

15.2.13: Chapter 14 Homework

1. A real estate agent uses a simple regression model to estimate the value of a home based on square size in which Y is the value of the home in dollars and X is the size in total square feet. The regression equation is $\hat{Y} = 253000 + 438X$.
 - a. Interpret the slope using the units of the problem.
 - b. Estimate the value of a home with 1347 square feet.
 - c. Will the correlation coefficient be positive or negative in this problem. Explain.
2. A car dealer uses a simple regression model to estimate the value of a used 2013 Toyota Prius basing the value on the car's mileage. Y is the value of the car in dollars and X is the total miles on the odometer. The regression equation is $\hat{Y} = 28000 - 0.048X$.
 - a. Interpret the slope by using the units of the problem.
 - b. Estimate the value of a car with an odometer reading of 143,282 miles.
 - c. Why would this model not work for a Prius that was driven 600,000 miles?
3. A manager is concerned that overtime (measured in hours) is contributing to more sickness (measured in sick days) among the employees. Data records for 20 employees were sampled with the MINITAB results shown at the end of the questions.
 - a. Identify the explanatory (Independent) Variable – include units.
 - b. Identify the response (dependent) variable – include units.
 - c. Find the least square line where Sick Days is dependent on Overtime. Interpret the slope using the appropriate units.
 - d. Test the hypothesis that the regression model is significant ($\alpha = .10$). Show all steps. Fill in the missing values on the ANOVA table.
 - e. Find and interpret the r^2 , coefficient of determination. (Blank Line)
 - f. Find the estimate of standard deviation of the residual error. (Blank Line)
 - g. Identify any residual that is more than two standard deviations from the regression line.

Regression Equation
Sickdays = 2.707 + 0.0425 Overtime

Predictor	Coef	SE Coef	T	P
Constant	0.537	1.321	0.41	0.695
Overtime	0.06205	0.01600	3.88	0.005

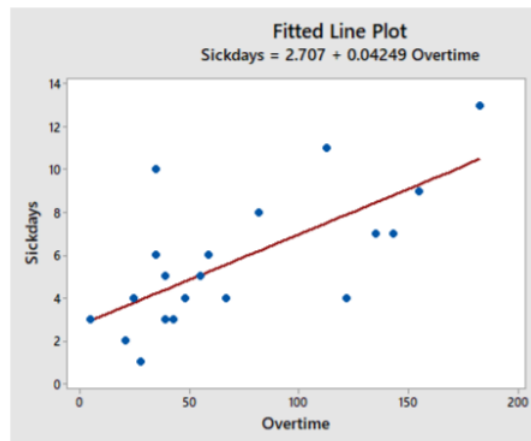
S = _____ R-Sq = _____

Analysis of Variance

Source	DF	SS	MS	F	P-value
Regression	_____	91.276	_____	_____	0.001
Residual Error	_____	98.474	_____		
Total	_____	189.750			

Fits and Diagnostics for All Observations

Row	Overtime	Sickdays	Fit	Resid	Std Resid
1	5	3	2.920	0.080	0.04
2	21	2	3.600	-1.600	-0.72
3	25	4	3.770	0.230	0.10
4	28	1	3.897	-2.897	-1.30
5	35	10	4.195	5.805	2.58
6	35	6	4.195	1.805	0.80
7	39	5	4.365	0.635	0.28
8	39	3	4.365	-1.365	-0.61
9	43	3	4.535	-1.535	-0.68
10	48	4	4.747	-0.747	-0.33
11	55	5	5.045	-0.045	-0.02
12	59	6	5.215	0.785	0.35
13	67	4	5.555	-1.555	-0.68
14	82	8	6.192	1.808	0.79
15	113	11	7.509	3.491	1.56
16	122	4	7.892	-3.892	-1.75
17	135	7	8.444	-1.444	-0.66
18	143	7	8.784	-1.784	-0.83
19	155	9	9.294	-0.294	-0.14
20	183	13	10.484	2.516	1.28



4. 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 are given below the questions.
 - 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 ($\alpha = .05$)
 - 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 20 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
 - i. 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.
 - ii. 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%).

Data for Exercise 4 Regression Analysis: BAC versus Beers

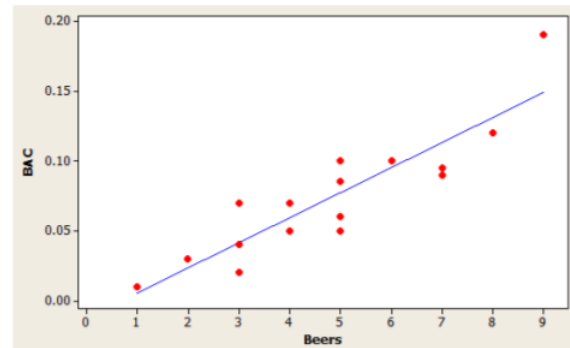
The regression equation is
 $BAC = -0.0127 + 0.0180 \text{ Beers}$

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
Regression	1	0.023375	0.023375	55.94
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.

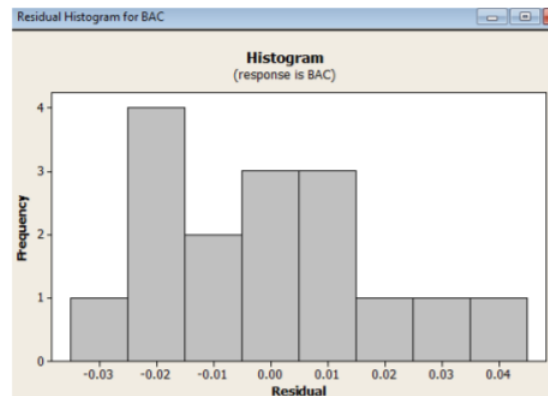
Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	0.07712	0.00513	(0.06612, 0.08812)	(0.03192, 0.12232)
2	0.34657	0.03683	(0.26758, 0.42557)	(0.25623, 0.43692)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	20.0



5. The following regression analysis was used to test Poverty (percentage of population living below the poverty line) as a predictor for Dropout (High School Dropout Percentage).

a. Five items have been blanked out been; find these missing which can be calculated based on other information in the output.

- r^2
- r
- Standard Error of the Residuals

- iv. F Test Statistic
- v. Predicted Value for Poverty = 15

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

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

- b. Write out the regression equation.
- c. Conduct the Hypothesis Test that Poverty and HSDropout are correlated with $\alpha = .01$ (Critical Value for F is 7.19 ($\alpha = .01$, DF-num=1,DF-den=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.

For the studies in questions 6 to 8:

- a. Identify the explanatory variable.
 - b. Identify the response variable
 - c. Choose the appropriate model from among these three:
 - i. Chi-square test of independence
 - ii. One factor ANOVA
 - iii. Simple Linear Regression
6. A golf course designer was studying types of grass to be used in a region that was susceptible to droughts. The designer studied 5 types of grass: Bent Grass, Fescue, Rye Grass, Bermuda Grass and Paspalum. Ten samples were taken of each grass and watered to keep the grass in prime condition for a month. For each sample, the daily water usage was calculated in liters per square meter. The designer wanted to know if there was a significant difference in mean water usage due to grass type.
7. A school psychologist believes students who have more homework will sleep less. 200 students participated in a study. For each of 14 consecutive days, students were asked to count how many minutes they spent doing their homework and how many minutes they slept that night.

8. Does smoking change the way someone tastes salt? A researcher sampled 200 smokers and 200 non-smokers. They were then given a bowl of soup and ask to classify the salt level into one of 3 categories: low salt, average salt and high salt. The researcher wanted to know if there was a significant difference in the saltiness classification due to whether the participant was a smoker.

This page titled [15.2.13: Chapter 14 Homework](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Maurice A. Geraghty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.