Technichal University of Munich
Department of Mathematics

MA4401 Applied Regression, Homework problem 4

Prof. Donna Ankerst, Stephan Haug (December 12, 2017)

Problem H.4

Consider the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with X an $n \times (p+1)$ matrix with rank $p+1$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ a vector of uncorrelated errors with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I_n$. Further let $\widehat{\boldsymbol{\mu}} = X\widehat{\boldsymbol{\beta}}$ be the fitted values, where $\widehat{\boldsymbol{\beta}}$ is the vector of least squares estimates, and $H = X(X'X)^{-1}X'$ denotes the hat matrix.

a) Find the mean vector and covariance matrix of $\widehat{\boldsymbol{\mu}}$.

b) Show that

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(\widehat{\boldsymbol{\mu}}_i) = \sigma^2 \frac{p+1}{n}$$

Hint: Find the trace of $\mathrm{Cov}(\widehat{\boldsymbol{\mu}})$ and use the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for matrices A and B, whenever the product is well-defined.

c) Show that H is a symmetric and idempotent matrix (https://en.wikipedia.org/wiki/Idempotent_matrix). Further show that the diagonal elements h_{ii} must lie between zero and one.

Hint: Consider $\mathbf{a}_i' H \mathbf{a}_i$, where $\mathbf{a}_i \in \mathbb{R}^n$ is a vector with all components equal to 0 except for the i-th, which is 1.

d) Assume that the linear model contains a constant term. Show that the diagonal elements h_{ii} of the hat matrix satisfy $h_{ii} \geq \frac{1}{n}$.

Hint: Parametrise the model by centering the predictor variables, i.e. consider $x_{ij} - \bar{x}_j, j = 1, \dots, p$, as predictor variables instead of x_{ij}.

e) Read in the weightloss data set available on moodle. The response variable is `Loss` (weight loss in pounds after 1 month of diet). The predictor variables are `Diet` (type of diet), and `Before` (weight in pounds before the diet).
Use `ggplot()` for a scatterplot of `Loss` against `Before`. Determine the hat matrix for the model `Loss ~ Before`. Based on the hat matrix, compute the leverage for all data points. Mark the data points with high leverage in a different colour in the scatterplot. Does this approach catch all outliers?

General grading instructions: For each part there exist also other approaches to solve the problem. Hence you have to check if solutions, which are different than the proposed one, are also correct.

Remark: If in the end the total number of points is not an integer (moodle only knows integers), you have to round up. So e.g. a result of 8.5 will be evaluated as 9 in moodle.

Solution

a) From Lecture 2a we know that $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(X'X)^{-1}$. This leads to

$$E(\widehat{\boldsymbol{\mu}}) = E(X\widehat{\boldsymbol{\beta}}) = XE(\widehat{\boldsymbol{\beta}}) = X\boldsymbol{\beta}$$
$$\mathrm{Cov}(\widehat{\boldsymbol{\mu}}) = \mathrm{Cov}(X\widehat{\boldsymbol{\beta}}) = X\mathrm{Cov}(\widehat{\boldsymbol{\beta}})X' = \sigma^2 X(X'X)^{-1}X'$$

Grading: Each result gives **0.5 points**.

b) Due to the hint, we start with computing the trace of the covariance matrix of $\widehat{\boldsymbol{\mu}}$, where we use the second part of the hint for $A := X$ and $B := (X'X)^{-1}X'$. We get the following

$$\mathrm{tr}\left(\mathrm{Cov}(\widehat{\boldsymbol{\mu}})\right) = \mathrm{tr}\left(\sigma^2 X(X'X)^{-1}X'\right) = \sigma^2 \mathrm{tr}\left((X'X)^{-1}X'X\right) = \sigma^2 \mathrm{tr}(I_{p+1}) = \sigma^2(p+1)$$

Since $\mathrm{tr}\left(\mathrm{Cov}(\widehat{\boldsymbol{\mu}})\right) = \sum_{i=1}^n \mathrm{Var}(\widehat{\boldsymbol{\mu}}_i)$ we get

$$\frac{1}{n}\sum_{i=1}^n \mathrm{Var}(\widehat{\boldsymbol{\mu}}_i) = \sigma^2 \frac{p+1}{n}$$

Grading: Computing the trace of the covariance matrix gives **0.5 point**. *Applying this result in the second step gives* **0.5 points**.

c) In the first step we show that H is a symmetric and idempotent matrix.

i. We have

$$H' = \left(X(X'X)^{-1}X'\right)' = \left((X'X)^{-1}X'\right)'X' = X\left((X'X)^{-1}\right)'X' = X(X'X)^{-1}X' = H\,,$$

which shows that H is symmetric.

ii. We have

$$HH = X(X'X)^{-1}X'X(X'X)^{-1}X' = XI_{p+1}(X'X)^{-1}X' = X(X'X)^{-1}X' = H\,,$$

which shows that H is idempotent.

In the next step we have to show that the diagonal elements h_{ii} of H must lie between zero and one. To show the first assumption we consider for any vector $\mathbf{a} \in \mathbb{R}^n$ the quadratic form

$$\mathbf{a}'H\mathbf{a} = \mathbf{a}'X(X'X)^{-1}X'\mathbf{a} =: \tilde{\mathbf{a}}'(X'X)^{-1}\tilde{\mathbf{a}}\,.$$

Since $(X'X)^{-1}$ is a positive semidefinite matrix, we have

$$0 \le \tilde{\mathbf{a}}'(X'X)^{-1}\tilde{\mathbf{a}} = \mathbf{a}'H\mathbf{a} \qquad \forall \mathbf{a} \in \mathbb{R}^n\,.$$

In particular we have

$$\mathbf{a}_i'H\mathbf{a}_i \ge 0 \qquad \forall i \in \{1, \dots, n\},$$

but since $\mathbf{a}_i'H\mathbf{a}_i = h_{ii}$, this shows the first part of the assumption. To show the second part, we use the fact that H is symmetric and idempotent. This implies that

$$h_{ii} = h_{ii}^2 + \sum_{\substack{j=1\\ j\neq i}}^n h_{ij}^2 \ge 0$$

and thus

$$\sum_{\substack{j=1\\ j\neq i}}^n h_{ij}^2 = h_{ii}(1-h_{ii}) \ge 0\,.$$

Since $h_{ii} \geq 0$, we need to have $1 - h_{ii} \geq 0$, and hence $h_{ii} \leq 1$.

Grading: Showing that H is symmetric **and** *idempotent gives* **1 point**. *For showing that $h_{ii} \geq 0$ and $h_{ii} \leq 1$ one gets* **1 point** *each.*

d) If we use centred predictor variables, we can parametrise the model like this

$$\mathbf{Y} = \begin{pmatrix} \mathbf{1}_n & V \end{pmatrix} \boldsymbol{\beta}^c + \boldsymbol{\varepsilon}\,,$$

where $\mathbf{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$, $V \in \mathbb{R}^{n \times (p+1)}$ with columns $\mathbf{v}_j = \mathbf{x}_j - \bar{x}_j \cdot \mathbf{1}_n$ and $\boldsymbol{\beta}^c = (\alpha, \beta_1, \ldots, \beta_p)'$ with

$$\alpha = \beta_0 + \beta_1 \bar{x}_1 + \cdots + \beta_p \bar{x}_p\ .$$

Now we have $X = \begin{pmatrix} \mathbf{1}_n & V \end{pmatrix}$ and

$$\mathbf{1}_n' \mathbf{v}_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j) = 0$$

for $j \in \{1, \ldots, p\}$. Hence

$$X'X = \begin{pmatrix} n & \mathbf{0}' \\ \mathbf{0} & V'V \end{pmatrix} \qquad \text{and} \qquad (X'X)^{-1} = \begin{pmatrix} n^{-1} & \mathbf{0}' \\ \mathbf{0} & (V'V)^{-1} \end{pmatrix}\,,$$

which leads to a hat matrix

$$H = \begin{pmatrix} \mathbf{1}_n & V \end{pmatrix} \begin{pmatrix} n^{-1} & \mathbf{0}' \\ \mathbf{0} & (V'V)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_n' \\ V' \end{pmatrix} = \begin{pmatrix} n^{-1} \mathbf{1}_n \mathbf{1}_n' + V(V'V)^{-1} V' \end{pmatrix}\ .$$

The matrix $H^c = V(V'V)^{-1}V'$ is symmetric and idempotent as we have shown in part c). Further we have shown that the diagonal elements h_{ii}^c are between zero and one. Hence, the diagonal elements of H satisfy

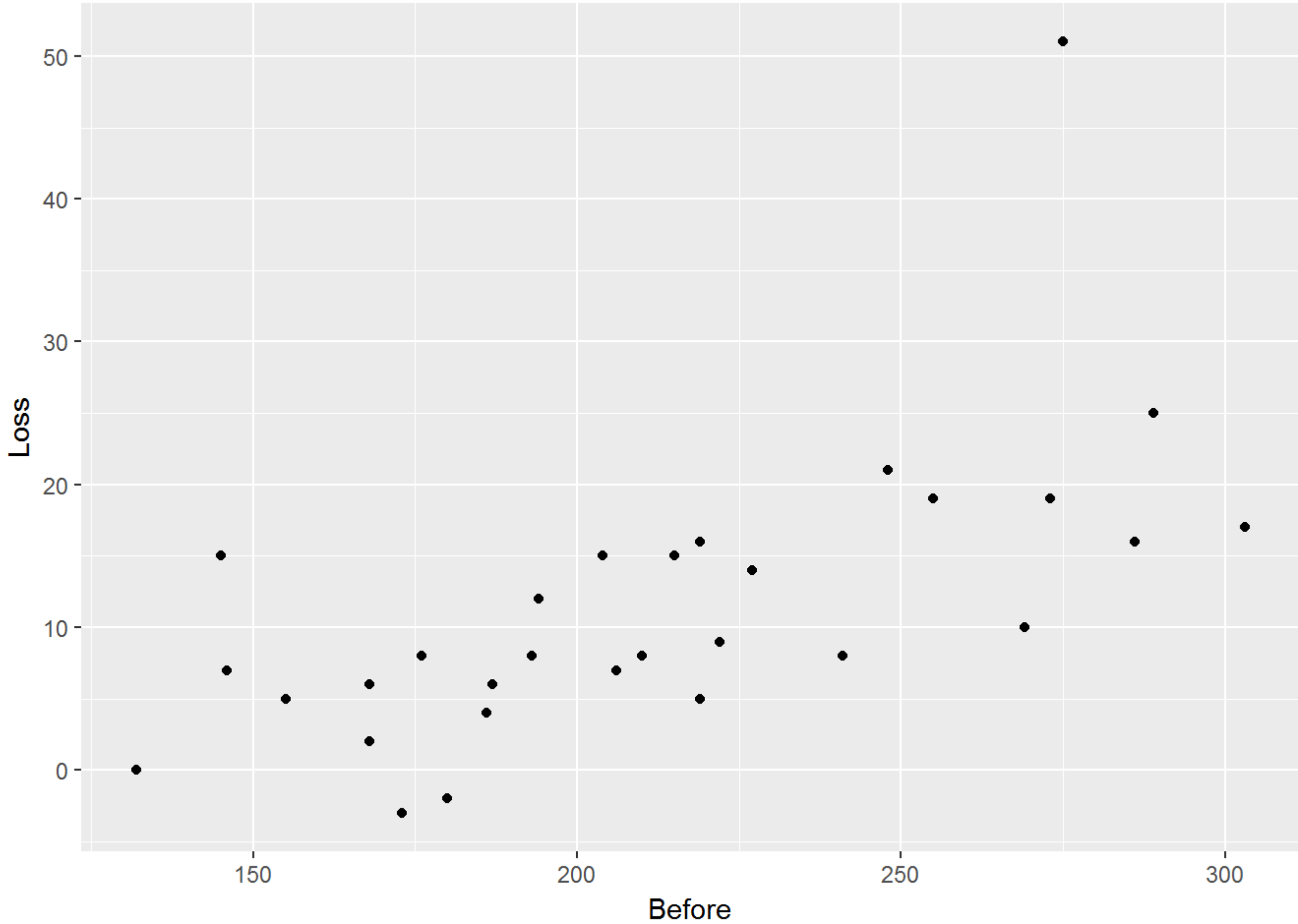
$$h_{ii} = n^{-1} + h_{ii}^c \geq n^{-1}\ .$$

Grading: The parametrisation of the linear model in terms of $\boldsymbol{\beta}^c$ gives **1 point.** *The representation of the hat matrix H gives* **1 point**. *Forthe last step in showing $h_{ii} \geq n^{-1}$ one gets* **1 point.**

e)

```
library(tidyverse)
weight <- read_csv("weightloss.csv")
```

```
ggplot(weight, aes(x = Before, y = Loss)) + geom_point()
```



To compute the hat matrix $H = X(X'X)^{-1}X'$, we first construct the design matrix X

```
X <- cbind(rep(1,ncol(weight)), weight$Before)
```

Afterwards we compute H and extract the diagonal elements with `diag()` to compute the leverage h_{ii} for each observation.

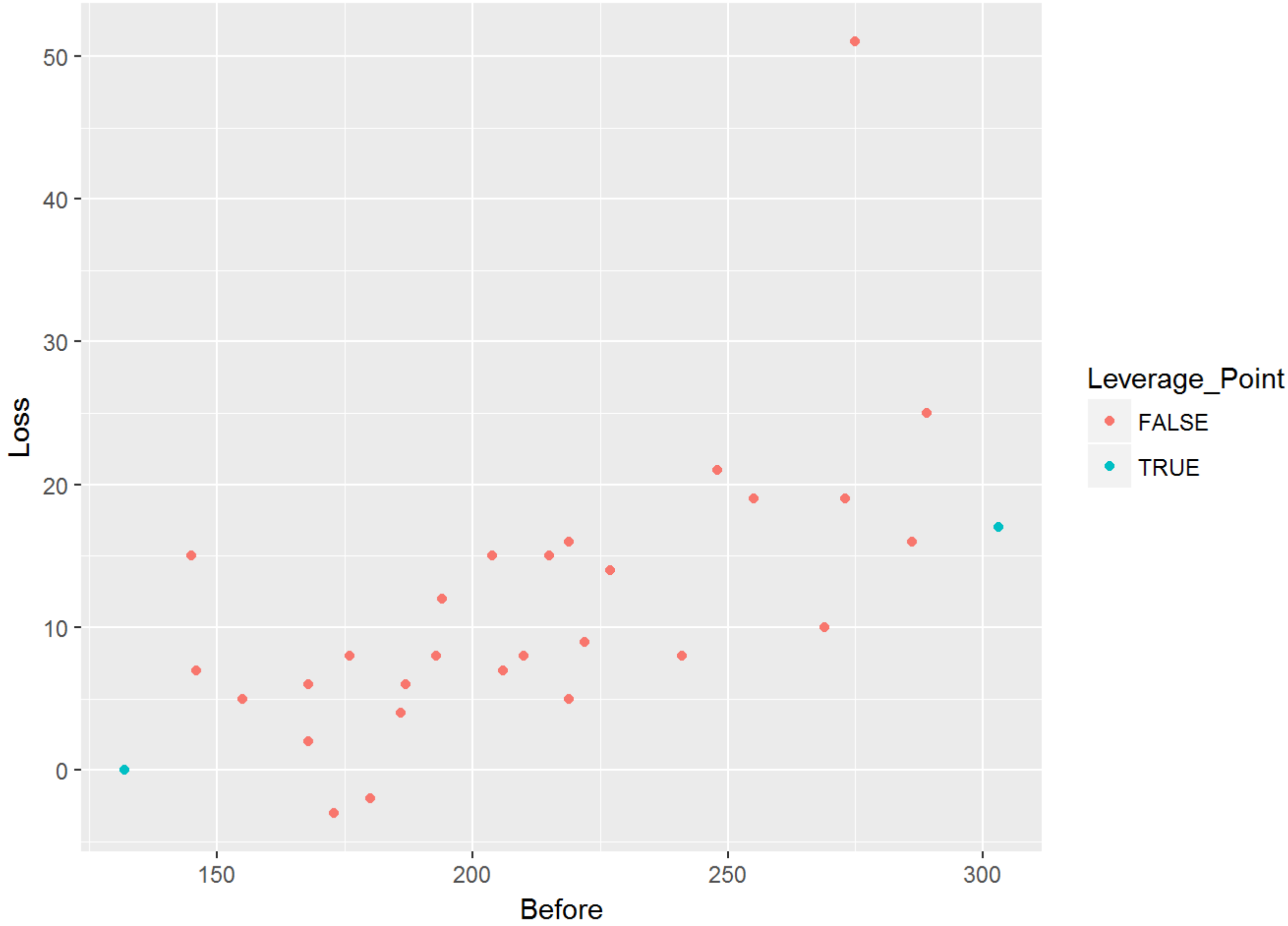
```
H <- X %*% solve(t(X)%*%X) %*% t(X)
h <- diag(H)  # leverage
```

High leverage observation are those, where $h_{ii} > 2\bar{h}$. Hence, we check this condition for all leverage values and store the information as a new column of `weight`

```
weight$Leverage_Point <- h > 2*mean(h)
```

Based on this new variable, we can then choose different colours in the scatterplot for high leverage and non-high leverage observations.

```
ggplot(weight, aes(x = Before, y = Loss, colour = Leverage_Point)) + geom_point()
```



The scatterplot shows an observation with an unusually high weight loss. This is an outlier in the response value y, but not in x. Leverage only catches outliers in x.

Grading: Computing the leverage values h_{ii} gives **0.5 points**. *The scatterplot with correctly coloured points gives* **1 point**. *If the colours are wrong (not there, or the wrong points are identified due to a different criteria), but a scatterplot of the data is shown one gets* **0.5 points** *instead*. *For a sentence, equivalent to the last one in the solution, one gets* **0.5 points**.