1. (25 points) Consider the multiple linear regression model

\[ Y_{n \times 1} = Z_{n \times (r+1)} \beta_{(r+1) \times 1} + \epsilon_{n \times 1}, \]

where \( E(\epsilon) = 0 \) and \( E(\epsilon' \epsilon) = \sigma^2 V \), where \( V \) is a known \( n \times n \) positive definite matrix. Assume that \( Z \) is full rank and \( n > (r+1) \). Briefly derive the following properties:

(a) The ordinary least squares estimator \( \hat{\beta} = (Z'Z)^{-1}Z'Y \) is an unbiased estimate of \( \beta \), i.e. show \( E(\hat{\beta}) = \beta \). Obtain the covariance matrix of \( \hat{\beta} \).

Answer: \( E(\hat{\beta}) = (Z'Z)^{-1}Z'E(Y) = (Z'Z)^{-1}Z'\beta = \beta \).

Next, \( \text{Cov}(\hat{\beta}) = (Z'Z)^{-1}Z'(\epsilon \epsilon')(Z'Z)^{-1} = \sigma^2 (Z'Z)^{-1}Z'VZ(Z'Z)^{-1} \).

(b) Show that the weighted least squares estimator is \( \tilde{\beta}_w = (Z'V^{-1}Z)^{-1}Z'V^{-1}Y \).

Answer: Let \( V^{1/2} \) be the positive definite square root matrix of \( V \). Premultiplying the equation by \( V^{-1/2} \), we have

\[ V^{-1/2}Y = (V^{-1/2}Z)\beta + e, \]

where \( e = V^{-1/2} \epsilon \) such that \( \text{Cov}(e) = \sigma^2 I \). This is the usual multivariate multiple linear regression model. Therefore,

\[ \tilde{\beta}_w = (Z'V^{-1/2}V^{-1/2}Z)^{-1}(Z'V^{-1/2}V^{-1/2}Y) = (Z'V^{-1}Z)^{-1}(Z'V^{-1}Y). \]

(c) Assume further that \( \epsilon \) are multivariate normal. What are the distributions of \( \hat{\beta} \) and \( \tilde{\beta}_w \)?

Answer: From part (a), \( \hat{\beta} \sim N[\beta, \sigma^2(Z'Z)^{-1}(Z'VZ)(Z'Z)^{-1}] \). By the same technique as before, we have \( \text{Cov}(\tilde{\beta}_w) = (Z'V^{-1}Z)^{-1}(ZV^{-1}E(\epsilon \epsilon')V^{-1})(Z'V^{-1}Z)^{-1} = \sigma^2 (Z'V^{-1}Z)^{-1} \). Therefore, \( \tilde{\beta}_w \sim N[\beta, \sigma^2(Z'V^{-1}Z)^{-1}] \).

(d) Consider the special case

\[ y_i = \beta z_i + \epsilon_i, \quad i = 1, \ldots, n, \]

where \( \text{Var}(\epsilon_i) = \sigma^2 z_i^2 \). Write down the variances of the ordinary and weighted least squares estimators of \( \beta \). That is, obtain \( \text{Var}(\hat{\beta}) \) and \( \text{Var}(\tilde{\beta}_w) \).

Answer: In this special case, \( V = \text{diag}\{z_1^2, z_2^2, \ldots, z_n^2\} \). It is then easy to see that \( \text{Var}(\hat{\beta}) = \sigma^2 (\sum_{i=1}^n z_i^4)(\sum_{i=1}^n z_i^2)^{-2} \) and \( \text{Var}(\tilde{\beta}_w) = \sigma^2 / n \).
(e) Focus on the simple model in part (d). Show that the weighted least squares estimator $\hat{\beta}_w$ is more efficient than the ordinary least squares estimator $\hat{\beta}$. In other words, prove $\text{Var}(\hat{\beta}) \geq \text{Var}(\hat{\beta}_w)$. This justifies the use of weighted least squares estimators.

**Answer:** Consider $\sigma^{-2}[\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_w)] = \frac{\sum_{i=1}^{n} z_i^4}{(\sum_{i=1}^{n} z_i^2)^2} - \frac{1}{n}$

$$= \frac{n}{(\sum_{i=1}^{n} z_i^2)^2} \left[ \frac{\sum_{i=1}^{n} z_i^4}{n} - \left( \frac{\sum_{i=1}^{n} z_i^2}{n} \right)^2 \right]$$

$$= \frac{n}{(\sum_{i=1}^{n} z_i^2)^2} \text{Var}(z_i^2) \geq 0,$$

where we have used $\text{Var}(X) = E(X^2) - [E(X)]^2$.

2. (20 points) Suppose that $\mathbf{X} = (X_1, X_2, X_3)'$ follows a multivariate normal distribution with mean $\mathbf{\mu} = (1, -1, 2)'$ and covariance matrix

$$\Sigma = \begin{bmatrix} 4 & 1 & -1 \\ 1 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$  

Answer the following questions:

(a) Find the distribution of $Y = X_1 + X_2 + X_3$.

**Answer:** $Y \sim N(2, 11)$

(b) Are $X_1$ and $X_2 + X_3$ independent? Why?

**Answer:** Yes, because $\text{Cov}(X_1, X_2 + X_3) = \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) = 1-1 = 0$.

(c) Drive the conditional distribution $X_1|X_2 = 0.5$.

**Answer:** The conditional distribution is normal with $E(X_1|X_2 = 0.5) = 1.3$ and $\text{Var}(X_1|X_2 = 0.5) = 3.8$.

(d) Derive the conditional distribution of $X_1|(X_2 = 0.5, X_3 = 3)$.

**Answer:** Normal with $E[X_1|(X_2 = 0.5, X_3 = 3)] = 0.8$ and $\text{Var}(X_1|(X_2 = 0.5, X_3 = 3))$.

(e) Obtain a linear combination $\mathbf{a}'X$ with $\mathbf{a}'\mathbf{a} = 1$ such that $\text{Var}(\mathbf{a}'X)$ is as small as possible.

**Answer:** The linear combination is the eigenvector corresponding to the smallest eigenvalue. $\mathbf{a}' = c(0.415, -0.12, 0.902)$.  

2
3. (20 points) Consider the course evaluation of two instructors of the same course at Chicago Booth. The sample size for each instructors is 7. We focus on the class averages of rating for Question 2 to Question 6. In other words, we have \( p = 5 \) and \( n_1 = n_2 = 7 \). Some relevant R output is attached. Answer the following questions.

(a) Suppose we are interested in the overall satisfaction of students with the course so that the data are combined over the instructors. Test the null hypothesis \( H_0 : \mu = (4.5, 4.5, 4.5, 4.5, 4.5)' \) versus \( H_a : \mu \neq (4.5, 4.5, 4.5, 4.5, 4.5)' \). Compute the test statistic, its \( p \)-value and draw the conclusion.

**Answer:** The Hotelling \( T^2 \) is \( 14 \times 49.73 = 696.22 \). The corresponding \( p \)-value is \( 1.55 \times 10^{-7} \) so that the null hypothesis is rejected. That is, the mean is not \((4.5, 4.5, 4.5, 4.5, 4.5)\).

(b) Suppose we are indeed interested in comparing the instructors. Assume that the evaluation scores are normally distributed, say \( X_1 \sim N(\mu_1, \Sigma_1) \) and \( X_2 \sim N(\mu_2, \Sigma_2) \), respectively, for the two instructors. Test \( H_0 : \Sigma_1 = \Sigma_2 \) versus the alternative hypothesis \( H_a : \Sigma_1 \neq \Sigma_2 \). What is the test statistic? What is the \( p \)-value? Draw a conclusion.

**Answer:** Use the Box-M statistic, which is 30.95 with \( p \)-value 0.0089. Thus, the equal covariance hypothesis is rejected.

(c) Perform the test \( H_0 : \mu_1 = \mu_2 \) versus \( H_a : \mu_1 \neq \mu_2 \). What is the test statistic? What is the \( p \)-value? Draw a conclusion.

**Answer:** Use the Behrens test, which is 9.78 with \( p \)-value 0.366. Thus, there is no evidence to suggest any difference between the two instructors.

(d) Alternatively, one can use the multivariate analysis of variance to test the difference in the means. Write down the model for comparing the two instructors and perform the MANOVA. Draw a conclusion.

**Answer:** The model is

\[
x_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2; j = 1, \ldots, 7,
\]

where \( \mu \) is the overall mean vector and \( \tau_i \) is the effect of instructor \( i \) such that \( \tau_1 + \tau_2 = 0 \). \{\epsilon_{1j}\} and \{\epsilon_{2j}\} are independent with mean zero and positive definite covariance matrix. Performing the MANOVA, the Wilks statistic is 0.551 with \( p \)-value 0.351. Thus, there is no evidence of any difference between the two instructors.

(e) Construct the 95% Bonferroni confidence intervals for the marginal differences in the means, i.e. intervals for \( \mu_{i1} - \mu_{i2} \) for \( i = 1, \ldots, 5 \).

**Answer:** Since the sample sizes are the same, we are in effect using the pooled sample covariance matrix. The Bonferroni 95% confidence intervals are

\[ > \text{sii=diag(S)} \]
4. (20 points) Consider the Pulp and Paper Properties Data of Table 7.7 of the textbook; see exercise 7.26 on page 427. There are four response variables (BL, EM, SF, BS), four explanatory variables (AFL, LFF, FFF, ZST) and 62 observations. We perform some multivariate multiple linear regression analysis and the output is attached. Answer the following questions.

(a) Based on the output, which is based on the ordinary least squares method, what is the LSE of $\Sigma$? What is the MLE estimate of $\Sigma$?

Answer: The LSE and MLE of $\Sigma$ are, respectively,

$$
\hat{\Sigma} = \begin{bmatrix}
2.24 & 0.40 & 0.91 & 0.51 \\
0.40 & 0.12 & 0.19 & 0.09 \\
0.91 & 0.19 & 0.42 & 0.21 \\
0.51 & 0.09 & 0.21 & 0.12
\end{bmatrix},
\tilde{\Sigma} = \frac{62-5}{62} \hat{\Sigma}.
$$

(b) Perform a principal component analysis of the explanatory variables. What is the percentage of variability explained by the first principal component? What is the percentage of variability explained by the first three principal components?

Answer: From the output, they are 83.94% and 97.69%, respectively.

(c) Re-run the regression using the four principal components as the regressors. What is the least squares estimate of the $\Sigma$? Explain why this estimate is identical to that of using the original four explanatory variables.

Answer: The LSE of $\tilde{\Sigma}$ is the same as that given in part (a). Since PCs are linear combinations of the original regressors, the LS regression provides the same fit. Therefore, the residual covariance matrix is the same.

(d) Let $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})'$ be the four principal components (they are in the proper order with $X_{i4}$ having the smallest variance.) To simplify the regression model, consider the model

$$
y_i = \beta'_1 Z_{1i} + \beta'_2 Z_{2i} + \epsilon_i,
$$
where $Z_{2i} = X_{4i}$ and $Z_{1i}$ contains a constant and the first three principal component. Test $H_0 : \beta_2 = 0$ versus $H_a : \beta_2 \neq 0$. Compute the test statistic and its $p$-value. Draw a conclusion.

**Answer:** Use the determinants to compute the likelihood ratio test with Bartlett’s adjustment. The test statistic is 7.05 with $p$-value 0.13. Therefore, the last PC can be removed from the regression.

5. (15 points) Briefly answer the following questions:

(a) Suppose that the $p$-dimensional random vector $X$ has mean $\mu$ and covariance matrix $\Sigma$, which is positive definite. Then $(X - \mu)' \Sigma^{-1} (X - \mu)$ is distributed as $\chi^2_p$. True or false? Why?

**Answer:** False, because $X$ may not be normal.

(b) Suppose that $X_1$ and $X_2$ are distributed as $X_1 \sim N(\mu_1, \sigma_{11})$ and $X_2 \sim N(\mu_2, \sigma_{22})$. Also, Cov($X_1, X_2$) = $\sigma_{12}$. Then $X = (X_1, X_2)'$ is bivariate normal with mean $\mu = (\mu_1, \mu_2)'$ and covariance matrix $\Sigma = [\sigma_{ij}]_{2 \times 2}$. True or false? Why?

**Answer:** False, because marginal normality does not imply joint normality. The statement is true if $\sigma_{12} = 0$.

(c) Describe a method to check the validity of the assumption of a $p$-dimensional normal distribution, where $p > 1$.

**Answer:** Use chi-square QQ plot.

(d) Let $X_1, X_2, \ldots, X_n, X_{n+1}$ be independently distributed as $N(\mu, \Sigma)$, where $\Sigma$ is positive definite. Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. What is the distribution of $X_{n+1} - \bar{X}$?

**Answer:** The mean is zero and the covariance matrix is $(1 + \frac{1}{n}) \Sigma$ because $X$ and $X_{n+1}$ are independent. Therefore, $X_{n+1} - \bar{X} \sim N(0, \frac{n+1}{n} \Sigma)$.

(e) Consider the multiple linear regression

$$Y_{n \times 1} = Z_{n \times (r+1)} \beta_{(r+1) \times 1} + \epsilon_{n \times 1}.$$

Write down the conditions under which the ordinary least squares estimator is the best linear unbiased estimator (BLUE).

**Answer:** (a) $Z$ is full rank, (b) $\{\epsilon_i\}$ are independent random variables with mean zero and Var($\epsilon_i$) = $\sigma$, which is a positive constant, where $\epsilon_i$ is the $i$th element of $\epsilon$, and (c) the linear regression model is the true model.