STAT 205

Exam 2 Name ________ANSWER KEY________

Fall 2006

$$P\{Y=j\} = {}_{n}C_{j}\, p^{j}(1-p)^{n-j}$$

$$Z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$$

$$\bar{Y} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$t = \frac{\bar{Y}-\mu}{s/\sqrt{n}}$$

$$(\bar{Y}_1-\bar{Y}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$t = \frac{(\bar{Y}_1-\bar{Y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

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 probability under H_0 of observing a test statistic as extreme or more extreme (in the direction of H_A) as that actually observed.

2) (Circle the correct answer) A hypothesis test has been conducted at the 0.05 significance level, resulting in a P-value of 0.003. Obviously, in this case, we reject H_0. If an error was made, it would be a **Type I / Type II / neither** error.

3) (Circle the correct answer) Suppose you have calculated a 95% confidence interval for the mean, μ. Now, you want to calculate a 99% confidence interval using the same sample. The 99% confidence interval will be **narrower / wider** than the 95% confidence interval.

4) Consider taking a random sample of size 4 from a population of persons who smoke and recording how many of them, if any, have lung cancer. Let $\hat{p}$ represent the proportion of persons in the sample with lung cancer. List the possible values of $\hat{p}$.

0, 0.25, 0.5, 0.75, 1

5) (Circle the correct answer) Trichotillomania is a psychiatric illness that causes its victims to have an irresistible compulsion to pull their own hair. Two drugs were compared as treatments for trichotillomania in a study involving 13 women. Each woman took clomipramine during one time period and desipramine during another time period in a double-blind experiment (the women nor their doctors knew which drug they were taking during which time period). We would use the **independent / dependent (paired)** samples method in order to conduct a test of hypothesis.

6) The Type II Error rate, $\beta = P\{\text{failing to reject } H_0 | H_0 \text{ is false}\}$, for a hypothesis test was calculated to be $\beta = 0.07$. What is the power $= P\{\text{rejecting } H_0 | H_0 \text{ is false}\}$ for this test?

Power $= 1 - \beta = 1 - 0.07 = 0.93$

7) State the assumptions required for validity of a *t* confidence interval on the population mean, μ.

• Data collected from a random sample from a large population
• Observations in the sample must be independent from each other
• n small, population distribution must be approximately normal
• n large, population need not be approximately normal (CLT kicks in)

8) Suppose you have computed a 99% confidence interval for the mean, μ. What is the probability that true mean, μ is in the interval you just computed?

Either 0 or 1 and we don't know which

9) The Central Limit Theorem says that for any i.i.d. random sample, $Y_1, Y_2, \ldots, Y_n$ where $E[Y_i] = \mu$ and $E[(Y_i-\mu)^2] = \sigma^2$, then as $n \to \infty$ the distribution of the sample mean is approximately normal with mean, μ, and variance $\frac{\sigma^2}{n}$.

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

1a) (25 pts) A scientist conducted a study to test whether a parakeet chirps more often if there is music playing. The scientist took a random sample of 28 parakeets. Each of the 28 parakeets were observed for two different 30 minute periods – one with music playing and one without. Using the 0.01 significance level, test whether the mean number of chirps (per 30 minutes) when music is playing is higher than when the room is silent.

I've numbered the steps for you on the next page. Please put the appropriate step next to the appropriate number.

(1) $\alpha = 0.01$

(2) H_0: $\mu_m - \mu_w = 0$ or $\mu_m = \mu_w$ or $\mu_d = 0$
H_A: $\mu_m - \mu_w > 0$ or $\mu_m > \mu_w$ or $\mu_d > 0$

TI-84 STAT -> TESTS -> Ttest (using the difference column)

(3) $t_s = 19.48$

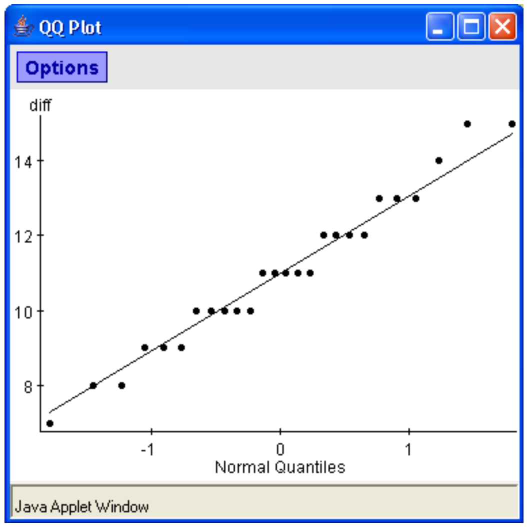
(4) $P = 9.87 \times 10^{-18}$

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

(6) There is significant evidence (at the $\alpha = 0.01$ significance level) that the true mean number of parakeet chirps (per 30 minutes) is larger when music is playing than when the room is silent.

Parakeet	Music	No Music	Difference
1	12	3	9
2	14	1	13
3	11	2	9
4	13	1	12
5	20	5	15
6	14	3	11
7	10	0	10
8	12	2	10
9	8	6	2
10	13	3	10
11	14	2	12
12	15	4	11
13	12	3	9
14	13	2	11
15	8	0	8
16	18	5	13
17	15	3	12
18	12	2	10
19	17	2	15
20	15	4	11
21	11	3	8
22	22	4	18
23	14	2	12
24	18	4	14
25	15	5	10
26	8	1	7
27	13	2	11
28	16	3	13
Mean	**13.7**	**2.8**	**10.929**
SD	**3.4**	**1.5**	**2.968**

1b) (5 pts) Comment on whether the assumption that can be checked using the QQplot below seems to be met for this analysis.



Notice, our data set is discrete, so it can't possibly come from a normal distribution. But, we still check the "shape" of it with a QQplot, making sure it's not too different from symmetric and bell-shaped, so that the CLT will give us that the distribution of the sample mean is approximately normal.

The points on this QQplot are making a linear pattern. The points that are grouped together horizontally are just repeated observations (this happens frequently with discrete data). There is the one point above the line in the upper tail, but I'm not too worried about this, since there is no systematic departure from the line. This QQplot shows that the distribution is shaped like a normal, so the moderate sample size of 28 will allow us to use the Central Limit Theorem, meeting our normality assumption.

2) A study was conducted to determine whether relaxation training, aided by biofeedback and meditation, could help in reducing high blood pressure. Subjects were randomly allocated to a biofeedback group or a control group. The biofeedback group received training for eight weeks. The table reports the reduction in systolic blood pressure (mm Hg) after eight weeks. *Note*: WS approximation yields 190 degrees of freedom for these data.
(You may proceed as though the assumptions have been checked and deemed acceptable.)

	Biofeedback	**Control**
n	99	93
$\bar{Y}$	13.8	4.0
s	13.3	12.54

a) (25 pts) Construct a 99% confidence interval for the difference in mean response (reduction in systolic blood pressure).

Note: This is an independent samples setting since the patients were randomly allocated into one group or the other (there was only one observation made on each participant). We want to construct a 99% CI, so use 2SampTInt to get

(4.9478, 14.652) or $4.9478 < \mu_B - \mu_C < 14.652$ mm Hg

b) (5 pts) Interpret the interval you computed in part (a).

We are 99% confident that the true mean reduction in systolic blood pressure over 8 weeks is bigger under the biofeedback condition than for the control condition by as little as 4.978 mm Hg or as much as 14.652 mm Hg.