STAT 205 **Name: ANSWER KEY**_____
Fall 2007
Exam 3

$$\tilde{p} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+Z^2_{\alpha/2}}} \quad \text{where} \quad \tilde{p} = \frac{Y + \frac{1}{2}Z^2_{\alpha/2}}{n+Z^2_{\alpha/2}}$$

$$(\tilde{p}_1 - \tilde{p}_2) \pm Z^2_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}} \quad \text{where} \quad \tilde{p}_1 - \tilde{p}_2 = \frac{Y_1+1}{n_1+2} - \frac{Y_2+1}{n_2+2}$$

$$n \geq \frac{Z^2_{\alpha/2}\, p_0(1-p_0)}{\varepsilon_0^2} - Z^2_{\alpha/2}$$

$$\sum \frac{(O - E)^2}{E}$$

Part I: Answer eight of the following nine questions. If you complete more than eight, I will grade only the first eight. Five points each.

1) State the definition of a P-value.
The P-value of a test is the probability under H_0, of observing a test statistic as extreme or more in the direction of H_A as that actually observed.

2) **(Fill in the blank)** Suppose we have a two dependent samples data set and the differences come from a normal population. We use a *t* distribution hypothesis test rather than a sign test because the *t* test is **more powerful** than the sign test.

3) The ability to taste the compound phenylthiocarbamide (PTC) is a genetically controlled trait in humans. In Europe and Asia, about 70% of people are "tasters". Suppose a study is being conducted to estimate the population proportion of tasters (via a 95% confidence interval). The researchers would like to keep the margin of error for the confidence interval less than or equal to 0.01. Write down the formula – **with the appropriate numbers in it (no letters, please)** – that will tell the researchers what sample size to take in order to meet this criterion. *You do not need to simplify this expression.*

$$n \geq \frac{1.96^2(0.7)(1-0.7)}{0.01^2} - 1.96^2$$

4) **(Circle the correct answer)** When we report a sample size calculation to control margin of error, we should **round down** / **round up** to the next integer value.

5) **(Circle the correct answer)** Marine biologists have noticed that the color of the outermost growth band on a clam tends to be related to the time of year in which the clam dies. A biologist collected a sample of the species *Protothaca staminea.* Each clam in the sample was observed for month the clam died (February or March) and color of the outermost growth band (clear, dark, in between). Which of the following is the correct H_0 to choose for the Chi-square test using the resulting 2 x 3 contingency table?

H_0: prevalence of growth band color is the same for clams who die in February or March

H_0: growth band color and month of death are not associated

6) What are the degrees of freedom for the Chi-square test of question (5)?

df = (r-1)(c-1) = (2-1)(3-1) = 2

7) An experiment was conducted to study the mutagenic potential of the chemical ethyl carbamate in fruit flies. A 90% Agresti-Coull confidence interval for the proportion of fruit flies exposed to ethyl carbamate exhibiting a lethal mutation was (0.046, 0.062). Interpret this interval.

We are 90% confident that the true proportion of fruit flies exposed to ethyl carbamate exhibiting a lethal mutation is no less than 0.046 and no more than 0.062.

8) What is the probability the true population proportion is in the interval computed for question (7)?

0 or 1 and we don't know which one

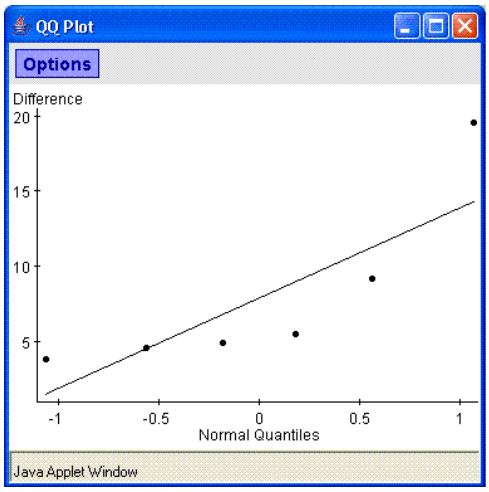
9) What is the probability the true population proportion will be in the next 90% confidence interval (or any future 90% confidence interval, for that matter) computed from a new sample from the same fruit fly population mentioned in question (7)?

90%

Part II: Answer every part of the next two problems. Read each question carefully, and show your work for full credit.

1) Six patients with renal disease underwent plasmapheresis. Urinary protein excretion (grams of protein per gram of creatinine) was measured for each patient before and after plasmapheresis. The data and QQplot of the differences are given below.

Patient	Before	After	Difference
1	20.3	.8	19.5
2	9.3	.1	9.2
3	7.6	3.0	4.6
4	6.1	.6	5.5
5	5.8	.9	4.9
6	4.0	.2	3.8
Mean	8.9	0.9	7.9
SD	5.9	1.1	6.0



1a) (10 points) Use the QQplot of the differences above to comment on whether the assumption of normality has been met. Two or three sentences should suffice.

There is a systematic departure from the line. The "U" shape is indicative of a skewed right distribution. This is a clear indication these data do not come from a normal population and with only 6 data points, we cannot invoke the CLT. The normality assumption has not been met.

1b) (25 points) Conduct a sign test to investigate at the 0.05 significance level whether or not urinary protein excretion tends to go down after plasmapheresis in patients with renal disease.

(1) $\alpha = 0.05$

(2) H_0: Urinary protein excretion is the same before and after plasmapheresis

H_A: Urinary protein excretion tends to go down after plasmapheresis

(3) H_A: "effect before" > "effect after", then $B_s = N_+ = 6$

(4) $P = Pr\{Bin(6,1/2) \geq 6\} = Pr\{Bin(6,1/2) = 6\} = {}_6C_6(.5)^6(.5)^0 = 0.5^6 = 0.015625$

Or, using the TI calculator

$P = 1 - Pr\{Bin(6,1/2) \leq 5\} = 1 - binomcdf(6, 0.5, 5) = 0.015625$

(5) $P < \alpha$, reject H_0

(6) We have significant evidence to conclude urinary protein excretion tends to go down after plasmapheresis.

2) (25 points) In an experiment to treat patients with anxiety disorders, the effectiveness of the drug hydroxyzine was studied. Subjects were randomly allocated to the hydroxyzine group or a control group. At the end of the study, patients were checked for improvement. The data are summarized in the table below.

	Hydroxyzine	Placebo
Improved	30 (25.177)	20 (24.822)
No improvement	41 (45.822)	50 (45.177)

Test whether there is significant evidence that the proportion of those improved is higher when taking hydroxyzine. Use a Chi-square test at the $\alpha = 0.05$ significance level.

(1) $\alpha = 0.05$

(2) H_0: $p_1 = p_2$

H_A: $p_1 > p_2$

(3) $\hat{p}_1 = \frac{30}{71} = 0.423 > 0.286 = \frac{20}{70} = \hat{p}_2$ Data deviate in the direction of H_A.

Then $X_s^2 = \frac{(30-25.177)^2}{25.177} + \frac{(20-24.822)^2}{24.822} + \frac{(41-45.822)^2}{45.822} + \frac{(50-45.177)^2}{45.177}$

$= 0.923911864 + 0.9367369269 + 0.5074349439 + 0.5148931757$

$= 2.883$

OR

After checking the data deviate in the direction specified by H_A, then entering our 2x2 matrix into Matrix[A], the TI calculator Chi-square test option returns $X_s^2 = 2.883$

(4) $P = \frac{1}{2}(\Pr\{\chi^2_{df=1} \geq 2.883\})$

$0.05 < 2P < 0.10 \Rightarrow 0.025 < P < 0.05$ (Table 9)

OR

TI calculator returns $P = 0.0895100557$ *for the two-sided test.* So, $P = \frac{1}{2}(0.0895100557) = 0.0448$

(5) $P < \alpha$, reject H_0

6) We have significant evidence to conclude the true proportion of anxiety disorder patients who improve is higher for patients taking hydroxyzine than that of anxiety disorder patients taking a placebo.