BIOS 6611 Homework 8

Due Thursday, November 7, 2019 by 11:59 pm to Canvas Assignment Basket

Submit your ***complete*** SAS code as an appendix to your answers and include only relevant SAS output with your answers to each part.

1. From lectures 15-16, show that $SS_{Total} = SS_{Model} + SS_{Error}$ using the notes below.

   Consider $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. We can take material from lecture and rewrite it in terms of the residual ($\hat{e}_i$):

   $$\frac{\partial}{\partial \beta_0} SS_{Error} = \frac{\partial}{\partial \beta_0}\left(\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2\right) = 0 \rightarrow \sum_{i=1}^{n} -2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = -2\sum_{i=1}^{n}\hat{e}_i = 0$$

   $$\frac{\partial}{\partial \beta_1} SS_{Error} = \frac{\partial}{\partial \beta_1}\left(\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2\right) = 0 \rightarrow \sum_{i=1}^{n} -2X_i\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = -2\sum_{i=1}^{n} X_i\hat{e}_i = 0$$

   Note 1: $\sum_{i=1}^{n}\hat{e}_i = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right) = 0$.
   Note 2: $\sum_{i=1}^{n} X_i\hat{e}_i = 0$
   Hint:

   $$SS_{Total} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left((Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 + 2\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)(\hat{Y}_i - \bar{Y})$$

2. The following table presents the gender (0=Female; 1=Male), weight (kg), age (years), and plasma levels of total cholesterol (mg/100ml) for a sample of 7 patients suffering from hyperlipoproteinemia. The investigator is interested in examining the effect of weight on cholesterol levels.

Patient	*Gender*	*Cholesterol (mg/100ml)*	*Weight (kg)*	*Age (yr)*
1	0	254	57	23
2	1	402	79	57
3	0	288	63	28
4	1	354	84	46
5	0	220	30	34
6	1	451	76	57
7	0	405	65	52

For the data use the following:

```
DATA hw1;
INPUT id gender chol wtkg age;
chol2 = chol*chol;
wt2 = wtkg*wtkg;
cholwt = chol*wtkg;

LABEL wtkg = 'Weight (kg)'
        chol = 'Cholesterol (mg/100mL)'
        ;

DATALINES;
1  0     254   57    23
2  1     402   79    57
3  0     288   63    28
4  1     354   84    46
5  0     220   30    34
6  1     451   76    57
7  0     405   65    52
;
RUN;
```

A. Performing the regression of cholesterol (Y) on weight (X) in SAS, provide an ANOVA table and a parameter estimate table (only those tables, do NOT give all of the SAS output).

```
PROC REG;
      MODEL chol=wtkg/CLB;
RUN;
```

B. Write down the least-squares regression equation that describes the relationship between cholesterol (dependent variable) and weight (independent variable).

C. Inference about the intercept:

   (I) What is the estimated intercept, and how would you interpret it?

   (II) Obtain a 95% confidence interval for the intercept and give its interpretation.

   (III) Test the hypothesis that the true intercept is 0.

   (IV) Is it scientifically interesting to test whether or not this intercept equals zero? Why or why not?

D. Inference about the slope:

   (I) What is the estimated slope, and how would you interpret it?

   (II) Obtain a 95% confidence interval for the slope and give its interpretation.

   (III) Test the hypothesis that the true slope is zero.

E. Write a brief, but complete (i.e., include the point estimate, p-value, 95% CI, and summary/decision), summary of the effect of weight on cholesterol.

F. Use SAS to produce a scatterplot of cholesterol and weight along with the least squares regression line, the 95% confidence interval, and the 95% prediction interval. (Note: Depending on SAS version, PROC REG may also provide this type of plot.)

```
PROC GPLOT data=hw1;
   PLOT chol*wtkg=1 chol*wtkg=2 / OVERLAY VAXIS=axis1;
   SYMBOL1 INTERPOL=rlcli COLOR=black VALUE=dot;
   SYMBOL2 INTERPOL=rlclm COLOR=black VALUE=dot;
   AXIS1 LABEL = (FONT=ARIAL HEIGHT= 1.5 ANGLE=90 POSITION=center );
RUN;
```

3. Read the paper by Moser and Stevens, “Homogeneity of Variance in the Two-Sample Means Test”, which is located on Canvas in the Paper Repository.

   In one paragraph, summarize the problem they studied, the methods they used, the results they obtained, their recommendations for statistical practice, and how you will apply the recommendations in the future.