STAT 515 – Fall 2003- Practice Final Exam Solutions

1) **The probability of observing a test statistic at least as extreme as the one observed if the null hypothesis is true.**

2) The automaker would likely prefer that the **<u>median</u>** fuel economy for the entire line be reported. Because the distribution is very skewed to the left, if we were given the mean and standard we would be able to say that **<u>at least 75%</u>** of the mileages should be within two standard deviations of the mean.

3) A biased coin (probability of a head on one flip = 0.4) is flipped 9 times. What are the mean and standard deviation of the number of heads that will be observed? **mean=np=3.6 standard deviation=square root of(np(1-p))= 1.469694**

4) **P(20<X<25)=P((20-20)/4 < (X-mean)/sd < (25-20)/4)= P(0<Z<1.25)=0.3944**

1A) **H_0: p=0.4 H_A: p<0.4 where p=proportion of consumers deterred by web-site complexity**

B) $$z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}=\frac{0.25-0.4}{\sqrt{\frac{0.4(1-0.4)}{60}}}=\frac{-0.15}{0.0632}=2.37$$

p-value=0.5-0.4911=0.0089.
We reject the null hypothesis because the p-value is less than α. The percentage who are deterred is less than 40.

C) You could use
$$\hat{p}\pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\Rightarrow 0.25\pm 1.96\sqrt{\frac{0.25(1-0.25)}{60}}\Rightarrow 0.25\pm 0.11\Rightarrow(0.14,0.36)$$
But it is better to use the Agresti and Coull correction... where we add 2 to *x* and 4 to *n*.
$$p^{*}=\frac{x+2}{n+4}=\frac{15+2}{60+4}=\frac{17}{64}=0.2656\qquad p^{*}\pm z_{\alpha/2}\sqrt{\frac{p^{*}(1-p^{*})}{n+4}}\Rightarrow 0.2656\pm 1.96\sqrt{\frac{0.2656(1-0.2656)}{60+4}}\Rightarrow 0.2656\pm 0.1082$$

D) The sample needs to be random and $n\hat{p}\geq 5$ and $n(1-\hat{p})\geq 5$.

2) A)

Summary of Fit			
Mean of Response	36.4545	R-Square	0.8900
Root MSE	6.7441	Adj R-Sq	0.8845

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	7361.7866	7361.7866	**161.8566**	<.0001
Error	**20**	**909.6679**	**45.48340**		
C Total	21	8271.4545			

B) **If x (# observed geese) = 100 then the estimated y (actual # geese)is 2.9101+1.1714(100)= 120.0501**

C) **Because it is extrapolating (we don't have any observations that big).**

D) On average, how far from the estimated regression line do you expect the observed values to be? **Square root of MSE is 6.7441.**

E) Residual vs. Predicted Plot is for checking: **Errors have mean 0 at each x and errors have equal variance at each x**
Q-Q plot of the Residuals is for checking: **Errors are normally distributed**

3A) **Test of independence because we know the total number of heart patients in advance, but not the row and column totals.**

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

	Abstain	7 or fewer	7 or more	
Con.	146 **281*896/1913=131.6**	106 **281*696/1913=102.2**	29 **281*321/1913=47.2**	281
Not	750 **1632*896/1913=764.4**	590 **1632*696/1913=593.8**	292 **1632*321/1913=273.8**	1632
	896	696	321	1913

C) Give the formula for X^2 for this problem (plugging the values in, but not needing to simplify).
$$\frac{(146-131.6)^2}{131.6}+\frac{(106-102.2)^2}{102.2}+\frac{(29-47.2)^2}{47.2}+\frac{(750-764.4)^2}{764.4}+\frac{(590-593.8)^2}{593.8}+\frac{(292-273.8)^2}{273.8}$$

D) What is the rejection region for conducting this test at α=0.01? **Checking df=(3-1)(2-1)=2 we get 9.21034 or greater.**

E) Why is, or why isn't, the sample size of this experiment large enough for performing this hypothesis test?
It is large enough because all of the expected values are greater than 5.