MAST90104: A First Course in Statistical Learning

Solutions to Week 6 Practical and Workshop questions

1 Practical questions

1. The data set `ufc.csv` contains forest inventory observations from the University of Idaho Experimental Forest. In the experiment, scientists randomly selected a number of plots and then from each plot selected a number of trees. For each tree they measured its height and diameter (which are numeric), and also the species of tree (which is a character string). Answer the following questions:

   (a) What are the species of the three tallest trees? Of the five fattest trees? (Use the `order` command.)

   Solution:

```
> ufc = read.csv("data sets/data/ufc.csv")
> str(ufc)
'data.frame': 336 obs. of  5 variables:
$ plot    : int  2 2 3 3 3 4 4 5 5 6 ...
$ tree    : int  1 2 2 5 8 1 2 2 4 1 ...
$ species : chr  "DF" "WL" "GF" "WC" ...
$ dbh.cm  : num  39 48 52 36 38 46 25 54.9 51.8 40.9 ...
$ height.m: num  20.5 33 30 20.7 22.5 18 17 29.3 29 26 ...
> # a. names of tallest/ fattest trees
> ufc_sort_by_height = ufc[order(ufc$height.m,decreasing = TRUE),]
> head(ufc_sort_by_height)
plot tree species dbh.cm height.m
272  110    4      GF   81.2     47.0
222   88    3      WL   70.8     42.5
141   55    2      DF   99.8     42.0
169   68    1      GF   78.0     42.0
194   78    2      GF   64.4     42.0
150   59    1      GF   86.1     40.2
>
> ufc_sort_by_diameter = ufc[order(ufc$dbh.cm,decreasing = TRUE),]
> head(ufc_sort_by_diameter)
plot tree species dbh.cm height.m
112   43    4      WC  101.5     39.0
141   55    2      DF   99.8     42.0
111   43    3      WC   94.0     36.0
33    16    3      WC   89.5     35.0
150   59    1      GF   86.1     40.2
38    16   10      WC   83.0     35.0
```

   (b) What are the mean diameters by species?

   Solution

```
> tapply(ufc$dbh.cm, ufc$species, mean)
DF           GF           WC           WL
39.905263158 35.211864407 38.844604317 33.727272727
```

   (c) What are the two species that have the largest third quartile diameters?

```
> tapply(ufc$dbh.cm, ufc$species, quantile, prob = c(0.75))
DF    GF    WC    WL
50.20 44.40 49.20 43.85
```

```
> sort(tapply(ufc$dbh.cm, ufc$species, quantile, prob = c(0.75)))
WL    GF    WC    DF
43.85 44.40 49.20 50.20
```

(d) What is the identity of the tallest tree of the species that was the fattest on average?

Solution

```
> # need to identify species with largest average diameter
> (ave_diameter_by_species <- tapply(ufc$dbh.cm, ufc$species, mean))
DF           GF           WC           WL
39.905263158 35.211864407 38.844604317 33.727272727
> which.max(ave_diameter_by_species)
DF
1
> all_DF_data = ufc[ufc$species== "DF",]
> all_DF_data[all_DF_data$height.m== max(all_DF_data$height.m),]
plot tree species dbh.cm height.m   slenderness
141   55    2      DF   99.8       42 0.42084168337
```

2. The following questions use the 'sleep' dataset, which you can download from the course website. This dataset contains (among other things) data on the body weight (kg) and brain weight (g) of 62 mammals. Use the following commands to read the data (make sure the data file is in your working directory, or change to the correct path):

```
mammals <- read.csv("sleep.csv")
```

This creates a data frame, `mammals`, with components (among others) named `BodyWt` and `BrainWt`We are interested in predicting brain weight from body weight.

(a) Plot the data. Fit the model of brain weight vs. body weight using the `lm` function. Plot the diagnostics plots and comment on the plots. Is the model appropriate ?

Solution:

```
par(mfrow = c(2,2))
hist(mammals$BodyWt)
hist(mammals$BrainWt)
plot(mammals$BodyWt,mammals$BrainWt,pch = 16)
```

We see that both brain weight and body weight have extremely right-skewed distributions (Figure 1). In addition, both variables are constrained to be positive. Therefore it may be reasonable to consider transforming the data.

Fit the model

```
> model_naive = lm(BrainWt~ BodyWt, data = mammals)
> summary(model_naive)

Call:
lm(formula = BrainWt ~ BodyWt, data = mammals)

Residuals:
Min      1Q  Median      3Q     Max
-810.07  -88.52  -79.64  -13.02 2050.33

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 91.00440   43.55258    2.09   0.0409 *
BodyWt       0.96650    0.04766   20.28   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

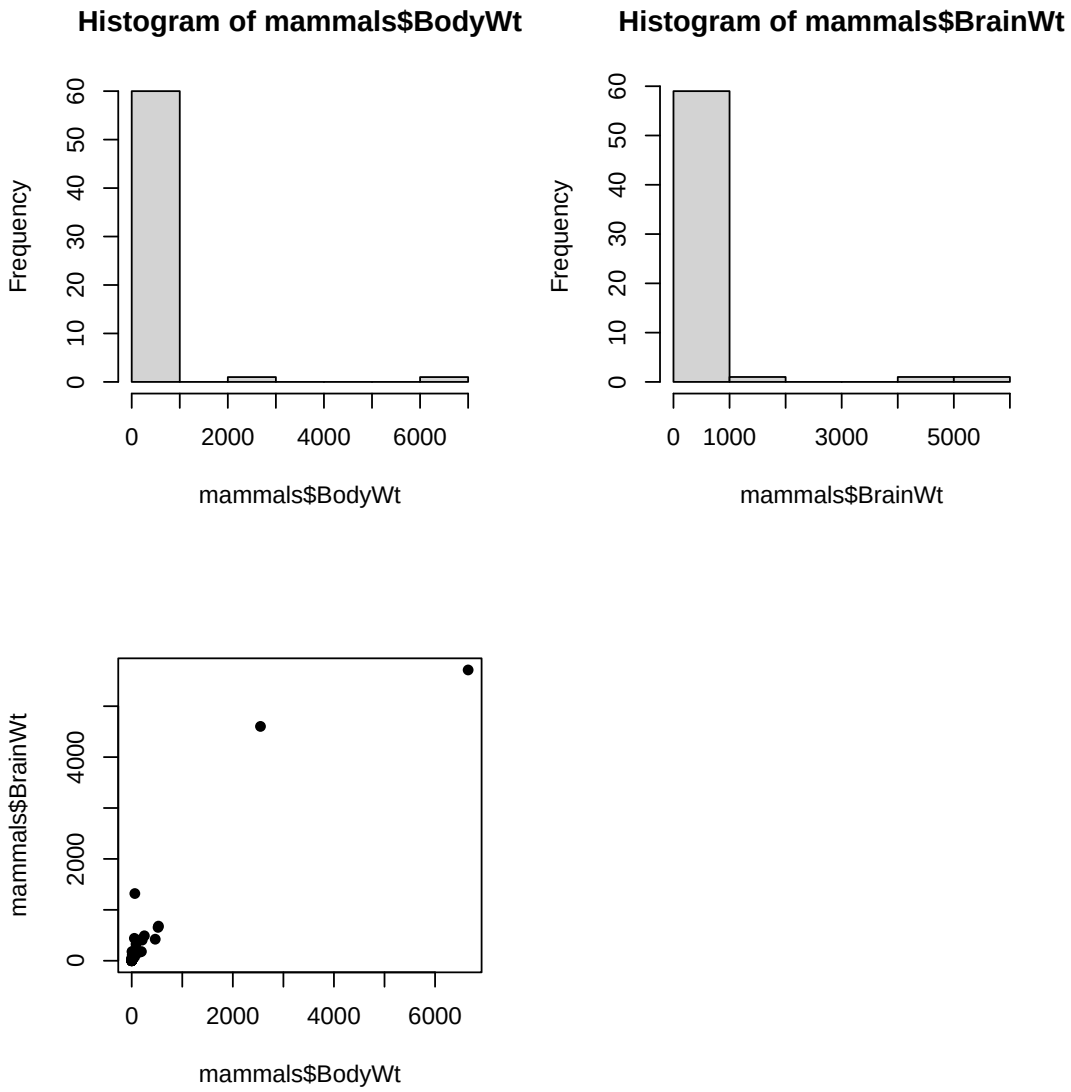

Figure 1: Plots of data in the original scale

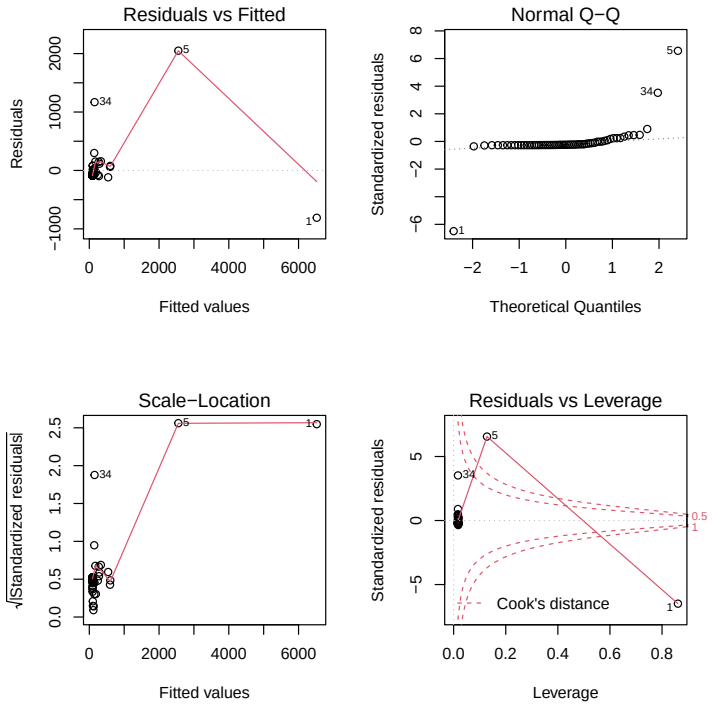

Figure 2: Diagnostic of the fist model

```
Residual standard error: 334.7 on 60 degrees of freedom
Multiple R-squared:  0.8727,Adjusted R-squared:  0.8705
F-statistic: 411.2 on 1 and 60 DF,  p-value: < 2.2e-16
```

Diagnostics plots of the first model

```
> plot(model_naive,which = 1)
> plot(model_naive,which = 2)
> plot(model_naive, which = 3)
> plot(model_naive,which = 5)
```

The diagnostics plots suggest that the naive model may not be appropriate. The plot of brain weight against body weight suggests a non-linear relationship (Figure 1). The Scale-Location plot shows some increase in the square root of the absolute values of the residuals with fitted values increasing, so a log transformation of the response variable (brain weight) may be appropriate (Figure 2).

Merely being right-skewed would not be a strong enough case to transform the predictor, although the extreme nature of the skew results in some points with extremely high leverage/Cook's distance. However, transforming the brain weight alone does not result in a linear relationship, while transforming both brain and body weight results in an obviously linear relationship (Figure 3).

(b) Apply a logarithmic transformation to both `BodyWt` and `BrainWt`.

```
mammals$BodyWt <- log(mammals$BodyWt)
mammals$BrainWt <- log(mammals$BrainWt)
```

Fit a linear model explaining (transformed) brain weight from body weight, using the `lm` command.

Display the summary of the fitted model, and then create a scatter plot of the data and superimpose the fitted regression line on it. Does it look like a reasonable fit?

Use diagnostic plots to assess if the model assumptions are satisfied.

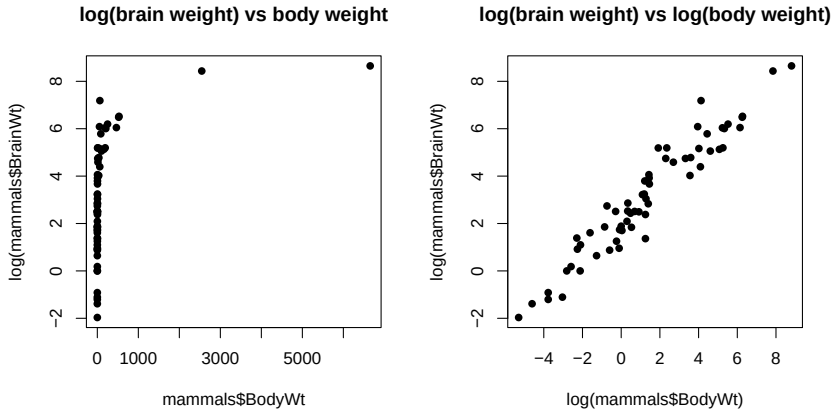

Figure 3:

Solution:

```
model <- lm(BrainWt ~ BodyWt, data = mammals)
summary(model)

##
## Call:
## lm(formula = BrainWt ~ BodyWt, data = mammals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71550 -0.49228 -0.06162  0.43597  1.94829
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.13479    0.09604   22.23   <2e-16 ***
## BodyWt       0.75169    0.02846   26.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6943 on 60 degrees of freedom
## Multiple R-squared:  0.9208,Adjusted R-squared:  0.9195
## F-statistic: 697.4 on 1 and 60 DF,  p-value: < 2.2e-16

plot(mammals$BodyWt, mammals$BrainWt)
abline(model, col="red")
```

The fit is good.

```
plot(model, which=1)
plot(model, which=2)
plot(model, which=3)
plot(model, which=5)
```

The residuals show a slight trend toward negativity as the fitted values increase, but not enough to be a problem. The Q-Q plot looks reasonably linear. The standardised residuals get smaller on both sides of the Scale-Location plot. This is not ideal, but the lack of a definite trend makes it difficult to correct. The residuals vs leverage plot is fine, there is no point with unusually large leverage and residuals.

(c) Find a 95% confidence interval for a mammal weighing 50 kg.

Solution:

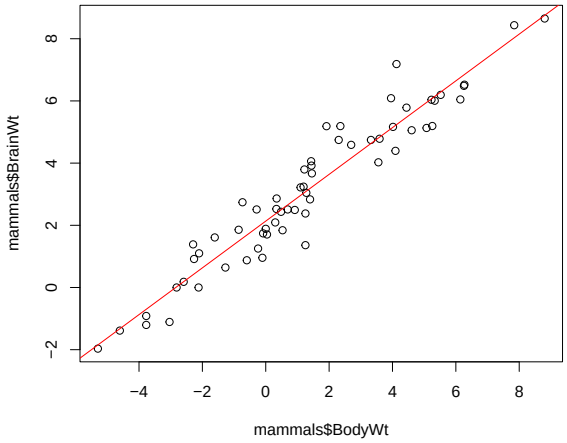

Figure 4: Transformed data and the fitted regression line

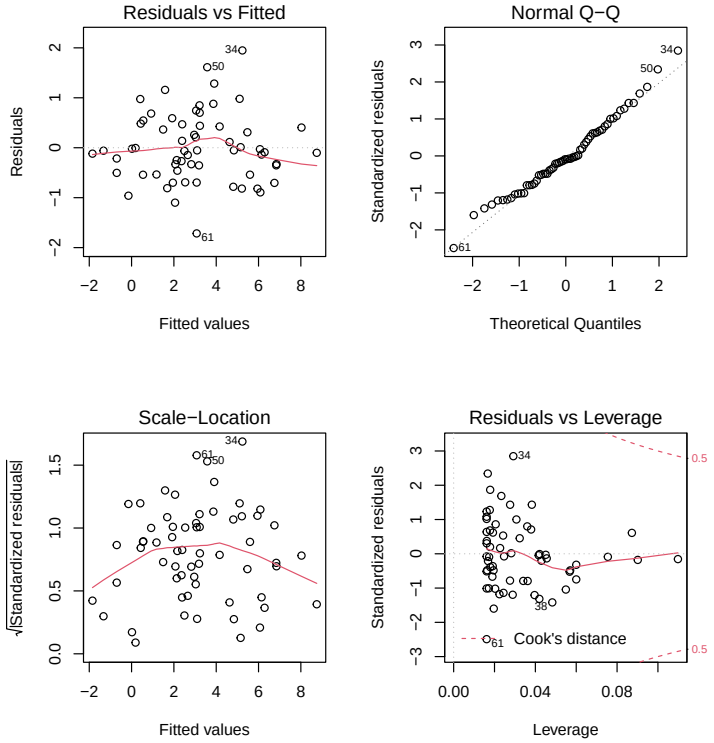

Figure 5: Diagnostic of the second model

```
predict(model, data.frame(BodyWt = log(50)), interval = "confidence", level = 0.95)
```

```
##        fit      lwr      upr
## 1 5.075401 4.846066 5.304736
```

(d) Find a 95% prediction interval for a mammal weighing 50 kg.
Solution:

```
predict(model, data.frame(BodyWt = log(50)), interval = "prediction", level = 0.95)
```

```
##        fit      lwr      upr
## 1 5.075401 3.667797 6.483006
```

(e) Test the following hypotheses, using the `anova` function.

i. $H_0 : \boldsymbol{\beta} = 0$
ii. $H_0 : \beta_1 = 0$
iii. $H_0 : \beta_0 = 0$
iv. $H_0 : \boldsymbol{\beta} = (2, 1)$

Solution:

```
null <- lm(BrainWt ~ 0, data = mammals)
anova(null, model)
```

```
## Analysis of Variance Table
##
## Model 1: BrainWt ~ 0
## Model 2: BrainWt ~ BodyWt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     62 976.48
## 2     60  28.92  2    947.56 982.85 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
null <- lm(BrainWt ~ 1, data = mammals)
anova(null, model)
```

```
## Analysis of Variance Table
##
## Model 1: BrainWt ~ 1
## Model 2: BrainWt ~ BodyWt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     61 365.11
## 2     60  28.92  1    336.19 697.42 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
null <- lm(BrainWt ~ 0 + BodyWt, data = mammals)
anova(null, model)
```

```
## Analysis of Variance Table
##
## Model 1: BrainWt ~ 0 + BodyWt
## Model 2: BrainWt ~ BodyWt
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     61 267.079
## 2     60  28.923  1    238.16 494.05 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(car)
linearHypothesis(model,diag(2),c(2,1))

## Linear hypothesis test
##
## Hypothesis:
## (Intercept) = 2
## BodyWt = 1
##
## Model 1: restricted model
## Model 2: BrainWt ~ BodyWt
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     62 68.024
## 2     60 28.923  2    39.101 40.558 7.199e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We reject all null hypotheses.

For the last part, we can also solve that using `anova` function by defining the null model with an `offset`, which gives the same result as using the `linearHypothesis` function.

```
> X = cbind(1,mammals$BodyWt)
> y = mammals$BrainWt
> h0 <- X %*% c(2,1)
> # The following model is equivalent to y = h0 + error
> basemodel <- lm(BrainWt ~ 0, data=mammals, offset=h0)
> anova(basemodel, model)
Analysis of Variance Table

Model 1: BrainWt ~ 0
Model 2: BrainWt ~ BodyWt
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     62 68.024
2     60 28.923  2    39.101 40.558 7.199e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The agurment `offset` specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector matching those of the response. The offset is like an extra column in your design matrix with coefficients fixed to 1.

(f) Write down the final model for the untransformed data. The final model is

$$\text{brain weight} = 8.46 \cdot (\text{body weight})^{0.75} \cdot \varepsilon.$$

2 Workshop questions

1. Suppose X is $n \times p, p \leq n$ of full rank and C is $r \times p, r \leq p$ also of full rank.

   (a) Show that X^TX is positive definite (hint: use the definition).

   Solution: Suppose β is a $p \times 1$ vector. Then $\beta^T(X^TX)\beta = (X\beta)^T(X\beta)$ wich is a sum of squares of the $n \times 1$ vector $X\beta$. Hence, it is non-negative showing that X^TX is positive semidefinite. To show it is positive definite, note that $\beta^T(X^TX)\beta = 0$, implies $X\beta = 0$, wich in turn implies that $X^TX\beta = 0$. But X^TX is rank p so its columns are linearly independent, and hence $\beta = 0$.

   (b) Show that $C(X^TX)^{-1}C^T$ is positive definite (hint: why does $(X^TX)^{-1}$ have a matrix square root?).

   Solution: Since (X^TX) is positive definite and symmetric, Slide 14 of Lecture 3 shows that it has a positive definite square root. The square root is of the form $P\Lambda^{1/2}P^T$ where P is orthonormal and the diagonal matrix Λ has the strictly positive eigenvalues(Theorem 2.7) of X^TX on the diagonal. Therefore $(X^TX)^{-1}$ has the positive definite and invertible square root $Q = P\Lambda^{-1/2}P^T$. Hence, since Q is symmetric, $C(X^TX)^{-1}C^T = CQQ^TC^T = Z^TZ$ where $Z = Q^TC^T$. Since the rank of a matrix is unaltered by premultiplication by a non-singular matrix, $r(Z) = r(Q^TC^T) = r(C^T) = r(C) = r$, so Z is full rank. By the previous part of this question applied to Z instead of X, $C(X^TX)^{-1}C^T$ is positive definite.

   (c) Show that $C(X^TX)^{-1}C^T$ is invertible.

   Solution: From facts about rank, $r(C(X^TX)^{-1}C^T) = r(Z^TZ) = r(Z) = r$ from the previous part of this question.

   (d) Show that $[C(X^TX)^{-1}C^T]^{-1}$ is positive definite.

   Solution: Since $(X^TX)^{-1}$ is symmetricc, $[C(X^TX)^{-1}C^T]^T = C(X^TX)^{-1}C^T$ is symmetric. The matrix $C(X^TX)^{-1}C^T$ is also positive definite from this question, so it has strictly positive eigenvalues. Hence, the matrix can be expressed as $R\Theta R^T$ where R is an orthogonal matrix of eigenvectors of $C(X^TX)^{-1}C^T$ and, by Theorem 2.7, Θ is a diagonal matrix with strictly positive eigenvalues on the diagonal. Then $R\Theta^{-1}R^T$ is the inverse of $C(X^TX)^{-1}C^T$ and, for any β, $\beta^TR\Theta^{-1}R^T\beta = (R^T\beta)^T\Theta R^T\beta$ is a sum of squares, some of whose entries must be positive because they are elements of the diagonal of Θ multiplied by elements of $(R^T\beta)^TR^T\beta = \beta^T\beta > 0$.

2. In this question we consider the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^*$. The test statistic for this hypothesis is

$$\frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^TX^TX(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)/p}{SS_{Res}/(n-p)}.$$

   (a) Show that

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^TX^TX(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = (\mathbf{y} - X\boldsymbol{\beta}^*)^T(\mathbf{y} - X\boldsymbol{\beta}^*) - (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}).$$

   That is, it is the SS_{Res} for the null model minus the SS_{Res} for the full model.
   Also show that, in general,

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^TX^TX(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \neq \mathbf{y}^TX(X^TX)^{-1}X^T\mathbf{y} - \boldsymbol{\beta}^{*T}X^TX\boldsymbol{\beta}^*.$$

   That is, in this case we can not write it as the SS_{Reg} for the full model minus the SS_{Reg} for the model under H_0.

   Solution: Letting $H = X(X^T)^{-1}X^T$, equation (1) on p.10 of Module 4 gives

$$\begin{aligned}
&(\mathbf{y} - X\boldsymbol{\beta}^*)^T(\mathbf{y} - X\boldsymbol{\beta}^*) \\
&= (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}) + (X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta}^*)^T(X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta}^*) \\
&= (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^T(\mathbf{y} - X\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^TX^TX(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*).
\end{aligned}$$

and rearranging gives the required equality. Note also that $X\widehat{\boldsymbol{\beta}} = H\mathbf{y}$ and H is idempotent so

$$\begin{aligned}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &= (X\widehat{\boldsymbol{\beta}})^T(X\widehat{\boldsymbol{\beta}}) + (X\boldsymbol{\beta}^*)^T(X\boldsymbol{\beta}^*) - 2(X\widehat{\boldsymbol{\beta}})^T X\boldsymbol{\beta}^* \\ &= (H\mathbf{y})^T H\mathbf{y} + \boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^* - 2(X\widehat{\boldsymbol{\beta}})^T X\boldsymbol{\beta}^* \\ &= \mathbf{y}^T H\mathbf{y} + \boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^* - 2(X\widehat{\boldsymbol{\beta}})^T X\boldsymbol{\beta}^* \\ &= y^T H\mathbf{y} - \boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^*\end{aligned}$$

only if $-2(X\widehat{\boldsymbol{\beta}})^T X\boldsymbol{\beta}^* = -2\boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^*$ which, in general, only occurs if $\boldsymbol{\beta}^* = 0$.

(b) Show directly that $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$ and SS_{Res} are independent, that is without using our existing results that $\widehat{\boldsymbol{\beta}}$ and SS_{Res} are independent.

Hint: set $\mathbf{q} = \mathbf{y} - X\boldsymbol{\beta}^*$ then

i. Show that $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \mathbf{q}^T X(X^T X)^{-1} X^T \mathbf{q}$.

ii. Show that $SS_{Res} = \mathbf{q}^T[I - X(X^T X)^{-1} X^T]\mathbf{q}$ and hence that these two quadratic forms are independent.

Solution: We express both quantities as quadratic forms in $\mathbf{q}$. Arguing as in the last part,

$$\begin{aligned}\mathbf{q}^T X(X^T X)^{-1} X^T \mathbf{q} &= \mathbf{y}^T H\mathbf{y} - 2(X\boldsymbol{\beta}^*)^T H\mathbf{y} + (X\boldsymbol{\beta}^*)^T HX\boldsymbol{\beta}^* \\ &= (X\widehat{\boldsymbol{\beta}})^T X\widehat{\boldsymbol{\beta}} - 2(X\boldsymbol{\beta}^*)^T H\mathbf{y} + (\boldsymbol{\beta}^*)^T X^T X\boldsymbol{\beta}^* \\ &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*).\end{aligned}$$

For the SS_{Res} note first that

$$\begin{aligned}\boldsymbol{\beta}^{*T} X^T[I - H]X\boldsymbol{\beta}^* &= \boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^* - \boldsymbol{\beta}^{*T} X^T X(X^T X)^{-1} X^T X\boldsymbol{\beta}^* \\ &= \boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^* - \boldsymbol{\beta}^{*T} X^T X\boldsymbol{\beta}^* \\ &= \mathbf{0}.\end{aligned}$$

Similarly $\mathbf{y}^T[I - H]^T X\boldsymbol{\beta}^* = \mathbf{0}$ and $\boldsymbol{\beta}^{*T} X^T[I - H]\mathbf{y} = \mathbf{0}$, so

$$\begin{aligned}&\mathbf{q}^T[I - H]\mathbf{q} \\ &\quad= \mathbf{y}^T[I - H]\mathbf{y} - \boldsymbol{\beta}^{*T} X^T[I - H]\mathbf{y} \\ &\qquad -\mathbf{y}^T[I - H]X\boldsymbol{\beta}^* + \boldsymbol{\beta}^{*T} X^T[I - H]X\boldsymbol{\beta}^* \\ &\quad= \mathbf{y}^T[I - H]\mathbf{y} \\ &\quad= SS_{Res}.\end{aligned}$$

Finally, we know that $\text{var}\,\mathbf{q} = \sigma^2 I$, so, using our theorem for the independence of quandratic forms

$$\begin{aligned}AVB &= X(X^T X)^{-1} X^T \sigma^2 I[I - H] \\ &= \sigma^2\left(X(X^T X)^{-1} X^T - X(X^T X)^{-1} X^T X(X^T X)^{-1} X^T\right) \\ &= \sigma^2\left(X(X^T X)^{-1} X^T - X(X^T X)^{-1} X^T\right) \\ &= \mathbf{0}\end{aligned}$$

as required.

3. Recall the joint confidence region for the parameters of a full rank linear model:

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T X^T X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq ps^2 f_\alpha.$$

Use this to derive a test for the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^*$. Show that this test is equivalent to the test for $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^*$ obtained using the general linear hypothesis.

Solution: We do not reject H_0 if and only if $\boldsymbol{\beta}^*$ lies in the joint confidence region, i.e., if and only if

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T \leq ps^2 f_\alpha.$$

But the general linear hypothesis test does not reject H_0 if and only if

$$\begin{aligned}
\frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)/p}{SS_{Res}/(n-p)} &\leq f_\alpha \\
(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &\leq p\frac{SS_{Res}}{n-p} f_\alpha \\
(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T X^T X (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &\leq p s^2 f_\alpha.
\end{aligned}$$

Therefore the two tests are equivalent.