

χ^2-Test: $\chi^2 = \sum_{k=1}^{m} \frac{(f_{beob,k} - f_{erw,k})^2}{f_{erw,k}}$

Aufgabe 8.1)
Peter hat 25 Punkte in einem Test A erzielt, Paul 24 Punkte in einem Test B. Beide Tests messen das gleiche Merkmal. Wer von beiden erzielte die bessere Leistung, wenn gilt

	$\overline{x}$	s
A	20	5
B	18	6

Lösung:

$z_{Peter} = \frac{25-20}{5} = \frac{5}{5} = 1$

$z_{Paul} = \frac{24-18}{6} = \frac{6}{6} = 1$

Daraus folgt, dass Peter und Paul gleich gut abgeschnitten haben.

Aufgabe 8.2)
Maria hat an einem Dynamometer 43 kg Zugleistung erbracht, Manfred 52 kg. Die geschlechtsspezifische Vergleichsgruppe für Maria hat einen Mittelwert von 31 kg und eine Standardabweichung von 6 kg. Die entsprechende männliche Norm liegt bei 40 kg Zug und 8 kg Standardabweichung. Wer von beiden ist relativ zu seiner Gruppennorm besser?

Lösung:

$z_{Maria} = \frac{43-31}{6} = \frac{12}{6} = 2$

$z_{Manfred} = \frac{52-40}{8} = \frac{12}{8} = 1.5$

Daraus folgt, dass Maria relativ zu ihrer Gruppennorm besser als Manfred abschneidet.

Aufgabe 8.3)
Um einen Kurs über umweltfreundliche Autofahrer zu benutzen, wurde der durchschnittliche wöchentliche Benzinverbrauch der Teilnehmer ermittelt

$$M_{vor} = 25;\ s_{vor} = 5;\ M_{nach} = 21;\ s_{nach} = 6$$

Der Benzinverbrauch wird als normalverteilt angenommen. Frau A verbrauchte vorher 15l und nachher 12l pro Woche. Hat sie ihren Verbrauch im Vergleich zu den durchschnittlichen Autofahrern gesenkt oder erhöht?

Lösung:

$z_{vor} = \frac{15-25}{5} = \frac{-10}{5} = -2$

$z_{nach} = \frac{12-21}{6} = \frac{9}{6} = -1.5$

Daraus folgt dass sich der Verbrauch der Frau erhöht hat.

Aufgabe 8.4)
In einem Entwicklungstest soll der kleine Fritz 120 Kugeln so in sechs gleichgroße Fächer verteilen, dass in jedem Fach gleich viele Kugeln sind. Zum Schluss klatscht er in die Hände und ruft “fertig!”. Der Versuchsleiter notiert folgendes Ergebnis:

18	19	22	20	23	17

Sind die Abweichungen von der zu erwartenden Gleichverteilung noch mit dem Zufall ($\alpha = 0.05$) zu erklären, oder ist Fritz dieser Aufgabe (noch) nicht gewachsen?

LÖSUNG:

k	$f_{beo,k}$	$f_{erw,k}$	$f_{beo,k} - f_{erw,k}$	$(f_{beo,k} - f_{erw,k})^2$	$(f_{beo,k} - f_{erw,k})^2/f_{erw,k}$
1	18	20	-2	4	0.2
2	19	20	-1	1	0.05
3	22	20	2	4	0.2
4	20	20	0	0	0
5	23	20	3	9	0.45
6	17	20	-3	9	0.45

$\chi^2 = \sum_{k=1}^{m}(f_{beo,k} - f_{erw,k})^2/f_{erw,k} = 1.35$

$df = m - 1 = 5$

$\alpha = 0.05 \Rightarrow 1 - \alpha = 0.95$

Aus der Chi-Quadrat-Test Tabelle folgt: $\chi^2(0.95; 5) = 11.07$

$\chi^2 = 1.35 < 11.07 \Rightarrow$Die Abweichungen von der zu erwartenden Gleichverteilung sind durch den Zufall zu erklären.

Aufgabe 8.5)
Sechs Bücher belegen die ersten Plätze der Bestenliste im Feuilleton einer überregionalen Zeitung. Dazu wurden 150 Buchhändler nach dem am häufigsten verkauften Buch befragt. Die Auszählung führt zu folgendem Ergebnis:

Buch	f_{beo}			
1	21			
2	23			
3	27			
4	25			
5	28			
6	26			

Bilden die Häufigkeiten eine Rangordnung, oder sind die Nennungen eigentlich gleichverteilt und die Abweichungen noch mit dem Zufall ($\alpha = 0.05$) zu erklären?
Tabellenwert: $\chi_{krit}(\alpha = 0.05, df = 5, einseitig) = 11.07$

Lösung:

Buch	f_{beo}	f_{erw}	$(f_{beo} - f_{erw})^2$	$(f_{beo} - f_{erw})^2/f_{erw}$
1	21	25	16	0.64
2	23	25	4	0.16
3	27	25	4	0.16
4	25	25	0	0
5	28	25	9	0.36
6	26	25	1	0.04
$\sum$				1.36

$\chi^2 = 1.36 < 11.07 \Rightarrow$ Die Abweichungen sind durch den Zufall zu erklären.

9 Statistische Inferenz

Formeln:

Konfidenzintervall: $\mu = \mu_M \pm z_\alpha * \sigma_M$

Aufgabe 9.1)
In einer Zufallsstichprobe von 64 Personen wurde ein durchschnittlicher Intelligenzquotient von 80 Punkten erreicht. Die Standardabweichung beträgt s = 12 Punkte. Geben Sie das 95% Vertrauensintervall an.

Lösung: $\sigma_M = \frac{s}{\sqrt{n}} = \frac{12}{\sqrt{64}} = 1.5$

$\mu_M = 80$

Um z_α zu bekommen lesen wir aus der "t-Verteilung" Tabelle ab:

$z_\alpha(\frac{1+\alpha}{2}; n-1) = z_\alpha(\frac{1+0.95}{2}; 64-1) = z_\alpha(0.975; 63)$ ab$\Rightarrow z_\alpha = 2.0$

$\Rightarrow \mu = [80 - 2.0 * 1.5; 80 + 2.0 * 1.5] = [77; 83]$

Das 95%-ige Vertrauensintervall geht von 77 bis 83.