STAT 515 – Spring 2001 - Final Exam

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.

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 **mean / median** of the class sizes. This distribution of class sizes is **skewed left / symmetric / skewed right**.

3) An IQ test is supposedly constructed to be normally distributed with mean 100 and standard deviation 16. What percent of people taking the test should have scores between 90 and 110?

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?

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.

B) Test the hypothesis in part a. Report the p-value.

C) At an $\alpha=0.05$ level what is your conclusion (i.e. do truckers suffer more obstructive sleep apnea)?

D) Why is, or isn't, the sample size in this problem large enough to trust the results of this test?

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.

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 _____   DF for total _____

      DF for errors _____   TSS   _____

B) What null and alternate hypothesis are being tested by the bold faced p-value on the SAS output?

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?

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?

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.

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
Aspirin	5	99	10,933

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.

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

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

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

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

```
DATA quitters;
INPUT quitrate wage;
CARDS;
1.4  8.20
 .7 10.35
2.6  6.18
3.4  5.37
1.7  9.94
1.7  9.11
1.0 10.59
 .5 13.29
2.0  7.99
3.8  5.54
2.3  7.50
1.9  6.43
1.4  8.83
1.8 10.93
2.0  8.80
;

PROC INSIGHT;
OPEN quitters;
FIT quitrate = wage;
RUN;
```

▶	4	Int	Int	Int	Int				
15		quitrate	wage	R_quitrate	P_quitrate				
■	1	1.4	8.20	-0.6198	2.0198				
■	2	0.7	10.35	-0.5747	1.2747				
■	3	2.6	6.18	-0.1198	2.7198				
■	4	3.4	5.37	0.3995	3.0005				
■	5	1.7	9.94	0.2832	1.4168				
■	6	1.7	9.11	-0.0044	1.7044				
■	7	1.0	10.59	-0.1915	1.1915				
■	8	0.5	13.29	0.2442	0.2558				
■	9	2.0	7.99	-0.0926	2.0926				
■	10	3.8	5.54	0.8584	2.9416				
■	11	2.3	7.50	0.0376	2.2624				
■	12	1.9	6.43	-0.7332	2.6332				
■	13	1.4	8.83	-0.4014	1.8014				
■	14	1.8	10.93	0.7263	1.0737				
■	15	2.0	8.80	0.1882	1.8118				

Model Equation				
quitrate	=	4.8615	-	0.3466	wage
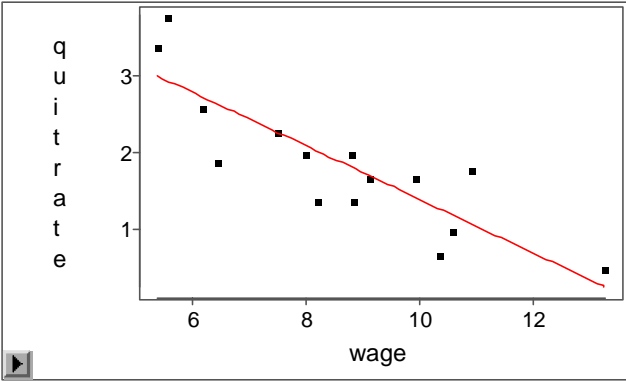
Summary of Fit			
Mean of Response	1.8800	R-Square	0.7286
Root MSE	0.4862	Adj R-Sq	0.7077

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model		8.2507	8.2507	34.90	**<.0001**
Error		3.0733	0.2364		
C Total					

Parameter Estimates							
Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	4.8615	0.5201	9.35	<.0001	.	0
wage	1	-0.3466	0.0587	-5.91	<.0001	1.0000	1.0000