STAT 608, Spring 2022 - Assignment 1
Solutions

1. Question 1, page 38 in our textbook. Notice that to answer these questions correctly, you should be thinking like a statistician and talking about population parameters, not only sample statistics. That is, do some inference in every part, using as much everyday layperson's terminology as possible. For example, part b should not just say "The intercept is (or is not) 10,000." What does an intercept mean in context to someone who is selling movie tickets? Use the discussion board as needed to get all your details correct.

   (a) The 95% confidence interval is $0.9821 \pm 2.119905 \times 0.01443$: $(0.9515, 1.013)$. This interval does include 1, so yes, 1 is one of the plausible values for the population slope. (Note that we did not prove that the population slope is equal to 1, to infinitely many decimal places. Our sample slope is not exactly 1. We only say that we don't have evidence that the slope is not 1.)

   (b) The test statistic for this hypothesis test is $(6,805 - 10,000)/9,929 = -0.322$, for a two-sided p-value of 0.752. Assuming all model assumptions are correct, the test of $H_0 : \beta_0 = 10,000$ vs. $H_a : \beta_0 \neq 10,000$ is not rejected, because our p-value is so large. That is, we don't have evidence that the intercept is not $10,000$, so it is one of the plausible values for the intercept. The interpretation is that when the previous week's gross box office results were \$0, our model predicts that \$10,000 would be a plausible box office result for the current week.

   (c) Using the predict() function in R, we have a prediction interval of (\$359,800, \$439,400). Because this interval does not include the value \$450,000, we can say that we are 95% confident that the current week will not have a gross box office result of \$450,000 if last week's result was \$400,000.

   (d) Because confidence intervals for the intercept and slope contain 0 and 1, respectively, we do not have confidence that the true relationship between last week's and this week's gross box office results are not equal to each other; this rule is plausible, based on our data. (Be sure not to accept the null hypothesis: we do not say that we have evidence that this rule works!)

2. Show that $\mathrm{Var}(Y_i|X_i = x_i) = \mathrm{Var}(\epsilon_i)$ in the simple linear regression model. Don't overthink this; the answer is simple. What did you assume in answering this?

$$\begin{aligned}\mathrm{Var}(Y_i|X_i = x_i) &= \mathrm{Var}(\beta_0 + \beta_1 x_i + \epsilon_i) \\ &= \mathrm{Var}(\epsilon_i)\end{aligned}$$

   This is true because the only thing that varies in the second expression is the error term, once we hold x_i fixed. Suppose we are interested in $x =$ square footage of a house and $y =$ price. The idea is that $\mathrm{Var}(Y_i|x_i)$ is the variability

IN PRICE OF HOUSES THAT ARE ALL THE SAME SIZE, WHILE $\text{VAR}(Y_i)$ IS THE VARIABILITY IN PRICE OF ALL HOUSES IN THE ENTIRE DATASET, OF ALL SIZES. THE SECOND NUMBER IS LARGER THAN THE FIRST IF THE MODEL DOES ANYTHING IN TERMS OF EXPLAINING A RELATIONSHIP BETWEEN THE TWO VARIABLES.

3. Define using only words what the least squares criterion is.

   TO FIND ESTIMATES FOR THE PARAMETERS SLOPE AND INTERCEPT, WE WISH TO MINIMIZE THE SQUARED VERTICAL DIFFERENCES BETWEEN THE POINTS AND THE REGRESSION LINE. THAT IS, WE WANT THE LEAST SQUARED ERROR POSSIBLE.

4. Question 4, page 40 in our textbook, except do:

   (a) Setup:

      i. Write down your design matrix $\mathbf{X}$.

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

      ii. Show, using matrix notation and starting with the principle of least squares, that the least squares estimate of β is

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y} \\ &= \left(\sum x_i^2\right)^{-1} \sum x_i y_i \\ &= \frac{\sum x_i y_i}{\sum x_i^2} \end{aligned}$$

   (b) As in text.

      i. The least squares estimate is unbiased:

$$\begin{aligned} \mathrm{E}(\hat{\beta}|\mathbf{X}) &= \mathrm{E}\left(\frac{\sum x_i y_i}{\sum x_i^2}|\mathbf{X}\right) \\ &= \frac{\sum x_i \mathrm{E}(y_i|x_i)}{\sum x_i^2} \\ &= \frac{\sum x_i \mathrm{E}(\beta x_i|x_i)}{\sum x_i^2} \\ &= \frac{\sum x_i^2 \beta}{\sum x_i^2} \\ &= \beta \end{aligned}$$

ii. Its variance is:

$$\begin{aligned}
\text{Var}\left(\hat{\beta}|\mathbf{X}\right) &= \text{Var}\left(\frac{\sum x_i y_i}{\sum x_i^2}|\mathbf{X}\right) \\
&= \frac{1}{\left(\sum x_i^2\right)^2}\text{Var}\left(\sum x_i y_i|\mathbf{X}\right) \\
&= \frac{1}{\left(\sum x_i^2\right)^2}\text{Var}\left(\sum x_i(\beta x_i + e_i)|\mathbf{X}\right) \\
&= \frac{1}{\left(\sum x_i^2\right)^2}\text{Var}\left(\sum x_i e_i|\mathbf{X}\right) \\
&= \frac{1}{\left(\sum x_i^2\right)^2}\left[\sum_i \text{Var}(x_i e_i|x_i) + 2\sum_{i\neq j}\text{Cov}(x_i e_i, x_j e_j|x_i, x_j)\right] \\
&= \frac{1}{\left(\sum x_i^2\right)^2}\left[\sum x_i^2\sigma^2\right] \\
&= \frac{\sigma^2}{\sum x_i^2}
\end{aligned}$$

iii. Finally, we use a theorem that all linear combinations of normal random variables are also normal. We can see that the least squares estimate of β above is a linear combination of the y_i's, which is itself a linear combination of the errors, since $y_i = \beta x_i + e_i$, and the x_i are held constant. Therefore the distribution of $\hat{\beta}$ is normal as long as our assumption that the errors are normal holds.

5. Show that the least-squares criterion applied to the "intercept-only" model, i.e.,

$$y_i = \beta_0 + \epsilon_i,\ i = 1, 2, \ldots, n$$

results in the least squares estimator of β_0: $\hat{\beta}_0 = \bar{y}$ by following these steps:

(a) Write down your design matrix $\mathbf{X}$. (It won't be the same as any we've used in class.) Double check: does $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ give the set of equations listed above? Notice this model has no predictor variable.

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

(b) Use the previously derived formula $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ to get the least squares estimator.

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\
&= n^{-1}\sum y_i \\
&= \bar{y}
\end{aligned}
$$

6. Prove that $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.

$$
\begin{aligned}
\text{Cov}(aX, bY) &= \text{E}\left[(aX - a\mu_X)(bY - b\mu_Y)\right] \\
&= \text{E}\left[ab(X - \mu_X)(Y - \mu_Y)\right] \\
&= ab\text{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= ab\text{Cov}(X, Y)
\end{aligned}
$$

7. Question 7, page 42 in our textbook.

   The confidence interval only gives an interval of plausible values for the mean at each x value. We can be rather precise about the location of the mean, especially for large sample sizes, but we need much wider intervals (and we call them prediction intervals) to determine where 95% of the individual values are.

8. Using $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, finish our algebra from class and show that $\hat{\beta}_0 = \bar{y} - \frac{SXY}{SXX}\bar{x}$ for the simple linear regression case.

   It is easiest to begin with the general task of deriving $\hat{\beta}_0$ and $\hat{\beta}_1$, which is to minimize

$$
\sum_i (y_i - \hat{y}_i)^2 = \sum_i \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right)^2
$$

   Taking the partial derivative with respect to $\hat{\beta}_0$ and setting equal to zero, we have

$$
2\sum_i \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right)(-1) = 0
$$

$$
\sum_i y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_i x_i = 0
$$

$$
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
$$

   Now taking the partial derivative with respect to $\hat{\beta}_1$ and setting equal to zero, we have

$$
2\sum_i \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right)(-x_i) = 0
$$

$$
\hat{\beta}_1 = \frac{\sum_i (y_i - \bar{y})\, x_i}{\sum_i (x_i - \bar{x})\, x_i} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{SXY}{SXX}
$$

   Thus, we have the claimed result.

9. Show that for the usual regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the usual regression assumptions from question 4 apply, $\text{Var}(\mathbf{a}'\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2 \mathbf{a}' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{a}$, where $\mathbf{a}$ is a constant vector.

   USING SLIDES 38 AND 44 FROM THE CHAPTER 2 NOTES:

$$\begin{aligned}
\textsc{Var}\left(\mathbf{a}'\hat{\boldsymbol{\beta}}|\mathbf{X}\right) &= \mathbf{a}'\textsc{Var}\left(\hat{\boldsymbol{\beta}}|\mathbf{X}\right)\mathbf{a} \\
&= \mathbf{a}'\sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{a} \\
&= \sigma^2 \mathbf{a}' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{a}
\end{aligned}$$