Math 10 – Homework 11 Name________________________________

Q1 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:

Partial Computer Output are attached on next page

a) Find the regression equation where Sick Days is dependent on Overtime. Interpret the slope.

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

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

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

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

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

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

Overtime	20	25	35	39	43	55	67	113	135	155
Sick Days	0	0	2	7	3	5	4	11	7	9

Regression Analysis

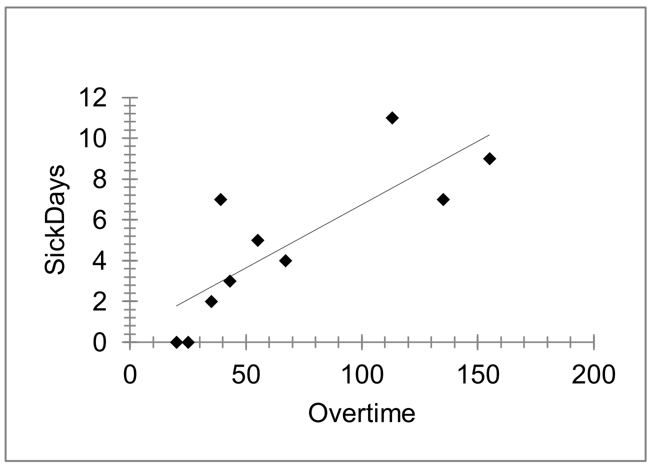
r^2		n	10
r		k	1
Std. Error		Dep. Var.	**SickDays**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	80.6944	1	80.6944	15.05	.0047
Residual	42.9056	8	5.3632		
Total	123.6000	9			

Regression output					confidence interval	
variables	*coefficients*	*std. error*	*t (df=8)*	*p-value*	*95% lower*	*95% upper*
Intercept	0.5369	1.3207	0.407	.6950	-2.5086	3.5824
Slope	0.0621	0.0160	3.879	.0047	0.0252	0.0989

Observation	*SickDays*	*Predicted*	*Residual*
1	0.0	1.8	-1.8
2	0.0	2.1	-2.1
3	2.0	2.7	-0.7
4	7.0	3.0	4.0
5	3.0	3.2	-0.2
6	5.0	3.9	1.1
7	4.0	4.7	-0.7
8	11.0	7.5	3.5
9	7.0	8.9	-1.9
10	9.0	10.2	-1.2



Predicted values for: SickDays

		95% Confidence Intervals		*95% Prediction Intervals*		
Overtime	*Predicted*	*lower*	*upper*	*lower*	*upper*	*Leverage*
100	6.742	4.696	8.788	1.023	12.461	0.147
500	31.564	15.563	47.564	14.696	48.432	8.977

Q2 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.

b) Find and Interpret the r-squared statistic.

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

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

e) Would this model be appropriate for a student who consumed 30 beers? Explain.

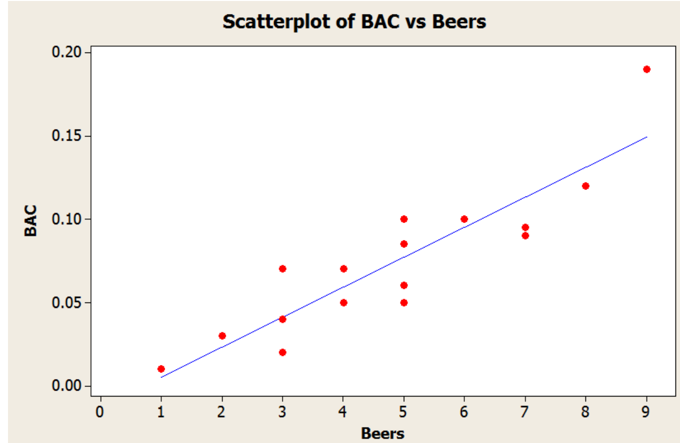
f) Joe claims that he can still legally drive after consuming 5 beers, The legal BAC limit is 0.08. Find a 95% Prediction interval for Joe's BAC. Do you think Joe can legally drive?

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.

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 "Std Resid" or "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%)

Regression Analysis: BAC versus Beers



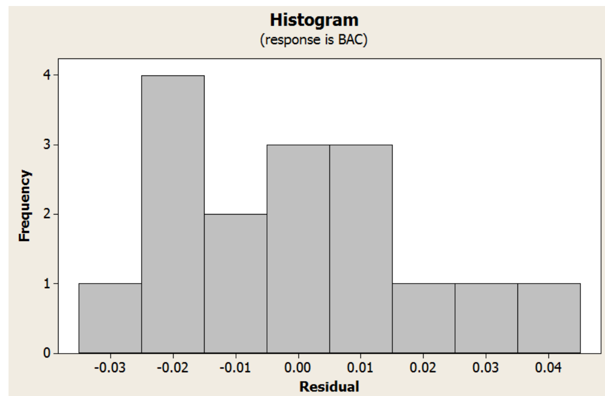
The regression equation is
BAC = - 0.0127 + 0.0180 Beers

Predictor	Coef	SE Coef	T	P
Constant	-0.01270	0.01264	-1.00	0.332
Beers	0.017964	0.002402	7.48	0.000

S = 0.0204410   R-Sq = 80.0%   R-Sq(adj) = 78.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.023375	0.023375	55.94	0.000
Residual Error	14	0.005850	0.000418		
Total	15	0.029225			



Obs	Beers	BAC	Fit	SE Fit	Residual	St Resid
1	5.00	0.10000	0.07712	0.00513	0.02288	1.16
2	2.00	0.03000	0.02323	0.00847	0.00677	0.36
3	9.00	0.19000	0.14897	0.01128	0.04103	2.41R
4	8.00	0.12000	0.13101	0.00920	-0.01101	-0.60
5	3.00	0.04000	0.04119	0.00671	-0.00119	-0.06
6	7.00	0.09500	0.11305	0.00733	-0.01805	-0.95
7	3.00	0.07000	0.04119	0.00671	0.02881	1.49
8	5.00	0.06000	0.07712	0.00513	-0.01712	-0.87
9	3.00	0.02000	0.04119	0.00671	-0.02119	-1.10
10	5.00	0.05000	0.07712	0.00513	-0.02712	-1.37
11	4.00	0.07000	0.05915	0.00547	0.01085	0.55
12	6.00	0.10000	0.09508	0.00585	0.00492	0.25
13	5.00	0.08500	0.07712	0.00513	0.00788	0.40
14	7.00	0.09000	0.11305	0.00733	-0.02305	-1.21
15	1.00	0.01000	0.00526	0.01049	0.00474	0.27
16	4.00	0.05000	0.05915	0.00547	-0.00915	-0.46

R denotes an observation with a large standardized residual.

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	0.07712	0.00513	(0.06612, 0.08812)	(0.03192, 0.12232)
2	0.52621	0.06071	(0.39601, 0.65642)	(0.38882, 0.66360)XX

XX denotes a point that is an extreme outlier in the predictors.

Values of Predictors for New Observations

New Obs	Beers
1	5.0
2	30.0

Q3 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

r^2		n	50
r		k	1
Std. Error		Dep. Var.	**HSDropouts**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	67.45	1	67.45		
Residual	216.18	48	4.50		
Total	283.62	49			

Regression output

variables	*coefficients*	*std. error*
Intercept	6.212	1.086
Poverty	0.291	0.075

Predicted values for: HSDropouts

		95% Confidence Intervals		*95% Prediction Intervals*	
Poverty	*Predicted*	*lower*	*upper*	*lower*	*upper*
10	9.117	8.273	9.961	4.767	13.466
15		9.944	11.195	6.257	14.882

i. r^2
ii. r
iii. Std. Error
iv. F Test Statistic
v. Predicted Value for Poverty = 15

b. Write out the regression equation.

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)).

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

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.

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.