

# Nachholklausur Statistik

## Lösungshinweise

Prüfungsdatum: 22. Januar 2015 – Prüfer: Etschberger, Heiden, Jansen  
Studiengang: IM und BW

### Aufgabe 1

10 Punkte

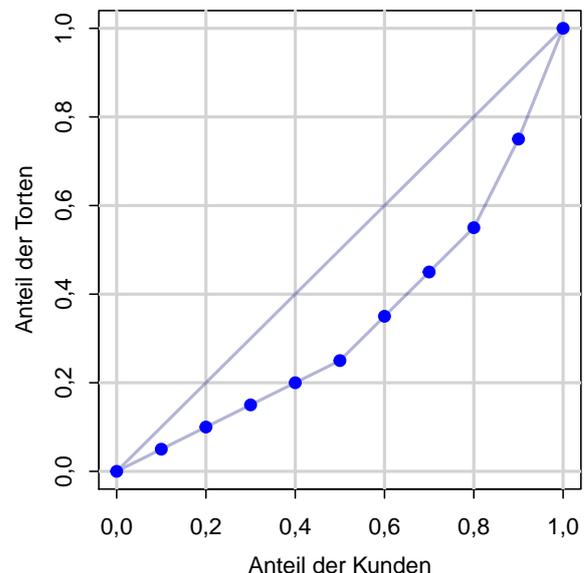
Die Firma CelebWedCake liefert zu einem Festpreis von 200.000 € eine exklusive Premium-Hochzeitstorte an Prominente. In den letzten 5 Jahren wurden insgesamt 20 von diesen Torten verkauft. Pro Kunde ist die Anzahl der verkauften Torten in dieser Zeitspanne mittels der verschiedenen Ausprägungen  $a_i$  und den zugehörigen absoluten Häufigkeiten  $h_i$  erfasst:

$i$	1	2	3	4
$a_i$	1	2	4	5
$h_i$	5	3	1	$h_4$

- Bestimmen Sie  $h_4$ .
- Zeichnen Sie die Lorenzkurve,
- berechnen Sie den normierten Gini-Koeffizienten sowie
- den Herfindahl- und
- den Exponentialindex der Anzahl der verkauften Torten pro Kunde.

Lösungshinweis:

- Insgesamt 20 Torten, laut Tabelle  $1 \cdot 5 + 2 \cdot 3 + 4 \cdot 1 = 15$  Torten für bis zu 4 Torten. Bleibt ein Kunde mit 5 Torten, also  $h_4 = 1$ .
- Lorenzkurve siehe rechts
- Zehn Kunden. Gini:  $G = 0,33$ , normiert:  $G_* = 0,3666667$
- Herfindahl: 0,145
- Exponentialindex: 0,1214601



## Aufgabe 2

10 Punkte

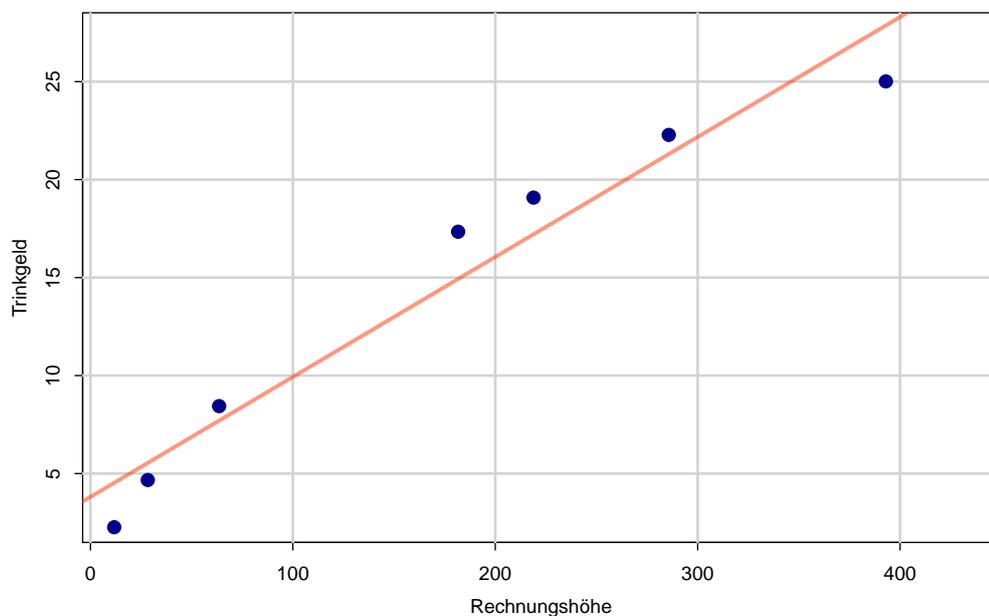
In einem Restaurant wird bei 7 Gästen die Rechnungshöhe  $R$  zusammen mit dem gegebenen Trinkgeld  $T$  erfasst. Es ergeben sich folgende Werte:

$i$	1	2	3	4	5	6	7
$R_i$	218,92	63,56	11,74	285,72	181,66	28,33	392,99
$T_i$	19,08	8,44	2,26	22,28	17,34	4,67	25,01

- Berechnen Sie einen geeigneten Korrelationskoeffizienten zwischen Rechnungshöhe und Trinkgeld. Interpretieren Sie das Ergebnis.
- Bestimmen Sie die Regressionsparameter eines linearen Modells, in dem die Höhe des Trinkgeldes in Abhängigkeit von der Rechnungshöhe beschrieben wird. Wieviel Trinkgeld würden Sie nach diesem Modell bei einer Rechnung von 200 € erwarten?
- Wie hoch ist der Determinationskoeffizient dieses Modells? Was sagt diese Größe aus?
- Zeichnen Sie die Werte zusammen mit der Regressionsgerade in ein geeignetes Koordinatensystem ein.

Lösungshinweis:

- Bravais-Pearson:  $r = 0,9738711$ .
- Modell:  $T = 0,0612153 \cdot R + 3,8095985$   
Prognose:  $T(200) = 16,05$
- Determinationskoeffizient:  $R^2 = 0,9484248$ , also stecken ca. 95 % der Information aus den Daten im Modell. (Aber (nicht bewertet): Verteilung der Residuen lässt vermuten, dass eventuell ein lineares Modell nicht optimal ist; eventuell Modell mit Krümmung besser geeignet)
- Streuplot:



### Aufgabe 3

10 Punkte

Johann wohnt bei seinen Eltern. Sein Studium und insbesondere die Begleiterscheinungen durch soziale Aktivitäten haben ihn in letzter Zeit ziemlich gefordert, so dass er eines Tages nach dem Aufwachen überhaupt nicht weiß, welcher Wochentag gerade ist.

Allerdings hat er folgendes beobachtet:

- ▶ Wenn er die Küche seiner Eltern an Sonntagen betritt erklingt mit 20 % Wahrscheinlichkeit Blasmusik aus dem Radio, an anderen Wochentagen nur mit 2 % Wahrscheinlichkeit.
  - ▶ Johanns Vater ist am Montag, Dienstag und Mittwoch immer schlecht gelaunt.
- a) Johann öffnet die Küchentüre. Es erklingt Blasmusik. Mit welcher Wahrscheinlichkeit ist gerade Sonntag?
- b) Johann stellt nach einer Minute am Küchentisch fest, dass sein Vater blendender Laune ist. Wie hoch ist jetzt die Wahrscheinlichkeit, dass Sonntag ist?

Lösungshinweis:

$B \equiv$  „Es erklingt Blasmusik“

$S \equiv$  „Es ist Sonntag“

$$\text{a) } P(S|B) = \frac{P(B|S)P(S)}{P(B|S)P(S) + P(B|\bar{S})P(\bar{S})} = \frac{0,2 \cdot 1/7}{0,2 \cdot 1/7 + 0,02 \cdot 6/7} = \frac{20}{20 + 2 \cdot 6} = 5/8 = 62,5\%$$

b) Genauso, nur ist jetzt  $P(S) = 1/4$ , denn Montag, Dienstag, Mittwoch scheiden aus. Also:

$$P(S|B) = \frac{0,2 \cdot 1/4}{0,2 \cdot 1/4 + 0,02 \cdot 3/4} = \frac{20}{20 + 2 \cdot 3} = 10/13 \approx 76,9\%$$

## Aufgabe 4

10 Punkte

Gegeben ist eine im Intervall  $[2; b]$  gleichverteilte Zufallsvariable  $X$ .

- Bestimmen Sie  $b$  so, dass für die Verteilungsfunktion  $F(x)$  gilt:  $F\left(\frac{b}{2}\right) = 0,4$ .
- Berechnen Sie  $F(4)$ .

Jetzt ist eine poissonverteilte Zufallsvariable  $Y$  gegeben, für die  $P(Y \geq 5) = 0,3$  gilt.

- Bestimmen Sie auch für  $Y$  den Wert von  $F(4)$ .

Lösungshinweis:

$$\text{a) } F\left(\frac{b}{2}\right) = \frac{\frac{b}{2} - 2}{b - 2} = 0,4 \Rightarrow b = 12$$

$$F(4) = \frac{4 - 2}{12 - 2} = \frac{2}{10} = 0,2$$

$$\text{b) } P(Y \geq 5) = 0,3 \Leftrightarrow F(4) = P(Y < 5) = 0,7$$

**Aufgabe 5****10 Punkte**

Es sei  $(X_1, \dots, X_n)$  eine einfache Stichprobe, wobei für die Stichprobenvariablen gelte:

$$E(X_i) = 1, \quad \text{Var}(X_i) = \sigma^2 \quad \text{für } i = 1, \dots, n$$

a) Berechnen Sie  $E(X_i^2)$ .

(Hinweis: Benutzen Sie den Verschiebungssatz der Varianz).

b) Nun soll die Varianz  $\sigma^2$  der Grundgesamtheit geschätzt werden. Ist die Funktion

$$\hat{\theta} = \frac{1}{n} (X_1^2 + X_2^2 + \dots + X_n^2) - 1$$

ein erwartungstreuer Schätzer für  $\sigma^2$ ?

Lösungshinweis:

a)  $\text{Var}(X_i) = \sigma^2 = E(X_i^2) - E^2(X_i) \Leftrightarrow \sigma^2 = E(X_i^2) - 1 \Leftrightarrow E(X_i^2) = \sigma^2 + 1$

b)  $E(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - 1 = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + 1) - 1 = \sigma^2$ , also ist  $\hat{\theta}$  erwartungstreu

Ein Getränkehersteller füllt Limonade in 700ml Flaschen ab. Die Anlage ist auf ein Abfüllvolumen von 702ml eingestellt. Dem Hersteller ist daran gelegen, dass das Abfüllvolumen in den Flaschen dem eingestellten Volumen entspricht. Bei einer Stichprobe von 17 zufällig ausgewählten Flaschen betrug das Abfüllvolumen 701,7ml, bei einer Standardabweichung von 1ml. Das Abfüllvolumen in der Grundgesamtheit kann als normalverteilt angesehen werden.

- Prüfen Sie mit einem Signifikanzniveau von 2,5%, ob das Abfüllvolumen  $\mu$  dem eingestellten Volumen entspricht.
- Prüfen Sie mit einem Signifikanzniveau von 5%, ob das Mindest-Abfüllvolumen von 700ml eingehalten wird. Der statistische Nachweis soll dabei vom Getränkehersteller erbracht werden!

Lösungshinweis:

- Schritt 1** Erstellen der Hypothese:  $H_0 : \mu = \mu_0 = 702\text{ml}$ ,  $H_1 : \mu \neq \mu_0$ .

**Schritt 2** Verteilungsform und Standardabweichung:

$X$  normalverteilt mit unbekanntem  $\sigma^2$ , daraus folgt  $t$ -verteilt mit  $n - 1$  Freiheitsgraden.  $\hat{\sigma}_{\bar{X}} = \frac{1\text{ml}}{\sqrt{17}} \approx 0,24\text{ml}$ .

**Schritt 3** Signifikanzniveau 2,5%

**Schritt 4** Beibehaltungsbereich

$$\begin{aligned} [702 - t \cdot \hat{\sigma}_{\bar{X}}, 702 + t \cdot \hat{\sigma}_{\bar{X}}] &= [702 - 2,473 \cdot 0,24, 702 + 2,473 \cdot 0,24] \\ &= [701,4; 702,6] \end{aligned}$$

**Schritt 5** Entscheidung:  $H_0$  wird beibehalten

- Schritt 1** Erstellen der Hypothese:  $H_0 : \mu < \mu_0 = 700\text{ml}$ ,  $H_1 : \mu \neq \mu_0$ .

**Schritt 2** siehe 1. Teil.

**Schritt 3** Signifikanzniveau 5%

**Schritt 4** Ermittlung Beibehaltungsbereich:

$$\begin{aligned} [0; 700 + t \cdot \sigma_{\bar{X}}] &= [0; 700 + 1,746 \cdot 0,24] \\ &= [0; 700,4] \end{aligned}$$

**Schritt 5** Ablehnung von  $H_0$ .

## Aufgabe 7

10 Punkte

Es ist jeweils genau eine Antwort pro Aufgabenteil richtig. Für jede richtige Antwort gibt es 2 Punkte, jede falsch beantwortete gibt  $-0,5$  Punkte, eine nicht beantwortete Frage 0 Punkte. Man kann nicht weniger als 0 Punkte in der Aufgabe bekommen. Tragen Sie ihre Lösungsbuchstaben in die folgenden Kästchen ein. Bitte schreiben sie deutlich!

Aufgabenteil	1	2	3	4	5
Lösung	<input type="checkbox"/>				

### Aufgabenteil 1

Was liefert die folgende R-Eingabe?

```
(2^3:1) + seq(10,17,by = 2) - 1
```

- (A) [1] 17 18 19 20 13 14 15 16
- (B) [1] 17 17 17 17 17 17 17 17
- (C) [1] 17 14 13 20 17 16 23 20
- (D) [1] 10 13 16 19 14 17 20 23
- (E) [1] 10 19 14 23

### Aufgabenteil 3: Matrizen

Gegeben sei eine Matrix M. Mit Hilfe von `cbind()` und `rbind()` vergrößern Sie die Matrix wie unten angegeben.

```
M <- matrix(1:9, nrow = 3, byrow = TRUE)
M <- cbind(M,1:3)
M <- rbind(c(2,3),M)
```

Welche Ausgabe erwarten Sie danach bei der Eingabe

```
M[1,]*M[,2]
```

- (A) [1] 1 2 3 4 2 4 6 8
- (B) [1] 40
- (C) [1] 2 10 24 4
- (D) [1] 6 6 10 24
- (E) [1] 2 3 8 21

### Aufgabenteil 2

Welche Ausgabe erwarten Sie wenn Sie die folgende Zeile eingeben?

```
sum(table(rep(1:5,3)))
```

- (A) [1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
- (B) [1] 3 3 3 3 3
- (C) [1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5
- (D) [1] 15
- (E) [1] 18

### Aufgabenteil 4: Funktionen

Gegeben ist folgende Funktion:

```
myFunction <- function(x, y){
  rval <- if(x>y) {
    (x-y)^2
  } else {
    1
  }
  return(rval)
}
```

Welche Ausgabe erwarten Sie für die beiden Aufrufe

```
myFunction(1,1)
myFunction(8,5)
```

- (A) [1] 0 und [1] 3
- (B) [1] 1 und [1] 1
- (C) [1] 0 und [1] 13
- (D) [1] 0 und [1] 9
- (E) [1] 1 und [1] 9

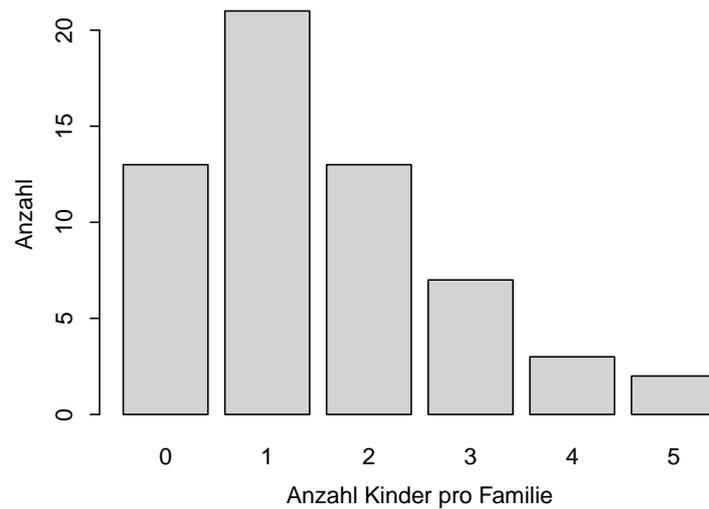
## Aufgabenteil 5: Graphik

Franzi hat eine Umfrage bei ihren alten Klassenkameraden gemacht, wie viele Kinder diese 20 Jahre nach ihrer Schulzeit haben:

```
table(x)
```

```
## x
## 0 1 2 3 4 5
## 13 21 13 7 3 2
```

Franzi stellt das Ergebnis graphisch dar.



Mit welcher Funktion hat Franzi dies gemacht?

- (A) `barplot(x, COL = "darkgrey", xlab="Anzahl der Kinder pro Familie")`
- (B) `hist(x, col = "lightgrey", xlab="Anzahl der Kinder pro Familie")`
- (C) `plot(table(x), xlab="Anzahl der Kinder pro Familie")`
- (D) `plot(x, col = grey, xlab = "Anzahl Kinder pro Familie", ylab = "Anzahl", bar = TRUE)`
- (E) `barplot(table(x), col = "lightgrey", ylab = "Anzahl", xlab = "Anzahl Kinder pro Familie")`