Statistics 2 Unit 6 Team 8

Nikolaos Kornilakis Rodrigo Viale Jakub Trnan Aleksandra Daneva
Luis Diego Pena Monge

2024-05-01

Exercise 119

First of all, notice that the $\hat{\beta} = (X'X)^{+}X'y$ provides a solution to the least squares problem as $X(X'X)^{+}X'$ is the orthogonal projector into the column space of X, and the above can be rewritten as $X\hat{\beta} = X(X'X)^{+}X'y$. Note that this column space is not spanned by all columns as otherwise X would have full rank.

Now, to see that this is the minimum norm estimator, let β be another solution to the minimization of the squared norm, then:

$$\beta = \hat{\beta} + (\beta - \hat{\beta})$$

Notice that:

$$\begin{aligned} X = U_r D_r V_r' &\implies X' = V_r D_r U_r' \\ &\implies (X'X)^{+}X' = V_r D_r^{-2} V_r' V_r D_r U_r = V_r D_r^{-1} U_r' \\ &\implies (X'X)^{+}X'X = I \end{aligned}$$

Now:

$$\hat{\beta}'(\beta - \hat{\beta}) = \hat{\beta}'(X'X)^{+}X'X(\beta - \hat{\beta}) = \hat{\beta}'0 = 0$$

This means that $\hat{\beta}$ and $(\beta - \hat{\beta})$ are orthogonal. In other words, due to orthogonality:

$$||\beta||^2 = ||\hat{\beta}||^2 + ||\beta - \hat{\beta}||^2$$

We therefore minimize the norm of β by minimizing the right hand side, which happens when $\beta = \hat{\beta}$. Therefore, $\hat{\beta}$ is the minimum norm estimator.

Exercise 120

Recall if X has full rank, then $X'X$ has full rank and $\hat{\beta} = (X'X)^{-1}X'y$. Let $X = QR$ be the QR decomposition of X, where Q is orthogonal and R upper triangular. Then we can calculate X''s QR decomposition as well as $X' = R'Q'$.

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (R'Q'QR)^{-1}R'Q'y \\
&= (R'R)^{-1}R'Q'y \\
&= R^{-1}(R')^{-1}R'Q'y \\
&= R^{-1}Q'y
\end{aligned}
$$

The above is completely equivalent to having:

$$R\hat{\beta} = Q'y$$

Notice R is known and upper triangular, and $Q'y$ is also known. This means that the system can be easily and efficiently solved for $\hat{\beta}$ using backward solve.

Exercise 121

Let $\tilde{\beta} = AY$ be an unbiased linear estimator of β and $\hat{\beta} = (X'X)^{-1}X'y$. Then notice that:

$$cov(\tilde{\beta}) = cov(Ay) = Acov(y)A' = \sigma^2 AA'$$

$$
\begin{aligned}
cov(\hat{\beta}) &= cov((X'X)^{-1}X'y) \\
&= (X'X)^{-1}X'cov(y)((X'X)^{-1}X')' \\
&= (X'X)^{-1}X'cov(y)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

This means then that:

$$cov(\tilde{\beta}) - cov(\hat{\beta}) = \sigma^2(AA' - (X'X)^{-1})$$

Since σ^2 is just a positive number we will ignore it and it won't change the results from now on. Let $z \in \mathbb{R}^p \backslash \{0\}$. Recall that, as seen in class, for any unbiased linear estimator we must have that $AX = I$. Then:

$$
\begin{aligned}
z'AA'z - z'(X'X)^{-1}z &= z'AA'z - Z'AX(X'X)^{-1}X'A'z \\
&= z'AA'z - z'AP_XA'z \\
&= z'A(1 - P_X)A'z
\end{aligned}
$$

Now from Lemma 8.1 from Abramovich and Ritov, $1 - P_X$ is non-negative definite, and therefore we have that:

$$z'AA'z - z'(X'X)^{-1}z \geq 0 \qquad \forall z \in \mathbb{R}^p$$

Which means it is non-negative definite. Let's check equality sufficiency and necessity:

($\Rightarrow$) Suppose $\tilde{\beta} = \hat{\beta}$, then the equality is trivial.

($\Leftarrow$) Suppose $cov(\tilde{\beta}) = cov(\hat{\beta})$. Then $AA' = (X'X)^{-1}$. Then we have that:

$$\tilde{\beta} = Ay = AA'X'y = (X'X)^{-1}X'y = \hat{\beta}$$

Exercise 122

Let $y \sim N(X\beta, \sigma^2 I_n)$. Then:

$$f(y|\theta) = (2\pi)^{-n/2} det(\sigma^2 I_n)^{-1/2} \exp\left(\frac{-1}{2}(y - X\beta)'(\sigma^2 I_n)^{-1}(y - X\beta)\right)$$

Recall $(\sigma^2 I_n)^{-1} = \frac{1}{\sigma^2} I_n$. Then:

$$\begin{aligned}(y - X\beta)(\sigma^2 I_n)^{-1}(y - X\beta) &= \frac{1}{\sigma^2}(y - X\beta)'(y - X\beta) \\ &= \frac{1}{\sigma^2}\left(y'y - 2\beta' X' y + (X\beta)'(X\beta)\right)\end{aligned}$$

Then:

$$f(y|\theta) = (2\pi)^{-n/2} det(\sigma^2 I_n)^{-1/2} \exp\left(\frac{-y'y}{2\sigma^2} + \frac{\beta'}{\sigma^2} X' y - \frac{(X\beta)'(X\beta)}{2\sigma^2}\right)$$

Now notice that:

$$\frac{\beta'}{\sigma^2} X' y = \sum_{j=1}^{p} \frac{\beta_j}{\sigma^2}\left(\sum_{i=1}^{n} X_{ij} y_i\right)$$

Then we can define:

$$c_j(\theta) = \begin{cases} \frac{\beta_j}{\sigma^2} & j \in \{1, ..., p\} \\ \frac{1}{\sigma^2} & j = p+1 \end{cases} \qquad T_j(y) = \begin{cases} (X'y)_j & j \in \{1, ..., p\} \\ \frac{-1}{2} y'y & j = p+1 \end{cases}$$

Then we have that:

$$\begin{aligned} f(y|\theta) &= (2\pi)^{-n/2} det(\sigma^2 I_n)^{-1/2} \exp\left(\frac{-y'y}{2\sigma^2} + \frac{\beta'}{\sigma^2} X' y - \frac{(X\beta)'(X\beta)}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(\frac{-y'y}{2\sigma^2} + \frac{\beta'}{\sigma^2} X' y - \frac{(X\beta)'(X\beta)}{2\sigma^2}\right) \\ &= \exp\left(\sum_{j=1}^{p+1} c_j(\theta) T_j(y) + \frac{(X\beta)' X\beta}{2\sigma^2} + n\ln\left(\frac{1}{\sqrt{2\pi}}\right) - n\ln(\sigma)\right)\end{aligned}$$

Define:

$$d(\theta) = \frac{(X\beta)' X\beta}{2\sigma^2} - n\ln(\sigma) \qquad s(y) = n\ln\left(\frac{1}{\sqrt{2\pi}}\right)$$

Then:

$$f(y|\theta) = \exp\left(\sum_{j=1}^{p+1} c_j(\theta) T_j(y) + d(\theta) + s(y)\right)$$

Which proves that the distribution of y belongs to a $(p+1)$-parameter exponential family with the desired natural parameters. To check that $(X'y, y'y)$ is sufficient, we first make a slight adjustment to the T_j defined earlier:

$$c_j(\theta) = \begin{cases} \frac{\beta_j}{\sigma^2} & j \in \{1, ..., p\} \\ \frac{-1}{2\sigma^2} & j = p+1 \end{cases} \qquad T_j(y) = \begin{cases} (X'y)_j & j \in \{1, ..., p\} \\ y'y & j = p+1 \end{cases}$$

Notice that all else remains the same as above, with only a slight change to the natural parameters. Then:

$$\begin{aligned} f(y|\theta) &= \exp\left(\sum_{j=1}^{p+1} c_j(\theta)T_j(y) + d(\theta)\right) \cdot \exp\left(s(y)\right) \\ &= g(y, \theta) \cdot h(y) \end{aligned}$$

By factorization theorem, T_j is a sufficient statistic for θ, and recall $T_j(y) = (X'y, y'y)$, $\theta = (\beta, \sigma^2)$.

Exercise 123

Let $y \sim N(X\beta, \sigma^2 I_n)$. Recall $\hat{\beta} = (X'X)^{-1}X'y$. Then $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y$ is normal, as it is a linear combination of a normal. Now:

$$E[\hat{y}] = E[X\hat{\beta}] = XE[\hat{\beta}] = X\beta$$

$$\begin{aligned} cov(\hat{y}) &= cov(X(X'X)^{-1}X'y) \\ &= X(X'X)^{-1}X'cov(y)X(X'X)^{-1}X' \\ &= \sigma^2 X(X'X)^{-1}X'X(X'X)^{-1}X'y \\ &= \sigma^2 X(X'X)^{-1}X' \\ &= \sigma^2 P_X \end{aligned}$$

Then $\hat{y} \sim N(X\beta, \sigma^2 P_X)$. Noe, $\hat{e} = y - \hat{y}$ is also normal again since it's a linear transformation of normals. In this case we have that:

$$E[\hat{e}] = E[y - \hat{y}] = E[y] - E[\hat{y}] = X\beta - X\beta = 0$$

$$\begin{aligned} cov(\hat{e}) &= cov(y - X(X'X)^{-1}X'y) \\ &= cov((I - X(X'X)^{-1}X')y) \\ &= cov((I - P_X)y) \\ &= cov(Q_X y) \\ &= Q_X cov(y) Q_X' \\ &\text{Since Q is symmetric and idempotent} \\ &= \sigma^2 Q_x \end{aligned}$$

Then $\hat{e} \sim N(0, \sigma^2 Q_x)$. To check independence, since both are normal, it's enough to check the covariance between them:

$$\begin{aligned}
cov(\hat{y}, \hat{e}) &= cov(P_X y, Q_x y) \\
&= P_X Q_X cov(y, y)(P_X Q_X)' \\
&= P_X (I - P_X) cov(y, y)(P_X (I - P_X))' \\
&= (P_X - P_X^2) cov(y, y)(P_X (I - P_X))' \\
&= 0 \cdot cov(y, y)(P_X (I - P_X))' \\
&= 0
\end{aligned}$$

Therefore $\hat{y}, \hat{e}$ are independent.

Exercise 124

Let's start defining our null and alternative hypotheses:

$$H_0 : \beta_2 = 0 \qquad H_A : \beta_2 \neq 0$$

Under the alternative hypothesis, we have simply a full model, and as seen in previous exercises, the likelihood function is given by:

$$L_A = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left(\frac{-||y - X\beta||^2}{2\sigma^2} \right)$$

Meanwhile, under the null hypothesis the likelihood function is:

$$L_0 = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left(\frac{-||y - X_1\beta||^2}{2\sigma^2} \right)$$

As we know from class, for a fixed β, the log-likelihood (and therefore the likelihood) is maximized when $\sigma^2 = \frac{||y - X\beta||^2}{n}$. Replacing this above we have that:

$$L_0 = \left(\frac{1}{\sigma_0\sqrt{2\pi}} \right)^n \exp\left(\frac{-n}{2} \right) L_A = \left(\frac{1}{\sigma_A\sqrt{2\pi}} \right)^n \exp\left(\frac{-n}{2} \right)$$

Then we can define out likelihood ratio as:

$$\Lambda = \frac{L_0}{L_A} = \frac{\left(\frac{1}{\sigma_0} \right)^n}{\left(\frac{1}{\sigma_A} \right)^n} = \left(\frac{\sigma_A^2}{\sigma_0^2} \right)^{n/2} = \left(\frac{RSS_A}{RSS_0} \right)^{n/2}$$

We reject the null if Λ is small, or equivalently if $\frac{RSS_0}{RSS_A}$ is large. Following the logic from Abramovich and Ritov:

$$
\begin{aligned}
\frac{RSS_0}{RSS_A} &= \frac{RSS_A + RSS_0 - RSS_A}{RSS_A} \\
&= 1 + \frac{RSS_0 - RSS_A}{RSS_A} \\
&= 1 + \frac{r}{n-p} \frac{\left[\frac{RSS_0 - RSS_A}{r}\right]}{\left[\frac{RSS_A}{n-p}\right]}
\end{aligned}
$$

Now, notice that $\frac{RSS_A}{\sigma^2} \sim \chi^2_{n-p}$ and $\frac{RSS_0 - RSS_A}{r} \sim \chi^2_r$, and therefore:

$$
T = \frac{\left[\frac{RSS_0 - RSS_A}{r}\right]}{\left[\frac{RSS_A}{n-p}\right]} \sim F_{r,n-p}
$$

Then given a level α, we reject whenever $T > F^{-1}_{r,n-p}(\alpha)$.

Now let's verify our results using R. We start calculating our theoretical F statistic:

```
german.data <- read.table("~/Documents/QFin/Q3/Statistics 2/Assignments/Data/german.data.txt")

german.data <- german.data[,c("V2", "V5", "V13")]

colnames(german.data) <- c("Duration","Amount","Age")

full_model <- lm(Amount ~ Duration + Age, data = german.data)
red_model <- lm(Amount ~ Duration, data = german.data)

RSS_0 <- sum(red_model$residuals^2)
RSS_A <- sum(full_model$residuals^2)

r <- 1 # We want to test the significance of 1 predictor
p <- 2 # Number of predictors in full model
n <- nrow(german.data) # Number of rows of X

F_stat <- ((RSS_0 - RSS_A)/r) / (RSS_A/(n-p))
F_stat
```

```
## [1] 5.040277
```

```
1 - pf(F_stat, df1 = r, df2 = n-p)
```

```
## [1] 0.02498285
```

```
anova(red_model,
      full_model)
```

```
## Analysis of Variance Table
##
## Model 1: Amount ~ Duration
## Model 2: Amount ~ Duration + Age
##   Res.Df        RSS Df Sum of Sq      F  Pr(>F)
## 1    998 4850706433
## 2    997 4826331637  1  24374796 5.0352 0.02506 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exercise 125

Let $y \sim N(X\beta, \sigma^2 I_n)$, where $X = [X_1, X_2]$ and $\beta = [\beta_1', \beta_2']'$. Under H_0, assume $\beta_2 = 0$.

Notice that we can rewrite our model in the following way:

$$\begin{aligned} y \sim N(X\beta, \sigma^2 I_n) &\equiv y = X\beta + \varepsilon_i \\ &\equiv y = X_1\beta_1 + X_2\beta_2 + \varepsilon_i \\ &\equiv y = X_1\beta_1 + \varepsilon_i \\ &\equiv y \sim N(X_1\beta_1, \sigma^2 I_n) \end{aligned}$$

This is completely equivalent to having a linear model but only on the predictors for which we don't assume the coefficients to be 0 under the null. That means that the restricted MLE for β is going to be given by $[\hat{\beta_1}', 0]'$, where:

$$\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' y$$

Exercise 126

Before we start the proof, we write an intermediate calculation that will be central to it. Notice that:

$$\begin{aligned} (\hat{y} - \bar{y}1)'(X\beta - \overline{X\beta}) &= (X(X'X)^{-1}X'y - X(X'X)^{-1}X'\bar{y}1)'(X\beta - \overline{X\beta}1) \\ &= X(X'X)^{-1}X'(y - \bar{y}1)'(X\beta - \overline{X\beta}1) \\ &= X(X'X)^{-1}X'(X\beta - \overline{X\beta}1)'(y - \bar{y}1) \\ &= (X(X'X)^{-1}X'X\beta - \overline{X(X'X)^{-1}X'X\beta}1)'(y - \bar{y}1) \\ &= (X\beta - \overline{X\beta}1)'(y - \bar{y}1) \\ &= (y - \bar{y}1)'(X\beta - \overline{X\beta}1) \end{aligned}$$

Using this, we have the following:

$$\begin{aligned} corr(y, X\beta) &= \frac{(y - \bar{y}1)'(X\beta - \overline{X\beta}1)}{\sqrt{var(y)var(X\beta)}} \\ &= \frac{(\hat{y} - \bar{\hat{y}}1)'(X\beta - \overline{X\beta}1)}{\sqrt{var(y)var(X\beta)}} \\ &= \sqrt{\frac{var(\hat{y})}{var(y)}} \frac{(\hat{y} - \bar{\hat{y}}1)'(X\beta - \overline{X\beta}1)}{\sqrt{var(\hat{y})var(X\beta)}} \\ &= \sqrt{\frac{var(\hat{y})}{var(y)}} corr(\hat{y}, X\beta) \\ &= \sqrt{R^2} \cdot corr(\hat{y}, X\beta) \\ &\text{Since } corr(\hat{y}, X\beta) \in [-1, 1] \\ &\leq \sqrt{R^2} \end{aligned}$$

Now, clearly if $\beta = \hat{\beta}$ we attain equality as $corr(\hat{y}, X\beta)$ is exactly 1. Therefore $\hat{y} = X\hat{\beta}$ maximizes the desired sample correlation.

Exercise 127

Consider the linear regression model $\mathbb{E}(y) = X\beta$, $\text{cov}(y) = \sigma^2 I_n$. Let L be an $r \times p$ matrix of rank $r \leq p$. Show that $L\hat{\beta}$ is a BLUE for $L\beta$.

We need to show that $L\hat{\beta}$ serves as the Best Linear Unbiased Estimator (BLUE) for $L\beta$. In other words, we need to show that $L\hat{\beta}$ is unbiased and it has minimum variance among all other unbiased estimators. First, note that since we have a estimator that is result of matrix multiplication, we can conclude that this estimator will be linear.

1) Unbiasedness

We need to show that the $L\hat{\beta}$ is unbiased, that is $E(L\hat{\beta}) = L\beta$. Note that the least squares estimator is defined as $\hat{\beta} = (X'X)^{-1}X'y$.

$$\begin{aligned}
\mathbb{E}[L\hat{\beta}] &= \mathbb{E}[L(X'X)^{-1}X'y] \\
&= L(X'X)^{-1}X'\mathbb{E}[y] \qquad (\mathbb{E}[y] = X\beta) \\
&= L(X'X)^{-1}X'X\beta \qquad (\text{note the identity}) \\
&= L\beta
\end{aligned}$$

Thus, we have shown that $L\hat{\beta}$ is unbiased.

Here, $L\epsilon$ is a linear combination of the errors, which have a mean of zero ($\mathbb{E}[\epsilon] = 0$) and are uncorrelated with X, so we have $\mathbb{E}[L\epsilon] = 0$. Thus, $L\hat{\beta}$, being a linear combination of y and $L\epsilon$, is also a linear function of y.

2) Variance Optimality

Finally, we show that $L\hat{\beta}$ is BLUE by showing its optimality in terms of variance. We can re-write variance of $L\hat{\beta}$ as:

$$\begin{aligned}
\text{Var}[L\hat{\beta}] &= \text{Var}\left[L(X'X)^{-1}X'y\right] \\
&= L(X'X)^{-1}X' \quad \text{Var}\,[y] \quad X(X'X)^{-1}L' \qquad (\text{Var}[y] = \sigma^2 I_n) \\
&= L(X'X)^{-1}X' \quad \sigma^2 I_n \quad X(X'X)^{-1}L' \\
&= \sigma^2 I_n \quad L(X'X)^{-1}X'\, X(X'X)^{-1}L' \qquad (\text{identity}) \\
&= \sigma^2 L(X'X)^{-1}L'.
\end{aligned}$$

Let $a'y$ be any unbiased linear estimator of $L\beta$, since we know that any linear estimator can be written in that form for some vector a. Now we calculate its variance,

$$\begin{aligned}
\text{Var}[a'y] &= a'\text{Var}[y]a \\
&= \sigma^2 a' I_n a \\
&= \sigma^2 a' a
\end{aligned}$$

Now, if $L\hat{\beta}$ has the smallest variance among all linear unbiased estimators of $L\beta$, then

$$\text{Var}[L\hat{\beta}] \leq \text{Var}[a'y]$$

Since we assume $a'y$ is unbiased and its given that $\mathbb{E}(y) = X\beta$,

$$\mathbb{E}[L\hat{\beta}] = \mathbb{E}[a'y] = a'\mathbb{E}[y] = a'X\beta = L\beta$$

keeping in mind what we have shown in part 1) of the exercise ($E[L\hat{\beta}] = L\beta$). Most importantly, from this relationship we can derive that

$$L = a'X.$$

Finally, to show that $\text{Var}[L\hat{\beta}] \leq \text{Var}[a'y]$ holds, we have

$$\begin{aligned}
\text{Var}[a'y] - \text{Var}[L\hat{\beta}] &= \sigma^2 a'a - \sigma^2 L(X'X)^{-1}L' \\
&= \sigma^2 \left(a'a - L(X'X)^{-1}L'\right) && \text{(substitute for } L = a'X\text{)} \\
&= \sigma^2 \left(a'a - a'X(X'X)^{-1}X'a\right) && \text{(from slides } Q_x = I - X(X'X)^{-1}X'\text{)} \\
&= \sigma^2 a'Q_x a && \text{(note Euclidian norm, } a'Q_x a = \|Q_x a\|^2 \geq 0\text{)} \\
&\geq 0
\end{aligned}$$

So, we have shown that the variance is smallest for $L\hat{\beta}$ and hence it is a BLUE for $L\beta$.

Exercise 128

Let $y \sim N(X\beta, \sigma^2 I_n)$ and $y_0 \sim N(x_0'\beta, \sigma^2)$ independent from y.

Since $y \sim N(X\beta, \sigma^2 I_n)$, then the MLE for β is $\hat{\beta} = (X'X)^{-1}X'y$. Since $\hat{\beta}$ is an unbiased estimator then we have that $E[x_0'\hat{\beta}] = x_0'\beta$, and:

$$\begin{aligned}
var(x_0\hat{\beta}) &= x_0' var(\hat{\beta}) x_0 \\
&= x_0' var((X'X)^{-1}X'y) x_0 \\
&= \sigma^2 x_0'(X'X)^{-1}X'X(X'X)^{-1}x_0 \\
&= \sigma^2 x_0'(X'X)^{-1}x_0
\end{aligned}$$

Then we have that:

$$\frac{x_0'\hat{\beta} - x_0'\beta}{\sqrt{\sigma^2 x_0'(X'X)^{-1}x_0}} \sim N(0,1)$$

Now, let $s^2 = \frac{RSS}{n-p}$. We know that $\hat{\beta}$ and RSS are independent, and that $\frac{RSS}{\sigma^2} \sim \chi^2_{n-p}$. Then:

$$\frac{\left(\frac{x_0'\hat{\beta} - x_0'\beta}{\sqrt{\sigma^2 x_0'(X'X)^{-1}x_0}}\right)}{\sqrt{\frac{RSS}{\sigma^2(n-p)}}} = \frac{x_0'\hat{\beta} - x_0'\beta}{s\sqrt{x_0'(X'X)^{-1}x_0}} \sim t_{n-p}$$

Then a $100(1-\alpha)\%$ confidence interval for $x_0'\beta$ is given by:

$$x_0'\hat{\beta} \pm t_{n-p,\alpha/2} \cdot s \cdot \sqrt{x_0'(X'X)^{-1}x_0}$$

Now assume we predict a new observation y_0. Since y_0 is normal sharing the same variance as $x_0'\hat{\beta}$ and due to independence we have that:

$$y_0 - x_0'\hat{\beta} \sim N(0, \sigma^2(1 + x_0'(X'X)^{-1}x_0))$$

Then similarly to above, we have that:

$$\frac{y_0 - x_0'\hat{\beta}}{\sqrt{\sigma^2(1 + x_0'(X'X)^{-1}x_0)}} \sim N(0,1)$$

Then due to the same arguments above on s^2 and RSS, we have that:

$$\frac{\left(\frac{y_0 - x_0'\hat{\beta}}{\sqrt{\sigma^2(1+x_0'(X'X)^{-1}x_0)}}\right)}{\sqrt{\frac{RSS}{\sigma^2(n-p)}}} = \frac{y_0 - x_0'\hat{\beta}}{s\sqrt{1+x_0'(X'X)^{-1}x_0}} \sim t_{n-p}$$

Then the prediction interval for y_0 is given by:

$$x_0'\hat{\beta} \pm t_{n-p,\alpha/2} \cdot s \cdot \sqrt{1+x_0'(X'X)^{-1}x_0}$$

Exercise 129

We need to derive a generalized likelihood ratio test (LRT) for $H_0 : \mu = 0$ versus $H_A : \mu \neq 0$. We start by noting that the LRT is

$$\Lambda(y) = \frac{\mathcal{L}(\hat{\theta}_0|y)}{\mathcal{L}(\hat{\theta_A}|y)}$$

Since we have a normal linear regression model, we derive our likelihood function based on the normal density function,

$$\begin{aligned}
\mathcal{L}(\theta|y) &= \prod_{i=1}^{n} f(y_i|\theta) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i-\mu)^2\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)
\end{aligned}$$

Given that the MLE of the parameters is the sample mean, let us denote it $\bar{y}$. Now, we can input both likelihoods into the expression for the LRT,

$$\begin{aligned}
\Lambda(y) &= \frac{\mathcal{L}(\hat{\theta}0|y)}{\mathcal{L}(\hat{\theta}A|y)} \\
&= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-0)^2\right)}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})^2\right)} \qquad \text{(recall that } H_0 : \mu = 0\text{)} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n}(y_i-\bar{y})^2\right]\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2)\right]\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} y_i^2 + 2n\bar{y}^2 - n\bar{y}^2\right]\right) \qquad \text{(substite for the mean)} \\
&= \exp\left(-\frac{1}{2\sigma^2}[n\bar{y}^2]\right) \\
&= \exp\left(-\frac{n}{2\sigma^2}\bar{y}^2\right)
\end{aligned}$$

Now we reject the null whenever $\Lambda(y) < c$ for some constant c under the null hyypothesis. Given a level α:

$$
\begin{aligned}
P(\Lambda(y) < c|H_0) = \alpha &\iff P\left(\exp\left(-\frac{n}{2\sigma^2}\bar{y}^2\right)|H_0\right) = \alpha \\
&\iff P\left(\sum_{i=1}^n y_i > -2\sigma^2\ln(c)|H_0\right) = \alpha \\
\text{Under the null } y_i &\overset{i.i.d}{\sim} N(0,\sigma^2) \implies \sum_{i=1}^n y_i \sim N(0, n\sigma^2) \\
&\iff 1 - P\left(\frac{\sum_{i=1}^n y_i}{\sqrt{n}\sigma} < \frac{-2\sigma\ln(c)}{\sqrt{n}}|H_0\right) = \alpha \\
&\iff 1 - \alpha = P\left(\frac{\sum_{i=1}^n y_i}{\sqrt{n}\sigma} < \frac{-2\sigma\ln(c)}{\sqrt{n}}|H_0\right) \\
&\iff c = \exp\left(\frac{-\sqrt{n}}{2\sigma}z_{1-\alpha}\right)
\end{aligned}
$$

Exercise 130

Notice that we have that $y_{j,k} = \mu_j + \varepsilon_{j,k}$, where $\varepsilon \sim N(0,\sigma^2)$. Since these are i.i.d., then we have essentially a linear model of the form:

$$y_i = X\mu + \varepsilon_i$$

Where:

$$X = \begin{pmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_p} & 0_{n_p} & \cdots & 1_{n_p} \end{pmatrix}$$

Here notice this is not the identity matrix, but rather each 1_{n_j} and 0_{n_j} represents a vector of ones and zeroes of size n_j.

We want to test the hypothesis $H_0 : \mu_1 = ... = \mu_p$. To solve this, we refer to the general linear hypothesis (slide 54 of lecture 6). We have that for the general null hypothesis $H_0 : A\beta = b$ we have that under H_0:

$$\frac{(A\hat{\beta} - b)'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b)}{\sigma^2} \sim \chi^2_r$$

And our test statistic is given by:

$$F = \frac{(A\hat{\beta} - b)'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b)/r}{s^2} \sim F_{r,n-p}$$

We then have to choose A and b fittingly to fit our null hypothesis. Consider:

$$A = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}_{p-1\times p} \qquad b = \vec{0}$$

This matrix sequentially guarantees $\mu_1 = \mu_2$, then $\mu_2 = \mu_3$, and so on for all μ's.

Now notice that

$$X'X = diag(n_1, ..., n_p)$$

$$X'y = \begin{pmatrix} 1'_{n_1} & 0'_{n_2} & \cdots & 0_{n'_p} \\ 0'_{n_1} & 1'_{n_2} & \cdots & 0_{n'_p} \\ \vdots & \vdots & \ddots & \vdots \\ 0'_{n_1} & 0'_{n_2} & \cdots & 1'_{n_p} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n_1} y_{1,k} \\ \sum_{k=1}^{n_2} y_{2,k} \\ \vdots \\ \sum_{k=1}^{n_p} y_{p,k} \end{pmatrix}$$

$$= \begin{pmatrix} n_1 \bar{y}_{1,\cdot} \\ n_2 \bar{y}_{2,\cdot} \\ \vdots \\ n_p \bar{y}_{p,\cdot} \end{pmatrix}$$

Then the MLE for each μ is given by:

$$\hat{\mu} = (X'X)^{-1}X'y = diag(1/n_1, ..., 1/n_p) \begin{pmatrix} n_1 \bar{y}_{1,\cdot} \\ n_2 \bar{y}_{2,\cdot} \\ \vdots \\ n_p \bar{y}_{p,\cdot} \end{pmatrix} = \begin{pmatrix} \bar{y}_{1,\cdot} \\ \bar{y}_{2,\cdot} \\ \vdots \\ \bar{y}_{p,\cdot} \end{pmatrix}$$

Now, under the null, if we assume all groups have the same mean then we must have that the MLE is given by:

$$\hat{\mu}_0 = \bar{y} = \frac{1}{\sum_{j=1}^{p} n_j} \sum_{j=1}^{p} \sum_{k=1}^{n_j} y_{j,k} = \bar{y}_{\cdot,\cdot}$$

Now we can finally use the given hint. We start by calculating the components:

$$X(\hat{\mu} - \hat{\mu_0}) = \begin{pmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_p} & 0_{n_p} & \cdots & 1_{n_p} \end{pmatrix} \begin{pmatrix} \bar{y}_{1,\cdot} - \bar{y}_{\cdot,\cdot} \\ \bar{y}_{2,\cdot} - \bar{y}_{\cdot,\cdot} \\ \vdots \\ \bar{y}_{p,\cdot} - \bar{y}_{\cdot,\cdot} \end{pmatrix}$$

$$= \begin{pmatrix} [\bar{y}_{1,\cdot} - \bar{y}_{\cdot,\cdot}]_{n_1 \times 1} \\ [\bar{y}_{2,\cdot} - \bar{y}_{\cdot,\cdot}]_{n_2 \times 1} \\ \vdots \\ [\bar{y}_{p,\cdot} - \bar{y}_{\cdot,\cdot}]_{n_p \times 1} \end{pmatrix}$$

Then:

$$||X(\hat{\mu} - \hat{\mu_0})||^2 = \sum_{j=1}^{p} n_j (\bar{y}_{j,\cdot} - \bar{y}_{\cdot,\cdot})^2$$

then we have that:

$$Q_X y = (I - X(X'X)^{-1}X')y = y - X\hat{\mu}$$

Which means that:

$$||Q_X y||^2 = \sum_{j=1}^{p}\sum_{k=1}^{n_j}(y_{j,k} - \bar{y}_{j,\cdot})^2$$

And finally all that's left is to find the value of r. Note that under the null hypothesis of all μ_j being equal, we basically reduce our full model from a p-variable one to a single variable one (where that single variable can be seen as the sum of all other variables, since they share a single β coefficient). Therefore, $r = p - 1$. And therefore by the hint:

$$F = \frac{\left(\sum_{j=1}^{p} n_j(\bar{y}_{j,\cdot} - \bar{y}_{\cdot,\cdot})^2\right)/(p-1)}{\left(\sum_{j=1}^{p}\sum_{k=1}^{n_j}(y_{j,k} - \bar{y}_{j,\cdot})^2\right)/(n-p)} \sim F_{p-1,n-p}$$

which, as seen in class has the desired F distribution and rejects when F is large.