Problem Set #12 Solutions: Simple Linear Regression

1. Answer the following question on the basis of the SPSS output below which was generated from a regression analysis examining the relationship between anxiety and mistakes made during an honors thesis defense (this was an example we looked at during the correlation chapter).
	1. What was the regression equation relating anxiety to mistakes?

$$Ÿ= 1.375+ .438(Anx)$$

* 1. Was anxiety a significant predictor of mistakes?

Anxiety was not a significant predictor of the number of mistakes made, F (1, 3) = 9.800, p = .052. OR

Anxiety was not a significant predictor of the number of mistakes made, t (3) = 3.130, p = .052.

In either case, the p-value is greater than .05.

* 1. How much of the variance in mistakes could be attributed to variability in anxiety?

Because r2 = .766, anxiety accounted for 76.6% of the variance in mistakes.



1. You and Buster were sitting around one weekend evening, sipping on some age-appropriate refreshments (i.e. “Beer”). You noticed that as you drank more and more refreshments, your need to visit the bathroom (BR) increased. What are you going to do? You decide to collect some data. The table below presents information on a randomly selected sample of students for whom two pieces of data were collected: the number of refreshing beverages consumed in an evening, and the number of trips to the BR. Use these data to perform the following calculations:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X (beers) | Y (trips to the BR) | X2 | Y2 | X\*Y |
| 3 | 0 | 9 | 0 | 0 |
| 4 | 2 | 16 | 4 | 8 |
| 5 | 5 | 25 | 25 | 25 |
| 5 | 5 | 25 | 25 | 25 |
| 4 | 6 | 16 | 36 | 24 |
| 7 | 8 | 49 | 64 | 56 |
| 4 | 2 | 16 | 4 | 8 |
| 6 | 4 | 36 | 16 | 24 |
| 5 | 3 | 25 | 9 | 15 |
| 7 | 5 | 49 | 25 | 35 |
| Sum = 50 | Sum = 40 | Sum = 266 | Sum = 208 | Sum = 220 |
| Mean = 5 | Mean = 4 |  |  |  |

* 1. What is the linear regression relating beer consumption and trips to the BR?

SSx = Σ(x2) - [(Σx)2/n] SP = Σ(x\*y) - [(Σx\*Σy)/n]

 = 266 - 502/10 = 220 - [(50\*40)/10]

 = 16 = 20

β1 = SP / SSx = 20/16 = **1.25**

β0 = mean(y) - β1\*mean(x) = 4 - 1.25\*(5) = **-2.25**

Regression Equation: **Ŷ = -2.25 + 1.25x**

* 1. How many trips to the BR would you predict someone would make if they consumed 5 beers?

For x = 5: Ŷ = -2.25 + 1.25(5) = **4**

* 1. Could you make a prediction for someone who drank a 12-pack? Explain.

You should not estimate the number of BR trips for someone who drank a 12-pack because that value (12) is not within the range of x's used to calculate the regression equation.

* 1. Is the y-intercept an interpretable value? If so, what does it suggest?

Yes and no…it does tell us the number of bathroom trips we predict when someone consumes zero beers, but it also is a negative value- which doesn’t make sense because you can’t have negative bathroom trips. This is a good example of how statistics can’t think!

* 1. Perform a hypothesis test to determine whether the number of beers consumed is a significant predictor of trips to the BR.

$$SS\_{y}=\sum\_{}^{}\left(y^{2}\right)-\frac{\left(\sum\_{}^{}y\right)^{2}}{n}=208- \frac{40^{2}}{10}=208-160=48$$

$$r= \frac{SP}{\sqrt{SS\_{x}SS\_{y}}}= \frac{20}{\sqrt{\left(16\right)}\left(48\right)}= \frac{20}{27.72}= .72; r^{2}= .72^{2}= .52$$

$$SS\_{regression}= \left(r^{2}\right)SS\_{y}= \left(.52\right)48=24.96 $$

$$SS\_{residual}= \left(1-r^{2}\right)SS\_{y}= \left(1-.52\right)48=23.04 $$

$$MS\_{regression}= \frac{SS\_{regression}}{df\_{regression}}=\frac{24.96}{1}$$

$$MS\_{residual}= \frac{SS\_{residual}}{df\_{residual}}=\frac{23.04}{10-2}= \frac{23.04}{8}=2.88$$

$$F= \frac{MS\_{regression}}{MS\_{residual}}=\frac{24.96}{2.88}=8.67$$

 Fcrit = (1, 8) = 5.32

Decision: reject the null; beer is a significant predictor of bathroom breaks F(1,8) = 8.67, p < .05. The correlation coefficient is positive which suggest a direct relationship. Therefore, the more beer people drink the more they go to the bathroom.

* 1. How much of the variance in trips to the BR can be explained by beer consumption?

52% of the variance

SPSS output on the next page…

