STAT 200 Midterm Exam II (Section 101)

(Time: **50 minutes**, *Friday, November 10, 2006, 9:00 - 9:50am*)

Name:

Student ID:

This exam is to be done with closed notes/books. One "cheat sheet" ($8.5' \times 11'$, two sided) is allowed. There are 5 problems, with a total of 100 points. Make sure you write your name and student ID. We prefer short answers without omitting key steps – you **must** *show the key steps for Problems 3 – 5.*

Problem 1 (15 pts). Circle the correct answer: true (T) or false (F).

(1) We must know the approximate distribution of a statistic, which is used to estimate a population mean, in order to computer a confidence interval for that population mean. T F

(2) A t-distribution with small degree of freedom has heavier tails than the standard normal distribution, but when the degree of freedom is large a t-distribution is very similar to the standard normal distribution. T F

(3) Since there is sampling error in estimation, the sampling distribution has a standard deviation which is larger than the population standard deviation. T F

(4) If population II has a larger true mean than population I, the sample mean from population II must be larger than the sample mean from population I. T F

(5) If a test is significant at 5% level, the probability of H_0 being true is 0.05. T F

Problem 2 (36 pts). Circle the answer which is the *closest* to the correct answer (please choose only one answer).

(1). An engineer designs an improved light bulb. The previous design had an average lifetime of 1200 hours. The new bulb had a lifetime of 1201 hours, using a sample of 2000 bulbs. Although the difference is quite small, the effect was statistically significant. The explanation is

a) that new designs typically have more variability than standard designs.
b) that the sample size is very large.
c) that the mean of 1200 is large.
d) all of the above.

(2) A simple random sample of size n is taken from a population with mean 10 and standard deviation 5. If we use only half the sample, the mean of the sampling distribution of $\bar{x}$ would be

a) 5 b) $10/\sqrt{2}$ c) 10 d) $2\sqrt{n/2}$

(3) The time it takes for a battery to fail is often not normally distributed. We select a simple random sample of 1000 batteries and compute the average of all the failure times. The sampling distribution of this average may be approximated by a

a) Binomial distribution b) uniform distribution c) normal distribution d) t distribution

(4). An agricultural researcher plants 25 plots with a new variety of corn. The average yield for these plots is $\bar{x} = 150$ bushels per acre. Assume that the yield per acre for the new variety of corn follows a normal distribution with unknown mean μ and that a 95% confidence interval for μ is found to be 150 ± 3.29. Which of the following is true?
a) A test of the hypotheses $H_0 : \mu = 150$ vs. $H_a : \mu \neq 150$ would be rejected at the 0.05 level.
b) A test of the hypotheses $H_0 : \mu = 150$ vs. $H_a : \mu > 150$ would be rejected at the 0.05 level.
c) A test of the hypotheses $H_0 : \mu = 160$ vs. $H_a : \mu \neq 160$ would be rejected at the 0.05 level.
d) All the above.

(5) In hypothesis testing, the term "significant difference" refers to
(a) the difference between the critical value and the test statistic.
(b) the difference between the sample standard deviation and the population mean.
(c) the difference between the sample mean and the hypothetical population mean that leads to the rejection of the null hypothesis.
(d) none of the above.

(6) For a two-sample testing problem, if the mean of the first sample is greater than the mean of the second sample,
(a) the null hypothesis must be false.
(b) the alternative hypothesis must be false.
(c) the mean of the first population must be greater than the mean of the second population.
(d) it is possible for the mean of the sampling distribution of differences to be equal to zero.

(7) If there is a statistically significant difference between the means of two populations,
(a) the means of the two populations must be different.
(b) the alternative hypothesis must be two-sided.
(c) the alternative hypothesis must be one-sided.
(d) only one alternative hypothesis is possible.
(e) none of the above.

(8) In a hypothesis testing problem, the p-value tells us
a) if the null hypothesis is true.
b) if the alternative hypothesis is true.
c) the evidence against the null hypothesis.
d) the evidence against the alternative hypothesis.

(9) Consider testing $H_0 : \mu_2 - \mu_1 = 0$ versus $H_a : \mu_2 - \mu_1 \neq 0$. The 90% confidence interval for $\mu_2 - \mu_1$ is 14.6 ± 7.80. Based on this confidence interval,
a) we would not reject the null hypothesis
b) we would reject the null hypothesis of no difference at the 0.10 level
c) we would reject the null hypothesis of no difference at the 0.05 level
d) none of the above.

This table is for instructor use only

Problem 1	Problem 2	Problem 3	Problem 4	Problem 5	Total Mark

Problem 3 (18 pts). In a large school, the time students spend on homework each week is normally distributed with a mean of 15 hours and a standard deviation of 2 hours.

(a) (6 pts) The probability that a student spends at least 14 hours on homework is $.6915$ ($P(X \geq 14)$)
For 100 randomly selected students, the average number of students who spend at least 14 hours on homework is $np = .69.15$, with a standard deviation of $\sqrt{np(1-p)} = .4.62$.

(b) (6 pts) For a group of 16 randomly selected students, what is the probability that the average time this group spends on homework in a given week is between 13 hours and 16 hours? (Assume that each student works independently.)

$n = 16$.

$P(13 < \bar{X} < 16)$

$= P\left(\frac{13-15}{2/\sqrt{16}} < Z < \frac{16-15}{2/\sqrt{16}} \right)$

$= P(-4 < Z < 2)$

$\cong 0.9772 - 0.0002$

$= 0.9770$

(c) (6 pts) For 100 randomly selected students, find an approximation to the probability that more than 40 of these students will need at least 14 hours on the homework (use continuity correction).

$P(X > 40)$

$= P(X > 40 + 0.5)$

$= P\left(Z > \frac{40.5 - 69.15}{4.62}\right)$

$= P(Z > -6.20)$

$\approx 1.$

Problem 4 (18 pts). A researcher is interested in evaluating a "diet+exercise" program for weight loss. He randomly selects 20 students who are over-weight, and randomly assigns these students into two groups, with 10 students in each group. Students in Group 1 will be on the "diet+exercise" program and students in Group 2 will be on the "diet only" program. For each subject, his/her weight is measured before he/she starts the program and after the end of the program, and the reduction in weight (before − after) is computed. The data approximately follow a normal distribution. Part of the data and some summary statistics are shown below

Group 1 (diet+exercise)				Group 2 (diet only)			
ID	Before	After	Reduction	ID	Before	After	Reduction
1	141	123	18	1	132	133	−1
2	129	125	4	2	128	125	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	139	128	11	10	130	125	5
sample mean	$\bar{x}_{11} = 135$	$\bar{x}_{12} = 126$	$\bar{x}_1 = 9$	sample mean	$\bar{x}_{21} = 133$	$\bar{x}_{22} = 128$	$\bar{x}_2 = 5$
sample SD	$s_{11} = 15$	$s_{12} = 11$	$s_1 = 10$	sample SD	$s_{21} = 12$	$s_{22} = 9$	$s_2 = 11$
pop. mean (SD)	μ_{11} (σ_{11})	μ_{12} (σ_{12})	μ_1 (σ_1)	pop. mean (SD)	μ_{21} (σ_{21})	μ_{22} (σ_{22})	μ_2 (σ_2)

(Note: ID = student ID number, SD = standard deviation, pop. = population. The sample means and sample standard deviations in the tables are computed from the corresponding columns in the table. In the following questions, use the notation defined in the above table.)

(a) (6 pts) The 95% confidence interval for the average reduction in weight for students in the "*diet+exercise*" program (Group 1) is

$\bar{X} \pm t^* \cdot s/\sqrt{n} = 9 \pm 2.262 * 10/\sqrt{10} = 9 \pm 7.15 \Rightarrow$ C.I.: $(1.85, 16.15)$

(b) (6 pts) Is there strong evidence that students in the "*diet only*" (Group 2) program lose weights? State your hypotheses and test the hypotheses.

$H_0: \mu_2 = 0$. vs. $H_a: \mu_2 > 0$

$t = \frac{5-0}{11/\sqrt{10}} = 1.44$. d.f $= 9$.

p-value $= P(t(9) > 1.44)$

$\Rightarrow 0.05 <$ p-value < 0.10

Some evidence support the deduction in weight ~~not~~ doesn't equal $\emptyset$, but not very strong.

(c) (6 pts) Is there strong evidence that students in the "diet+exercise" program (Group 1) lose more weights than those in the "diet only" program (Group 2)?

The hypotheses to be tested are: $H_0: \mu_1 = \mu_2$, vs. $H_a: \mu_1 > \mu_2$.

The value of the test statistic and its degrees of freedom are $\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 0.8508$. $df = 9$.

Is the above result statistically significant at 1% level? No (Yes or No)

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Problem 5 (13 pts). The average lifetime of type A batteries is 1000 hours, with a standard deviation of 100 hours. A researcher wants to know if type B batteries have shorter average lifetime than type A batteries. Assume that the standard deviation of type B battery lifetime is the same as type A batteries. Also assume that the lifetimes of both types of batteries follow normal distributions.

(a) (5 pts) The researcher plans to obtain a simple random sample and perform a hypothesis test. If the researcher wants the power of the test to be at least 90% when the true average lifetime of type B batteries is 950 hours, how many batteries are needed in the sample? (assume that the significance level is $\alpha = 0.05$.)

$H_0: \mu = 1000$, vs. $H_a: \mu < 1000$. $\sigma = 100$.

$$\text{power} = P\left(\frac{\bar{X}-1000}{100/\sqrt{n}} < -1.645 \mid \mu = 950\right)$$

$$= P\left(\bar{X} < 1000 - 1.645 \cdot 100/\sqrt{n} \mid \mu = 950\right)$$

$$= P\left(Z < \frac{1000-950-1.645 \cdot 100/\sqrt{n}}{100/\sqrt{n}} \mid \mu = 950\right)$$

$$= P\left(Z < \frac{50}{100/\sqrt{n}} - 1.645\right) \geq 0.90 \text{ (as required)}$$

$$\therefore \frac{50}{100/\sqrt{n}} - 1.645 \geq 1.28.$$

$$\therefore 0.5 \cdot \sqrt{n} \geq 1.28 + 1.645$$

$$\therefore n \geq \left(\frac{1.28+1.645}{0.5}\right)^2 = 34.2.$$ Need at least 35 batteries

(b) (2 pts) If the true average lifetime of type B batteries is 950, but the researcher fails to reject the null hypothesis that the true average lifetime of type B batteries is not shorter than type A batteries, then a type __II__ error has been committed.

(c) (6 pts) The researcher also wants to test if more than half (i.e., more than 50%) type B batteries have lifetimes shorter than 1000 hours. The null and alternative hypotheses should be

$H_0: p \leq 0.5$. vs. $H_a\ p > 0.5$.

Suggest a possible test statistic for this test:

$Z = (\hat{p} - 0.5)/\sqrt{0.5 \cdot (1-0.5)/n}$.

The approximate distribution of this test statistic under the null hypothesis should be

$N(0, 1)$.