Solutions of the exercises of Chapter 9 Mathematical Statistics

Exercise 1

a. $\beta_1 = 0$, if $\beta_1 = 0$ then the model equation $Y = \beta_0 + \beta_1 x + \varepsilon$ changes into $Y = \beta_0 + \varepsilon$ with as consequence that Y does not depend on x anymore.

b. If we test $H_0: \beta_1 = 0$ against $H_1: \beta_1 \neq 0$ and the null hypothesis is not rejected then we have lack of proof for the alternative hypothesis H_1 . So we can NOT prove that there is some relation between the dependent variable y and the explanatory variable x. The practical consequence is that we skip x out off the model because we prefer simple models. If we claim some relation according to our model then it should be proven.
Remember that in statistics we NEVER prove a null hypothesis. We prove or we can't prove the alternative hypothesis.

c. Formulas for the estimators: $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2}$ and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$.
We are interested in the changes in the estimates (outcomes of the estimators) in case of the indicated unit change (from guilder to euro).
Change in the Y_i: $\quad Y_i \quad \rightarrow \quad \frac{1}{2.20} \times Y_i$
As a consequence $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2}$ decreases by a factor 2.20.
Also $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$ decreases by a factor 2.20.

d. If we change the unit of x from meter to centimeter then the values x_i increase by a factor 100 (e.g. 1.23 meter becomes 123 centimeter).
Change in the x_i : $x_i \rightarrow 100 \times x_i$.
We conclude that $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2}$ decreases by a factor 100.
Consider now $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$:
Both parts, $\bar{Y}$ and $-\hat{\beta}_1 \bar{x}$, don't change, hence $\hat{\beta}_0$ does not change.

e. Researcher 2 does not have more reason for rejecting the null hypothesis.
Why not?
A formal reason: not $\hat{\beta}_1$ but $T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ is the appropriate test statistic, and we don't have any information about the standard error $se(\hat{\beta}_1)$.
A more practical reason: in this exercise we saw already that the value of $\hat{\beta}_1$ depends on the units chosen, this is an undesirable feature of a test statistic. So the value $\hat{\beta}_1$ alone is not informative
It turns out: the test statistic $T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ does not depend on the chosen units, see next elaboration.

f. Let us firstly investigate the implications of the unit change for Y_i from guilder tot euro.
We found: $\quad Y_i \quad \rightarrow \quad \frac{1}{2.20} \times Y_i \,,\ \hat{\beta}_1 \ \rightarrow \ \frac{1}{2.20} \times \hat{\beta}_1$
Consider now $se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$. Note S_{xx} does not change. According to the formula
$\hat{\sigma}^2 = \frac{\sum_i \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{n-2}$

$\hat{\sigma}^2$ decreases by a factor $(2.20)^2$ as Y_i , $\hat{\beta}_0$ and $\hat{\beta}_1$ all decrease by factor 2.20.

We conclude that both $\hat{\beta}_1$ and $se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$ decrease by a factor 2.20, hence $T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ does not change.

Now we investigate the implications of the unit change for x_i from meter to centimeter.

We found: $x_i \rightarrow 100 \times x_i$, $\hat{\beta}_1 \rightarrow \frac{1}{100} \times \hat{\beta}_1$

We have to investigate the (possible) change in $se(\hat{\beta}_1) = \hat{\sigma}/\sqrt{S_{xx}}$.

Because Y_i, $\hat{\beta}_0$ and $\hat{\beta}_1 x_i$ all don't change, $\hat{\sigma}$ does not change as well but $\sqrt{S_{xx}} = \sqrt{\sum_i (x_i - \bar{x})^2}$ increases by factor 100.

We conclude that $se(\hat{\beta}_1) = \hat{\sigma}/\sqrt{S_{xx}}$ decreases by a factor 100.

Since both $\hat{\beta}_1$ and $se(\hat{\beta}_1)$ decrease by a factor 100, $T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ does not change.

Exercise 2

a. Estimates: $\hat{\beta}_0 = 162.281$, $\hat{\beta}_1 = -81.304$.

b. $R^2 = 1 - \frac{SS_{error}}{SS_{total}} = 1 - \frac{72.998}{1203.369} = 93.9\%$

This means that 93,9% of the variance (spread) has been explained by the predictor/explanatory variable x.

c. From the formulas we can conclude that $\hat{\beta}_1$ and the sample correlation coefficient r share the same sign, we read $\hat{\beta}_1 = -81304$ so r has to be negative.

Furthermore $r^2 = R^2 = 0.939$, so $r = -\sqrt{0.939} = -0.969$.

d. We are estimating σ^2 by means of $S^2 = \frac{\sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$: we have to divide the residual sum of squares (numerator) by the corresponding degrees of freedom $(n-2)$: we get 9.125 .

e. 95% confidence interval for β_1 : boundaries are $\hat{\beta}_1 \pm c \times se(\hat{\beta}_1)$

Elaboration: $\hat{\beta}_1 = -81.304$ and $se(\hat{\beta}_1) = 7.305$ (from the output)

$c = 2.306$ (t-distribution, $n-2 = 8$ degrees of freedom, 95% probability between $-c$ and c)

95% confidence interval becomes:

$$(-81.304 - 2.306 \times 7.305, -81.304 + 2.306 \times 7.305) \approx (-98.1\,, -64.5)$$

f. Let Y denote the hardness and let denote x the corresponding annealing temperature. The eight steps of the test:

1. $Y = \beta_0 + \beta_1 x + \varepsilon$, With independent disturbances ε being $N(0, \sigma^2)$-distributed.
2. We test $H_0: \beta_1 = 0$ against $H_1: \beta_1 \neq 0$
3. Test statistic: $T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$
4. Under $H_0 : T \sim t_{n-2} = t_8$
5. Outcome of T: $T = \frac{-81.304}{7.305} = -11.130$ (from output)
6. We reject H_0 if $T \leq -c$ or $T \geq c$, where for $\alpha = 5\%$ $c = 2.306$.

7. The Rejection region contains outcome -11.130 so reject H_0 .
8. Using level of significance 5% we have proven that the hardness depends on the annealing temperature.

g. Looking at the scatter plot we should see only ‘chaos’, no pattern whatsoever.

We have to learn to judge residual plots. At first glance there is no pattern, perhaps caused by the fact that there are only a few points. Having a closer look you can distinguish a banana-shaped or V-shaped cloud of points vaguely.

This has its origin in the plot of y versus x: a slight curvature of the cloud of points can be distinguished as soon as you draw a straight line in the middle of the points. So a close examination of the plots may give rise to some doubt about the fit of the model..

We return to this data set in another exercise. Then we investigate whether quadratic regression fits the data better.

Exercise 3

a. The model equation becomes: $Y = \beta_1 x + \varepsilon$ (‘regression through the origin’)
We estimate the parameter β_1 by means of least squares:
Minimize $\sum_i (y_i - \beta_1 x_i)^2$ as function of β_1.
Taking the derivative: $-2\sum_i (y_i - \beta_1 x_i)x_i$
Equating the derivative to zero gives an equation for the estimate $\hat{\beta}_1$: $-2\sum_i \left(y_i - \hat{\beta}_1 x_i\right)x_i = 0$
Solving: $\sum_i \left(y_i - \hat{\beta}_1 x_i\right)x_i = \sum_i x_i y_i - \hat{\beta}_1 \sum_i x_i^2 = 0$ and hence $\hat{\beta}_1 = \sum_i x_i y_i / \sum_i x_i^2$.
You can check that indeed a minimum has been attained.

b. Now we have to study the new estimator $\hat{\beta}_1 = \sum_i x_i Y_i / \sum_i x_i^2$ with independent random variables Y_i distributed according to a normal distribution with expecation $\beta_1 x_i$ and (common) variance σ^2.
We compute its variance: $var\left(\hat{\beta}_1\right) = var\left(\sum_i x_i Y_i / \sum_i x_i^2\right)$

$$= var\left(\sum_i x_i Y_i\right) / \left(\sum_i x_i^2\right)^2$$
$$= \sum_i x_i^2 var(Y_i) / \left(\sum_i x_i^2\right)^2$$
$$= \sigma^2 \sum_i x_i^2 / \left(\sum_i x_i^2\right)^2 = \sigma^2 / \sum_i x_i^2$$

In general $\sum_i x_i^2 > \sum_i (x_i - \bar{x})^2$ so the new variance tends to be smaller than the old variance $\sigma^2 / \sum_i (x_i - \bar{x})^2$.

Exercise 4

We apply simple linear regression with y as dependent variable and $x = \dfrac{1}{z}$ as predictor variable.

a. Boundaries of 95% confidence interval for β_1: $\hat{\beta}_1 \pm c \times se\left(\hat{\beta}_1\right)$
Elaboration:
$\hat{\beta}_1 = 9.011$ and $se\left(\hat{\beta}_1\right) = 0.064$ (see output), $c = 2.306$ (8 degrees of freedom)
$c \times se\left(\hat{\beta}_1\right) = 2.306 \times 0.064 = 0.148$
95% confidence interval becomes: $(9.011 - 0.148, 9.011 + 0.148) = (8.86, 9.16)$

b. 95% confidence interval for β_0: $\hat{\beta}_0 \pm c \times se\left(\hat{\beta}_0\right)$
$\hat{\beta}_0 = 1.122$ and $se\left(\hat{\beta}_0\right) = 0.024$ (see output), $c = 2.306$ (8 degrees of freedom, unchanged)
$c \times se\left(\hat{\beta}_0\right) = 2.306 \times 0.024 = 0.055$

95% confidence interval for β_0 becomes: $(1.122 - 0.055, 1.122 + 0.055) = (1.07, 1.18)$

c. 95% confidence interval for $\beta_0 + \beta_1 x_0$ ($x_0 = 1/15$): $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0)$
Elaboration:
$\hat{\beta}_0 + \hat{\beta}_1 x_0 = 1.122 + 9.011 \times \left(\frac{1}{15}\right) = 1.723$ (c unchanged)
For $se(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$ we calculate using a simple desk calculator (no GR!):
$\bar{x} = 0.2251$ and $s_X = 0.3177$ (sample standard deviation of x)
So $S_{xx} = \sum_i (x_i - \bar{x})^2 = 9 \times 0.0.1009$,
$$se(\hat{\beta}_0 + \hat{\beta}_1 x_0) = 0.0637 \times \sqrt{\frac{1}{10} + \frac{(1/15 - 0.2251)^2}{9 \times 0.1009}} = 0.02253$$,
$c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0) = 2.306 \times 0.02253 = 0.052$.
95% confidence interval becomes: $(1.723 - 0.052, 1.723 + 0.052) = (1.67, 1.78)$

Exercise 5

We have to choose for the 95% prediction interval, note only one person is involved.
Prediction interval: $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm c \times se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)$
with $se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0) = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$
Elaboration:
$\hat{\beta}_0 + \hat{\beta}_1 x_0 = -14.380 + 0.769 \times 122 = 79.438$
$c = 2.179$ (12 degrees of freedom)
$\hat{\sigma} = \sqrt{68.688} = 8.288$
Using simple desk calculator: $\bar{x} = 133.93$ and $s_X^2 = (9.0423)^2 = 81.76$
So we get:
$S_{xx} = \sum_i (x_i - \bar{x})^2 = 13 \times 81.76$
$$c \times \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = 2.179 \times 8.288 \times \sqrt{1 + \frac{1}{14} + \frac{(122 - 133.93)^2}{13 \times 81.76}} = 19.82$$

95% prediction interval: $(79.44 - 19.82, 79.44 + 19.82) = (59.6, 99.3)$

Exercise 6
The joint density of observations $y_1, y_2, \ldots, y_n$ is
$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-n/2} exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right)$,
So the likelihood function is
$L(\beta_0, \beta_1, \sigma^2) = (2\pi\sigma^2)^{-n/2} exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right)$
which should maximized as function of β_0, β_1 and σ^2.
It is easy to see:
$L(\beta_0, \beta_1, \sigma^2) \le L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)$
with $\hat{\beta}_0$ and $\hat{\beta}_1$being the least squares estimates of β_0 and β_1, irrespective of the (positive) value of σ^2.
Formulas for the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$:
Minimize $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ as function of β_0 and β_1.
Equating derivatives to zero we get the following equations:

$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ en $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)\, x_i = 0$
From the first one we easily get: $\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$, hence $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.
Substituting this in the second equation we get:

$$\sum_{i=1}^{n}(y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)\, x_i = \sum_{i=1}^{n}\left(x_i(y_i - \bar{y}) - \hat{\beta}_1 x_i (x_i - \bar{x})\right) = 0$$

From which follows:

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} x_i(x_i - \bar{x}), \text{ hence } \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})}.$$

It can be justified that indeed a minimum is attained, but we don't pursue on this.
(Note furthermore there exist several formulas for $\hat{\beta}_1$.)

Now it remains to maximize $L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)$ as function of σ^2.

We maximize $ln\left(L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)\right)$ instead of $L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)$, this gives the same solution.

Note

$$ln\left(L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)\right) = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Its derivative with respect to σ^2 is

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \times \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

This derivative is zero when $\sigma^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 / n$,

it is positive when $\sigma^2 > \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 / n$ and

it is negative when $\sigma^2 < \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 / n$. We conclude that $ln\left(L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)\right)$ and hence $L(\hat{\beta}_0, \hat{\beta}_1, \sigma^2)$ is maximized for $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$.

We conclude $L(\beta_0, \beta_1, \sigma^2) \leq L(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$ with the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ presented and with $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$, hence $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ are the maximum likelihood estimates. Hence indeed the estimators presented in the text of the exercise are the maximum likelihood estimators.