Assignment 4 - Matrix Notation

your name goes here

Due: Wednesday, February 14, 2018, noon, to Sakai

Summary

The tasks in this homework assignment focus first on making multiple intervals with known The tasks in this homework provide practice in using matrices to write down the linear model information we've derived through the first few weeks of class. Additionally, the last problem allows consideration of correlated statistics and lack of independence in testing.

Assignment

1. Show how the following expressions are written in terms of matrices. Assume $i = 1, \ldots, 4$. This question is simply asking you to write down the new notation (writing it out by hand with pencil is fine). Assume X_i is the i^{th} observation of the single $\underline{X}$ variable (without the column of 1s yet appended). [If you are curious about the mathematical properties of the residuals, fitted values, and explanatory variables, see pages 23-24 in *Applied Linear Statistical Models*.]

   (a) $Y_i - \hat{Y}_i = e_i$

   (b) $\sum X_i e_i = 0$

2. Consider the Wii Mario Kart data. Recall that we are still only doing simple linear regression (i.e., one explanatory variable). Use auction price as the response variable and number of bids as the explanatory variable.

```
require(openintro)
data("marioKart")
marioKart <- marioKart %>% mutate(aucPr = totalPr - shipPr)
```

   Using matrix methods only (no `lm` function), do the following in R:

   (a) Find and print $\mathbf{Y}'\mathbf{Y}$.

   (b) Find and print $\mathbf{X}'\mathbf{X}$.

   (c) Find and print $\mathbf{X}'\mathbf{Y}$.

   (d) Find and print $(\mathbf{X}'\mathbf{X})^{-1}$.

   (e) Find and print $\mathbf{b}$.

   (f) Find and print $\hat{\mathbf{Y}}$. Only print the first 5 values.

   (g) Find and print SSE.

   (h) Find and print $s^2(\mathbf{b})$.

   (i) Find and print $s^2(pred)$ when $X_h = 12$.

3. Check your answers to $\mathbf{b}$, $\hat{\mathbf{Y}}$, and the diagonal of $s^2(\mathbf{b})$ using the `lm` function.

4. In class, we have discussed the relationship between the sample intercept and the sample slope.

   (a) Explain what it means for b_0 and b_1 to have a sampling distribution.

   (b) What is the intuition that says b_0 and b_1 are correlated? If one made a plot with b_0 on the x-axis and b_1 on the y-axis, what would a single point represent (that is, what is the observational unit of the scatterplot)? Do you expect b_0 and b_1 to be positively or negatively correlated? Explain.

   (c) For the `marioKart` data, find the correlation between b_0 and b_1. [Recall that the correlation is the covariance divided by the square root of the product of the variances.]

$$cor(b_0, b_1) = \frac{cov(b_0, b_1)}{\sqrt{var(b_0)var(b_1)}}$$

   (d) Consider 90% confidence intervals for β_0 and β_1, computed separately (no simultaneous inference adjustments). Do the intervals imply that:

   - in 10% of samples the confidence interval for β_0 will be incorrect?
   - in 20% of samples the confidence interval for at least one of β_0 or β_1 will be incorrect?
   - Discuss.