Name _____

MAT 201 Quiz

November 19, 2010

Problem 1 (5 Points Each) ***Circle True or False for each of the following***

A. In order to conduct a T or Z test for a difference between means with independent samples, if you don't know about the underlying distribution of the population, you need to have both samples' sizes greater than 30.

True **False**

True, in order to apply the central limit theorem for means the sample sizes must be greater than 30.

B. In order to conduct a Z test for a difference between proportions, you need to have the number of successes and the number of failures for both samples to be greater than 5.

True **False**

True, in order to apply the central limit theorem for np and nq must be greater than 5 for both samples.

C. A T-Test was conducted for dependent samples asking 35 people how many jackets they owned and how many hats they owned (H_0: $\mu_d = 0$, H_1: $\mu_d > 0$) and the P-Value was found to be 0.02 with $\alpha = 0.05$. Then the conclusion is that there is sufficient evidence to suggest that these 35 people average more jackets than hats. **True** **False**

False, the conclusion of a hypothesis test is always for the entire population not just the sample.

D. The slope of the regression line and the correlation always have the same sign.

True **False**

True, the sign of the correlation indicates the sign of the slope.

E. If the P-Value for the hypothesis test for correlation is less than 0.01, then it can be concluded that there is a strong correlation between the two variables.

True **False**

False, we can say that there is strong evidence to suggest that there is a nonzero correlation, but this nonzero correlation may be strong or weak.

F. A study was done to investigate the relationship between the amount of alcohol people drink per week and the number of traffic accidents they are involved in during their lifetimes. The correlation was found to be 0.993. Then it can be concluded that drinking a large quantity of alcohol results in a person having more traffic accidents during their lives. **True** **False**

False, correlation never shows causation.

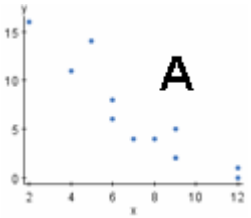
G. A study of many acres of land looked at the relationship between the number of squirrels on the ground and the number of coyotes on the ground produced a correlation of r = -0.91. Then areas with more coyotes on the ground tend to have fewer squirrels on the ground. **True** **False**

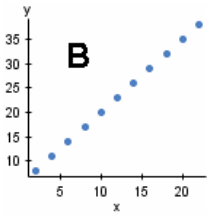
True, there is a strong negative correlation, so as x increases y decreases.

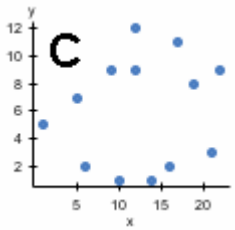
H. The regression line $\hat{y} = 1500 + 30x$ was found for the number of inches of base (x) that Heavenly has and the number of skiers ($\hat{y}$) that ski at Heavenly. Then if there is 40 inches of base there will be 2700 skiers that day. **True** **False**

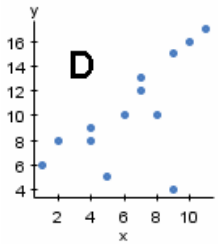
False, we can predict 2700 skiers, but we do not know for sure exactly how many skiers there will be.

Problem 2 (15 Points) Match the scatter plot with its correlation









0.15: _C_____ -0.92: __A____ 0.64: __D____ 1: _B_____

Notice that B is perfectly linear with positive slope, so the correlation is 1. A is not perfectly linear, but follows a negatively sloped line well, so it's correlation is close to -1. D has a positive correlation and certainly follows a linear pattern better than C, so D have the larger correlation compared to C.

Problem 3 (20 Points) A study was done to see student who take statistics (MAT 201) have a higher GPA on average compared to students who take the Survey of Math (MAT 102) class. The data is shown in the table below:

MAT 201	3.45	2.87	3.15	4.00	3.91	3.05	2.78	3.85	3.22
MAT 102	3.91	2.44	2.86	3.25	2.78	3.69	3.00	3.14	

A. State the null and alternative hypotheses.
Since these are dependent samples, there are two of them and the survey question is quantitative, and we see the word "higher", we have
H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 > \mu_2$

B. Find the test statistics and P-Value and state your conclusion using a complete sentence in the context of the question.
We use a 2SampTTest with data to get
t = 1.006110088 and p = 0.1653873663.
Since p is large (larger than any level of significance that is used) we fail to reject the null hypothesis and state that there is insufficient evidence to make a conclusion about the mean GPA for statistics students being higher than the mean GPA for Survey of Math Students.

C. Do you need to make any assumptions about the underlying distributions? Explain.
Yes, since the sample sizes are under 30, we need to assume that both underlying distributions are approximately normal.

Problem 4 (20 Points) A study was done to see if men and women differ in their likelihood of passing out while giving blood. Only 14 of the 2000 male blood donors and 15 of the 4800 female blood donors passed out.

A. State the null and alternative hypotheses.
H_0: $p_1 = p_2$ H_1: $p_1 \neq p_2$

B. Find the test statistics and P-Value and state your conclusion using a complete sentence in the context of the question. Use $\alpha = 0.05$
z = 2.234273702, p = 0.0254649849
Since p < 0.05 we reject the null hypothesis and accept the alternative hypothesis and conclude that there is statistically significant evidence t conclude that there is a difference between the proportion of men who pass out while giving blood and the proportion of women who pass out while giving blood.

C. Find the appropriate 90% confidence interval for the difference and use a complete sentence to interpret this confidence interval in the context of the question.
Using a 2-PropZInt, we get: (0.00053,0.00722). We are 90% confident that the men are between 0.053% and 0.722% more likely than women to pass out while giving blood

D. Explain why it was or why it was not appropriate to use the normal distribution for your calculations.
Since for men, np = 14 and nq = 1986 and for women np = 15 and nq = 4785 are all greater than 5, we may use the normal distribution for these calculations.

Problem 5 (25 Points) A study was done to see the relationship between the number of hours (x) students study per week and the percent (y) they received on their final. The equation of the regression line was $y = 62 + 3x$, $r^2 = 0.82$, and the p-value for the correlation was $p = 0.0001$.

A. Use the linear regression line to give a point estimate for the mean percent score students achieve on the final if they study five hours per week.
Just plug in: 62 + 3(5) = 77

B. Use a complete sentence to interpret the y-intercept in the context of the study.
Our best prediction for the score on the final exam for a person who does not study at all is 62%.

C. Use a complete sentence to interpret the slope in the context of the study.
For every additional hour of studying done, we can predict that the exam scores are increased by 3%.

D. Use a complete sentence to interpret r^2 in the context of the study.
82% of the variation in exam scores is attributable to the amount of studying that students do. The other 18% is attributable to other factors.

E. State the null and alternative hypothesis and state your conclusion in the context of the study.
H_0: $\rho = 0$ H_1: $\rho \neq 0$
Since the P-Value is so small (smaller than any level of significance) we has strong evidence to conclude that there is a correlation between hours students study and their exam scores.