GSE Math Camp
Lecture 5 HW

1. The average August temperatures (y) and geographic latitudes (x) of 20 cities in the United States were gathered. The regression equation for that data is:

Temperature = 113.6 – 1.01(Latitude)

a. What is the slope of the line? Interpret the slope in the context of the problem.

$\beta_1 = -1.01$; For every one unit increase in degree latitude, the regression model predicts a $-1.01°$ decrease in temperature.

b. Estimate the mean August temperature for a city with a latitude of 32.

$\widehat{temp} = 113.6 - 1.01(32)$
$113.6 - 32.32 = \underline{81.28°}$

c. San Francisco has a mean August temperature of 64 and its latitude is 38. Use the regression equation to estimate the mean August temperature in San Francisco, and then calculate the prediction error (residual) for San Francisco.

$\widehat{temp} = 113.6 - 1.01(38)$
$113.6 - 38.38 = \underline{75.22°}$

$e = y - \hat{y}$
$64 - 75.22 = -11.22°$

Model is overpredicting

d. Why should we not use this regression equation to estimate the mean August temperature at the equator? (latitude=0)

The x values of the observations in the data set likely range from about 20°N to 65°N (for the US). 0° is outside this range, so predicting a temperature for 0° latitude would be extrapolation.

2. Highway planners investigated the relationship between traffic density (number of automobiles per mile) and the average speed of the traffic on a moderately large city thoroughfare. The data were collected at the same location at 10 different times over a span of 3 months. They found a mean traffic density of 68.6 cars per mile (CPM) with standard deviation of 27.07 cpm. Overall, the cars' average speed was 26.38 mph with standard deviation of 9.68 mph. These researches found the correlation between speed and density to be -0.984

a. Identify the explanatory (independent) and response (dependent) variables for this situation.

explanatory (x) – traffic density response (y) – avg speed of traffic

b. Using the information given, what is the equation of the regression line? Show your work!

$\bar{x} = 68.6$ cpm $\bar{y} = 26.38$ cpm
$S_x = 27.07$ cpm $S_y = 9.68$ cpm
$r = -0.984$

$\beta_1 = r\left(\frac{S_y}{S_x}\right) = -.984\left(\frac{9.68}{27.07}\right) = \underline{-.352}$

$\beta_0 \rightarrow \bar{y} = \beta_0 - .352(\bar{x})$
$26.38 = \beta_0 - .352(68.6)$
$\beta_0 = \underline{50.527}$

$\boxed{\widehat{\text{avg speed}} = 50.527 - .352(\text{traffic density})}$

c. Interpret the coefficient of determination (R^2) in the context of this problem.

$(-.984)^2 = .968$

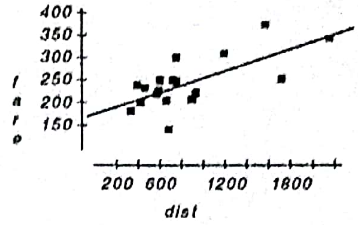
96.8% of the variability in the average speed of traffic is explained by traffic density.

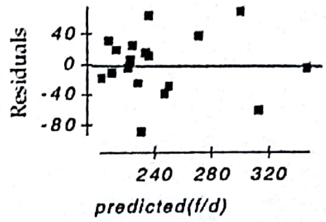
3.

Dependent variable is: fare
No Selector
R squared = 48.2% R squared (adjusted) = 45.0%
s = 41.82 with 18 - 2 = 16 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	26037.4	1	26037.4	14.9
Residual	27980.6	16	1748.79	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	177.215	19.99	8.86	≤ 0.0001
dist	0.078619	0.0204	3.86	0.0014





This is a residual plot. 400B

b. Describe the relationship from the scatterplot.

The scatterplot of fare vs. distance appears reasonably linear, positive, and moderately strong.

c. What is the correlation coefficient? Interpret this value.

- $r = \sqrt{R^2}$ $\sqrt{.482} = .694$
- take sign of slope $r = .694$ → The correlation is moderately strong and positive.

d. Explain what R^2 means in this context.

48.2% of the variability in fares is explained by distance.

e. Write the equation of the model. (Remember to put it in context!)

$\widehat{fare} = 177.215 + .079(distance)$

f. Explain what the slope means in context.

For every 1 mile increase in distance, the regression model predicts a .079 cent increase in fare.

g. Explain what the y-intercept means in context.

When distance is 0 mi, the regression model predicts the fare to be \$177.215.

h. Predict the airfare for a 1000-mile flight.

$\widehat{fare} = 177.215 + .079(1000)$ $\widehat{fare} = \$256.216$

i. The fare to fly to New York, 1358 miles from Tulsa, is \$455. Find the residual for this point.

$\widehat{fare} = 177.215 + .079(1358)$

$\widehat{fare}$ or $\hat{y} = 284.497$

$y = 455$

$e = y - \hat{y}$

$= 455 - 284.497$

$= 170.503$

↳ Model is under-predicting.