Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4267 Linear statistical models
Recommended exercises 5 – solutions

Problem 1 Simple linear regression

b) X has two non-zero columns (corresponding to intercept, and pressure), and neither is a multiple of the other. So the two columns are linearly independent, and $\operatorname{rank} X = \dim \operatorname{Col} X = 2$. We can check it in R by verifying that both eigenvalues of $X^{\mathrm{T}}X$ are positive ($\operatorname{rank}(X^{\mathrm{T}}X) = \operatorname{rank} X$ in general): `eigen(t(X) %*% X)`

c) The relationship between boiling point and pressure looks linear.

d) To calculate $\hat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}$:

```
betahat <- solve(t(X) %*% X) %*% t(X) %*% Y.
```

The estimated regression parameters are $\hat{\beta}_0 = 68.5$ and $\hat{\beta}_1 = 32.2$. We predict boiling point (or estimate its expected value) at pressure x by $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. thus, $\hat{\beta}_0$ is the boiling point the model would predict for $x = 0$ bar, i.e. in a vacuum, and $\hat{\beta}_1$ is the estimated rate of change of the boiling point with pressure: If the pressure increases by 1 bar, we expect the boiling point to increase by $\hat{\beta}_1$ kelvin.

e)
```
ehat <- Y - X %*% betahat
plot(X[, 2], ehat, pch = 20)
```

The points follow an inverted U-shape.

f) A linear model seems appropriate from the first plot. It is difficult to assess homoscedasticity from the second plot – it would have been easier with more data points. The second plot clearly shows a non-linear relationship, but it is difficult to judge whether the plot highlights a deviation from the linearity assumption of the model, or the error terms are somehow correlated. (For simple linear regression, a plot of residuals versus covariate is the same as the plot of response versus covariate, but with the estimated regression line rotated to be a horizontal axis, and the origin set at the intersection of the new horizontal axis and the vertical axis.) We see no specific evidence of non-additivity of errors, although one might consider several explanations for the look of the second plot.

We can use a normal Q–Q plot of the residuals to assess normality:

```
qqnorm(ehat, pch = 20)
qqline(ehat)
```

But bear in mind that the residuals are (slightly) correlated (their sum is zero) and only approximately normally distributed.

Problem 2 Results on $\hat{\boldsymbol{\beta}}$ and SSE in multiple linear regression

a) H is symmetric and idempotent, which means that $H\boldsymbol{y}$ is the orthogonal projection of $\boldsymbol{y}$ onto the column space of X (and H) for all $\boldsymbol{y} \in \mathbb{R}^n$. It is symmetric since

$$\begin{aligned} H^{\mathrm{T}} &= (X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})^{\mathrm{T}} = (X^{\mathrm{T}})^{\mathrm{T}}((X^{\mathrm{T}}X)^{-1})^{\mathrm{T}}X^{\mathrm{T}} \\ &= X((X^{\mathrm{T}}X)^{\mathrm{T}})^{-1}X^{\mathrm{T}} = X((X^{\mathrm{T}}(X^{\mathrm{T}})^{\mathrm{T}})^{-1}X^{\mathrm{T}} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = H \end{aligned}$$

and idempotent since

$$\begin{aligned} H^2 &= (X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})(X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}) \\ &= X((X^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}}X))(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = XI(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = H. \end{aligned}$$

$\operatorname{rank} H = \operatorname{rank} X = p$ since H and X have the same column spaces. It can also be seen by using $\operatorname{tr} R = \operatorname{rank} R$ for idempontent matrices R, and a property of the trace: If A is an $m \times n$ and B an $n \times m$ matrix, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. So $\operatorname{tr} H = \operatorname{tr}(X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}) = \operatorname{tr}((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X) = \operatorname{tr} I = p$, where I is an $p \times p$ identity matrix.

$H\boldsymbol{Y}$ is the orthogonal projection of $\boldsymbol{Y}$ onto the column space of X.

Also $I - H$ is symmetric and idempotent, where I is now an $n \times n$ identity matrix: $(I-H)^{\mathrm{T}} = I^{\mathrm{T}} - H^{\mathrm{T}} = I - H$, and $(I-H)^2 = (I-H)I - (I-H)H = I - H - IH + H^2 = I - H - H + H = I - H$.

$\operatorname{rank}(I - H) = \operatorname{tr}(I - H) = \operatorname{tr} I - \operatorname{tr} H = n - \operatorname{rank} H = n - p$.

By the orthogonal decomposition theorem (also known as the projection theorem) of linear algebra, if $\hat{\boldsymbol{Y}}$ is the orthogonal projection of $\boldsymbol{Y}$ onto a subspace, then $\boldsymbol{Y} - \hat{\boldsymbol{Y}}$ is the orthogonal projection of $\boldsymbol{Y}$ onto the orthogonal complement of the subspace. In our case, $H\boldsymbol{Y}$ is the orthogonal projection of $\boldsymbol{Y}$ onto the column space of X, so $\boldsymbol{Y} - H\boldsymbol{Y} = (I-H)\boldsymbol{Y}$ is the orthogonal projection of $\boldsymbol{Y}$ onto the orthogonal complement of the column space of X. (This provides another way to prove that $\operatorname{rank}(I-H) = n-p$.)

b) One of the key theorems of this course (Theorem B.8.2 on p. 651 of Fahrmeir et al.) states that if D is a symmetric and idempotent matrix with rank r, and $\boldsymbol{Z} \sim N(\boldsymbol{0}, I)$, then $\boldsymbol{Z}^{\mathrm{T}}D\boldsymbol{Z} \sim \chi^2_r$.

Applied to $D = I - H$ and $\boldsymbol{Z} = \frac{1}{\sigma^2}(\boldsymbol{Y} - X\boldsymbol{\beta}) \sim N(\boldsymbol{0}, I)$, we get

$$\frac{1}{\sigma^2}(\boldsymbol{Y} - X\boldsymbol{\beta})^{\mathrm{T}}(I - H)(\boldsymbol{Y} - X\boldsymbol{\beta}) \sim \chi^2_{n-p}.$$

But $(I - H)X\boldsymbol{\beta} = \boldsymbol{0}$, since $HX = X$, and the above simplifies to

$$\frac{1}{\sigma^2}\mathrm{SSE} = \frac{1}{\sigma^2}\boldsymbol{Y}^{\mathrm{T}}(I - H)\boldsymbol{Y} \sim \chi^2_{n-p},$$

which is the distributional result for SSE that is asked for.

In the following, we use that the expected value of a χ^2_r-variable is r and its variance $2r$. Then $E(\mathrm{SSE}/\sigma^2) = n - p$, and $\sigma^2 = E(\mathrm{SSE})/(n-p)$, suggesting the unbiased estimator $\mathrm{SSE}/(n-p)$ for σ^2. Its variance is

$$\operatorname{Var}\frac{\mathrm{SSE}}{n-p} = \operatorname{Var}\Big(\frac{\sigma^2}{n-p}\frac{\mathrm{SSE}}{\sigma^2}\Big) = \frac{\sigma^4}{(n-p)^2}\operatorname{Var}\frac{\mathrm{SSE}}{\sigma^2} = \frac{\sigma^4}{(n-p)^2}\cdot 2(n-p) = \frac{2\sigma^4}{n-p}.$$

c) Since X^{T}, and thus $(X^{\mathrm{T}}X)^{-1}$, has p rows, and X^{T} has n columns, $A = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ is a $p \times n$ matrix. $B = I - H$ is an $n \times n$ matrix.

$\operatorname{Cov}(A\boldsymbol{Y}, B\boldsymbol{Y}) = A(\operatorname{Cov}\boldsymbol{Y})B^{\mathrm{T}} = A \cdot \sigma^2 I B^{\mathrm{T}} = \sigma^2 AB^{\mathrm{T}} = \sigma^2(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(I-H) = \sigma^2(X^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}} - X^{\mathrm{T}}H) = O$, since $X^{\mathrm{T}}H = (HX)^{\mathrm{T}} = X^{\mathrm{T}}$. Since $(A^{\mathrm{T}}\ B^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{Y}$ is multivariate normal, $\operatorname{Cov}(A\boldsymbol{Y}, B\boldsymbol{Y}) = O$ implies $A\boldsymbol{Y}$ and $B\boldsymbol{Y}$ independent. Then $A\boldsymbol{Y} = \hat{\boldsymbol{\beta}}$ and $(B\boldsymbol{Y})^{\mathrm{T}}(B\boldsymbol{Y}) = \boldsymbol{Y}^{\mathrm{T}}(I-H)\boldsymbol{Y} = \mathrm{SSE}$ are independent. This independence is used in the construction of t-tests for the components β_j of $\boldsymbol{\beta}$.