STAT 608, Spring 2022 - Assignment 4
SOLUTIONS

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 $\text{Var}(e_i) = \sigma^2$, $i = 1, 2, \ldots, n_1$ and $\text{Var}(e_i) = \sigma^2/2$, $i = n_1 + 1, n_1 + 2, \ldots, n$.

   (a) Ignore the fact that the errors have different variances, and derive the least squares estimator for $\hat{\alpha}$ using matrix notation and $\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

   NOTICE FIRST THAT THE DESIGN MATRIX IS SIMPLY A COLUMN OF ONES SINCE WE'RE FITTING AN INTERCEPT-ONLY MODEL. IGNORING ANY WEIGHTING, WE HAVE $\hat{\alpha} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \bar{y}$, THE MEAN OF THE y-VALUES.

   (b) Derive the weighted least squares estimator for α, $\hat{\alpha}_{\text{WLS}}$.

   TAKING $w_i = 1/\sigma_i^2$, WE HAVE A DIAGONAL MATRIX WITH n_1 ELEMENTS EQUAL TO 1 AND n_2 ELEMENTS EQUAL TO 2:

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{bmatrix}$$

   THEN

$$\hat{\alpha}_{\text{WLS}} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} = \frac{1}{n_1 + 2n_2}\left(\sum_{i=1}^{n_1} y_i + 2\sum_{i=n_1+1}^{n} y_i\right)$$

   (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.")

   I'LL DEFINE $n_1 = n_2 = k$ SO THAT $n = 2k$.

$$\begin{aligned} \text{E}(\hat{\alpha}) &= \text{E}\left(\bar{Y}\right) \\ &= \frac{1}{n}\sum \text{E}(Y_i) \\ &= \frac{1}{n}\sum \text{E}(\alpha + e_i) \\ &= \frac{1}{n} n\alpha \\ &= \alpha \end{aligned}$$

$$
\begin{aligned}
\textsc{Var}\left(\hat{\alpha}\right) &= \textsc{Var}\left(\bar{Y}\right) \\
&= \textsc{Var}\left(\frac{1}{n}\sum Y_i\right) \\
&= \frac{1}{n^2}\sum \textsc{Var}\left(Y_i\right) \\
&= \frac{1}{(2k)^2}\sum \textsc{Var}\left(\alpha + e_i\right) \\
&= \frac{1}{(2k)^2}\left(k\sigma^2 + k\sigma^2/2\right) \\
&= \frac{k\sigma^2}{(2k)^2}\left(1 + \frac{1}{2}\right) \\
&= \frac{\sigma^2}{4k}\left(\frac{3}{2}\right) \\
&= \frac{3\sigma^2}{8k}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\hat{\alpha}_{\mathrm{WLS}}\right) &= \mathrm{E}\left[\frac{1}{k+2k}\left(\sum_{i=1}^{k} y_i + 2\sum_{i=k+1}^{2k} y_i\right)\right] \\
&= \frac{1}{3k}\left(\sum_{i=1}^{k}\mathrm{E}\left(\alpha + e_i\right) + 2\sum_{i=k+1}^{2k}\mathrm{E}\left(\alpha + e_i\right)\right) \\
&= \frac{1}{3k}\left(k\alpha + 2k\alpha\right) \\
&= \alpha
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\left(\hat{\alpha}_{\mathrm{WLS}}\right) &= \mathrm{Var}\left(\frac{1}{k+2k}\left(\sum_{i=1}^{k} y_i + 2\sum_{i=k+1}^{2k} y_i\right)\right) \\
&= \frac{1}{(3k)^2}\left(\sum_{i=1}^{k}\mathrm{Var}\left(\alpha + e_i\right) + 4\sum_{i=k+1}^{2k}\mathrm{Var}\left(\alpha + e_i\right)\right) \\
&= \frac{1}{(9k^2)}\left(k\sigma^2 + 4k\sigma^2/2\right) \\
&= \frac{k\sigma^2}{9k^2}\left(1 + 2\right) \\
&= \frac{\sigma^2}{3k} < \frac{3\sigma^2}{8k}
\end{aligned}
$$

Clearly, both estimators are unbiased. The variance of $\hat{\alpha}_{\mathrm{WLS}}$ is smaller, so we prefer the weighted least squares model; on average, its estimate will be closer to the true parameter value.

2. Question 2, Chapter 4

   First, we find that $w_i = 1/x_i^2$, so that $\text{Var}\left(\sqrt{w_i}e_i\right) = x_i^2\sigma^2/x_i^2 = \sigma^2$ is constant. Thus our weight matrix $\mathbf{W}$ is diag$(1/x_i^2)$. The design matrix here is simply the vector of values x_i, so we see, using $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$, that:

$$\mathbf{X}'\mathbf{W}\mathbf{X} = [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} 1/x_1^2 & 0 & \cdots & 0 \\ 0 & 1/x_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/x_n^2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum 1 = n$$

   and

$$\mathbf{X}'\mathbf{W}\mathbf{Y} = [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} 1/x_1^2 & 0 & \cdots & 0 \\ 0 & 1/x_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/x_n^2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum \frac{y_i}{x_i}$$

   So, $\hat{\beta} = \frac{\sum y_i/x_i}{n}$.

3. Question 3, Chapter 4

   (a) First, we see from the plot that the variances of the response variable are not equal at different values of either predictor variable. That is one assumption of a valid model, so any model assuming equal variances is not valid. Since the response variable is a median price per square foot, and we know that the variance of a median is proportional to $1/n_i$, where n_i is the number of observations from which the median is calculated, n_i is the appropriate choice for the weights.

   (b) We still do not have a valid model (even after using these weights) because the residuals do not have constant variance (indicated by the increasing slope in the last plot) across the different fitted values for the model. Variances of the residuals should be constant across any linear combination of the x values, include the y-hat values. Also, the relationships between y and the x variables do not appear to be linear.

   (c) First, I would consider adding predictors, such as indicator variables for when each of the current predictors equal zero, or a categorical variable for being located in an urban or suburban area of Houston, or being located inside or outside the loop or the beltway. Some of the largest home values may be explained better using one of those variables. Perhaps the average income of the neighborhood (possibly available from census data) could also be a predictor. Also, it is occasionally the case that a 0 is used to indicate that the information is actually missing. Next, I would try again to use the weights discussed here, and consider transformations of the predictors and response until I had the functional form of the model correct and constant variance of the residuals. If I still had an outlier, I might double-check the values of the observation for accuracy and then consider an indicator for that outlier to prevent it from being a bad leverage point in the model.

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(`$y \sim x$`, 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 $\mathbf{W}$.

   Since the first two measurements have one coin and the last two have two coins, the variances can be written as $\sigma^2, \sigma^2, 2\sigma^2, 2\sigma^2$. The appropriate weight matrix is thus:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}$$

   (b) Calculate the new least-squares estimates of the weights of the coins using weighted least squares.

   This time, we have

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_{\text{WLS}} &= \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} \\
&= \left(\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} y_1 + y_3/2 + y_4/2 \\ y_2 + y_3/2 + y_4/2 \end{bmatrix} \\
&= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 + y_3/2 + y_4/2 \\ y_2 + y_3/2 + y_4/2 \end{bmatrix} \\
&= \frac{1}{3} \begin{bmatrix} 2y_1 - y_2 + y_3/2 + y_4/2 \\ -y_1 + 2y_2 + y_3/2 + y_4/2 \end{bmatrix}
\end{aligned}$$

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.

   Yes, it is unbiased, since

$$\begin{aligned}
\mathrm{E}\left(\tilde{\beta}_1\right) &= \mathrm{E}\left[(y_4 + 2y_3 - 2y_2 - y_1)/5\right] \\
&= (1/5)\left[(\beta_0 + 4\beta_1) + 2(\beta_0 + 3\beta_1) - 2(\beta_0 + 2\beta_1) - (\beta_0 + \beta_1)\right] \\
&= (1/5)\left[4\beta_1 + 6\beta_1 - 4\beta_1 - \beta_1\right] \\
&= \beta_1
\end{aligned}$$

(b) What is the sampling variance of $\tilde{\beta}_1$?

$$\text{Var}\left(\left(y_4 + 2y_3 - 2y_2 - y_1\right)/5\right) = (1/25)\left(\sigma^2 + 4\sigma^2 + 4\sigma^2 + \sigma^2\right) = (10/25)\sigma^2 = (2/5)\sigma^2$$

(c) Get the usual least squares estimator of β_1 and calculate its sampling variance.
First, we calculate $\bar{x} = 2.5$, $\sum x_i^2 = 30$, and SXX $= 5$. Then we have

$$\begin{aligned}
\hat{\beta}_1 &= (1/5)\left(\sum x_i y_i - n\bar{x}\bar{y}\right) \\
&= (1/5)\left(y_1 + 2y_2 + 3y_3 + 4y_4 - 10\bar{y}\right) \\
&= (1/5)\left(y_1 + 2y_2 + 3y_3 + 4y_4 - 10 \times \frac{y_1 + y_2 + y_3 + y_4}{4}\right) \\
&= (1/5)\left(-\frac{3}{2}y_1 - \frac{1}{2}y_2 + \frac{1}{2}y_3 + \frac{3}{2}y_4\right)
\end{aligned}$$

Thus

$$\begin{aligned}
\text{Var}\left(\hat{\beta}_1\right) &= (1/25)\left(\frac{9}{4}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 + \frac{9}{4}\sigma^2\right) \\
&= (1/5)\sigma^2
\end{aligned}$$

(d) Compare the sampling variance of $\tilde{\beta}_1$ with the sampling variance of the least squares estimator of β_1.
The variance of $\hat{\beta}_1$ is smaller than that of $\tilde{\beta}_1$. So we find another linear unbiased estimator ($\hat{\beta}_1$) whose variance is smaller than $\tilde{\beta}_1$. Thus $\tilde{\beta}_1$ cannot be a BLUE. (Actually $\hat{\beta}_1$ is a BLUE and has the smallest variance among all linear unbiased estimators.)

6. A food manufacturing company is interested in modeling whether people prefer x_1 = Type A or Type B hotdog buns with their hot dogs. They also want to control for x_2 = 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 = 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?
The full answer is that it depends. As a first pass, we generally think of an interaction-free situation: that is, there are parallel lines, and as sodium increases, perceived taste might increase linearly, and we'll just be interested in whether Type A or Type B has a higher intercept. In that case, no, we don't need an interaction. To visually check whether an interaction is necessary, we could plot y vs. x_2 and color code the points by x_1. If it appears that the different colored points have different slopes, we might use an interaction term in the model.

(b) Develop a linear model for this study, interpreting all parameters in the context of the problem. You can assume that there is no interaction in the model. Write down your hypothesis to be testing in terms of your model parameters. (You don't have any data to conduct the test; just write down the hypotheses.)

Let's assume there is no interaction for this model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$. In this case, if x_1 is an indicator of being Type A, then

- β_0 is the mean perceived taste of bun Type B when the hot dog has no sodium.
- β_1 is the mean difference in taste between Types A and B, **holding sodium constant**.
- β_2 is the slope: the amount of increase or decrease in mean taste when sodium increases by one unit (probably mg), **holding the bun type constant**.

The relevant null hypothesis is $H_0 : \beta_1 = 0$, and we would compare this to the alternative hypothesis $H_1 : \beta_1 \neq 0$.

(c) Assume now that an interaction term was added and it is statistically significant. How should we interpret this interaction in context?

If there is an interaction, it means the effect of sodium in the hot dog is different depending on whether Type A or B buns are used; to put it a different way, it may be that the type of bun that is preferred differs depending on the amount of sodium in the hot dog (crossed lines interaction).

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.)

Our null hypothesis is $H_0 : \frac{\mu_1}{2} + \frac{\mu_2}{2} - \mu_3 = 0$, or that $\mathbf{A}\boldsymbol{\mu} = 0$ where $\mathbf{A} = [0.5\ 0.5\ -1]_{r\times(p+1)}$. The design matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Also,

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_{(p+1)\times 1}$$

So, $r = 1$ and $p = 2$. We can therefore write the F-statistic as follows:

$$\begin{aligned} F &= \frac{\left(\mathbf{A}\hat{\boldsymbol{\beta}} - h\right)' \left(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\right)^{-1} \left(\mathbf{A}\hat{\boldsymbol{\beta}} - h\right)/r}{\text{SSE}/(n-p-1)} \\ &= \frac{\left(\mathbf{A}\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix}\right)' \left(\mathbf{A}\begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}\mathbf{A}'\right)^{-1} \left(\mathbf{A}\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix}\right)}{\text{SSE}/(12-2-1)} \\ &= \frac{24\left(0.5\hat{\mu}_1 + 0.5\hat{\mu}_2 - \hat{\mu}_3\right)^2}{\text{SSE}} \end{aligned}$$

(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 $n - p - 1$.

Plugging in the provided values, we have

$$F = \frac{24(0.5 \times 5.6 + 0.5 \times 7.9 - 6.1)^2}{12.8} = 0.792 < F_{1,9}(0.05) = 5.12$$

Because the F-statistic falls below the critical value, we fail to reject the null hypothesis. There is not enough evidence to suggest that the average of the first two means is different from the third.