

# STAT 665 Final

Maggie Buffum

March 21, 2019

## Problem 1

Let

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

be a random matrix where the elements  $m_{11}$ ,  $m_{12}$ ,  $m_{21}$ , and  $m_{22}$  are independent  $N(0, 1)$  random variables. Show that the probability that both eigenvalues of  $M$  are real is equal to  $P_r(W \geq \frac{1}{2})$ , where  $W \sim F_{(2,1)}$ .

---

We know that the eigenvalues of  $M$  are found by solving the following equation for  $\lambda_i$ :

$$|M - \lambda I| = 0$$

Solving for the determinant of  $M - \lambda I$ , we have

$$\begin{aligned} |M - \lambda I| &= \left| \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} m_{11} - \lambda & m_{12} \\ m_{21} & m_{22} - \lambda \end{bmatrix} \right| \\ &= (m_{11} - \lambda)(m_{22} - \lambda) - m_{12}m_{21} = 0 \text{ (set)} \\ &\implies m_{11}m_{22} - m_{11}\lambda - m_{22}\lambda + \lambda^2 - m_{12}m_{21} = 0 \\ &\implies \lambda^2 - \lambda(m_{11} + m_{22}) + (m_{11}m_{22} - m_{12}m_{21}) = 0 \end{aligned}$$

We recognize this as a second order polynomial, and we can solve for solutions for  $\lambda$  using the quadratic formula:

$$\lambda = \frac{(m_{11} + m_{22}) \pm \sqrt{(m_{11} + m_{22})^2 - 4(m_{11}m_{22} - m_{12}m_{21})}}{2}$$

The solutions for  $\lambda$  are real only when we take the square root of a nonnegative value, that is, when

$$\begin{aligned} (m_{11} + m_{22})^2 - 4(m_{11}m_{22} - m_{12}m_{21}) &\geq 0 \\ \implies (m_{11} + m_{22})^2 &\geq 4(m_{11}m_{22} - m_{12}m_{21}) \\ \implies m_{11}^2 + m_{22}^2 + 2m_{11}m_{22} &\geq 4m_{11}m_{22} - 4m_{12}m_{21} \\ \implies m_{11}^2 + m_{22}^2 + 2m_{11}m_{22} - 4m_{11}m_{22} &\geq -4m_{12}m_{21} \\ \implies m_{11}^2 + m_{22}^2 - 2m_{11}m_{22} &\geq -4m_{12}m_{21} \\ \implies (m_{11} - m_{22})^2 &\geq -4m_{12}m_{21} \end{aligned}$$

Therefore, we are looking for the probability that  $(m_{11} - m_{22})^2 \geq -4m_{12}m_{21}$ .

Let's first consider the distribution of  $m_{11} - m_{22}$ . Both random variables are independent and have standard normal distributions, so to find the distribution of their difference we simply take the differences of the means as the new mean, and the sum of the variances as the new variance. So we have that  $m_{11} - m_{22} \sim N(0, 2)$ . Consider now  $(m_{11} - m_{22})^2$ : if we normalize  $m_{11} - m_{22} \sim N(0, 2) = \sqrt{2}N(0, 1)$ , we find that  $(m_{11} - m_{22})^2 \sim 2\chi_1^2$ .

Now let's look at the distribution of  $m_{12}m_{21}$ . Again, these two random variables are independent and follow the standard normal distribution. To find their joint distribution, consider the following:

$$\begin{aligned}
m_{12}m_{21} &= \frac{1}{4}(4m_{12}m_{21}) \\
&= \frac{1}{4}(4m_{12}m_{21} + (m_{12}^2 - m_{12}^2) + (m_{21}^2 - m_{21}^2)) \\
&= \frac{1}{4}(m_{12}^2 + m_{12}^2 + 2m_{12}m_{21} - m_{12}^2 - m_{21}^2 + 2m_{12}m_{21}) \\
&= \frac{1}{4}((m_{12} + m_{21})^2 - (m_{12} - m_{21})^2)
\end{aligned}$$

Since  $m_{12}$  and  $m_{21}$  are independent standard normal random variables, we know that their sum and differences are  $N(0, 2)$  random variables. Standardizing the resulting normal random variables and squaring, we have the difference in two scaled chi-square random variables such that

$$\begin{aligned}
m_{12}m_{21} &= \frac{1}{4}((m_{12} + m_{21})^2 - (m_{12} - m_{21})^2) \\
&= \frac{1}{4}(\sqrt{2}N(0, 1))^2 - (\sqrt{2}N(0, 1))^2 \\
&= \frac{1}{4}(2\chi_{(1)}^2 - 2\chi_{(1)}^2) \\
&= \frac{1}{2}(\chi_{(1)}^2 - \chi_{(1)}^2)
\end{aligned}$$

Note that both chi-squared random variables are independent: from Casella and Berger we know that the sum and difference of independent standard normal random variables are independent. Also from Casella and Berger, we know that functions of independent random variables are independent.

Let's go back to the probability we are trying to find:

$$\begin{aligned}
(m_{11} - m_{22})^2 &\geq -4m_{12}m_{21} \\
\implies 2\chi_{(1)}^2 &\geq -4\left(\frac{1}{2}(\chi_{(1)}^2 - \chi_{(1)}^2)\right) \\
\implies 2\chi_1^2 &\geq -4\left(\frac{1}{2}(\chi_2^2 - \chi_3^2)\right) \text{ (changing notation)} \\
\implies 2\chi_1^2 &\geq -2\chi_2^2 + 2\chi_3^2 \\
\implies \frac{2\chi_1^2 + 2\chi_2^2}{2\chi_3^2} &\geq 1 \\
\implies \frac{\chi_1^2 + \chi_2^2}{2\chi_3^2} &\geq \frac{1}{2}
\end{aligned}$$

Again, we know that  $\chi_1^2$  and  $\chi_2^2$  are independent since the sum and difference of standard normal variables (and functions of those random variables) are independent. Therefore, using Cochran's Theorem, the sum of these two independent chi-square random variables is chi-squared with two degrees of freedom:

$$\frac{\chi_1^2 + \chi_2^2}{2\chi_3^2} \geq \frac{1}{2} \implies \frac{\chi_{(2)}^2/2}{\chi_{(1)}^2/1} \geq \frac{1}{2}$$

This, of course, is the ratio of two independent chi-squared random variables, which has an F-distribution with 2 and 1 degrees of freedom. Therefore,

$$P_r\left(W \geq \frac{1}{2}\right) = P_r\left((m_{11} + m_{22})^2 \geq -4m_{12}m_{21}\right),$$

where  $W \sim F_{(2,1)}$ , is the probability that both eigenvalues of  $M$  are real.

---

## Problem 2

(Testing for outliers) Consider the following model,

$$\begin{bmatrix} y_1 \\ \underline{y}_{-1} \end{bmatrix} = \begin{bmatrix} \underline{x}' \\ X_{-1}' \end{bmatrix} \beta + \begin{bmatrix} d_1 \\ \underline{d}_{-1} \end{bmatrix}.$$

That is, we partition  $y$ ,  $X$ , and  $\underline{d}$  into two parts: the first part corresponds to the first observation, and the second part corresponds to the remaining  $n - 1$  observations. Furthermore, under this model,

$$\begin{bmatrix} d_1 \\ \underline{d}_{-1} \end{bmatrix} \sim N \left( \begin{bmatrix} \Delta \\ \underline{0} \end{bmatrix}, \sigma^2 I \right).$$

Here,  $\Delta$  is a known, possibly nonzero, scalar parameter. Assume that  $\text{rank}(X) = p$  (so  $X$  has the full column rank). Define

$$\hat{d}_{1,-1} = y_1 - \underline{x}'_1 \hat{\beta}_{-1},$$

where  $\hat{\beta}_{-1} = (X'_{-1} X_{-1})^{-1} X'_{-1} \underline{y}_{-1}$  and  $u = \underline{x}'_1 (X'_{-1} X_{-1})^{-1} \underline{x}_1$ .

---

(a) Determine the distribution of  $\hat{d}_{1,-1}$ .

In this problem, we are looking at outliers. Specifically, we are re-estimating our  $p$  parameters,  $\hat{\beta}$ s, after removing the first observation from the data, and looking at the residual of predicting the “new” observation (the one we removed). We are given that the vector of  $d_i$ s is normal, and so by the reproducibility of the normal distribution, we know that  $\hat{d}_{1,-1}$  is normal. Let’s find its expected value and variance, but first we need to rewrite it:

$$\begin{aligned} \hat{d}_{1,-1} &= y_1 - \mathbf{x}'_1 \hat{\beta}_{-1} \\ &= (\mathbf{x}'_1 \beta + d_1) - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \mathbf{y}_{-1} \\ &= \mathbf{x}'_1 \beta + d_1 - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} (\mathbf{X}_{-1} \beta + \underline{d}_{-1}) \\ &= \mathbf{x}'_1 \beta + d_1 - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \mathbf{X}_{-1} \beta - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \underline{d}_{-1} \\ &= \mathbf{x}'_1 \beta + d_1 - \mathbf{x}'_1 \beta - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \underline{d}_{-1} \\ &= d_1 - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \underline{d}_{-1} \end{aligned}$$

Take the expected value of  $\hat{d}_{1,-1}$ :

$$\begin{aligned} E[\hat{d}_{1,-1}] &= E[d_1 - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \underline{d}_{-1}] \\ &= E[d_1] - E[\mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \underline{d}_{-1}] \\ &= E[d_1] - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} E[\underline{d}_{-1}] \\ &= E[d_1] \text{ (second term drops because } E[\underline{d}_{-1}] = 0) \\ &= \Delta \end{aligned}$$

Now let’s find the variance:

$$V[\hat{d}_{1,-1}] = V[d_1 - \mathbf{x}'_1 (\mathbf{X}'_{-1} \mathbf{X}_{-1})^{-1} \mathbf{X}'_{-1} \underline{d}_{-1}]$$

The covariance term is zero because we are given that the  $d_i$ s are uncorrelated. Continuing,

$$\begin{aligned}
V[\hat{d}_{1,-1}] &= V[d_1] + V[\mathbf{x}'_1(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}d_{-1}] \\
&= V[d_1] + [\mathbf{x}'_1(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}]V[d_{-1}][\mathbf{x}'_1(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}]' \\
&= \sigma^2 + [\mathbf{x}'_1(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}]\sigma^2 I[\mathbf{X}_{-1}(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{x}_1] \\
&= \sigma^2 + \sigma^2 \mathbf{x}'_1(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}\mathbf{X}_{-1}(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{x}_1 \\
&= \sigma^2 + \sigma^2 \mathbf{x}'_1(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{x}_1 \\
&= \sigma^2 + \sigma^2 u \\
&= \sigma^2(1+u)
\end{aligned}$$

Therefore, the distribution of  $\hat{d}_{1,-1}$  is

$$\hat{d}_{1,-1} \sim N(\Delta, \sigma^2(1+u))$$

---

(b) Let

$$s_{-1}^2 = \frac{(y_{-1} - X_{-1}\hat{\beta}_{-1})'(y_{-1} - X_{-1}\hat{\beta}_{-1})}{n-p-1}.$$

Determine the distribution of

$$\frac{\hat{d}_{1,-1}}{s_{-1}\sqrt{1+u}}$$

We know that  $\hat{d}_{1,-1}$  is normally distributed from part (a). Note that we can create a  $T$  random variable as follows:

$$T = \frac{X}{\sqrt{U/k}} \sim t(k, \delta)$$

where  $X \sim N(\mu, \sigma^2)$ ,  $U/\sigma^2 \sim \chi_k^2$ , with  $X$  independent from  $U$ , and  $\delta = \mu/\sigma$ . We would like to show that  $s_{-1}\sqrt{1+u} = \sqrt{U/k}$ . First, we recognize that  $y_{-1} \sim N(X_{-1}\hat{\beta}_{-1}, \sigma^2 I)$  such that

$$\frac{y_{-1} - X_{-1}\hat{\beta}_{-1}}{\sigma^2 I} \sim Z = N(0, 1)$$

which gives us that

$$\frac{(y_{-1} - X_{-1}\hat{\beta}_{-1})'(y_{-1} - X_{-1}\hat{\beta}_{-1})}{\sigma^2 I} \sim \chi_{n-p-1}^2$$

Let's now consider the following:

$$\frac{\frac{\hat{d}_{1,-1} - \Delta}{\sigma I \sqrt{1+u}}}{\left( \frac{(y_{-1} - X_{-1}\hat{\beta}_{-1})'(y_{-1} - X_{-1}\hat{\beta}_{-1})}{\sigma^2 I(n-p-1)} \right)^{1/2}} = \frac{Z}{\sqrt{\chi_{(n-p-1)}^2/(n-p-1)}} \sim T_{(n-p-1), \Delta/\sqrt{\sigma^2 I(1+u)}}$$

Simplifying, we see that

$$\frac{\frac{\hat{d}_{1,-1} - \Delta}{\sigma I \sqrt{1+u}}}{\left( \frac{(y_{-1} - X_{-1}\hat{\beta}_{-1})'(y_{-1} - X_{-1}\hat{\beta}_{-1})}{\sigma^2 I(n-p-1)} \right)^{1/2}} = \frac{\frac{\hat{d}_{1,-1} - \Delta}{\sigma I \sqrt{1+u}}}{\frac{s_{-1}}{\sigma I}} = \frac{\hat{d}_{1,-1} - \Delta}{s_{-1}\sqrt{1+u}} \sim T_{((n-p-1), \Delta/\sqrt{\sigma^2 I(1+u)})}$$

This implies that

$$\frac{\hat{d}_{1,-1}}{s_{-1}\sqrt{1+u}} \sim T$$

has a central T distribution with  $n-p-1$  degrees of freedom.

- 
- (c) Explain how the hypothesis  $H_0 : \Delta = 0$  vs.  $H_1 : \Delta \neq 0$  could be tested using the statistic in part (b), and discuss how the value of  $u$  affects the power of this test.

As demonstrated in part (b), since we know the distribution of that random variable (the test statistic), if we observe  $\hat{d}_{1,-1}$  such that the test statistic has a high probability of occurring given the distribution of the test statistic, then we fail to reject the null hypothesis that  $\Delta = 0$ . Alternatively, if we observed a test statistic with a low probability of occurring, then we have evidence to reject the hypothesis that  $\Delta = 0$  in favor of the alternative hypothesis,  $H_1 : \Delta \neq 0$ . Let's consider the effect of  $u$ : as  $u$  increases, the size of the test statistic decreases, and we are less likely to reject the null hypothesis that  $\Delta = 0$ .

---

### Problem 3

Consider the two-part (unordered) model

$$\underline{y} = X_1 \underline{\alpha}_1 + X_2 \underline{\alpha}_2 + \underline{d},$$

where  $\underline{\alpha}_1$  and  $\underline{\alpha}_2$  are unrestricted unknown parameter vectors and  $\underline{d}$  is a vector of disturbances.

---

- (a) Give, in tabular form, the two ANOVA's in terms of orthogonal projection matrices. Define all orthogonal matrices you use in terms of  $X_1$  and/or  $X_2$ .

Consider the partitioned matrix and parameters  $[X_1|X_2]$  and  $[\alpha_1|\alpha_2]'$  such that

$$Y = [X_1|X_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + d$$

The two ANOVA tables are

| Source    | Df                           | SS   |
|-----------|------------------------------|--|
| $X_1$     | $rank(X_1)$                  | $y'(I - P_{X_1})y$                           |
| $X_2 X_1$ | $rank(X_1, X_2) - rank(X_1)$ | $y'(I - P_{[X_1, X_2]})y - y'(I - P_{X_1})y$ |
| Error     | $n - rank(X_1, X_2)$         | $y'(I - P_{[X_1, X_2]})y$                    |
| Total     | $n$                          | $y'y$  |

and

| Source    | Df                           | SS   |
|-----------|------------------------------|--|
| $X_2$     | $rank(X_2)$                  | $y'(I - P_{X_2})y$                           |
| $X_1 X_2$ | $rank(X_1, X_2) - rank(X_2)$ | $y'(I - P_{[X_1, X_2]})y - y'(I - P_{X_2})y$ |
| Error     | $n - rank(X_1, X_2)$         | $y'(I - P_{[X_1, X_2]})y$                    |
| Total     | $n$                          | $y'y$  |

---

- (b) Under what circumstances are the two ANOVA's in part (a) the same except for order?

The projection matrices for  $X_1$  and  $X_2$  must be equal. This implies that the ranks of  $X_1$  and  $X_2$  are equal, which satisfies the degrees of freedoms.

---

- (c) Derive the solutions  $\hat{\alpha}_1$  and  $\hat{\alpha}_2$  for normal equations.

The least squares solutions of  $\alpha_1$  and  $\alpha_2$  are obtained by simultaneously solving the normal equations

$$\begin{aligned} X_1' X_1 \hat{\alpha}_1 + X_1' X_2 \hat{\alpha}_2 &= X_1' y \\ X_2' X_1 \hat{\alpha}_1 + X_2' X_2 \hat{\alpha}_2 &= X_2' y \end{aligned}$$

for  $\hat{\alpha}_1$  and  $\hat{\alpha}_2$ . Let's first solve for  $\hat{\alpha}_1$ :

$$\begin{aligned} X_1'X_1\hat{\alpha}_1 + X_1'X_2\hat{\alpha}_2 &= X_1'\mathbf{y} \\ \implies X_1'X_1\hat{\alpha}_1 &= X_1'\mathbf{y} - X_1'X_2\hat{\alpha}_2 \\ \implies X_1'X_1\hat{\alpha}_1 &= X_1'(\mathbf{y} - X_2\hat{\alpha}_2) \\ \implies \hat{\alpha}_1 &= (X_1'X_1)^{-1}X_1'(\mathbf{y} - X_2\hat{\alpha}_2) \end{aligned}$$

Now, let's consider

$$\begin{aligned} X_2'X_1\hat{\alpha}_1 + X_2'X_2\hat{\alpha}_2 &= X_2'\mathbf{y} \\ \implies X_2'X_1((X_1'X_1)^{-1}X_1'(\mathbf{y} - X_2\hat{\alpha}_2)) + X_2'X_2\hat{\alpha}_2 &= X_2'\mathbf{y} \\ \implies X_2'P_{X_1}(\mathbf{y} - X_2\hat{\alpha}_2) + X_2'X_2\hat{\alpha}_2 &= X_2'\mathbf{y} \\ \implies X_2'P_{X_1}\mathbf{y} - X_2'P_{X_1}X_2\hat{\alpha}_2 + X_2'X_2\hat{\alpha}_2 - X_2'\mathbf{y} &= 0 \\ \implies X_2'(P_{X_1} - I)\mathbf{y} + X_2'(I - P_{X_1})X_2\hat{\alpha}_2 &= 0 \\ \implies X_2'(I - P_{X_1})X_2\hat{\alpha}_2 &= X_2'(I - P_{X_1})\mathbf{y} \\ \implies \hat{\alpha}_2 &= [X_2'(I - P_{X_1})X_2]^{-1}X_2'(I - P_{X_1})\mathbf{y} \end{aligned}$$


---

## Problem 4

Do exercises 6.10 and 7.18 from the textbook.

### Problem 6.10

Let  $\mathbf{x} \sim N(\mu, \Sigma)$ . Show that  $b_{1,k}$  and  $b_{2,k}$  defined in (6.3.3) and (6.3.4) are invariant under a nonsingular linear transformation  $\mathbf{y} = \mathbf{P}\mathbf{x} + \mathbf{q}$  (where  $\mathbf{P}$  is nonsingular). Discuss whether these coefficients depend on  $\mu$  and  $\Sigma$ .

From (6.3.3) and (6.3.4) we have that

$$b_{1,k} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N r_{ij}^3$$

where

$$r_{ij} = (N-1)(\mathbf{x}_i - \bar{\mathbf{x}}_N)' \mathbf{S}_N^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_N)$$

and

$$b_{2,k} = \frac{N-1}{N} \sum_{i=1}^N \{(\mathbf{x}_i - \bar{\mathbf{x}}_N)' \mathbf{S}_N^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_N)\}^2$$

Let's consider  $Y_i = Px_i + q$  where  $x_1, \dots, x_k \sim N_k(\mu, \Sigma)$ . To show that  $b_{1,k}$  and  $b_{2,k}$  are invariant under the nonsingular transformation  $\mathbf{y} = \mathbf{P}\mathbf{x} + \mathbf{q}$ , where  $P$  is nonsingular, we will see if using  $y$  in place of  $x$  in the skewness and kurtosis equations changes their values. For simplicity, let's first look at  $\bar{y}$ :

$$\begin{aligned} \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (Px_i + q) = \frac{1}{n} \left[ \sum_{i=1}^n Px_i + \sum_{i=1}^n q \right] \\ &= \frac{1}{n} \sum_{i=1}^n Px_i + \frac{1}{n} (nq) = \frac{1}{n} P \sum_{i=1}^n x_i + q \\ &= P\bar{x} + q \end{aligned}$$

Now we can plug this into the equation for  $S_N = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$  by first looking at  $x_i - \bar{x}$ :

$$x_i - \bar{x} = Px_i + q - (P\bar{x} + q) = Px_i + q - P\bar{x} - q = Px_i - P\bar{x} = P(x_i - \bar{x})$$

Now,

$$\begin{aligned}
S_N &= \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \\
&= \sum_{i=1}^n (P(x_i - \bar{x}))(P(x_i - \bar{x}))' \\
&= \sum_{i=1}^n P(x_i - \bar{x})(x_i - \bar{x})' P' \\
&= P \left[ \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \right] P' \\
&= P S_n P'
\end{aligned}$$

Now we can use this to calculate part of  $r_{ij}$ :

$$\begin{aligned}
(y_i - \bar{y})'[S_{N_y}]^{-1}(y_j - \bar{y}) &= (P(x_i - \bar{x}))'[P S_{N_x} P']^{-1}(P(x_j - \bar{x})) \\
&= (x_i - \bar{x})' P' [P S_{N_x} P']^{-1} P(x_j - \bar{x}) \\
&= (x_i - \bar{x})' P' P'^{-1} S_{N_x}^{-1} P^{-1} P(x_j - \bar{x}) \\
&= (x_i - \bar{x})' I S_{N_x}^{-1} I(x_j - \bar{x}) \\
&= (x_i - \bar{x})' S_{N_x}^{-1}(x_j - \bar{x})
\end{aligned}$$

From this we can determine that  $b_{1,k}$  remained the same and is thus invariant to the nonsingular matrix  $P$ . Similarly,  $b_{2,k}$  is also invariant.

### Problem 7.18

Based on  $N$  observations  $(Y_i, X_{ij})$ ,  $j = 1, \dots, p$ ,  $i = 1, \dots, N$  following the model (4.1.1) with  $\epsilon_i \sim N(0, \sigma^2)$ , suppose we wish to predict the value of a future observation,  $Y_0$ , say. Let  $\hat{\beta}$  and  $\hat{\sigma}^2$  denote the least squares estimates of  $\beta$  and  $\sigma^2$  respectively.

- (a) Show that the mean of  $Y_0$  and  $Var(Y_0 - \hat{Y}_0)$  are respectively  $\mathbf{x}'_0 \beta$  and  $\sigma_{Y_0}^2 = \hat{\sigma}^2 \{1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0\}$ . Find the distribution of  $(Y_0 - \mathbf{x}'_0 \hat{\beta}) / \hat{\sigma} \{1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0\}^{1/2}$ .

The mean of  $Y_0$  is

$$E[Y_0] = E[\mathbf{x}'_0 \beta + \epsilon] = E[\mathbf{x}'_0 \beta] + E[\epsilon] = \mathbf{x}'_0 \beta$$

and the mean of  $Var(Y_0 - \hat{Y}_0)$  is

$$\begin{aligned}
Var(Y_0 - \hat{Y}_0) &= Var(Y_0) + Var(\hat{Y}_0) - 2Cov(Y_0, \hat{Y}_0) \\
&= Var(\mathbf{x}'_0 \beta + \epsilon) + Var(\mathbf{x}'_0 \hat{\beta}) \\
&= Var(\epsilon) + \mathbf{x}'_0 Var(\hat{\beta}) \mathbf{x}_0 \\
&= \sigma^2 + \mathbf{x}'_0 \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0 \\
&= \sigma^2 (1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0) \\
&= \sigma_{Y_0}^2
\end{aligned}$$

Let's find the distribution of  $(Y_0 - \mathbf{x}'_0 \hat{\beta}) / \hat{\sigma} \{1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0\}^{1/2}$ . First, we already know that

$$\frac{(Y_0 - \hat{Y}_0) \sqrt{N}}{\hat{\sigma} (1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0)^{1/2}} \sim t_{(N-2)}$$

Therefore, it follows that

$$\frac{(Y_0 - \mathbf{x}_0' \hat{\beta})}{\hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2}} \sim \frac{1}{\sqrt{N}} t_{N-2}$$


---

(b) Find the 95% symmetric two-sided prediction interval for  $Y_0$ .

From part (a) we know that

$$\frac{(Y_0 - \hat{Y}_0) \sqrt{N}}{\hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2}} \sim t_{N-2}$$

Since we know the distribution of the random variable above, we can use it to build a confidence interval:

$$\begin{aligned} t_{(0.975, N-2)} &\leq \frac{(Y_0 - \hat{Y}_0) \sqrt{N}}{\hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2}} \leq t_{(0.025, N-2)} \\ \implies t_{(0.975, N-2)} \hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2} \frac{1}{\sqrt{N}} &\leq (Y_0 - \hat{Y}_0) \leq t_{(0.025, N-2)} \hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2} \frac{1}{\sqrt{N}} \\ \implies \hat{Y}_0 + t_{(0.975, N-2)} \hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2} \frac{1}{\sqrt{N}} &\leq Y_0 \leq \hat{Y}_0 + t_{(0.025, N-2)} \hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2} \frac{1}{\sqrt{N}} \end{aligned}$$


---

(c) Let  $\eta \in (0, 0.5]$ , and suppose the  $100\eta$ th percentile of the distribution of  $Y_0$  is  $\gamma(\eta) = \mathbf{x}_0' \beta + z_\eta \sigma_{Y_0}$ . Find a  $100(1 - \alpha)\%$  lower confidence bound for  $\gamma(\eta)$ .

Consider  $\gamma(\eta) = \mathbf{x}_0' \beta + z_\eta \sigma_{Y_0}$ , where  $\eta$  defines the area to the left of some critical value  $z$  on a standard normal distribution. Because  $E[Y_0]$  and  $\sigma_{Y_0}$  are unknown, we need to estimate them using  $\hat{y}$  and  $\hat{\sigma}$ , that is,  $\hat{\gamma}(\eta) = \hat{Y}_0 + z_\eta \hat{\sigma}_{Y_0}$ . It looks like we are being asked to find a tolerance interval, that is, provide a lower bound such that at least  $100\eta\%$  of the sampled population falls above it, with confidence  $1 - \alpha$ . We know from part (b) that with confidence  $100((1 - \alpha)/2)\%$ , the true population mean  $Y_0$  falls within  $\hat{Y}_0 \pm t_{(\alpha/2, N-2)} \hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2} \frac{1}{\sqrt{N}}$ . Writing this in terms of  $\hat{\gamma}(\eta)$ , we have that with confidence  $100((1 - \alpha)/2)\%$ , the true population mean  $Y_0$  falls within

$$\hat{\gamma}(\eta) + z_\eta \hat{\sigma} \pm t_{(\alpha/2, N-2)} \hat{\sigma} \{1 + \mathbf{x}_0' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\}^{1/2} \frac{1}{\sqrt{N}}$$