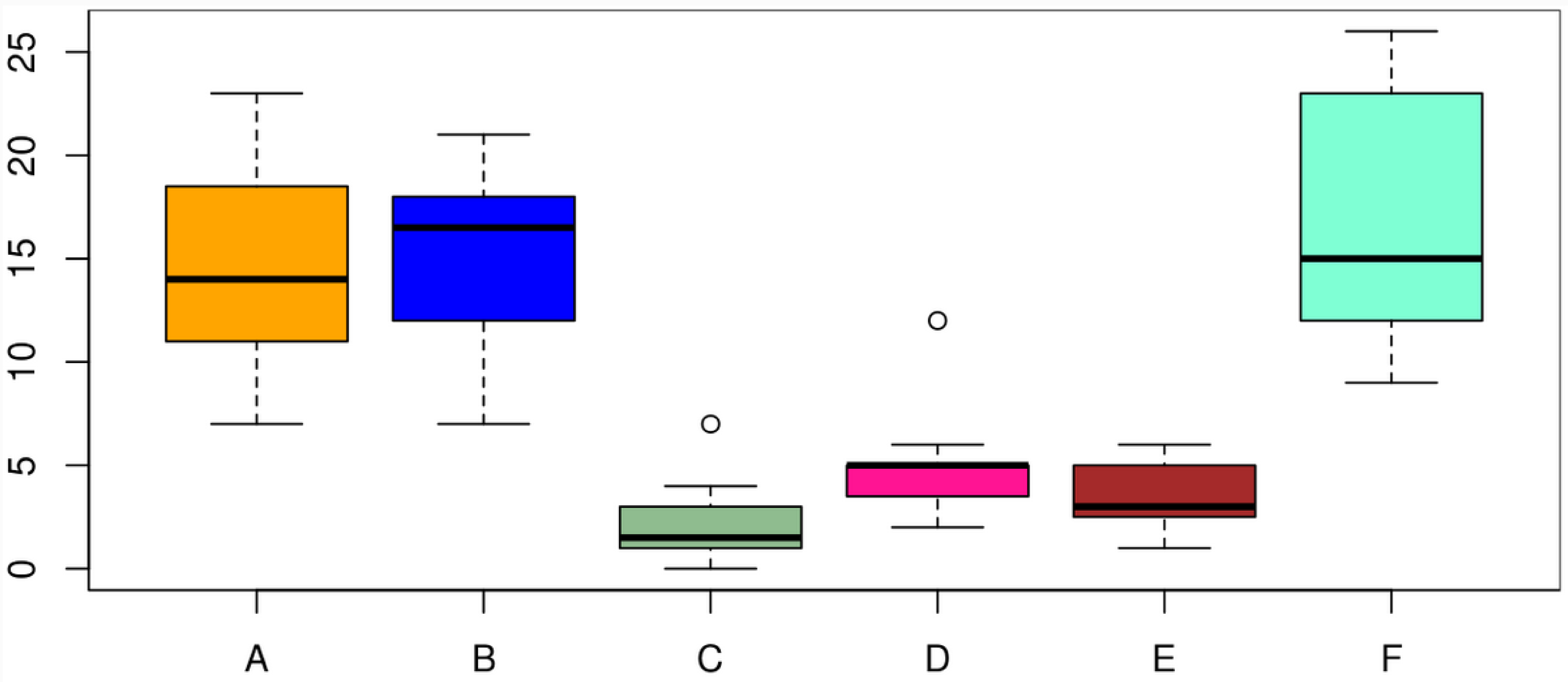
Testat 1

Aufgabe 1

Wir testen Insektensprays. Dabei wurden 6 verschiedene Insektensprays verwendet, die auf verschiedenen Fel-dern versprüht wurden. Danach wurde die Anzahl Insekten gezählt, die sich auf dem
entsprechenden Feld nach dem Besprühen befanden. Je kleiner die Anzahl, umso wirksamer der Spray.



Beachten Sie: Falsche Antworten geben Punkteabzug.

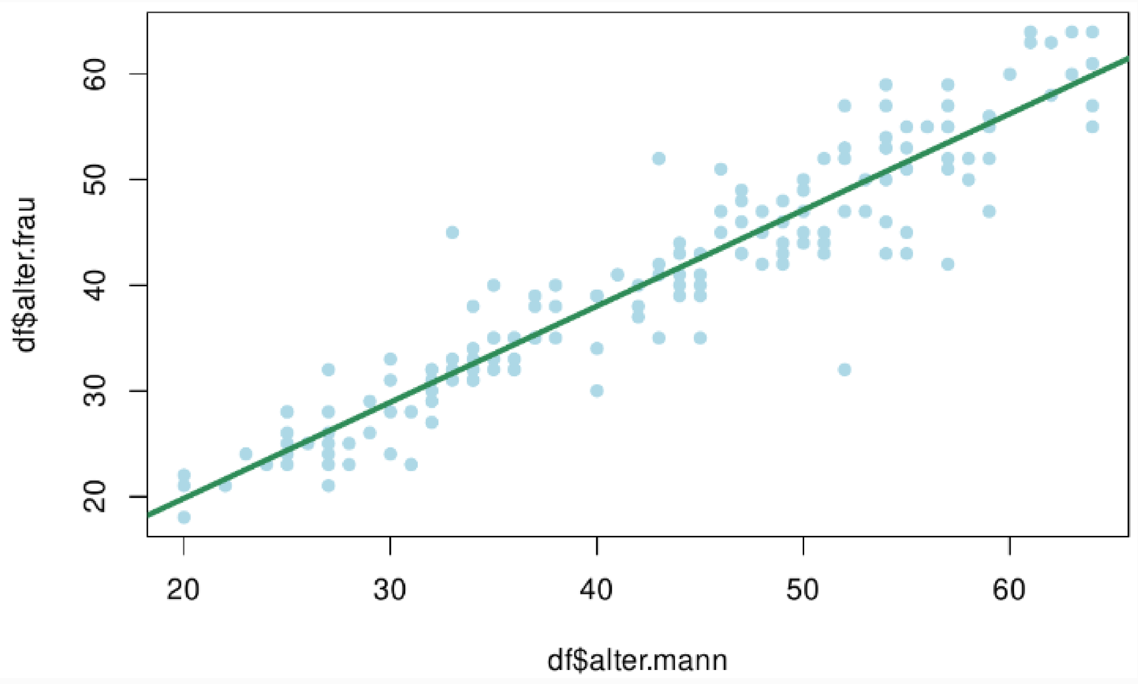
Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
◉ ✓	○	Für Spray B sind 25% der Messwerte ungefähr 17 oder grösser
○ ✓	◉	Ungefähr 25% der Messwerte von Spray F sind ungefähr 12 oder grösser
○ ✓	◉	Spray C hat die grössere Streuung bezüglich Interguartilsdifferenz als Spray F
◉ ✓	○	Ungefähr 50% der Messwerte von Spray F sind zwischen ungefähr 12 und 23

Aufgabe 2

Wir haben aus eigener Erfahrung das Gefühl, dass bei Ehepaaren der Mann eher älter als die Frau ist. Nun wollen wir statistisch untersuchen, ob dem so ist.
In einer Untersuchung in England wurden das Alter (in Jahren) und die Körpergrösse (in cm) von 170 Ehepaaren untersucht.

Das Streudiagramm sieht wie folgt aus:



Die Regressionsgerade wird wie folgt bestimmt:

```
lm(df$alter.frau ~ df$alter.mann)

##
## Call:
## lm(formula = df$alter.frau ~ df$alter.mann)
##
## Coefficients:
##   (Intercept)  df$alter.mann
##        1.5740         0.9112
```

Beachten Sie: Falsche Antworten ergeben Punkteabzug.

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
○ ✓	◉	Aus dem Streudiagramm ist ein quadratischer Zusammenhang erkennbar.
○ ✓	◉	Die Regressionsgerade lautet $y = 0.9112 + 1.574x$
○ ✓	◉	Für jedes Jahr das der Ehefrau älter ist, ist er Mann 0.911 Jahre älter
◉ ✓	○	Der Korrelationskoeffizient ist annähernd 1

Aufgabe 3

Bei einem Zufallsexperiment werden ein roter und ein blauer Würfel gleichzeitig geworfen. Wir nehmen an, dass sie „fair“ sind, d. h. die Augenzahlen 1 bis 6 eines Würfels treten mit gleicher Wahrscheinlichkeit auf.

Beachten Sie: Falsche Antworten ergeben Punkteabzug.

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
○ ✓	◉	7 ist ein mögliches Elementarereignis
◉ ✗	○	Die Wahrscheinlichkeit, dass das Produkt der Augenzahlen 7 ist, ist 6/36
◉ ✓	○	Die Wahrscheinlichkeit, dass die Augensumme kleiner oder gleich 11 ist, ist 35/36
◉ ✓	○	Die Wahrscheinlichkeit, dass der rote Würfel 5 ist, ist 1/6

Aufgabe 4

Ein Multiple-Choice-Test besteht aus 15 Fragen, mit jeweils 5 Antwortmöglichkeiten, von denen genau eine richtig ist. Die Wahrscheinlichkeit dafür, eine Aufgabe richtig zu beantworten, ist also 0.2. Die Wahrscheinlichkeits- und Verteilungsfunktion sind gegeben durch:

k	8	9	10	11	12	13	14	15
$P(X \leq k)$	0.711	0.939	0.969	0.982	0.989	0.992	0.999	1

Beachten Sie: Es handelt sich hier um die kumulierten Wahrscheinlichkeiten $P(X \leq k)$ und nicht [tex]P (X = k)[tex].

Beachten Sie: Falsche Antworten geben Punkteabzug.

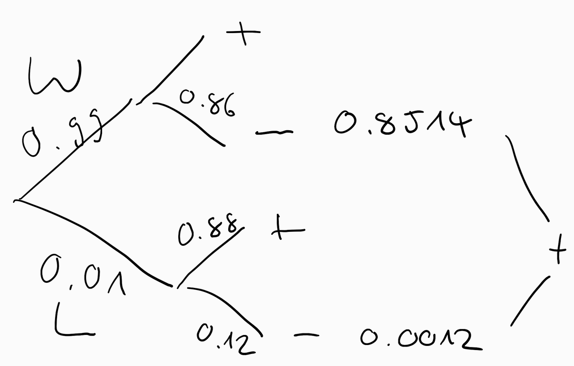
Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
○ ✓	◉	Die Wahrscheinlichkeit $P(X \leq 9)$ ist 0.228
○ ✓	◉	Die Wahrscheinlichkeit $P(X \geq 12)$ ist 0.989
◉ ✓	○	Die Wahrscheinlichkeit, dass genau 13 Fragen richtig beantwortet werden, ist 0.003
◉ ✓	○	Die Wahrscheinlichkeit, dass höchstens 10 Fragen richtig beantwortet werden, ist 0.969

Aufgabe 5

Ein Lügendetektor wird routinemässig bei Mitarbeitern durchgeführt, die in sensible Positionen arbeiten. Benützen wir nun +, um das Ereignis zu bezeichnen, dass der Test positiv ist, d.h. dass der Lügendetektor anzeigt, dass der Mitarbeiter gelogen hat. Mit W bezeichnen wir das Ereignis dass der Mitarbeiter die Wahrheit gesagt hat und mit L, dass der Mitarbeiter gelogen hat.
Aus Untersuchungen von Lügendetektoren wissen wir, dass

$P(+|L) = 0,88$ und $P(-|W) = 0,86$



+ = 0.8526 = P(−)

Gesetz der totalen W'keit

Darüber hinaus wissen wir, dass Folgendes gilt
$P(W) = 0,99$

Bei einer Person zeigt der Detektor an, dass die Wahrheit gesagt wurde. Was ist die effektive die Wahrscheinlichkeit, dass diese Person die Wahrheit gesagt hat?

$$P(W|-) = \frac{P(-|W) \cdot P(W)}{P(-)}$$

$$= \frac{0.86 \cdot 0.99}{} = \frac{0.8514}{0.8526} = \underline{\underline{0.99859}} \checkmark$$

(Gerundet auf 5 Dezimalstellen)

- ◉ 0.99859 ✓
- ○ 0.00141
- ○ 0.99898
- ○ 0.00102

Testat 2

Aufgabe 1

Für die Körpergrösse von 18-20jährigen Männern ergibt sich ein Mittelwert von 1.80 m bei einer Standardabweichung von 7.4 cm. Die Körpergrösse kann als normalverteilt angesehen werden.

Mit welcher Wahrscheinlichkeit ist ein zufällig ausgewählter Mann dieser Altersgruppe kleiner oder gleich als 1.90cm?

X sei die Zufallsvariable für die Körpergrösse eines zufällig ausgewählten Mannes.

Welche der folgenden Aussagen beschreibt die gesuchte Wahrscheinlichkeit?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
○	◉	1 - $P(X < 190)$
◉	○	pnorm(q=190, mean=180, sd=7.4)
○	◉	1 - qnorm(p=190, mean=180, sd=7.4)
◉	○	$P(X \leq 190)$

Aufgabe 2

Die Körpertemperatur von 10 Patienten wird zum Zeitpunkt der Verabreichung eines Medikaments (T 1) und 2 Stunden später (T 2) gemessen. Es soll geprüft werden, ob dieses Medikament eine fiebersenkende Wirkung hat.

Patient-Nr.	1	2	3	4	5	6	7	8	9	10
Temp. 1 in °C	39.1	39.3	38.9	40.6	39.5	38.4	38.6	39.0	38.6	39.2
Temp. 2 in °C	38.1	38.3	38.8	37.8	38.2	37.3	37.6	37.8	37.4	38.1

Wir führen einen Hypothesentest auf 5% durch um zu überprüfen, ob das Medikament fiebersenkend ist. Der R-Output zeigt:

```
##
##  Paired t-test
##
## data:  t.1 and t.2
## t = 5.6569, df = 9, p-value = 0.0001554
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.7976252       Inf
## sample estimates:
## mean of the differences
##                    1.18
```

Welche der folgenden Aussagen ist richtig?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
○	◉	Wir führen einen zweiseitigen Test durch.
○	◉	Die Nullhypothese ist, dass das Medikament Wirkung hat.
○	◉	Da der p-Wert unter dem Signifikanzniveau liegt wird die Alternativhypothese angenommen.
◉	○	Der Wert 1.18 in der letzten Linie bedeutet, dass die durchschnittliche Temperatur um 1.18 °C nach zwei Stunden tiefer war.

Aufgabe 3

Ein U.S. Magazin, Consumer Reports, führte eine Untersuchung des Kalorien- und Salzgehaltes von verschiedenen Hotdog-Marken durch. Es gab drei verschiedene Typen von Hotdogs: Rind, „Fleisch" (Rind, Schwein, Geflügel gemischt) und Geflügel.

Die Resultate unten führen den Kaloriengehalt verschiedener Marken von Rind- und Geflügel-Hotdogs auf.

Rinds-Hotdog: 186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132
Geflügel-Hotdog: 129, 132, 102, 106, 94, 102, 87, 99, 170, 113, 135, 142, 86, 143, 152, 146, 144

Haben die beiden Hotdog-Arten verschiedenen Kaloriengehalt? Wir führen einen Hypothesentest auf 5% Signifikanzniveau durch und erhalten folgenden Output:

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 285.5, p-value = 0.0004549
## alternative hypothesis: true location shift is not equal to 0
```

x enthält die Rindsdaten und y die Geflügeldaten.

Welche der folgenden Aussagen sind richtig?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
◉ ✔	○	Die Nullhypothese wird verworfen, da der p-Wert unter dem Signifikanzniveau liegt.
◉ ✔	○	Wir führen einen zweiseitigen Test durch.
○ ✔	◉	Die Alternativhypothese ist $\mu_x > \mu_y$
◉ ✘	○	Der Unterschied zwischen x und y ist nicht statistisch signifikant.

Aufgabe 4

Die MASS-Bibliothek enthält den Boston-Datensatz, der *medv* (median house value in $1000) für 506 Stadtviertel um Boston herum erfasst. Wir werden versuchen, medv mit 13 Prädiktoren wie *rm* (durchschnittliche Anzahl von Zimmern pro Haus) , *age* (Durchschnittsalter der Häuser) und *crim* (Kriminalitätsrate) vorherzusagen.

Wir werden damit beginnen, die lm()-Funktion zu verwenden, um ein einfaches lineares Regressionsmodell mit *medv* als Zielvariable und *crim* als Prädiktor anzupassen. Wir erhalten folgenden Output:

```
lm.fit <- lm(medv ~ crim)
summary(lm.fit)

##
## Call:
## lm(formula = medv ~ crim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.957  -5.449  -2.007   2.512  29.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.03311    0.40914   58.74   <2e-16 ***
## crim        -0.41519    0.04389   -9.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.484 on 504 degrees of freedom
## Multiple R-squared:  0.1508,Adjusted R-squared:  0.1491
## F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

Welche der folgenden Aussagen sind richtig?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
○ ✔	◉	Das Modell lautet $\mathbf{crim} = \beta_0 + \beta_1 \cdot \mathbf{medv}$.
○ ✔	◉	Der Wert von $R^2 = 0.1508$ bedeutet, dass die Daten gut zum Modell passen.
○ ✔	◉	Der y-Achsenabschnitt ist -0.4152
○ ✘	◉	Die Steigung ist statistisch signifikant ungleich 0.

Aufgabe 5

Die MASS-Bibliothek enthält den Boston-Datensatz, der medv (median house value in $1000)
für 506 Stadtviertel um Boston herum erfasst. Wir werden versuchen, medv mit 13 Prädiktoren wie rm (durchschnittliche Anzahl von Zimmern pro Haus) , dis (Distanz vom Zentrum von Boston) und crim (Anteil der Kriminalität) vorherzusagen.

Wir passen ein multiples lineares Regressionsmodell mit der Zielvariable medv und den Prädiktoren crim, dis und rm. Das Signifikanzniveau ist 5%. Wir erhalten folgenden Output:

```
fit <- lm(medv ~ crim + rm + dis, data = Boston)

summary(fit)

##
## Call:
## lm(formula = medv ~ crim + rm + dis, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.247  -2.930  -0.572   2.390  39.072
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -29.45838    2.60010 -11.330  < 2e-16 ***
## crim         -0.25405    0.03532  -7.193 2.32e-12 ***
## rm            8.34257    0.40870  20.413  < 2e-16 ***
## dis           0.12627    0.14382   0.878     0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.238 on 502 degrees of freedom
## Multiple R-squared:  0.5427,Adjusted R-squared:  0.5399
## F-statistic: 198.6 on 3 and 502 DF,  p-value: < 2.2e-16
```

Welche der folgenden Aussagen sind richtig?

Für jede Aussage muss entschieden werden: [richtig] oder [falsch]

richtig	falsch	
◉ ✔	○	Der p-Wert für rm ist kleiner als das Signifikanzniveau und somit wird die Nullhypothese $\beta_2 = 0$ verworfen.
◉ ✔	○	Der p-Wert 0.38 für dis bedeutet, dass es keinen statistisch signifkanten Unterschied im Mediumwert gibt, ob man nahe oder weit vom Zentrum entfernt lebt.
○ ✔	◉	Der Koeffizient für dis bedeutet, dass pro Mile Abstand mehr vom Zentrum von Boston, der Medianpreis um etwa $126 fällt.
○ ✔	◉	Der Wert $R^2 = 0.54$ ist statistisch signifikant.