1. (a) Argue that the in the case of simple linear regression, the least square regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ always passes through the point $(\bar{x}, \bar{y})$.

   (b) Consider multiple linear regression with a intercept

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

   Argue that we have the same result: the least square regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x + \cdots + \hat{\beta}_p x_p$ always passes through the point $(\bar{x}_{\cdot 1}, \bar{x}_{\cdot 2}, \ldots, \bar{x}_{\cdot p}, \bar{y})$, where $\bar{x}_{\cdot j} = \sum_{i=1}^{n} x_{ij}$.

   (c) For multiple linear regression without a intercept, i.e., $\beta_0 = 0$, argue that the same result may not hold. (Hint: consider the geometric interpretation and the fact that $\bar{\boldsymbol{y}} = \frac{1}{n}\mathbf{1}^T \boldsymbol{y}$, and consider

$$\frac{1}{n}\mathbf{1}^T X(X^T X)^{-1} X^T = \frac{1}{n}\mathbf{1}^T P = \frac{1}{n}(P\mathbf{1})^T.$$

   Does this equal to $\frac{1}{n}\mathbf{1}$?)

2. Consider three linear regression models

$$\boldsymbol{y} = X_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} = \boldsymbol{x}_1 \beta_1 + \boldsymbol{x}_2 \beta_2 + \cdots + \boldsymbol{x}_{p_1} \beta_{p_1} + \boldsymbol{\varepsilon}, \tag{1}$$
$$\boldsymbol{y} = X_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = \boldsymbol{x}_1 \beta_1 + \boldsymbol{x}_2 \beta_2 + \cdots + \boldsymbol{x}_{p_1} \beta_{p_1} + \cdots + \boldsymbol{x}_{p_2} \beta_{p_2} + \boldsymbol{\varepsilon}, \tag{2}$$
$$\boldsymbol{y} = X_3 \boldsymbol{\beta}_3 + \boldsymbol{\varepsilon} = \boldsymbol{x}_1 \beta_1 + \boldsymbol{x}_2 \beta_2 + \cdots + \boldsymbol{x}_{p_1} \beta_{p_1} + \cdots + \boldsymbol{x}_{p_2} \beta_{p_2} + \cdots + \boldsymbol{x}_{p_3} \beta_{p_3} + \boldsymbol{\varepsilon}, \tag{3}$$

   where $0 < p_1 < p_2 < p_3$. Assume that $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p_3}\}$ are linearly independent, so that the column spaces satisfy $\mathcal{C}(X_1) \subsetneq \mathcal{C}(X_2) \subsetneq \mathcal{C}(X_3)$, where $\subsetneq$ denotes strict subset. To estimate the variance σ^2 of the error ε, let

$$\widehat{\sigma}_i^2 = \frac{\|\widehat{\varepsilon}_i\|_2^2}{n - p_i}, \quad i = 1, 2, 3,$$

   denote the usual estimator of σ^2 based on fitting the i-th model, so that $\widehat{\varepsilon}_i$ is the vector of residuals after fitting the i-th model. Now suppose that the true model is model (2). As a result, model (1) is *under-specified* and model (3) is *over-specified*. It is often claimed that under-specification leads to biased estimators and that over-specification leads to estimators that potentially have high variability.

   Under the assumption that model (2) is the true model, answer the following questions

   (a) Derive the bias of $\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\sigma}_3^2$. Which of them are biased?

   (b) Derive the variance of $\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\sigma}_3^2$. Compare the variance of $\widehat{\sigma}_2^2$ and $\widehat{\sigma}_3^2$.

3. It is often seen in textbooks/paper that the author claims "without loss of generality (WLOG)." These claims are not necessarily trivial. Some examples are listed below. For each, show how the simpler case can indeed be assumed WLOG by constructing a transformation that will put the more general form into the simpler form.

   (a) The linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

   is estimated using ordinary least squares. *WLOG, we can assume that all the variables are in the deviation form (deviation from the sample mean, i.e., replace x_{ik} by $x_{ik} - \bar{x}_{\cdot k}$ and Y_i by $Y_i - \bar{Y}$) and that a no-intercept model is fitted.*

(b) The likelihood ratio test of $H_0 : \boldsymbol{\beta} = \mathbf{0}$ *v.s.* $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$ is to be developed for the regression model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where X has full rank. *WLOG, assume that* $X^T X = I$.

(c) Suppose $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\text{Var}(\varepsilon) = \sigma^2 V$, where V is a known positive definite matrix. *WLOG, assume that* $V = I$.

4. When we do regressions using time series variables, it is common for the errors (residuals) to have a time series structure. This violates the usual assumption of independent errors made in ordinary least squares regression. The consequence is that the estimates of coefficients and their standard errors will be wrong if the time series structure of the errors is ignored.

   It is possible, though, to adjust estimated regression coefficients and standard errors when the errors have an *autoregressive* structure.

   Consider the linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + u_i,\ i = 1, 2, \ldots, n,$$

   where $u_0 = 0$,

$$u_i = \rho u_{i-1} + \varepsilon_i,\ i = 1, 2, \ldots, n,$$

   with $|\rho| < 1$, and $\{\varepsilon_i\}$ are i.i.d. noise with mean zero and $\text{Var}(\varepsilon_i) = \sigma^2$.

   (a) Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^\top$. Express u_i in terms of the ε_i's and then find the expression for $\mathsf{E}[\boldsymbol{u}]$ and the variance-covariance matrix $\text{Var}(\boldsymbol{u})$.

   (b) Suppose we fit the above model using ordinary least squares, i.e.,

$$\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{Y},$$

   where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

   Show that $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. Find an expression for the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$.

   (c) Let $\widehat{u}_i$ be the i-th residual from the ordinary least squares fit of the above model. Based on the residuals, give an estimator of ρ.

   (d) Using the estimate of ρ constructed in part (c), given an alternative estimator of $\boldsymbol{\beta}$.

   [Hint: Use generalized least square. Explicit decomposition of the variance-covariance matrix may not be straightforward. It is okay to assume existence of such a decomposition. Find the explicit decomposition if you can.]