Statistics 206

Homework 1

Due : October 2, 2019, In Class

1. Review Appendix A3 and the matrix notes on canvas under "Files/Reading".

2. $\mathbf{Z}$ is an n-dimensional random vector with expectation $\mathbf{E}(\mathbf{Z})$ and variance-covariance matrix:
$$\mathbf{Var}(\mathbf{Z}) = \mathbf{Cov}(\mathbf{Z}, \mathbf{Z}) = \Sigma.$$
A is an $s \times n$ nonrandom matrix and B is a $t \times n$ nonrandom matrix. Show the following:
   (a) $\mathbf{E}(A\mathbf{Z}) = A\mathbf{E}(\mathbf{Z})$.
   (b) $\mathbf{Cov}(A\mathbf{Z}, B\mathbf{Z}) = A\Sigma B^T$. In particular, $\mathbf{Var}(A\mathbf{Z}) = A\Sigma A^T$.

3. Derive the following.
   (a) $\sum_{i=1}^n (X_i - \overline{X}) = 0, \quad \sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n (X_i - \overline{X})X_i = \sum_{i=1}^n X_i^2 - n(\overline{X})^2$.
   (b) $\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^n (X_i - \overline{X})Y_i = \sum_{i=1}^n X_iY_i - n\overline{X}\ \overline{Y}$.

4. Least-squares principle.
   (a) State the least-squares principle.
   (b) Derive the LS estimators for simple linear regression model.
   (c) Assume the observations follow:
$$Y_i = \exp(a + bX_i) + \epsilon_i, \quad i = 1, \cdots, n,$$
where $a, b \in \mathbb{R}$ are unknown parameters and ϵ_is are uncorrelated random variables with $E(\epsilon_i) = 0, Var(\epsilon_i) = \sigma^2$. Describe how to estimate the regression function (equivalently, a, b) by least-squares principle. (Notes: You only need to provide a description. This is an example of a nonlinear regression model.)

5. Tell true or false (with a brief explanation) of the following statements with regard to simple linear regression.
   (a) The least squares line always passes the center of the data $(\overline{X}, \overline{Y})$.
   (b) The true regression line fits the data the best.
   (c) If $\overline{X} = 0, \ \ \overline{Y} = 0$, then $\hat{\beta}_0 = 0$ no matter what is $\hat{\beta}_1$.
   (d) Given the sample size, the larger the range of X_is, the smaller the standard errors of $\hat{\beta}_0, \hat{\beta}_1$ tend to be.

(e) If the correlation coefficient r_{XY} between X_is and Y_is is such that $|r_{XY}| < 1$, then we will observe regression effect.

6. Properties of the residuals under simple linear regression model. Recall that

$$\begin{aligned} e_i &= Y_i - \widehat{Y_i} = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i), \quad i = 1, \cdots n. \\ &= (Y_i - \overline{Y}) - \hat{\beta}_1 (X_i - \overline{X}). \end{aligned}$$

Show that

(a) $\sum_{i=1}^n e_i = 0$.

(b) $\sum_{i=1}^n X_i e_i = 0$.

(c) $\sum_{i=1}^n \widehat{Y}_i e_i = 0$.

7. Under the simple linear regression model, show the following.

(a) The LS estimator $\hat{\beta}_0$ is an unbiased estimator of β_0 and derive the formula for its variance.

(b) $\overline{Y}$ and $\hat{\beta}_1$ are uncorrelated. (Hint: Write them as linear combinations of Y_is.)

8. Under the simple linear regression model:

(a) Show that

$$SSE = \sum_{i=1}^n (Y_i - \overline{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \overline{X})^2.$$

(b) Derive $E(\hat{\beta}_1^2)$.

(c) Show that

$$E((Y_i - \bar{Y})^2) = \beta_1^2 (X_i - \bar{X})^2 + E((\epsilon_i - \bar{\epsilon})^2),$$

where $\bar{\epsilon} = \frac{1}{n} \sum_{i=1}^n \epsilon_i$.

(d) Show that $E(SSE) = (n-2)\sigma^2$. (Hint: What is $E(\sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2)$?)