Assignment 2

[1] Let $X \in \mathbb{R}^{n\times p}$ with $p \le n$. Show that $r(X) = p$ if and only if $X^T X \in \mathbb{R}^{p\times p}$ is invertible.

[2] Consider the linear model

$$\mathbf{Y} = X\beta + \epsilon \tag{1}$$

where $\mathbf{Y}$ and ϵ are in $\mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, and $\beta \in \mathbb{R}^p$, with $p \le n$ and $r(X) = p$.

Show, using matrix algebra only, that

$$\|\mathbf{Y} - X\widehat{\beta}\|^2 = \inf_{\beta\in\mathbb{R}^p} \|\mathbf{Y} - X\beta\|^2 \quad \text{for} \quad \widehat{\beta} = (X^T X)^{-1} X^T Y. \tag{2}$$

[3] For the linear model in (1), prove (2) using only calculus. Compare to the method used in item ([2])

[4] When $r(X) < p$ prove a statement parallel to (2) for the quantity $\|\widehat{\mathbf{Y}} - \mathbf{Y}\|^2$ for a suitably defined $\widehat{\mathbf{Y}}$. State in what ways this situation differs from the one where X has full rank.

[5] For the linear model (1) and $\widehat{\beta}$ as given in (2) with $E[\epsilon] = 0$ and $\text{Cov}(\epsilon) = \sigma^2 I$, consider the variance estimator

$$S^2 = \frac{1}{n-p} \sum_{i=1}^{n} \widehat{\epsilon_i}^2$$

where the 'residual' $\widehat{\epsilon_i}$ is given by

$$\widehat{\epsilon_i} = y_i - \widehat{\beta_1} x_{i,1} - \cdots - \widehat{\beta_p} x_{i,p}.$$

Write out S^2 in terms only of vector and matrices, and prove that it is unbiased for σ^2 using only the results of the previous relevant exercises, and also identities and properties such as $PX = X$ for the projection matrices P and N.

[6] In the setting of the previous exercise, prove that the sum of the residuals is zero when the vector $\mathbf{1}_n = (1, \ldots, 1)^T \in \mathbb{R}^n$ is an element of $\mathcal{R}(X)$.

[7] Apply the results developed to derive estimates of β_1, β_2 and σ^2 in the basic 'Galton' linear regression. Can one interpret the estimate of the slope parameter in terms of the estimated correlation between X and Y, and their estimated standard deviations?

[8] For the Galton linear regression, with x_2 restricted to take values in the interval $[-1, 1]$, at which values should n observations be taken so as to minimize the variance of the least squares estimate of the slope parameter β_2? Generalize the result to an arbitrary interval $[a, b]$.

[9] Suppose that $\mathbf{X} \sim \mathcal{N}_n(\mu, \Sigma)$ where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{3}$$

with $\mathbf{X}_i \in \mathbb{R}^{n_i}, n_1 + n_2 = n$, and the mean μ and covariance matrix Σ similarly partitioned.

a) Show that $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\Sigma_{12} = 0$. Show one direction using the definition of independent, and the other using the moment generating function.

b) Show, without using any density or conditional density functions, that if Σ_{22} is invertible, then

$$\mathcal{L}(\mathbf{X}_1|\mathbf{X}_2) \sim \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Hint: Consider

$$\mathbf{W} = (\mathbf{X}_1 - \mu_1) - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2)$$

and use part a).

[10] For $\mathbf{Y} \in \mathbb{R}^n$, and partitioning as in (3), prove that

$$\inf_{\phi} E\|\mathbf{Y}_1 - \phi(\mathbf{Y}_2)\|^2 = E\|\mathbf{Y}_1 - E[\mathbf{Y}_1|\mathbf{Y}_2]\|^2,$$

where the infimum is over all (measurable) functions $\phi : \mathbb{R}^{n_2} \to \mathbb{R}^{n_1}$. Hint: Use properties such as the 'tower property' of conditional expectations.

[11] Suppose $\mathbf{Y} \in \mathbb{R}^n$ has mean μ and covariance matrix Σ. Partitioning as in (3), and assuming Σ_{22} invertible, prove that

$$\inf_{\mathbf{a} \in \mathbb{R}^{n_1}, B \in \mathbb{R}^{n_2 \times n_1}} E\|\mathbf{Y}_1 - \mathbf{a} - B(\mathbf{Y}_2 - \mu_2)\|^2 = E\|\mathbf{Y}_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Y}_2 - \mu_2)]\|^2.$$

Note that no assumption is being made regarding the distribution of $\mathbf{Y}$, other than the existence of first and second moments. Prove this result using only the facts provided by the previous two exercises.

[12] Find a non-Gaussian family of joint distributions $\mathbb{P}_\rho, \rho \in [-1, 1]$ such that when $(X, Y) \sim \mathbb{P}_\rho$ then X and Y have mean zero and variance 1, satisfy $\text{Corr}(X, Y) = \rho$, and have the property that

$$X \text{ and } Y \text{ are independent if and only if } \text{Cov}(X, Y) = 0. \tag{4}$$

The existence of such an example shows that the multivariate normal family is not the unique one with property (4).(Hint: Consider constructing $\mathbb{P}_\rho$ as a mixture of one joint distribution where X and Y are independent, taken with probability $1 - |\rho|$ and a second joint distribution where X and Y are maximally dependent, taken with probability $|\rho|$.)

[13] Find a pair of random variables (X, Y) such that, marginally, both X and Y have a $\mathcal{N}(0, 1)$ distribution, but the pair is not bivariate normal.

[14] In the linear model with assumptions as in [2], and also assuming that $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, show that

$$(n - p)S^2/\sigma^2 \sim \chi^2_{n-p}.$$

[15] Recall that we say X_n has the χ^2_n distribution if it has the same distribution as $Z_1^2 + \cdots + Z_n^2$, for $Z_1, \ldots, Z_n$ independent standard normal variables.

(a) Find the moment generating function of X_n.

(b) For $\mu_1, \ldots, \mu_n$, any real numbers, find the moment generating function of $X_{n,\nu} = (Z_1 + \mu_1)^2 + \cdots + (Z_n + \mu_n)^2$. We say $X_{n,\nu}$ has a non-central chi squared distribution with parameter ν. What function of the means $\mu_1, \ldots, \mu_n$ should we take for the non-centrality parameter ν?

[16] For the linear model under the assumptions that $X \in \mathbb{R}^{n \times p}$ with $r(X) = p \leq n$, and $\epsilon \sim \mathcal{N}_n(0, \sigma^2 I)$, using matrix methods and the properties of the multivariate normal distribution, show that the least squares estimate $\widehat{\beta}$ and the variance estimate

$$S^2 = \frac{1}{n-p} \sum_{i=1}^{n} \|Y - X\widehat{\beta}\|^2$$

are independent.