

STAT 608, Spring 2022 - Assignment 4

1. Suppose that for the model $y_i = \alpha + e_i$, the errors are independent with mean 0. Also suppose that measurements are taken using one device for the first n_1 measurements, and then a more precise instrument was used for the next n_2 measurements. Thus $\text{Var}(e_i) = \sigma^2$, $i = 1, 2, \dots, n_1$ and $\text{Var}(e_i) = \sigma^2/2$, $i = n_1 + 1, n_1 + 2, \dots, n$.
 - (a) Ignore the fact that the errors have different variances, and derive the least squares estimator for $\hat{\alpha}$ using matrix notation and $\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$.
 - (b) Derive the weighted least squares estimator for α , $\hat{\alpha}_{\text{WLS}}$.
 - (c) Suppose that $n_1 = n_2$. Compute the expected values and variances of the two estimators above. Which is a better estimator and why? (Use theoretical $\text{MSE}(\hat{\alpha}) = \text{Bias}^2(\hat{\alpha}) + \text{Var}(\hat{\alpha})$ as your definition of “better.”)
2. Question 2, Chapter 4
3. Question 3, Chapter 4
4. Return to Question 4 from Homework 2, about coins being put on a scale. Now suppose that the variance in Y is proportional to the number of coins put on the scale. I recommend double-checking using both linear algebra and (if you're working in R) the linear model function `lm(y ~ x, weight = w)`, where w is a vector of weights (the diagonal of the weight matrix) and you invent your own y .
 - (a) Design an appropriate matrix of weights \mathbf{W} .
 - (b) Calculate the new least-squares estimates of the weights of the coins using weighted least squares.
5. For the model $y_i = \beta_0 + \beta_1 x_i + e_i$, the errors are iid with mean 0. The four observed values of x_i are $\mathbf{x} = [1 \ 2 \ 3 \ 4]'$. The estimator of β_1 is $\tilde{\beta}_1 = (y_4 + 2y_3 - 2y_2 - y_1) / 5$. For this model, do the following:
 - (a) Is $\tilde{\beta}_1$ unbiased? Show why or why not.
 - (b) What is the sampling variance of $\tilde{\beta}_1$?
 - (c) Get the usual least squares estimator of β_1 and calculate its sampling variance.
 - (d) Compare the sampling variance of $\tilde{\beta}_1$ with the sampling variance of the least squares estimator of β_1 .
6. A food manufacturing company is interested in modeling whether people prefer $x_1 = \text{Type A}$ or Type B hotdog buns with their hot dogs. They also want to control for $x_2 = \text{different amounts of sodium in the hot dogs themselves}$ and are testing the hot dog buns at a variety of sodium contents, giving each taster both a hot dog and a bun with no condiments. The response variable is $y = \text{perceived taste of the bun, on a scale of 1 to 10}$.
 - (a) In order to find out whether Type A or Type B is preferred, is it necessary to have an interaction term? Why or why not?

- (b) Develop a linear model for this study, interpreting all parameters in the context of the problem. Write down your hypotheses to be testing in terms of your model parameters. (You don't have any data to conduct the test; just write down the hypotheses.)
 - (c) Whether or not you added an interaction term above, assume now that it was added and it is statistically significant. How should we interpret this interaction in context?
7. In a one-way ANOVA model with $k = 3$ groups and 4 observations per group:
- (a) Use the F-statistic in Model Reduction Method 2 to derive a statistic for testing whether the average of the means of the first two groups is the same as the mean of the third group. That is, create the F-statistic for testing $H_0 : (\mu_1 + \mu_2) / 2 = \mu_3$. (Hint: Don't fit a model with a y -intercept. It makes everything easier.)
 - (b) Where $\hat{\mu}_1 = 5.6$, $\hat{\mu}_2 = 7.9$, $\hat{\mu}_3 = 6.1$, and $SSE = 12.8$, test your hypothesis. Use $\alpha = 0.05$. Note that the degrees of freedom for the F-statistic are r (the number of rows of your \mathbf{A} matrix) and $n - p - 1$.