Aufgabe 1 (20 Punkte)

In dieser Aufgabe betrachten wir die individuelle Ersparnis pro Jahr (gemessen in Tausend USD) und welchen Einfluss der prozentuale Anteil der Unter-15-Jährigen in der Bevölkerung darauf hat; die letztere Variable heißt pop15, ein Wert von 3 bedeutet hier 3%. Betrachten Sie den R-Output und beantworten Sie dazu die folgenden Fragen. Der R-Output ist zur besseren Lesbarkeit leicht editiert worden.

1. Betrachten Sie den Einfluss des Anteils der Unter-15-Jährigen auf die Ersparnis in der Regression fit1. Interpretieren Sie den geschätzten Regressionskoeffizienten, dessen statistische Signifikanz, und das R^2. Interpretieren Sie in analoger Weise die Regressionsergebnisse in fit2 und fit3.

2. Welche Regression passt am besten? Begründung.

```
> summary(ersparnis)
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  0.600    6.970   10.510    9.671   12.617   21.100
> summary(pop15)
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  21.44    26.21    32.58    35.09    44.06    47.64
> fit1 = lm(ersparnis ~ pop15)
> summary(fit1)

Call:
lm(formula = ersparnis ~ pop15)

Residuals:
    Min      1Q  Median      3Q     Max
 -8.637  -2.374   0.349   2.022  11.155

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.49660    2.27972   7.675 6.85e-10 ***
pop15       -0.22302    0.06291  -3.545 0.000887 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.03 on 48 degrees of freedom
Multiple R-squared:  0.2075,    Adjusted R-squared:  0.191
```

```
F-statistic: 12.57 on 1 and 48 DF,  p-value: 0.0008866

> fit2 = lm(log(ersparnis) ~ pop15)
> summary(fit2)

Call:
lm(formula = log(ersparnis) ~ pop15)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4722 -0.1809  0.1286  0.3044  1.1279

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.174606   0.359303   8.835 1.24e-11 ***
pop15       -0.030530   0.009914  -3.079  0.00343 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6351 on 48 degrees of freedom
Multiple R-squared:  0.165,     Adjusted R-squared:  0.1476
F-statistic: 9.482 on 1 and 48 DF,  p-value: 0.003427

> fit3 = lm(log(ersparnis) ~ log(pop15))
> summary(fit3)

Call:
lm(formula = log(ersparnis) ~ log(pop15))

Residuals:
    Min      1Q  Median      3Q     Max
-2.4468 -0.2074  0.1303  0.2985  1.1213

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.7999     1.1816   4.909  1.1e-05 ***
log(pop15)   -1.0493     0.3344  -3.138  0.00291 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6332 on 48 degrees of freedom
```

```
Multiple R-squared:  0.1702,    Adjusted R-squared:  0.1529
F-statistic: 9.844 on 1 and 48 DF,  p-value: 0.00291
```

Aufgabe 2 (18 Punkte)

Ein Buchhändler notiert über 24 Monate hinweg jeden Monat seinen erzielten Gewinn (in Tausend EUR) aus dem Betrieb seiner Buchhandlung. Nun fragt er sich, wie sicher er sein kann, mit seiner Buchhandlung auf die Dauer Gewinn erzielen zu können. Betrachten Sie den R-Output und beantworten Sie bitte die folgenden Fragen.

1. Viermal wird der Befehl t.test aufgerufen. Interpretieren Sie bitte jeweils die Konfidenzintervalle aus der Sicht des Buchhändlers. Formulieren Sie bitte außerdem die Hypothesen der statistischen Tests und interpretieren Sie das Testergebnis aus der Sicht des Buchhändlers.

2. Interpretieren Sie bitte das berechnete Quantil.

```
> Gewinn = c(10,4,6,-12,-6,-3,4,9,12,-3,5,7,4, 7,
               6,8,5,-3,-8,5,8,6,7,5)
> length(Gewinn)
[1] 24
### Test 1 ###
> t.test(Gewinn)

        One Sample t-test

data:  Gewinn
t = 2.8035, df = 23, p-value = 0.01009
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.9065256 6.0101411
sample estimates:
mean of x
 3.458333

### Test 2 ###
> t.test(Gewinn, conf.level = 0.99)

        One Sample t-test

data:  Gewinn
t = 2.8035, df = 23, p-value = 0.01009
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
```

```
 -0.004676133   6.921342799
sample estimates:
mean of x
 3.458333

### Test 3 ###
> t.test(Gewinn, alternative="less")

        One Sample t-test

data:  Gewinn
t = 2.8035, df = 23, p-value = 0.995
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
     -Inf 5.572492
sample estimates:
mean of x
 3.458333

### Test 4 ###
> t.test(Gewinn, alternative="greater")

        One Sample t-test

data:  Gewinn
t = 2.8035, df = 23, p-value = 0.005044
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 1.344175      Inf
sample estimates:
mean of x
 3.458333
> quantile(Gewinn, 0.2)
20%
 -3
```

Aufgabe 3

Die freiwillige Feuerwehr Garfeld muss durchschnittlich 3 mal im Monat mit einer Drehleiter ausrücken, um eine Katze von einem Baum zu holen.

1. Wie lautet die Verteilung der Zufallsvariable

   ‚Anzahl der Einsätze zur Rettung einer Katze pro Monat'?

   Bitte geben Sie den Namen der Verteilung und den Wert der Parameter an.

2. Für diese Verteilung ergibt sich für ausgewählte Werte der Dichtefunktion f

x	0	1	2	3
f(x)	0.05	0.15	0.22	0.22

   Mit welcher Wahrscheinlichkeit gibt es

   (a) genau einen Einsatz,

   (b) höchstens zwei Einsätze,

   (c) drei Einsätze oder mehr.

3.* Der Verein der KatzenbesitzerInnen macht der freiwilligen Feuerwehr zwei Angebote: entweder zahlt der Verein jeden Monat fest 3.000 GE oder beteiligt sich variabel nach Anzahl der getätigten Einsätze:

   a) bei zwei Einsätzen oder weniger zahlt der Verein 2.000 GE in diesem Monat,

   b) ab drei Einsätzen zahlt der Verein 4.000 GE in diesem Monat.

   Welches Angebot ist langfristig für die freiwillige Feuerwehr besser?

Aufgabe 4

Drei Dozenten diskutieren darüber, wie sie mit Abmeldungen in ihren Fortbildungsseminaren umgehen. Dozent A hat 10 Anmeldungen und rechnet damit, dass eine beliebige angemeldete Person mit Wahrscheinlichkeit 20% zurücktritt. Kollege B hat 400 Anmeldungen und geht ebenfalls davon aus, dass eine beliebige angemeldete Person mit Wahrscheinlichkeit 20% absagt. Kollegin C hat auch 400 Anmeldungen, rechnet aber damit, dass die Anzahl X der endgültig Teilnehmenden normalverteilt mit Parametern $\mu = 320$ und $\sigma^2 = 64$ ist.

1. Rechnen Sie die Wahrscheinlichkeit aus, dass bei Dozent A mindestens 8 Personen dabei bleiben.

2. Welche Verteilung (Name, Parameter) hat die Zufallsvariable Y, die die Anzahl der **Teilnehmenden** bei Dozent B zählt? Berechnen Sie $P(Y \geq 329)$. Nutzen Sie wenn nötig eine geeignete Approximation für Ihre Verteilung (Voraussetzungen erfüllt?).

3.* Für Dozentin C sind zwischen 320 und 336 Teilnehmende optimal. Mit welcher Wahrscheinlichkeit geschieht das?

4.* Welche Teilnehmerzahl wird bei Dozentin C nur in höchstens 5% der Fälle überschritten?

Nachfolgend einige Werte der Verteilung und der Quantile der Standardnormalverteilung (fehlende durch den nächstbesten nähern, etwa Stelle $z = 1.006$ durch $z = 1$):

```
> pnorm(1)
[1] 0.8413447
> pnorm(1.125)
[1] 0.8697055
> pnorm(0)
[1] 0.5
> pnorm(2)
[1] 0.9772499
> pnorm(-1)
[1] 0.1586553
> pnorm(-2)
[1] 0.02275013
> qnorm(0.05)
[1] -1.644854
> qnorm(0.95)
```

```
[1] 1.644854
```

Aufgabe 5

Herr Urban fährt mit der U-Bahn ins Büro. Seine Haltestellen werden von drei Linien U_1, U_2, U_3 befahren. Die Tabelle gibt den jeweiligen Anteil der Linien an der Gesamtheit aller Fahrten an sowie die Wahrscheinlichkeit für eine Verspätung (Anteile werden als Wahrscheinlichkeiten interpretiert).

U-Bahnlinie	Anteil an den Gesamtfahrten	Anteil der verspäteten U-Bahnen auf dieser Linie
U_1	35%	5%
U_2	50%	10%
U_3	15%	40%

1. Wie wahrscheinlich ist es, dass eine U-Bahn im Allgemeinen pünktlich ist?

2. Wie wahrscheinlich ist es, dass eine beliebige U-Bahn verspätet ist **und** der Linie U_3 gehört?

3. Mit welcher Wahrscheinlichkeit gehört eine U-Bahn, die verspätet kommt, der Linie U_3?

4.* Mit welcher Wahrscheinlichkeit gehört eine U-Bahn, die gerade pünktlich ist, tatsächlich der Linie U_1?

Aufgabe 1

In dieser Aufgabe untersuchen wir den Einfluss der Variablen Einkommen (in USD) auf die Lebenszufriedenheit (Skala 1-10). Die Daten liegen für 500 Personen vor und sind frei erfunden. Betrachten Sie den R-Output und beantworten Sie dazu die folgenden Fragen. Der R-Output ist zur besseren Lesbarkeit leicht editiert worden.

1. Betrachten Sie den Einfluss des Einkommens auf die Lebenszufriedenheit in der Regression fit1. Interpretieren Sie den geschätzten Regressionskoeffizienten, dessen statistische Signifikanz, und das R^2. Interpretieren Sie in analoger Weise die Regressionsergebnisse in fit2.

2.* Welche Regression passt am besten? Begründung.

```
> head(x) # Uebersicht zum Datensatz
  Person Einkommen Lebenszufriedenheit
1      1  38626.47            2.314489
2      2  49793.81            3.433490
3      3  49239.57            4.599373
4      4  32143.72            2.791114
5      5  71964.09            5.596398
6      6  37296.43            2.458556
> fit1 = lm(x$Lebenszufriedenheit~x$Einkommen)
> summary(fit1)

Call:
lm(formula = x$Lebenszufriedenheit ~ x$Einkommen)

Residuals:
     Min       1Q   Median       3Q      Max
-2.02479 -0.48526  0.04078  0.45898  2.37805

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.043e-01  8.884e-02   2.299   0.0219 *
x$Einkommen 7.138e-05  1.854e-06  38.505   <2e-16 ***
---
Signif. codes: 0  ***  0.001  **  0.01
                 *  0.05  .  0.1       1
Residual standard error: 0.7181 on 496 degrees of freedom
Multiple R-squared: 0.7493, Adjusted R-squared: 0.7488
```

```
F-statistic:  1483 on 1 and 496 DF,  p-value: < 2.2e-16

> fit2 = lm(x$Lebenszufriedenheit~log(x$Einkommen))
> summary(fit2)

Call:
lm(formula = x$Lebenszufriedenheit ~ log(x$Einkommen))

Residuals:
     Min       1Q   Median       3Q      Max
-2.24886 -0.52853  0.00531  0.51613  2.66532

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -25.98525    0.82622  -31.45   <2e-16 ***
log(x$Einkommen)   2.76654    0.07774   35.59   <2e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01
                 *  0.05  .  0.1  1
Residual standard error: 0.7609 on 496 degrees of freedom
Multiple R-squared: 0.7186, Adjusted R-squared: 0.718
F-statistic:  1266 on 1 and 496 DF,  p-value: < 2.2e-16
```

Aufgabe 2

Die Variable 'Rendite' enthält wöchentliche Renditen des US-Aktienindizes S&P 500. Ein Wert von 3.15 bedeutet zum Beispiel, dass der S&P 500 in einer Woche um 3.15% gestiegen ist. Betrachten Sie den R-Output und beantworten Sie die folgenden Fragen.

1. Dreimal wird der Befehl t.test aufgerufen. Interpretieren Sie bitte jeweils die Konfidenzintervalle. Formulieren Sie bitte außerdem die Hypothesen der statistischen Tests und interpretieren Sie das Testergebnis.

2.* Betrachten Sie Histogramm und Boxplot. Gibt es eine auffallende statistische Eigenschaft in den Daten, die man dort erkennen kann? Erläutern Sie kurz aus der Sicht eines Statistikers, nicht eines Ökonomen.

3.* Interpretieren Sie bitte das berechnete Quantil.

```
> head(Rendite)
[1] -0.270 -2.576  3.514  0.712  1.178 -1.372
> t.test(Rendite)

        One Sample t-test

data:  Rendite
t = 2.0988, df = 1088, p-value = 0.03607
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.009758461 0.290039519
sample estimates:
mean of x
 0.149899

> t.test(Rendite, alternative = "less")

        One Sample t-test

data:  Rendite
t = 2.0988, df = 1088, p-value = 0.982
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
      -Inf 0.2674779
sample estimates:
```

```
mean of x
 0.149899

> t.test(Rendite, alternative = "greater")

        One Sample t-test

data:  Rendite
t = 2.0988, df = 1088, p-value = 0.01803
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 0.03232007        Inf
sample estimates:
mean of x
 0.149899

> boxplot(Rendite, range=0)
> hist(Rendite, nclass=30)
> quantile(Rendite, 0.1)
    10%
-2.4304
```

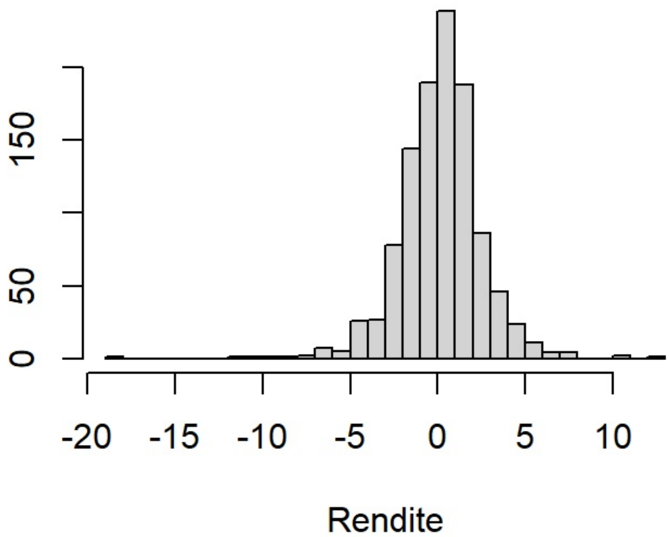

Abbildung 1: Histogramm der Variablen Rendite. hist(Rendite, nclass=30).

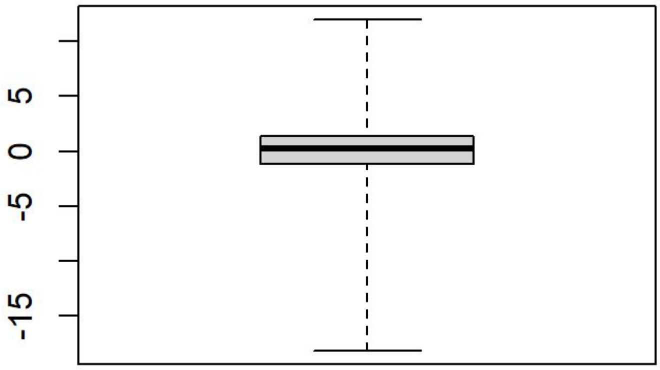

Abbildung 2: Boxplot der Variablen Rendite. boxplot(Rendite, range=0).

Aufgabe 3

Die freiwillige Feuerwehr Garfeld muss durchschnittlich 3 mal im Monat mit einer Drehleiter ausrücken, um eine Katze von einem Baum zu holen.

1. Wie lautet die Verteilung der Zufallsvariable

   ‚Anzahl der Einsätze zur Rettung einer Katze pro Monat'?

   Bitte geben Sie den Namen der Verteilung und den Wert der Parameter an.

2. Für diese Verteilung ergibt sich für ausgewählte Werte der Dichtefunktion f

x	0	1	2	3
f(x)	0.05	0.15	0.22	0.22

   Mit welcher Wahrscheinlichkeit gibt es

   (a) genau einen Einsatz,

   (b) höchstens zwei Einsätze,

   (c) drei Einsätze oder mehr.

3.* Der Verein der KatzenbesitzerInnen macht der freiwilligen Feuerwehr zwei Angebote: entweder zahlt der Verein jeden Monat fest 3.000 GE oder beteiligt sich variabel nach Anzahl der getätigten Einsätze:

   a) bei zwei Einsätzen oder weniger zahlt der Verein 2.000 GE in diesem Monat,

   b) ab drei Einsätzen zahlt der Verein 4.000 GE in diesem Monat.

   Welches Angebot ist langfristig für die freiwillige Feuerwehr besser?

Aufgabe 4

Drei Dozenten diskutieren darüber, wie sie mit Abmeldungen in ihren Fortbildungsseminaren umgehen. Dozent A hat 10 Anmeldungen und rechnet damit, dass eine beliebige angemeldete Person mit Wahrscheinlichkeit 20% zurücktritt. Kollege B hat 400 Anmeldungen und geht ebenfalls davon aus, dass eine beliebige angemeldete Person mit Wahrscheinlichkeit 20% absagt. Kollegin C hat auch 400 Anmeldungen, rechnet aber damit, dass die Anzahl X der endgültig Teilnehmenden normalverteilt mit Parametern $\mu = 320$ und $\sigma^2 = 64$ ist.

1. Rechnen Sie die Wahrscheinlichkeit aus, dass bei Dozent A mindestens 8 Personen dabei bleiben.

2. Welche Verteilung (Name, Parameter) hat die Zufallsvariable Y, die die Anzahl der **Teilnehmenden** bei Dozent B zählt? Berechnen Sie $P(Y \geq 329)$. Nutzen Sie wenn nötig eine geeignete Approximation für Ihre Verteilung (Voraussetzungen erfüllt?).

3.* Für Dozentin C sind zwischen 320 und 336 Teilnehmende optimal. Mit welcher Wahrscheinlichkeit geschieht das?

4.* Welche Teilnehmerzahl wird bei Dozentin C nur in höchstens 5% der Fälle überschritten?

Nachfolgend einige Werte der Verteilung und der Quantile der Standardnormalverteilung (fehlende durch den nächstbesten nähern, etwa Stelle $z = 1.006$ durch $z = 1$):

```
> pnorm(1)
[1] 0.8413447
> pnorm(1.125)
[1] 0.8697055
> pnorm(0)
[1] 0.5
> pnorm(2)
[1] 0.9772499
> pnorm(-1)
[1] 0.1586553
> pnorm(-2)
[1] 0.02275013
> qnorm(0.05)
[1] -1.644854
> qnorm(0.95)
```

```
[1] 1.644854
```

Aufgabe 5

Herr Urban fährt mit der U-Bahn ins Büro. Seine Haltestellen werden von drei Linien U_1, U_2, U_3 befahren. Die Tabelle gibt den jeweiligen Anteil der Linien an der Gesamtheit aller Fahrten an sowie die Wahrscheinlichkeit für eine Verspätung (Anteile werden als Wahrscheinlichkeiten interpretiert).

U-Bahnlinie	Anteil an den Gesamtfahrten	Anteil der verspäteten U-Bahnen auf dieser Linie
U_1	35%	5%
U_2	50%	10%
U_3	15%	40%

1. Wie wahrscheinlich ist es, dass eine U-Bahn im Allgemeinen pünktlich ist?
2. Wie wahrscheinlich ist es, dass eine beliebige U-Bahn verspätet ist **und** der Linie U_3 gehört?
3. Mit welcher Wahrscheinlichkeit gehört eine U-Bahn, die verspätet kommt, der Linie U_3?

4.* Mit welcher Wahrscheinlichkeit gehört eine U-Bahn, die gerade pünktlich ist, tatsächlich der Linie U_1?

Aufgabe 1 (20 Punkte)

Wir untersuchen den Einfluss der Lebenszufriedenheit auf die Lebenserwartung. Dazu analysieren wir einen (frei erfundenen) Datensatz mit Angaben zu 100 Personen. Betrachten Sie den R-Output und beantworten Sie dazu die folgenden Fragen. Der R-Output ist zur besseren Lesbarkeit leicht editiert worden.

1. Betrachten Sie den Einfluss der Lebenszufriedenheit (Skala $0-100$, höhere Werte drücken höhere Lebenszufriedenheit aus) auf die Lebenserwartung (in Jahren) in der Regression fit1. Interpretieren Sie den geschätzten Regressionskoeffizienten für die unabhängige Variable, dessen statistische Signifikanz, und das R^2. Interpretieren Sie in analoger Weise die Regressionsergebnisse in fit2 und fit3.

2. Welche Regression passt am besten? Begründung.

```
> summary(Lebenszufriedenheit)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 0.09445 22.86803 43.28927 47.90287 76.16149 99.25370
> summary(Lebenserwartung)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  41.54   64.92   74.17   74.52   85.04  106.90
> fit1 = lm(Lebenserwartung~Lebenszufriedenheit)
> summary(fit1)

Call:
lm(formula = Lebenserwartung ~ Lebenszufriedenheit)

Residuals:
    Min      1Q  Median      3Q     Max
-14.128  -5.715  -0.397   4.863  16.339

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         55.6     1.369      40.61   <2e-16 ***
Lebenszufriedenheit 0.39     0.024      16.23   <2e-16 ***

Residual standard error: 7.236 on 98 degrees of freedom
Multiple R-squared:  0.729,    Adjusted R-squared:  0.7262
F-statistic: 263.6 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
> fit2 = lm(Lebenserwartung~log(Lebenszufriedenheit))
> summary(fit2)

Call:
lm(formula = Lebenserwartung ~ log(Lebenszufriedenheit))

Residuals:
     Min        1Q    Median        3Q       Max
-19.6688   -7.2487    0.4162    6.8219   23.4701

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                45.5      2.73    16.66   <2e-16 ***
log(Lebenszufriedenheit)    8.3      0.74    11.23   <2e-16 ***

Residual standard error: 9.192 on 98 degrees of freedom
Multiple R-squared:  0.5626,    Adjusted R-squared:  0.5581
F-statistic:   126 on 1 and 98 DF,  p-value: < 2.2e-16

> fit3 = lm(log(Lebenserwartung)~log(Lebenszufriedenheit))
> summary(fit3)

Call:
lm(formula = log(Lebenserwartung) ~ log(Lebenszufriedenheit))

Residuals:
     Min        1Q    Median        3Q       Max
-0.29303  -0.08587   0.01616   0.09314   0.24755

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.866    0.0358   107.97   <2e-16 ***
log(Lebenszufriedenheit) 0.122    0.0097    12.64   <2e-16 ***

Residual standard error: 0.1203 on 98 degrees of freedom
Multiple R-squared:  0.6199,    Adjusted R-squared:  0.616
F-statistic: 159.8 on 1 and 98 DF,  p-value: < 2.2e-16
```

Aufgabe 2 (16 Punkte)

Ein Förster lässt bei 100 Bäumen den Gehalt eines bestimmten Schadstoffs messen. Wenn der mittlere Schadstoffgehalt in den Bäumen den Grenzwert von $100\mu g$ statistisch signifikant übersteigt dann sollen die Bäume im Folgejahr in einem aufwendigen und kostspieligen Verfahren genauer untersucht werden. Übersteigt der mittlere Schadstoffgehalt statistisch signifikant die Grenze von $120\mu g$ dann müssen diese teuren Verfahren sogar sofort veranlasst werden.

1. Wie sollte sich der Förster anhand der statistischen Tests im R-Code entscheiden wenn er als Irrtumswahrscheinlichkeit 5% wählt? Formulieren Sie bitte die Hypothesen der statistischen Tests und begründen Sie Ihre Entscheidung.
2. Wie entscheidet sich der Förster wenn er aufgrund der hohen Kosten der Maßnahmen nur handeln möchte wenn die Ergebnisse eine besonders hohe statistische Sicherheit aufweisen?
3. Interpretieren Sie bitte das berechnete Quantil.

```
> t.test(Schadstoffgehalt, mu = 100,
         alternative = "greater")

        One Sample t-test

data:  Schadstoffgehalt
t = 10.825, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 120.8527      Inf
sample estimates:
mean of x
 124.6306

> t.test(Schadstoffgehalt, mu = 120,
         alternative = "greater")

        One Sample t-test

data:  Schadstoffgehalt
t = 2.0352, df = 99, p-value = 0.02225
```

```
alternative hypothesis: true mean is greater than 120
95 percent confidence interval:
 120.8527      Inf
sample estimates:
mean of x
 124.6306

> quantile(Schadstoffgehalt, 0.85)
     85%
148.1529
```

Aufgabe 3 (ca. 20 Punkte) **Teile unabhängig voneinander lösbar**

Aufgabe 3 Teil 1

Die Dauer von Telefongesprächen sei exponentialverteilt und betrage im Durchschnitt 4 Minuten.

1. Mit welcher Wahrscheinlichkeit hat ein zufällig ausgewählter Anruf eine Dauer von mindestens 8 Minuten?
2. Mit welcher Wahrscheinlichkeit hat ein zufällig ausgewählter Anruf eine Dauer zwischen 3 und 5 Minuten?
3. Ab welcher Gesprächszeit gehört ein Anruf zu den 2% längsten?

Hinweis zur Exponentialverteilung mit Parameter λ: Die Verteilungsfunktion lautet

$$F(x) = 1 - \exp(-\lambda x) \text{ für } x \geq 0.$$

Der Erwartungswert einer entsprechend verteilten Zufallsvariablen lautet

$$\mathrm{E}(X) = 1/\lambda.$$

(Platz für Aufgabe 3 Teil 1)

Aufgabe 3 Teil 2

Die Dauer eines Handgriffs in einem Herstellungsprozess sei normalverteilt mit Parameter $\mu = 4$ Min und $\sigma^2 = 6,25$ Min^2.

1. Mit welcher Wahrscheinlichkeit dauert ein Handgriff mindestens 8 Minuten?

2. Mit welcher Wahrscheinlichkeit dauert ein Handgriff zwischen 3 und 5 Minuten?

3. Ab welcher Zeitdauer gehört ein Handgriff zu den 2% längsten?

Rechenhilfe: nachfolgend einige Werte der Verteilung und der Quantile der Standardnormalverteilung:
(Platz für Aufgabe 3 Teil 2)

```
> pnorm(-0.4)
[1] 0.3445783
> pnorm(0.4)
[1] 0.6554217
> pnorm(0.8)
[1] 0.7881446
> pnorm(1.2)
[1] 0.8849303
> pnorm(1.6)
[1] 0.9452007
> pnorm(2)
[1] 0.9772499
> pnorm(-2)
[1] 0.02275013
> qnorm(0.02)
[1] -2.053749
> qnorm(0.98)
[1] 2.053749
```

Aufgabe 4 (ca. 18 Punkte) **Teile unabhängig voneinander lösbar**

Aufgabe 4 Teil 1

Was für ein Chaos! In der Schublade liegen die 20 Socken einzeln und die Hälfte davon hat Löcher... Bevor er zum Fußballturnier fährt, holt Francis noch schnell 5 Socken und hofft dabei, mindestens ein Paar ohne Loch anziehen zu können.

1. Welche Verteilung (Name, Parameter) hat die Zufallsvariable X, Anzahl der Socken ohne Loch in der Sporttasche?

2. Mit welcher Wahrscheinlichkeit sind tatsächlich mindestens 2 Socken ohne Loch dabei? Als Rechenhilfe nutzen Sie bitte folgendes R-Code, wobei der Vektor socken1 die Wahrscheinlichkeitsfunktion (= diskrete Dichtefunktion) der Variablen X an der Stellen 0 bis 5 angibt.

```
> socken1
[1] 0.01625387 0.13544892 0.34829721 0.34829721 0.13544892
0.01625387
```

(Platz für Aufgabe 4 Teil 1)

Aufgabe 4 Teil 2

In der Socken-Produktion entdeckt man einen Fehler in der Maschineneinstellung. So haben 5% der produzierten Socken eine Laufmasche. Es werden 40 Stück als Stichprobe entnommen (aufgrund der großen Gesamtheit kann man den Unterschied zwischen 'mit' und 'ohne Zurücklegen' vernachlässigen).

1. Welche Verteilung (Name, Parameter) hat die Zufallsvariable Y, Anzahl der Socken **mit Laufmasche** in der Stichprobe?

2. Mit welcher Wahrscheinlichkeit haben in dieser Stichprobe höchstens 2 Socken eine Laufmasche?

(Platz für Aufgabe 4 Teil 2)

Aufgabe 5 (ca. 12 Punkte)

Einem Parkhauswächter ist aufgefallen, dass Autos aus dem Landkreis häufiger bunt (rot, blau, grün usw.) und Autos aus der Stadt eher neutral (weiß, Silber, schwarz) sind. Nach einigen Monaten intensiver Zählungen steht fest:

- 60% der Autos im Parkhaus haben ein Landkreis Kennzeichen,
- 45% der Autos mit Landkreis Kennzeichen sind bunt, die Übrigen 55% neutral,
- 8 % der Autos mit Stadt Kennzeichen sind bunt, die Übrigen neutral.

Folgende Fragen sind zu beantworten, dabei werden Anteile als Wahrscheinlichkeiten interpretiert. Wenn die Antwort nicht bereits im Text zu finden ist, ist die Berechnung anzugeben.

1. Eine Kundin erzählt, dass sie ein neues Auto bestellt hat. Das Auto hat ein Stadt Kennzeichen. Mit welcher Wahrscheinlichkeit P_1 ist es bunt?
2. Wie wahrscheinlich (P_2) ist es, dass ein beliebiges Auto (etwa: das nächste einfahrende Auto) neutral ist und ein Stadt Kennzeichen hat?
3. Mit welcher Wahrscheinlichkeit P_3 ist ein Auto in diesem Parkhaus neutral?
4. Ein buntes Auto biegt in die Einfahrt zum Parkhaus ein. Mit welcher Wahrscheinlichkeit P_4 ist es aus dem Landkreis?

(Platz für Aufgabe 5)