Math 10 – Homework 11 **Answers**

1. A manager is concerned that overtime (measured in hours) is contributing to more sickness (measured in sick days) among the employees. Data records for 10 employees were sampled with the following results:

a) Find the least square line where Sick Days is dependent on Overtime. Interpret the slope.

$\hat{Y} = 0.5369 + 0.0621X$

Each hour of overtime increases sick days by 0.0621.

b) Test the hypothesis that the regression model is significant (α = .10)

Ho: Sick days and overtime are not correlated
Ha: Sick days and overtime are correlated

Model: Simple Linear Regression ANOVA F = MSRegression/MSError

Pvalue=.0047<.10 -→ Reject Ho

Sick days and overtime are positively correlated.

c) Find and interpret the r^2, coefficient of determination. (Blank Box)

r^2 = 80.6944/123.60 = .655 65.5% of the variability of sick days can be explained by overtime.

d) Find he estimate of standard deviation of the residual error. (Blank Box)

$s_e = \sqrt{5.3632.} = 2.32$

e) What would your prediction of sick days be for an employee who works 100 hours overtime.

6.742 sick days

f) Analyze the residuals and determine which pair of data is the most unusual.

Observation 4 (39,7) has the highest residual error (+4)

g) Explain why this model would not be appropriate for an employee who works 500 hours overtime.

This would be extrapolation (choosing a value of X outside the range of data.) and leads to erroneous results.

Math 10 **Hypothesis Testing – Regression** Name_____ (25)

16 student volunteers drank a randomly assigned number of cans of beer. Thirty minutes later a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood. **Data and computer output attached on next page.**

a) Find the least square line where BAC is dependent on Beers consumed. Interpret the slope.

$\hat{y} = -.0127 + 0.018x$ Each beer increases BAC by .018

b) Find and Interpret the r-squared statistic.

89% of the variability of BAC is explained by variability of Beers consumed.

c) Test the hypothesis that the beers consumed and BAC are correlated ($\alpha = .05$)

H_0: Beers and BAC are uncorrelated $H_0: \beta_1 = 0$

H_a: Beers and BAC are correlated. $H_a: \beta_1 \neq 0$

$\alpha = .05$ $F = \frac{MS\,Reg}{MSE}$

p-value = .00000297 . Reject the

Beer and BAC are positively correlated

d) Find a 95% Confidence Interval for the mean BAC for a student who consumes 5 beers.

$x = 5$ $\hat{y} = -.0127 + .018(5) = .0771$

(.066, .088) ← confidence interval for mean BAC when 5 Beers are consumed.

e) Would this model be appropriate for a student who consumed 20 beers? Explain.

Not reasonable because 20 is outside range of data (extrapolation)

f) Joe claims that he can still legally drive after consuming 5 beers, The legal BAC limit is 0.08. Find a 95% Prediction interval for Joe's BAC. Do you think Joe can legally drive?

$\hat{y} = 0.0771$

(.032, .122) I dunno!

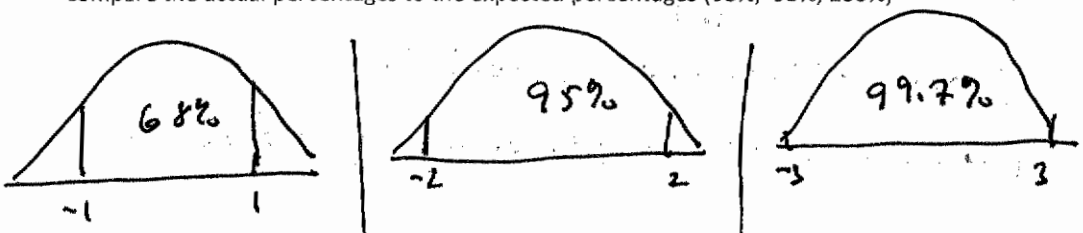
g) Residual Analysis

1. We would expect the residuals to be random, about half would be positive and half would be negative. Check the actual residuals and compare the actual percentages to the expected percentages.

+	8
−	8

perfect!

2. The assumption for regression is that the residuals have a Normal Distribution. This means about 68% of the residuals would have a Z-score between -1 and 1, 95% of the residuals would have a Z-score between -2 and 2 and all the residuals would have a Z-score between -3 and 3. The Column labeled "Standardized Residual" is the Z-score for each residual. Check to see what percentage of the data has Z-scores in each of these three intervals and compare the actual percentages to the expected percentages (68%, 95%, 100%)



$\frac{10}{16} = 62.5\%$

$\frac{15}{16} = 93.7\%$

$\frac{16}{16} = 100\%$

little low

about right

3. The following regression analysis was used to test Poverty (percentage living below the poverty line) as a predictor for Dropout (High School Dropout Percentage. Five items have been blanked out been can be calculated based on other information in the output.

a. Fill in the missing information from the output
   i. r^2 **r^2 = 67.45/283.62 = 0.238**
   ii. r **r = sqrt(0.238) = 0.488**
   iii. Std. Error **s_e = sqrt(4.50) = 2.12**
   iv. F Test Statistic **F = 67.45/4.50 = 14.99**
   v. Predicted Value for Poverty=15 **$Yhat_{15}$ = 6.212+0.291(15)=10.577**

b. Write out the regression equation. **Yhat = 6.212+0.291X**

c. Conduct the Hypothesis Test that Poverty and HSDropout are correlated with α =.01 (Critical Value for F is 7.19 (α =.01, DFnum=1,DF=48)).

   Ho: Poverty and HS Dropout are not correlated.
   Ha: Poverty and HS Dropout are correlated.
   Model: Simple Linear Regression ANOVA F = MSRegression/MSError

   14.99>7.19 Reject Ho
   Conclusion: Poverty and HS Dropout are positively correlated.

d. What percentage of the variability of High School Dropout Rates can be explained by Poverty?

   r^2 = 0.238 or 23.8%

e. North Dakota has a Poverty Rate of 11.9 percent and a HS Dropout Rate of 4.6 percent.

   i. Calculate the predicted HS Dropout Rate for North Dakota from the regression equation.

      $Yhat_{11.9}$ = 6.212+0.291(11.9)=9.675

   ii. The Standard Error (from part a-iii) is the standard deviation with respect to the regression line. Calculate the Z-score for the actual North Dakota HS Dropout Rate of 4.6 (Subtract the predicted value and divide by the Standard Error). Do you think that the North Dakota HS Dropout Rate is unusual? Explain

      Z = (4.6-9.675)/2.12 = -2.39

      North Dakota's Actual Dropout rate is unusually low.