STAT 205
Spring 2007
Exam 2

Name:________________________________

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

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

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

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

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

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.

2) (Circle the correct answer) A hypothesis test has been conducted at the 0.05 significance level, resulting in a P-value of 0.06. Obviously, in this case, we fail to reject H_0. If an error was made, it would be a **Type I** / **Type II** / **neither** error.

3) Suppose you want to analyze data with a sample of size, n=10 and your QQplot shows a violation of normality. In practice, a ________________ is commonly attempted before resorting to a non-parametric test.

4) (Circle one.) **True** / **False** A *t*-based confidence interval for the mean, μ, and a *t*-based hypothesis test for the mean, μ, are different forms of the same inference.

5) (Circle the correct answer) Twenty institutionalized epileptic patients participated in a study of a new anticonvulsant drug, Valproate. Ten of the patients (chosen at random) were started on the daily Valproate and the other ten received an identical placebo pill. During an eight-week observation period, the number of major and minor seizures were counted. After this, the patients were "crossed over" to the other treatment, and seizure counts were made during a second eight week observation period. We would use the **independent** / **dependent (paired)** samples method in order to conduct a test of hypothesis.

6) We learned several standard errors of means (SEM's) for use in making inference on the population mean. In the independent samples setting, we have been applying one type of SE for the difference between two means, but you also learned of another type of SE which can be used in the independent samples setting called SE_{pool}. Describe the special situation for which we can employ its use.

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

8) A geneticist weighed 28 female lambs at birth. The lambs were all born in April, were all the same breed (Rambouillet), and were all single births (no twins).The diet and all other environmental conditions were the same for all the parents. The geneticist reported a *t* distribution based confidence interval for the mean birth weight. If twin births had been included, would the confidence interval still be valid? Why or why not?

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 _______________ with mean, ____, and variance, _______(note, I'm asking for variance here – not standard deviation).

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

1) A classical method of researching the nature vs. nurture question is using identical twins to conduct research. A book by Susan Farber (*Identical Twins Reared Apart*, New York: Basic Books, 1981) contains a chronicle and reanalysis of identical twins reared apart. One question of interest is whether there are significant differences in the IQ scores of identical twins where one member of the pair is reared by natural parents and the other is not.

1a) (25 points) A random sample of 32 sets of twins under the above circumstances was taken. The sample mean of the difference (IQ twin reared by natural parent – IQ not reared by natural parent) in IQ scores is -2.906, the standard deviation of the differences is 8.895. Conduct a test of hypothesis to see whether there is a significant difference in the mean IQ scores of these two groups. Operate at the 0.05 significance level.

(1)

(2)

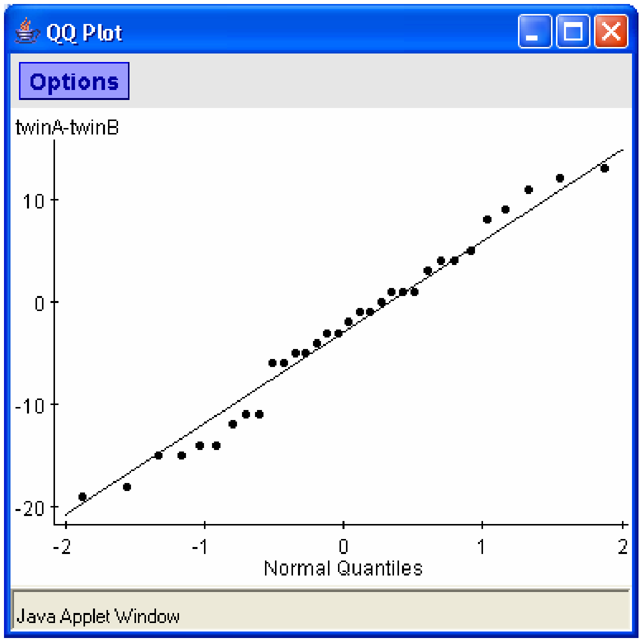
(3)

(4)

(5)

(6)

1b) (5 points) Using the QQplot of the differences below, comment on whether the assumptions appear to be met for the *t* procedure.



2) Researchers were interested in the short-term effect that caffeine has on heart rate. They took a random sample of individuals and measured each person's resting heart rate. Then they had each subject drink six ounces of coffee. Nine of the subjects were given coffee containing caffeine and eleven were given decaffeinated coffee. After ten minutes each person's heart rate was measured again. The data in the following table show the change in heart rate; a positive number means that heart rate went up and a negative number means that heart rate went down.

	Caffeine	**Decaf**
	28	26
	11	1
	−3	0
	14	−4
	−2	−4
	−4	14
	18	16
	2	8
	2	0
		18
		−10
n	9	11
$\bar{y}$	7.3	5.9
s	11.1	11.2
SE	3.7	3.4

2a) (25 points) Use these data to construct a 90% confidence interval for the difference in mean heart rate between the caffeinated and decaffeinated coffee groups. You may proceed as though the assumptions have been checked and deemed acceptable. *Note*: Formula (7.1) yields 17.3 degrees of freedom for these data.

2b) (5 pts) Interpret the interval you computed in part (a).