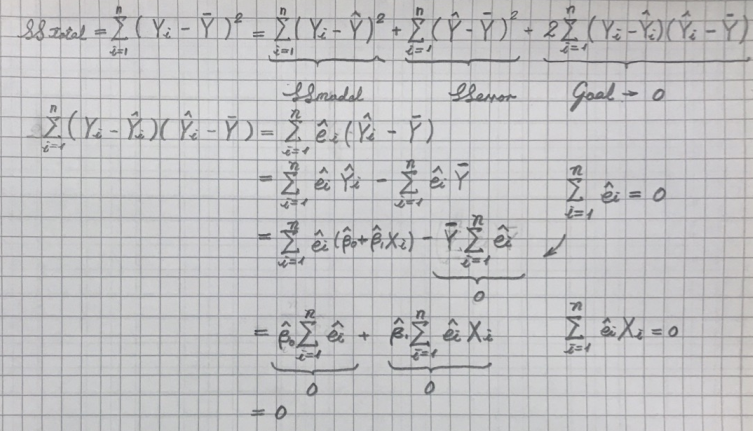
BIOS 6611 Homework 8

## Question1. From lectures 15-16, show that SSTotal = SSModel + SSError



SSTotal = SSModel + SSError

## Question 2A. Performing the regression of cholesterol (Y) on weight (X) in SAS, provide an ANOVA table and a parameter estimate table

**Table 1: ANOVA results for the effect of weight on cholesterol levels**

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 27382 | 27382 | 7.74 | 0.0388 |
| **Error** | 5 | 17699 | 3539.79126 |  |  |
| **Corrected Total** | 6 | 45081 |  |  |  |

**Table 2: parameter estimate for the effect of weight on cholesterol levels**

| **Parameter Estimates** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t  Value** | **Pr > |t|** | **95% Confidence Limits** | |
| **Intercept** | Intercept | 1 | 97.395 | 89.782 | 1.08 | 0.328 | -133.396 | 328.186 |
| **wtkg** | Weight (kg) | 1 | 3.727 | 1.340 | 2.78 | 0.039 | 0.282 | 7.172 |

## Question 2B. Write down the least-squares regression equation that describes the relationship between cholesterol (dependent variable) and weight (independent variable).

**Cholesterol (mg/100ml) = 97.395 + 3.727 Weight (kg)**

## Question 2C. Inference about the intercept:

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

Intercept = 97.395; the intercept is the theoretical value of cholesterol level when the **Weight (kg)** is near to 0.

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

The 95% Confidence Interval for intercept : (-133.396, 328.186)

There will be 95% of the chance that the 95% confidence the range between -133.396 and 328.186mg/mL of **Cholesterol (mg/100mL)**) covers the real estimated value of **Cholesterol (mg/100mL)** when the **Weight (kg)** is near to zero.

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

Because the p-value is 0.328, larger than the critical value 0.05, we cannot reject the null hypothesis H0: 

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

This hypothesis is rarely of interest in most of the studies. We are not interested in knowing what happens at **Weight (kg)** near to 0kg, and we rarely choose the values of **Weight (kg)**  near 0kg.

Because we cannot reject the null hypothesis H0: it might be appropriate to remove the constant from the model, provided that previous experience or a relevant theory suggests that the line may go through the origin and provided that observations are taken around the origin to improve the estimate of Forcing the fitted line through the original point merely because the null hypothesis cannot be rejected may give a spurious appearance to the regression line.

## D. Inference about the slope:

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

Slope = 3.727; on average, the **Cholesterol (mg/100mL)** will increase 3.727mg/L for every 1kg increase of **Weight (kg).**

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

We are 95% confident that the **Cholesterol (mg/100mL)**increases by between 0.282 and 7.172 mg/mL for every 1kg increase in **Weight (kg).**

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

Because the p-value is 0.039, less than the critical value 0.05. We can reject the null hypothesis H0: the slope = 0; this implies that a straight-line model with **Weight (kg)** is better than a model that does not include **Weight (kg)** at all, although it may well represent only a linear approximation to a truly nonlinear relationship.

## 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.

There is a significant increase in the **Cholesterol (mg/100mL)** with increasing levels of **Weight (kg)** (F1, 5, 0.05=7.74, p<0.0388). On average, the **Cholesterol (mg/100mL)** decreases by 3.727mg/mL (95% confidence interval: between 0.282 and 7.172 mg/mL) for every 1kg increase in **Weight (kg)**.

## F. Use SAS to produce a scatterplot of cholesterol and weight along with the least square regression line, the 95% confidence interval, and the 95% prediction interval.

## 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.

The problem: is the current practice of preliminary variance test appropriate and is this emphasis on variance homogeneity warranted? The methods: Hypothesis is tested through F test on the ratio of sample variance is calculated; the results they obtained: the current practice of variance tests is not appropriate, and the variance of homogeneity is not warranted. Their recommendations: when sample sizes are unequal and the variance ratio is known to be near one, then the t-test is appropriate; when the sample sizes are equal, either t-test or the SWS test is appropriate; and when the sample sizes are unequal an the variance ratio is unknown or differ from one, SWS test is appropriate. how you will apply the recommendations in the future: first, check the variance ratio; and SWS test is always appropriate in all cases for unknown variance other than testing the assumption.