STAT 515 – Spring 2001 - Final Exam Answers

Part I: Answer three of the following four questions. If you answer more than three I will grade only the first three. Five points each.

1) Define what is meant by the p-value (or observed significance level) of a test. **The probability of observing a test statistic as extreme as the one observed, or more extreme, if the null hypothesis is true. -or- The smallest a-level at which we reject H_0.**

2) (Circle the correct answers) At a large university, most of the classes have fewer than 32 students. A few, however, have over 1,000 students. If the university wants it to sound like the students receive more individual attention, they should report the **median** of the class sizes. This distribution of class sizes is **skewed right**.

3) An IQ test is supposedly constructed to normal with mean 100 and standard deviation 16. What percent of people taking the test should have scores between 90 and 110? **(110-100)/16 = 0.625 so 2*0.2324 = 0.4648**

4) A certain shortstop in major league baseball has a 98.2 percent chance of successfully making plays in the field. If we assume that each play is independent of the others, what is the chance that this shortstop will successfully make all of the next 100 plays? **$(0.982)^{100}$= 0.1626 = 16.26%**

Part II: Answer every part of the next three problems. Read each problem carefully, and show your work for full credit. Twenty points each.

1) A study in the medical journal *Chest* in 1995 looked at the occurrence of obstructive sleep apnea in commercial truck drivers. It is believed that 25% of the general population suffers from obstructive sleep apnea. Of the 159 commercial truck drivers examined however, 124 of them suffered from the disorder.

A) State the appropriate null and alternate hypothesis if the researchers suspect that the stress of truck driving might increase the likelihood of suffering sleep apnea. **H_0: p=0.25 H_A: p>0.25**

B) Test the hypothesis in part a. Report the p-value. **p-hat=124/159=0.780 so z=(0.780-0.25)/sqrt(.25*.75/159) = 15.43… comparing to a normal table gives a p-value of 0.**

C) At an α=0.05 level what is your conclusion (i.e. do truckers suffer more obstructive sleep apnea)? **Yes, p-value < α.**

D) Why is, or isn't, the sample size in this problem large enough to trust the results of this test? **np=39.75 and n(1-p)=119.25 are both greater than or equal to 5.**

E) It is also possible that only the population of truckers is of interest. Construct a 90% confidence interval for the percentage of truck drivers that suffer from obstructive sleep apnea. **0.780 +/- 1.645*sqrt(.780*.220/159) = 0.780 +/- 0.054 so (0.726,0.834)**

2) A study was conducted to examine the relationship between wages and the percent of workers who quit a particular job. The attached data set `quitters` gives the quitting rate for each of 15 industries (`quitrate`) and the average hourly wage in dollars for those industries (`wage`). A simple linear regression is used to predict the quitting rate from the hourly wages.

A) The three degrees of freedom and the TSS have been deleted from the ANOVA table.

DF for model	**1 (note SS=MS)**	DF for total	**n-1=14**
DF for errors	**total-model=14-1=13**	TSS	**SSR+SSE=8.2507+3.0733=11.324**

B) What null and alternate hypothesis are being tested by the bold faced p-value on the SAS output? **H_0: β_1=0 H_A: $\beta_1 \neq 0$**

C) Assuming the assumptions for this simple linear regression are met, what percentage of the variation in quit rate is determined by the hourly wages? **R-squared = 0.7286, so 72.86%**

D) Assuming the assumptions for this simple linear regression are met, how would the quitting rate be predicted to change if the hourly wage increased by 50 cents? **0.5*slope=0.5*(-.3466)=-.1733**

E) Assuming the assumptions for this simple linear regression are met, why wouldn't you trust this regression equation to predict the quit rate for an industry whose hourly wage was $20.00. **It would be extrapolating, the highest amount we have for data is under 14.**

3) The data below is from the 1988 Harvard Study on the relationship between Aspirin and Heart Attacks. The study was done so that each person was randomly assigned to either aspirin or placebo (they didn't know which they were taking) and was then observed to see if they eventually had a fatal heart attack, non-fatal heart attack, or no heart attack.

	Heart Attack			
	Fatal	Non-fatal	No Attack	
Placebo	18	171	10,845	**11034**
Aspirin	5	99	10,933	**11037**
	23	**270**	**21,778**	**22071**

It is desired to test the null hypothesis that the probabilities of having fatal, non-fatal, and no heart attack are the same for those taking aspirin and not taking aspirin.

A) Determine the degrees of freedom for conducting this test. **(r-1)(c-1)=2*1=2**

B) Write out the tables of expected values for conducting this test.

23*11034/22071=11.5	**270*11034/22071=135.0**	**21778*11034/22071=10887.5**
23*11037/22071=11.5	**270*11037/22071=135.0**	**21778*11037/22071=10890.5**

C) Give the formula for X^2 for this problem (plugging the values in, but not needing to simplify).

$$(18-11.5)^2/11.5 + (171-135)^2/135 + (10845-10887.5)^2/10887.5 + (5-11.5)^2/11.5 + (99-135)^2/135 + (10933-10890.5)^2/10890.5$$

D) What is the rejection region (critical region) for conducting this test at α=0.05?

Reject if greater than or equal to 5.99147

E) Why is, or why isn't, the sample size of this experiment large enough for performing this hypothesis test?

All expected values are greater than or equal to 5.