Solutions of the exercises of Chapter 10 Mathematical Statistics

Exercise 1

a. Let Y denote the mortality rate per 100 000 males. Let us first investigate whether the variable *North* really affects the mortality rate Y, in addition to the variable *Calcium.* The eight steps of the test:

(1) $Y = \beta_0 + \beta_1 Calcium + \beta_2 North + \varepsilon$

With independent disturbances ε which are $N(0,\sigma^2)$-distributed.

(2) We test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$

(3) Test statistic: $T = \hat{\beta}_2 / se(\hat{\beta}_2)$

(4) Under $H_0 : T \sim t_{61-3} = t_{58}$

(5) Outcome of T: $\dfrac{176.711}{36.891} = 4.79$

(6) We reject H_0 if $T \leq -c$ or $T \geq c$. $\alpha = 5\%$, $t_{58} \Rightarrow c = 2.00$ (interpolation, not strictly necessary)

(7) The rejection region contains the outcome 4.79 so reject H_0

(8) Using level of significance 5% we have proven that the variable *North* really affects the mortality rate *Y*, in addition to the variable *Calcium.*

The test for investigating whether C*alcium* really affects mortality rate *Y*, in addition to *North,* is as follows:

(1) $Y = \beta_0 + \beta_1 Calcium + \beta_2 North + \varepsilon$

With independent disturbances ε which are $N(0,\sigma^2)$-distributed.

(2) We test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$

(3) Test statistics: $T = \hat{\beta}_1 / se(\hat{\beta}_1)$

(4) Under $H_0 : T \sim t_{61-3} = t_{58}$

(5) Outcome of T: $\dfrac{-2.034}{0.483} = -4.21$

(6) We reject H_0 if $T \leq -c$ or , where for $\alpha = 5\%$, $t_{58} \Rightarrow c = 2.00$ (interpolation)

(7) The Rejection region contains the outcome -4.21 so reject H_0.

(8) Using level of significance 5% we have proven that the variable *Calcium* really affects the mortality rate *Y*, in addition to the variable *North.*

b. We shall test whether it is useful to use the predictors *Calcium* and *North* for prediction of *Y*. (According to some textbooks this should be the first step in a statistical analysis.) The eight steps of the test:

(1) $Y = \beta_0 + \beta_1 Calcium + \beta_2 North + \varepsilon$, with independent disturbances ε which are $N(0,\sigma^2)$-distributed.

(2) We test $H_0 : \beta_1 = \beta_2 = 0$ against $H_1 : \beta_1 \neq 0 \vee \beta_2 \neq 0$

(3) Test statistic: $F = \dfrac{SS(regression)/k}{SS(error)/(n-k-1)} = \dfrac{SS(regression)/2}{SS(error)/58}$

(4) Under $H_0 : F \sim F_{58}^2$

(5) Outcome of F: $\dfrac{1248317.8/2}{864855.86/58} = \dfrac{624158.905}{14911.308} = 41.86$

(6) We reject H_0 if $F \geq c$

$\alpha = 5\%$, $F_{58}^2 \Rightarrow c = \dfrac{2}{20} \times 3.23 + \dfrac{18}{20} \times 3.15 = 3.16$ (interpolation between F_{40}^2 and F_{60}^2)

(7) Rejection Region contains outcome 41.86 so reject H_0.

(8) Using level of significance 5% we have proven that at least one of the predictor variables is useful for the prediction of the mortality rate Y.

c. We are searching for a pattern in the plot. There is no pattern, only chaos, this OK. No doubt about the fit.
Note we are looking in the vertical direction, judging the residuals.
The points are not evenly spread in the horizontal direction. This makes the judgement about the residuals more difficult.
Furthermore we can search for outliers. Nearly all standardized residuals are contained by the interval $(-2,2)$. On the average only 5% of the standardized residuals should be lying outside $(-1.96,1.96) \approx (-2.2)$. To our opinion the largest residual is not extreme enough to worry about.

d. $R^2_{adj} = 1 - \frac{n-1}{n-k-1} \times (SS_E/SS_T) = 1 - \frac{60}{58} \times \frac{864855.86}{2113173.7} = 57.7\%$
This means that 57.7% of the spread of the dependent variable is explained by the two predictor variables. This is not a good value, weak relationship, if we want predict the dependent variable because it is far away from the optimal value 100%. But this time it is natural: It is a bad situation if you could predict the mortality rate perfectly by means of only the calcium concentration and geographic variable *North.*

Exercise 2

a. The eight steps of the test:

(1) $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$, with independent disturbances ε which are $N(0,\sigma^2)$-distributed.

(2) We test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$

(3) Test statistic: $T = \hat{\beta}_2 / se(\hat{\beta}_2)$

(4) Under $H_0 : T \sim t_{10-3} = t_7$

(5) Outcome of T : $\frac{131.898}{40.648} = 3.25$

(6) We reject H_0 if $T \leq -c$ or $T \geq c$. $\alpha = 5\% \Rightarrow c = 2.365$

(7) The rejection region contains outcome 3.25 so reject H_0.

(8) Using level of significance 5% we have proven that the quadratic term has to be added to the model. We thus have proven that quadratic regression is better than simple linear regression.

b. We have to judge whether the residual plot shows chaos. That seems to be the case.
It is a little bit troublesome that there are more points in the left part of the plot, suggesting unequal spread of the residuals. I guess we need more data for assessing such a pattern.

c. Modified scheme of eight steps:

(1) $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$, with independent disturbances ε which are $N(0,\sigma^2)$-distributed.

(2) We test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$

(3) Test statistic: $T = \hat{\beta}_2 / se(\hat{\beta}_2)$

(4) Under $H_0 : T \sim t_{10-3} = t_7$

(5) Outcome of T : $\frac{131.898}{40.648} = 3.25$

(6) The (two-sided) p-value is 0.014

(7) Since p-value $\leq \alpha = 5\%$ we have to reject the null hypothesis.

(8) Using level of significance 5% we have proven that the quadratic term has to be added to the model. We thus have proven that quadratic regression is better than simple linear regression.

If we choose $\alpha = 1\%$ then p-value$= 0.014 > \alpha$, and hence we don't reject H_0.

Exercise 3

a. For answering the question we have to rearrange the model equation:

$Y = (\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_3 x_2) \times x_1 + \varepsilon$

If we keep x_2 fixed then Y depends on x_1 in a linear way, $\beta_0 + \beta_2 x_2$ is the (new) constant (intercept) and $\beta_1 + \beta_3 x_2$ is the (new) slope. Note that this slope is affected by x_2 , this disappears when we skip the interaction term.

b. We copy the output and replace all signs '?'.

Analysis of Variance				
	df	SS	MS	F
Regression	3	92.110	30.70	67.71
Residual	8	3.627	0.4534	
total	11	95.737		

Variables in the equation			
Variables	coefficient	Std. error	T
Intercept	14.9600		
x_1	1.5321	0.5910	2.59
x_2	-0.4323	1.7964	-0.24
$x_1 \times x_2$	-0.0553	0.1554	-0.36

c. The estimated equation:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2$$

We get from the output:

$$y = 14.9600 + 1.5321 x_1 - 0.4323 x_2 - 0.0553 x_1 x_2$$

We calculate a (point)prediction by using the estimated equation with $x_1 = 12$ and $x_2 = 3$. This renders the following prediction:

$$\hat{y} = 14.9600 + 1.5321 \times 12 - 0.4323 \times 3 - 0.0553 \times 12 \times 3 = 30.06$$

d. Prediction for $x_1 = 12$ en $x_2 = 4$:

$$\hat{y} = 14.9600 + 1.5321 \times 12 - 0.4323 \times 4 - 0.0553 \times 12 \times 4 = 28.96$$

This is strange because the last prediction is smaller than the first prediction.

We should think about the interpretation of the individual parameters β_i:

The parameter β_i in general reflects the mean change in the dependent variable Y when we increase x_i with 1 unit and fix the other predictor variables.

Fixing the other predictor variables while changing one specific predictor variable does not reflect reality often, because many times predictor variables are changing simultaneously:

Here raising advertisement costs and increasing the number of sales representatives may be simultaneous actions. Then distinguishing the respective causes of the predictor variables may be difficult because of the dependence the predictors.

In statistics it is a famous phenomenon that estimates $\hat{\beta}_i$ turn out to have the wrong sign (you expected a positive value but you get a negative value or vice versa). Many times this phenomenon can be explained by strong relationships between predictor variables, to some extent a number of predictor variables share the same information. Sometimes the 'strangeness' can be solved by simplifying the model.

e. Boundaries for 95% confidence interval for β_2 : $\hat{\beta}_2 \pm c \times se(\hat{\beta}_2)$ with
-0.4323 and 1.7964 for $\hat{\beta}_2$ and $se(\hat{\beta}_2)$ respectively, and $c = 2.31$ (t_8-distribution).We get:
$(-0.4323 - 2.31 \times 1.7964, -0.4323 + 2.31 \times 1.7964) = (-4.582, 3.717)$

f. The eight steps of the test:
1. Model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2 + \varepsilon$,
with independent disturbances ε which are $N(0, \sigma^2)$-distributed.
2. We test $H_0: \beta_3 = 0$ against $H_0: \beta_3 \neq 0$
3. Test statistic: $T = \hat{\beta}_3 / se(\hat{\beta}_3)$
4. Under H_0 $\mathrm{T} \sim \mathrm{t}_8$
5. Outcome of T: $\dfrac{-0.0553}{0.1554} = -0.36$
6. We reject H_0 if $T \leq -c$ or $T \geq c$. Level of significance 5%, t-table $\Rightarrow c = 2.31$
7. The rejection region does not contain -0.36, hence we don't reject H_0.
8. Using level of significance 5% we did not prove that the interaction term should be part of the model.

g. When you apply again the t-test for testing $H_0: \beta_2 = 0$ against $H_0: \beta_2 \neq 0$ then again you need not reject the null hypothesis.
This does not mean that you have to skip both terms $\beta_2 x_2$ and $\beta_3 x_1 \times x_2$. Skip first the interaction term (a model with the interaction term and without the term $\beta_2 x_2$ is weird). Continue with the model without interaction and test $H_0: \beta_2 = 0$ against $H_0: \beta_2 \neq 0$ in order to investigate whether we can skip x_2 completely. For this procedure we thus need more output.

h. Scatter plots of the data is always helpful for appreciating the model applied. Furthermore a model check by means of a scatter plot of residuals is missing.

Exercise 4

a. We test whether the three groups differ systematically. Differences between the expectations are represented by the parameters β_1 and β_2. We thus test $H_0: \beta_1 = \beta_2 = 0$ against $H_1: \beta_1 \neq 0 \vee \beta_2 \neq 0$.
The eight steps of the test:
(1) $Y = \beta_0 + \beta_1 Control + \beta_2 Family + \varepsilon$, with independent disturbances ε which are $N(0, \sigma^2)$-distributed.
(2) We test $H_0: \beta_1 = \beta_2 = 0$ against $H_1: \beta_1 \neq 0 \vee \beta_2 \neq 0$
(3) Test statistic: $F = \dfrac{SS_{Regr}/k}{SS_{Error}/(n-k-1)} = \dfrac{SS_{Regr}/2}{SS_{Error}/69}$
(4) Under $H_0: F \sim F_{69}^2$
(5) Outcome of F : 5.422 (from output)
(6) We reject H_0 if $F \geq c$

$\alpha = 5\%$, $F_{69}^2 \Rightarrow c = \dfrac{11}{20} 3.15 + \dfrac{9}{20} 3.11 = 3.13$ (interpolation between F_{60}^2 and F_{80}^2)

(7) The rejection region contains the outcome 5.422, hence we reject H_0.
(8) Using level of significance 5% we have proven that the three groups differ systematically.

b. From the medical point of view it is better to take the control group as reference group.
Then you need indicator variables for the cognitive behavioral treatment group and the family treatment group:

$x_1 = 1$ if the girl belongs to the cognitive behavioral treatment group, otherwise $x_1 = 0$,
$x_2 = 1$ if the girl belongs to the family treatment group, otherwise $x_2 = 0$.

c. Boundaries for the 99% confidence interval for β_1: $\hat{\beta}_1 \pm c \times se(\hat{\beta}_1)$

$c = \frac{51}{60} \times 2.660 + \frac{9}{60} \times 2.617 = 2.65$ ($df = 69$, interpolation)

The 99% confidence interval for β_1 becomes thus:

$(-3.457 - 2.65 \times 2.033, -3.457 + 2.65 \times 2.033) = (-8.84, 1.93)$

β_1 is the difference in the expectation of Y, if we compare the control group with the reference group being the cognitive behavioral treatment.

d. Note that the increase in weight is just the dependent variable. The estimated expectations are as follows:.

Cognitive behavioral treatment: $\hat{\beta}_0 = 3.007 = 3.01$

Control: $\hat{\beta}_0 + \hat{\beta}_1 = 3.007 - 3.457 = -0.450$

Family therapy: $\hat{\beta}_0 + \hat{\beta}_2 = 3.007 + 4.258 = 7.27$

Exercise 5

a. The residual plot resulting form the model with both predictors *diameter* and *height*, shows a pattern clearly, that is wrong. From left to right the spread (varance) is increasing, and a (slight) curvature is present
We don't see outliers.

b. The residual plot resulting from the model with dependent variable *lnvolume* and predictor *lndiameter*, shows no pattern any more, to our opinion. There are no outliers.

c. The residual plot resulting from the model with dependent variable *lnvolume* and predictors *lndiameter*, and lnheight shows no pattern any more, to our opinion. There are no outliers.

d. Note that we changed our data by transforming the variables, so the first two values of R^2_{adj} (before transformation) are not comparable with the last two values of R^2_{adj} (after transformation).

e. Note that in principle we are able to construct a prediction interval for the volume of Patrick's tree using both predictors *lndiameter* and *lnheight*. Extend the theory of section 9.8 using linear algebra and the theoretical results of this chapter. But we need a computer program to calculate such a prediction interval. So we don't do this.
The boundaries of the prediction of the *lnvolume* of the tree of Patrick:

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm cS\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

Elaboration: $x_0 = \ln(16.0) = 2.773$

$\hat{\beta}_0 + \hat{\beta}_1 x_0 = -2.219 + 2.150 \times 2.773 = 3.743$

$c = 2.056$

$S = \sqrt{S^2} = \sqrt{0.365/26} = 0.1185$

$S_{xx} = \sum_i (x_i - \bar{x})^2 = (n-1) \times sample\ varince\ of\ lndiameter = 27 \times (0.19983)^2 = 1.078$

$cS\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = 2.056 \times 0.1185\sqrt{1 + \frac{1}{28} + \frac{(2.773 - 2.6012)^2}{1.078}} = 0.251$

95% prediction interval for $lnvolume$: $(3.743 - 0.251, 3.743 + 0.251) = (3.492, 3.994)$

95% prediction interval for $volume$: $(32.85, 54.27)$

f. Note that the confidence interval with boundaries $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm cS\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$ is an confidence interval for $\beta_0 + \beta_1 x_0$, this is the average value for $lnvolume$ given a value x_0.
We don't want to know an average value but the value for Patrick's tree. Hence we computed the prediction interval for $lnvolume$ for Patrick's tree and transformed it to interval for the $volume$ of the tree.

Exercise 6

a. We rewrite the formula $F = \frac{SS_R/k}{SS_E/(n-k-1)}$

Dividing both numerator and denominator by SS_T and using $R^2 = SS_R/SS_T$ we get:

$$F = \frac{\frac{R^2}{k}}{\frac{(SS_E/SS_T)}{n-k-1}}$$

Noting $SS_E/SS_T = 1 - R^2$ we finally get: $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$.

b. We have to invert the formula $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$ as function of R^2:: $\frac{R^2}{1-R^2} = kF/(n-k-1)$

$R^2 = \frac{kF}{n-k-1} - \frac{kF}{n-k-1} \times R^2$ or $\left(1 + \frac{kF}{n-k-1}\right) \times R^2 = \frac{kF}{n-k-1}$ or $R^2 = (kF/(n-k-1))/(1 + kF/(n-k-1))$

Note that R^2 is strictly increasing function of F, and vice versa. You may check this by taking the derivative, or recognizing that we are dealing with a rescaled version of the well known function $g(x) = \frac{x}{1+x}$.

c. F table: the test statistic F has the F distribution with $k = 3$ and $n - k - 1 = 16$ degrees of freedom. We get critical value 3.24, we hence reject the null hypothesis if $F \geq 3.24$.

The corresponding critical value for 'test statistic' R^2 would be equal to

$(k \times 3.24/(n-k-1))/(1 + k \times 3.24/(n-k-1)) = 0.6075/1.6075 = 37.8\%$.

Regarding the strictly increasing function for R^2 as function of F, an equivalent procedure for the rejection of the null hypothesis is as follows: reject the null hypothesis if $R^2 \geq 37.8\%$.

Exercise 7

Assuming the model $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$ we have to show that the sum $\sum_i Y_i - \hat{Y}_i$ equals zero always, with $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}$.

Consider the vectors $Y = (Y_1, Y_2, \ldots, Y_n)^T$ and $\hat{Y} = \left(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n\right)^T$. From chapter 10 we learned that the vector of all residuals $Y - \hat{Y}$ is orthogonal to each column of the design matrix X when we rewrite the model as $Y = X\beta + \varepsilon$.

The first column of the matrix X is the n dimensional vector $a = (1,1,\ldots,1)^T$, so we get $a^T\left(Y - \hat{Y}\right) = 0$, and hence $\sum_i Y_i - \hat{Y}_i = 0$

Exercise 8

a. Simple linear regression is a special case of multiple regression, it is multiple regression with $k = 1$. Therefore:

(1) $\hat{\beta} = \begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix} \sim N\left(\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix}, \sigma^2(X^TX)^{-1}\right)$ with $X = \begin{pmatrix}1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n\end{pmatrix}$,

(2) $V = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$.

(3) $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

Hence $\hat{\beta}_0 + \hat{\beta}_1 x_0 = (1 \quad x_0)\hat{\beta} \sim N\left((1 \quad x_0)\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix}, (1 \quad x_0)\sigma^2(X^TX)^{-1}\begin{pmatrix}1 \\ x_0\end{pmatrix}\right) = N(\beta_0 + \beta_1 x_0, d\sigma^2)$

with $d = (1 \quad x_0)(X^TX)^{-1}\begin{pmatrix}1 \\ x_0\end{pmatrix}$.

Consulting the formula of $se\left(\hat{\beta}_0 + \hat{\beta}_1 x_0\right)$ we conclude $d = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}$.

Define $Z = \frac{\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0}{d\sigma}$, note $Z \sim N(0,1)$ and

$(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0)/se(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \frac{\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0}{d\hat{\sigma}} = \frac{Z}{\hat{\sigma}/\sigma} = \frac{Z}{\sqrt{V/(n-2)}}$.

Note furthermore Z and V are independent and (2): we conclude that $(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0)/se(\hat{\beta}_0 + \hat{\beta}_1 x_0)$ has the t distribution with $n - 2$ degrees of freedom.

b. Because of the t distribution we can find a constant c such that the event

$$-c < (\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0)/se(\hat{\beta}_0 + \hat{\beta}_1 x_0) < c$$

has probability 95% (or another confidence level).

Equivalent inequalities are:

$$c > (\beta_0 + \beta_1 x_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)/se(\hat{\beta}_0 + \hat{\beta}_1 x_0) > -c$$
$$-c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0) < \beta_0 + \beta_1 x_0 - (\hat{\beta}_0 + \hat{\beta}_1 x_0) < c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0)$$
$$\hat{\beta}_0 + \hat{\beta}_1 x_0 - c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0) < \beta_0 + \beta_1 x_0 < \hat{\beta}_0 + \hat{\beta}_1 x_0 + c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0).$$

So the probability of the last inequality is also 95%. The boundaries of the 95% confidence interval of $\beta_0 + \beta_1 x_0$ are thus: $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm c \times se(\hat{\beta}_0 + \hat{\beta}_1 x_0)$.

In case of $n = 20$ we get $n - 2 = 18$ degrees of freedom, so $c = 2.101$

c. Consider the random variables Y_0 and $\hat{\beta}_0 + \hat{\beta}_1 x_0$. Note that $Y_0, Y_1, Y_2, \ldots, Y_n$ are independent and that $\hat{\beta}_0 + \hat{\beta}_1 x_0$ is a function of $Y_1, Y_2, \ldots, Y_n$.

We conclude that Y_0 and $\hat{\beta}_0 + \hat{\beta}_1 x_0$ are independent. Let us determine the distribution of

$$D = Y_0 - (\hat{\beta}_0 + \hat{\beta}_1 x_0).$$

Noting that Y_0 and $\hat{\beta}_0 + \hat{\beta}_1 x_0$ are independent and that each has a normal distribution, we conclude that D has a normal distribution, using standard probability theory.

We compute expectation and variance of its distribution:

$$E(D) = E(Y_0) - E(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \beta_0 + \beta_1 x_0 - (\beta_0 + \beta_1 x_0) = 0$$
$$var(D) = var(Y_0) + var(\hat{\beta}_0 + \hat{\beta}_1 x_0) \quad \text{because of independence}$$
$$= \sigma^2 + d\sigma^2 = \sigma^2(1 + d) \quad \text{see } a \text{ for the constant } d$$

Define $Z = \frac{D}{\sigma\sqrt{1+d}}$. Note $Z \sim N(0,1)$ and

$$(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)/se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0) = \frac{D}{\hat{\sigma}\sqrt{1+d}} = \frac{Z}{\hat{\sigma}/\sigma} = \frac{Z}{\sqrt{V/(n-2)}}$$

Note that $Y_0, \hat{\beta}_0 + \hat{\beta}_1 x_0$ and V are independent and that hence D and V are independent, and that hence Z and V are independent. Noting $Z \sim N(0,1)$ and (2) we conclude that

$(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)/se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)$ has a t distribution with $n - 2$ degrees of freedom.

d. Thanks to the t distribution we can find a constant c such that the inequality

$$-c < (Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)/se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0) < c$$

holds with probability 95% (or another level of confidence). Equivalent inequalities are:

$$-c \times se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0) < Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0 < c \times se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)$$
$$\hat{\beta}_0 + \hat{\beta}_1 x_0 - c \times se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0) < Y_0 < \hat{\beta}_0 + \hat{\beta}_1 x_0 + c \times se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)$$

So the last inequality holds with probability 95% as well.

So the boundaries of 95% prediction interval are: $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm c \times se(Y_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0)$

In case of $n = 30$ we get $n - 2 = 28$ degrees freedom.

In case of 99% confidence we get (use column '0.005'): $c = 2.763$