Norwegian University of Science and Technology
Department of Mathematical Sciences

Correction 26 May
and *addition 28 May*

TMA4267 2015V – Solutions

Problem 1

a) The fitted regression model is $\widehat{T} = 0.44 + 0.0197L + 0.045\,\theta + 0.023\,m$, where T is period (in s), L length (cm), θ amplitude (radians) and m mass (kg).

The model explains 98% of the variation of the data. The hypothesis that all regression coefficients are zero is rejected. The intercept and the coefficients of length and mass are significantly different from zero at the 1% level.

The residual plot shows low residuals for small and large values of fitted periods and high residuals for medium values of fitted residuals, suggesting that the model is wrong. Many wrote in their papers that the appearance of the residual plot were due to residuals not being independent, but in fact they may be independent but the linear model wrong.

The Box–Cox plot suggests a square transform of the response variable T.

b) I would prefer the new model, since the residual plot shows no clear structure.

The best subset selection suggests that L should always be present as a covariate. With two covariates, also θ should be present. Based on Mallows' C_P, the overall "best" model is the one including L and θ only.

c) The estimate of the coefficient of $\ln L$ agrees well with the theory, which states that it should be $\frac{1}{2}$. The estimate of the coefficient of the term involving θ agrees less well – it should be 1; however, the standard error of the estimate is large.

The intercept β_0 of the regression model corresponds to $\ln(2\pi/\sqrt{g})$ of the physical model, $\beta_0 = \ln(2\pi/\sqrt{g})$. Thus, $\hat{\beta}_0 = \ln(2\pi/\sqrt{\hat{g}})$ defines an estimate of g. Solving, we get $\hat{g} = 4\pi^2 e^{-2\hat{\beta}_0} = 4\pi^2 e^{-2(-1.62)} = 1.0\cdot 10^3$, that is, $1.0\cdot 10^3$ cm/s^2 = 10 m/s^2 (the units of the data were s and cm).

$(\beta_0 - \hat{\beta}_0)/\operatorname{se}\hat{\beta}_0$ has the t distribution with $100-3=97$ degrees of freedom, where $\operatorname{se}\hat{\beta}_0$ denotes the standard error of $\hat{\beta}_0$ (the denominator in Corollary 3.33 of Bingham and Fry). Let t denote the upper 0.025 critical value of this distribution. Then

$$
\begin{aligned}
0.95 &= P\left(-t < \frac{\beta_0 - \hat{\beta}_0}{\operatorname{se}\hat{\beta}_0} < t\right) = P\left(-t < \frac{\ln(2\pi/\sqrt{g}) - \hat{\beta}_0}{\operatorname{se}\hat{\beta}_0} < t\right)\\
&= P\big(\hat{\beta}_0 - t\operatorname{se}\hat{\beta}_0 < \ln(2\pi/\sqrt{g}) < \hat{\beta}_0 + t\operatorname{se}\hat{\beta}_0\big) = P\big(e^{\hat{\beta}_0 - t\operatorname{se}\hat{\beta}_0} < 2\pi/\sqrt{g} < e^{\hat{\beta}_0 + t\operatorname{se}\hat{\beta}_0}\big)\\
&= P\big(4\pi^2 e^{-2(\hat{\beta}_0 + t\operatorname{se}\hat{\beta}_0)} < g < 4\pi^2 e^{-2(\hat{\beta}_0 - t\operatorname{se}\hat{\beta}_0)}\big).
\end{aligned}
$$

The statistical tables give $t = 1.98$, and from the R output we have $\hat{\beta}_0 = -1.62$ and $\operatorname{se}\hat{\beta}_0 = 0.0160$ for our data. Inserting these values in the inequalities above, we get $4\pi^2 e^{-2(-1.618+1.98\cdot 0.0160)} = 0.94\cdot 10^3$ and $4\pi^2 e^{-2(-1.618-1.98\cdot 0.0160)} = 1.07\cdot 10^3$ as bounds of the confidence interval, that is, the interval is $(9.4, 10.7)$ with unit m/s^2.

Problem 2

a) The least squares estimator of $\boldsymbol{\beta}$ is in general $(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{Y}$. Since the columns of X are orthogonal, $X^{\mathrm{T}}X$ is diagonal with $\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{x}_j$ as entry (j,j), where $\boldsymbol{x}_j$ denotes the jth column of X. So $(X^{\mathrm{T}}X)^{-1}$ is diagonal with $1/(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{x}_j)$ as entry (j,j). The jth row of $(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ is then $\boldsymbol{x}_j^{\mathrm{T}}/(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{x}_j)$, and the jth entry of the estimator $\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{Y}/(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{x}_j)$.

b) The interaction vector is $(1 \quad -1 \quad -1 \quad 1)^{\mathrm{T}}$. By the above, the coefficient estimate is $(1 \quad -1 \quad -1 \quad 1)(6 \;\; 4 \;\; 10 \;\; 7)^{\mathrm{T}}/4 = (6-4-10+7)/4 = -1/4$. The estimate of the effect is $2\cdot(-1/4) = -1/2$.

Problem 3

a) In general, $\mathrm{Cov}(A\boldsymbol{Y}) = A(\mathrm{Cov}\,\boldsymbol{Y})A^{\mathrm{T}}$. With $A = (X_0^{\mathrm{T}}X_0)^{-1}X_0^{\mathrm{T}}$, we get $\mathrm{Cov}\,\hat{\boldsymbol{\beta}}_{(0)} = (X_0^{\mathrm{T}}X_0)^{-1}X_0^{\mathrm{T}}(\mathrm{Cov}\,Y)X_0(X_0^{\mathrm{T}}X_0)^{-1} = (X_0^{\mathrm{T}}X_0)^{-1}X_0^{\mathrm{T}}(\sigma^2 I)X_0(X_0^{\mathrm{T}}X_0)^{-1} = \sigma^2(X_0^{\mathrm{T}}X_0)^{-1}$.

$$\mathrm{Cov}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0) = \mathrm{Cov}(-\hat{\boldsymbol{\beta}}_0) = \mathrm{Cov}\,\hat{\boldsymbol{\beta}}_0 = \sigma^2\begin{pmatrix}(X_0^{\mathrm{T}}X_0)^{-1} & O\\ O & O\end{pmatrix},$$

where the O denote zero matrices of various dimensions, making the size of the covariance matrix $p\times p$.

Write $X = \begin{pmatrix}X_0 & X_1\end{pmatrix}$. Then

$$\frac{1}{\sigma^2}X^{\mathrm{T}}X\,\mathrm{Cov}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0) = \begin{pmatrix}X_0^{\mathrm{T}}\\ X_1^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}X_0 & X_1\end{pmatrix}\begin{pmatrix}(X_0^{\mathrm{T}}X_0)^{-1} & O\\ O & O\end{pmatrix}$$
$$= \begin{pmatrix}X_0^{\mathrm{T}}\\ X_1^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}X_0(X_0^{\mathrm{T}}X_0)^{-1} & O\end{pmatrix} = \begin{pmatrix}X_0^{\mathrm{T}}X_0(X_0^{\mathrm{T}}X_0)^{-1} & O\\ X_1^{\mathrm{T}}X_0(X_0^{\mathrm{T}}X_0)^{-1} & O\end{pmatrix} = \begin{pmatrix}I & O\\ X_1^{\mathrm{T}}X_0(X_0^{\mathrm{T}}X_0)^{-1} & O\end{pmatrix},$$

where I is an $r\times r$ identity matrix and the lower O a $(p-r)\times(p-r)$ zero matrix. The trace of the matrix is r.

b)

$$E(X\hat{\boldsymbol{\beta}}_0) = E\left(\begin{pmatrix}X_0 & X_1\end{pmatrix}\begin{pmatrix}(X_0^{\mathrm{T}}X_0)^{-1}X_0^{\mathrm{T}}\boldsymbol{Y}\\ \mathbf{0}\end{pmatrix}\right) = E(H_0\boldsymbol{Y}) = H_0E(\boldsymbol{Y}) = H_0X\boldsymbol{\beta},$$

so that $E(X(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0)) = (I-H_0)X\boldsymbol{\beta}$ (which is equal to $\begin{pmatrix}O & (I-H_0)X_1\end{pmatrix}\boldsymbol{\beta} = (I-H_0)X_1\boldsymbol{\beta}_1$, where $\boldsymbol{\beta}_1$ is $\boldsymbol{\beta}$ with the first r entries replaced by 0.)

Note that $XE(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0) = E(X(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0))$. By the trace formula,

$$EJ_0 = \mathrm{tr}\left(\frac{1}{\sigma^2}X^{\mathrm{T}}X\,\mathrm{Cov}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0)\right) + \frac{1}{\sigma^2}E(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0)^{\mathrm{T}}X^{\mathrm{T}}XE(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_0)$$
$$= r + \frac{1}{\sigma^2}\boldsymbol{\beta}^{\mathrm{T}}X^{\mathrm{T}}(I-H_0)^{\mathrm{T}}(I-H_0)X\boldsymbol{\beta} = r + \frac{1}{\sigma^2}\boldsymbol{\beta}^{\mathrm{T}}X^{T}(I-H_0)X\boldsymbol{\beta}.$$

c) Again by the trace formula, $E\,\mathrm{SSE}_0 = \mathrm{tr}((I-H_0)\,\mathrm{Cov}\,\boldsymbol{Y}) + E\boldsymbol{Y}^{\mathrm{T}}(I-H_0)E\boldsymbol{Y} = (n-r)\sigma^2 + \boldsymbol{\beta}^{\mathrm{T}}X^{\mathrm{T}}(I-H_0)X\boldsymbol{\beta}$.

Now EJ_0 can be rewritten $EJ_0 = r + (E\,\mathrm{SSE}_0 - (n-r)\sigma^2)/\sigma^2 = \frac{1}{\sigma^2}E\,\mathrm{SSE}_0 - n + 2r$.

Mallows' $C_P = \frac{1}{\widehat{\sigma^2}}\mathrm{SSE}_0 - n + 2r$, where $\widehat{\sigma^2}$ is an estimate from the "full" model, showing that C_P is an estimate of EJ_0.