Klausur Statistik
Lösungshinweise

Prüfungsdatum: 21. Januar 2016 – Prüfer: Etschberger, Heiden, Jansen
Studiengang: IM und BW
Punkte: 15, 15, 12, 14, 16, 18 ; Summe der Punkte: 90

Aufgabe 1 **15 Punkte**

Bei einer Umfrage unter den Kindern einer Schulklasse nach der Anzahl ihrer Geschwister ergaben sich die folgenden Daten:

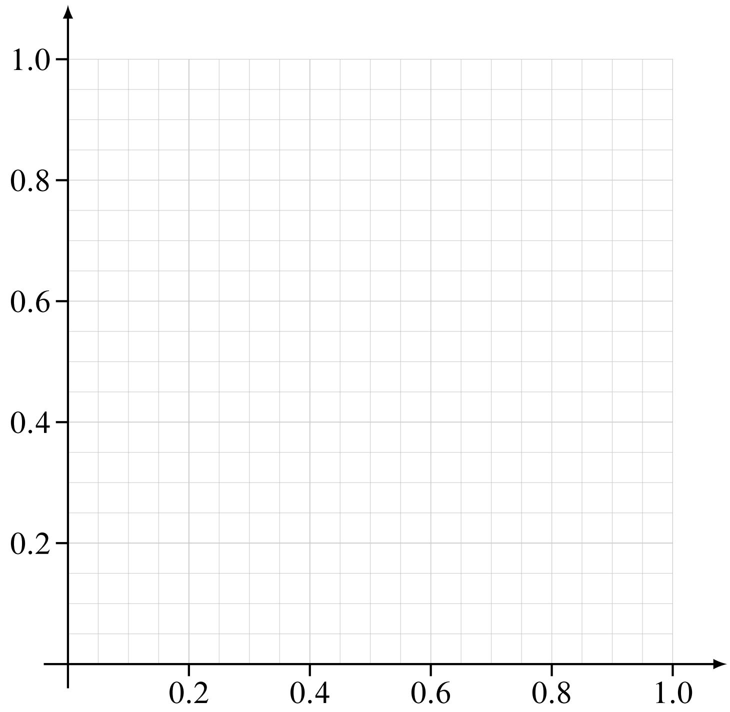
Anzahl Geschwister (a_i)	0	1	2	3	5	6
Häufigkeit (h_i)	6	4	5	2	2	1

a) Bestimmen Sie den Mittelwert, den Median, die Standardabweichung, $f(4)$ und $F(4)$ der Verteilung.

b) Erstellen Sie eine Tabelle, die die Koordinaten der Lorenzkurve enthält. Zeichnen Sie die Lorenzkurve in nebenstehendes Diagramm.

c) Angenommen, die R-Variable `x` enthält die Urliste der Daten. Wie generiert man damit die Tabelle der Häufigkeiten mit R?

d) Wie erzeugt man ein Balkendiagramm der Daten mit R?

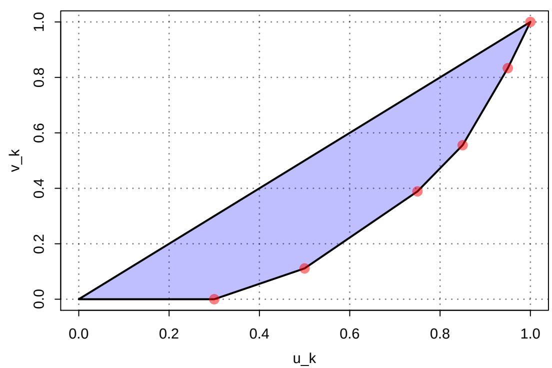


Lösungshinweis:

a) Für das arithmetische Mittel ergibt sich $\overline{x} = 1.8$. Der Median ist $x_{\text{med}} = 1.5$. Die Standardabweichung beträgt $s \approx 1.7776$, außerdem gilt $f(4) = 0$ und $F(4) = 0.85$.

b) Tabelle:

i	1	2	3	4	5	6
u_k	0.30	0.50	0.75	0.85	0.95	1.00
v_k	0.00	0.11	0.39	0.56	0.83	1.00



c)
```
table(x)
```

d)
```
barplot(table(x))
```

Aufgabe 2 **15 Punkte**

In einer Statistikklausur können die Noten 1 (entspricht einer sehr guten Leistung) bis 5 (nicht bestanden) erreicht werden. Von 15 Teilnehmern dieser Prüfung ist außer der Note noch die Anzahl des Klausurversuches in diesem Fach bekannt. Es ergeben sich folgende Daten:

Versuch	2	2	3	1	2	2	1	1	1	3	3	2	3	1	1
Note	3	5	3	1	4	5	2	3	2	2	3	2	1	2	3

a) Schreiben Sie in den folgenden Kästchen die Kontingenztabelle inklusive der Randhäufigkeiten zu den beiden Merkmalen auf.

b) Geben Sie vier R-Befehle an, mit denen man
   (1) das Merkmal Versuch in der Variablen `V` sowie
   (2) das Merkmal Note in der Variablen `N` einliest (Befehle können abgekürzt werden),
   (3) beide Merkmale in einem data-frame zusammenfasst und
   (4) eine Kontingenztabelle ausgibt.

c) Bestimmen Sie den normierten Kontingenzkoeffizienten nach Pearson.

Lösungshinweis:

a)
```
##         Note
## Versuch  1  2  3  4  5 Sum
##     1    1  3  2  0  0   6
##     2    0  1  1  1  2   5
##     3    1  1  2  0  0   4
##     Sum  2  5  5  1  2  15
```

b)
```r
V = c(2,1,1, ...) # abgekürzt
N = c(4,3,1, ...) # abgekürzt
A = data.frame(Versuch = V, Note= N)
table(A) # Kontingenztabelle
```

c) Bei Unabhängigkeit erwartete Häufigkeiten

```
##         Note
## Versuch     1     2     3     4     5
##       1 0.800 2.000 2.000 0.400 0.800
##       2 0.667 1.667 1.667 0.333 0.667
##       3 0.533 1.333 1.333 0.267 0.533
```

Beiträge zu χ^2:

```
##         Note
## Versuch     1     2     3     4     5
##       1 0.050 0.500 0.000 0.400 0.800
##       2 0.667 0.267 0.267 1.333 2.667
##       3 0.408 0.083 0.333 0.267 0.533
```

Dadurch ergibt sich $\chi^2 = 8.575$, damit $K = 0.6031$ und die normierte Variante $K^* = 0.7386$.

Aufgabe 3 **12 Punkte**

An einer Kohortenstudie zur Wirksamkeit eines Medikaments nahmen insgesamt 11 101 Personen teil. Von den Teilnehmern bekamen 6399 Personen das zu testende Medikament, der Rest bekam ein Placebo. Von allen Teilnehmern gesundeten 6624 Personen. Von den Personen, die das Medikament bekamen, gesundeten 87 nicht.

a) Stellen Sie die absoluten und relativen Häufigkeiten in einer Vierfeldtafel auf und zeichnen Sie zu den relativen Häufigkeiten das zugehörige Baumdiagramm.

b) Wie groß ist die Wahrscheinlichkeit bei einer zufällig aus den Teilnehmern der Studie ausgewählten Person, von der man nicht weiß, ob sie das Medikament eingenommen hat zu gesunden?

c) Wie groß ist die Wahrscheinlichkeit bei einer zufällig aus den Teilnehmern der Studie ausgewählten Person, von der man weiß, dass sie das Placebo eingenommen hat, nicht zu gesunden?

Lösungshinweis:

a)

	M	$\overline{\text{M}}$	Summe
G	6312	312	6624
$\overline{\text{G}}$	87	4390	4477
Summe	6399	4702	11 101

	M	$\overline{M}$	Summe
G	0,5686	0,0281	0,5967
$\overline{\text{G}}$	0,0078	0,3955	0,4033
Summe	0,5764	0,4236	1,0000

b) Die Wahrscheinlichkeit einer Person zu gesunden ist $P(G) = 0{,}5967$.

c) Die Wahrscheinlichkeit $P(\overline{G}|\overline{M}) = \frac{P(\overline{G}\cap\overline{M})}{P(\overline{M})} = \frac{4390}{4702} = 0{,}9336$

Aufgabe 4 **14 Punkte**

Es sei die Zufallsvariable X gegeben. Berechnen Sie in den Teilaufgaben a) und b) die Wahrscheinlichkeit $P(-1 \leqq X \leqq 7.6)$.

a) X ist binomialverteilt nach $B(15; 0.3)$.

b) X ist poissonverteilt mit folgender Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} \frac{\left(\frac{3}{2}\right)^x}{x!} \cdot e^{-\frac{3}{2}} & \text{falls } x = 0, 1, 2, \ldots \\ 0 & \text{sonst} \end{cases}$$

Gegeben sei nun die Zufallsvariable Y. Geben Sie für die Teilaufgaben c), d) und e) bitte jeweils ein R-Kommando an, welches die Wahrscheinlichkeit $P(Y < 5)$ berechnet.
(Hinweis: Sie müssen die Wahrscheinlichkeiten nicht berechnen, die Angabe des R-Befehls genügt)

c) Y ist binomialverteilt nach $B(60; 0.4)$. R

d) Y ist hypergeometrisch verteilt mit $N = 30$, $M = 10$ und $n = 8$. R

e) Y ist poissonverteilt mit Parameter $\lambda = 2$. R

Im Folgenden beschreibe die Zufallsvariable Z das Gewicht von Schokoladenosterhasen. Z sei normalverteilt mit den Parametern μ und σ. Es sei bekannt, dass nur 2.275 % der Osterhasen mehr als 500 g und nur 2.275 % weniger als 450 g wiegen.

f) Bestimmen Sie die Parameter μ und σ der Zufallsvariablen Z.

Lösungshinweis:

a) $P(-1 \leqq X \leqq 7.6) = P(X \leqq 7) = 0.95$

b) mit $\lambda = 1.5$ ergibt sich $F_x(7) = 0.9998$

```
# Teilaufgaben c), d), e)
pbinom(q=4, size=60, prob = 0.4) #  0,000000005244
phyper(q=4, m=10, n=30-10, k=8)  #  0.943682
ppois(4, 2)                      #  0.947347
```

f) $\mu = 475$ aus Symmetrie der Normalverteilung; Berechnung von σ: Gegeben: $\Phi(\frac{500-475}{\sigma}) = 0.97725$
$\frac{25}{\sigma} = 2 \iff \sigma = 12.5$, also $Z \sim N(475; 12.5)$

Aufgabe 5 **16 Punkte**

Die Zufallsvariable X beschreibe die Anzahl der erzielten Tore des FC Augsburg in der Saison des Jahres 2017. Es kommen nur die folgenden Ergebnisse mit den angegebenen Wahrscheinlichkeiten vor:

x	0	1	2	3	4
$P(X = x)$	0.2	0.3	0.3	0.17	0.03

Berechnen Sie damit

a) den Erwartungswert von X,

b) den Modus von X,

c) den Median von X sowie

d) den Erwartungswert von X^3.

Gegeben sei im Folgenden die gemeinsame (unvollständige) Wahrscheinlichkeitsverteilung der Zufallsvariablen Z und Y:

Z	Y			
	1	2	3	
0	**0.1**	0.2	**0.2**	**0.5**
1	0.1	0.3	**0.1**	0.5
	0.2	**0.5**	**0.3**	**1.0**

e) Vervollständigen Sie die fehlenden Werte in der Tabelle.

Bestimmen Sie jeweils den Erwartungswert und die Varianz

f) der Zufallsvariablen Y,

g) der Zufallsvariable W, welche durch $W = 3 + Z \cdot Y$ definiert sei.

Lösungshinweis:

a) $\sum_{i=0}^{4} x_i \cdot f(x) = 0 \cdot 0.2 + 1 \cdot 0.4 + 2 \cdot 0.3 + 3 \cdot 0.07 + 4 \cdot 0.03 = 1.53$

b) $x_{\text{mod}} \in 1{,}2$ Modus ist nicht eindeutig, 2 Punkte sind Modus.

c) $x_{\text{med}} = 1$ da $F(1) \geqq 0.5$.

d) $E(X^3) = 9.21$

e) Tabelle siehe oben $E(Y) = 2.1$, $Var(Y) = E(Y^2) - E(Y)^2 = (1^2 \cdot 0.2 + 2^2 \cdot 0.5 + 3^2 \cdot 0.3) - 2.1^2 = 0.49$

f)

$Z \cdot Y$	0	1	2	3
$P(Z \cdot Y)$	0.5	0.1	0.3	0.1

$E(W) = 3 + E(Z \cdot Y) = 3 + 0 \cdot 0.5 + 1 \cdot 0.1 + 2 \cdot 0.3 + 3 \cdot 0.1 = 4$

Aufgabe 6 — 18 Punkte

In einer Statistikklausur können die Noten 1 (entspricht einer sehr guten Leistung) bis 5 (nicht bestanden) erreicht werden. Von 300 Teilnehmern dieser Prüfung ist außer der Note noch die Anzahl des Klausurversuches in diesem Fach bekannt. Es ergibt sich folgende Kontingenztabelle:

	Note					
Versuch	1	2	3	4	5	Summe
1	8	42	36	26	8	120
2	3	14	23	32	12	84
3	14	42	25	14	1	96
Summe	25	98	84	72	21	300

Damit ergibt sich die Tabelle der bei Unabhängigkeit erwarteten Häufigkeiten zu:

	Note				
Versuch	1	2	3	4	5
1	**10**	39.2	33.6	28.8	8.4
2	7	27.44	**23.52**	20.16	5.88
3	8	31.36	26.88	23.04	**6.72**

a) Berechnen Sie die fehlenden Werte und tragen Sie diese in die obige Tabelle ein.

b) Testen Sie zum Signifikanzniveau von 1 %, ob die Nummer des Versuches von der Note in der Statistikklausur abhängt.

c) Wie lautet die Nullhypothese bei diesem Test?

d) Was bedeutet der Fehler zweiter Art in diesem Test?

e) Führt man den Test in R aus, ergibt sich:

R

```
chisq.test(table(A))

##
##  Pearson's Chi-squared test
##
## data:  table(A)
## X-squared = 39.92, df = 8, p-value = 0.000003311
```

Was bedeutet `X-squared`, `df`, `p-value`?

Lösungshinweis:

a) Siehe Tabelle.

b) Nullhypothese wird verworfen (Testfunktionswert $v = 39.923$, Verwerfungsbereich $B = (20.09, \infty)$).

c) H_0 : Note und Versuch sind unabhängig.

d) H_0 wird verworfen, obwohl sie stimmt. Man geht also durch die Stichprobe davon aus, dass die beiden Merkmale abhängig sind, obwohl sie nicht voneinander abhängen.

e) `X-squared`: Testfunktionswert v, `df`: Degrees of freedom (Freiheitsgrade), `p-value`: Bei jedem Signifikanzniveau, das größer als dieser p-value ist, wird die Nullhypothese abgelehnt (hier also auch).