James Young

Problem 1

**1. State the model**

We will use the simple linear regression model.

,

Where and represent the values of independent and dependent variables, respectively. represents the intercept, represents the slope of the regression line, and , is variables coming from a normal distribution with mean zero and variance 2.

**2. State the four assumptions related to the model**

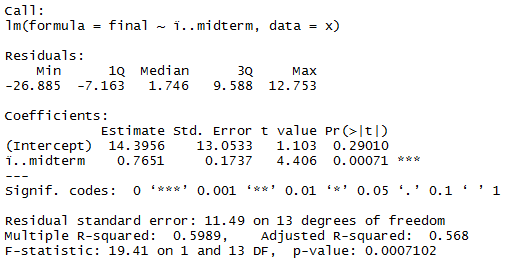
1. The linear regression is appropriate

2. The error terms are independent

3. The error terms are normally distributed

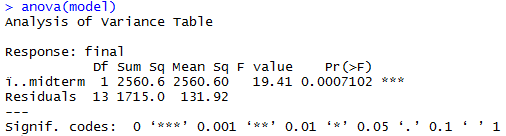
4. The error terms have common variance

**3. Fit the regression using R and provide the snip of the coefficients**



**4. Fit the estimated regression function (use the ‘hat’ function) using your results from step 3**

= 14.3956 + 0.7651

**5. Use the anova (F test, page 338) to test if β1 is equal to zero. State the hypothesis, calculate the test statistic, and provide the decision rule, and conclusion. (Use the code: anova(model) to obtain the numerator and denominator for F statistic )**

**Hypothesis:**

H0: = 0

H1: 0

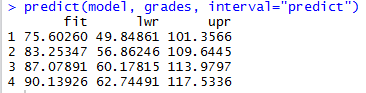
F = = 2560.6/131.92 = 19.41

There is 1 df in the numerator and 13 df

In the denominator, so critical F value is 4.67. Calculated F value is 19.41. 19.41>4.67, therefore we reject H0 and conclude ……

**6. Use the code below and provide snip of the prediction intervals for when midterm scores are 80,90,95,99. State in one sentence why this shows whether there is a need for final exam.**

There is need for a final exam because of the relatively low accuracy of the predictive model.



**7. Plot the fitted regression line with the scatterplot and comment in 1-2 sentences how strong the linear relationship**

The linear relationship is visible but there appears to be noticeable difference from the fitted line.

**A close up of a map

Description generated with very high confidence**

**8. Modify the predict function above to predict only when midterm=82 and provide the snip. In one sentence interpret what the predicted final exam score tells you about the model fit.**



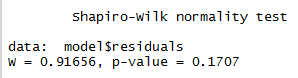
The model fit is not very precise.

**A screenshot of a cell phone

Description generated with very high confidence9. Provide the only the qqplot with the qqline. Examine the residuals distribution from the qqplot and comment on the distribution**

Distribution appears to be normal.

**10. Perform the Shapiro-Wilks Test to test the normality. You need hypothesis test, R snip with the result and the decision rule and conclusion**

Shapiro-Wilk Test (with )

H0 : All residuals came from normal distribution

H1: Any residuals did not come from a normal distribution

Decision: Calculated p-value = 0.1707. 0.1707 > 0.01 therefore

we fail to reject H0 and conclude the residuals are normal.

**11. Do this by hand using the example on page 340-341. Use the appropriate equations and symbols using the equation editor.**

**==**0.7651 0.3750

We can state with 95 % confidence that true slope lies between 0.3901 to 1.14

**12. Finally comment on the fit of the model in regard to the data**

The model fulfills the assumptions and shows that for each point increase in midterm score, the final increases bu 0.7651.

**Problem 2**

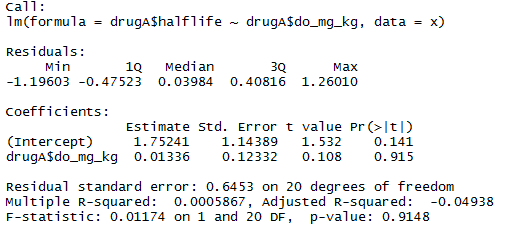
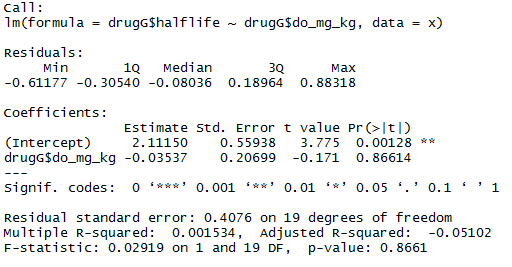
1. State the model. NOTE: since you stated the assumption in #3 you can omit this step here

We will use the simple linear regression model.

,

Where and represent the values of independent and dependent variables, respectively. represents the intercept, represents the slope of the regression line, and , is variables coming from a normal distribution with mean zero and variance 2.

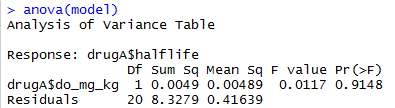
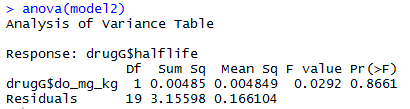
2. Fit the BOTH regressions for drug A and G using R and provide the snip of the coefficients side by side and label both.



3. Fit BOTH the estimated regression function (use the ‘hat’ function) using your results from step 2 Example: Drug A: put fitted regression function A here Drug G: put fitted regression function G here

Drug A : Drug G:

4. Use the R to obtain the anova table to test if β1 is equal to zero. Place the anova table snips side-by-side and label them(see below). State the hypothesis, use the p value (not F test), and provide the decision rule, and conclusion FOR EACH.



**Drug A Drug G**

**H0** : = 0 **H0** : = 0

**H1** : 0 **H1** : 0

Calculated p-val = 0.91. 0.91>0.05, therefore Calculated p-val = 0.87. 0.087>0.05, therefore

we fail to reject H0 and conclude dose is not an we fail to reject H0 and conclude dose is not an

accurate predictor of half-life. accurate predictor of half-life.

**5. Plot the fitted regression lines with the scatterplot and comment in 1-2 sentences how about each fitted line and comment on if the lines are parallel. Why is this important?**

A close up of a map

Description generated with very high confidenceThe lines are not parallel, showing a difference in slope. It

Shows the 2 drugs are metabolized differently.

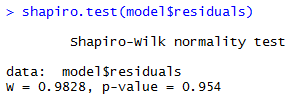
A close up of a map

Description generated with very high confidence6. Provide the only the qqplot with the qqline for A and G side by side. Examine the residuals distribution from the qqplot and comment on the distribution FOR EACH A AND G

A close up of a map

Description generated with very high confidence

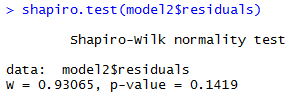
**Drug A Drug G**

7. Perform the Shapiro-Wilks Test to test the normality FOR EACH. You need hypothesis test, R snip with the result and the decision rule and conclusion

Shapiro-Wilk Test (with )

H0 : All residuals came from normal distribution

H1: Any residuals did not come from a normal distribution

Decision (DrugA): Calculated p-value = 0.954. 0.954 > 0.01

therefore we fail to reject H0 and conclude the residuals are normal.

Decision (DrugG): Calculate p-value = 0.0.1419. 0.1419 > 0.01

therefore we fail to reject H0 and conclude the residuals are normal.

**8. Do this by hand FOR EACH using the example on page 340-341. Use the appropriate equations and symbols using the equation editor.**

=(-0.244, 0.271)

We are 95% confident that the true slope for drug A is found between -0.244 and 0.271

=(-0.4685, 0.3978)

We are 95% confident that the true slope for drug G is found between -0.4685 and 0.3978)

9. Finally comment on the fit of each model in regard to the half-life.

Looking at the models, Drug A half-life increased with dose and Drug B half-life decreased with dose.

