Applied Statistics I **Spring 2009**

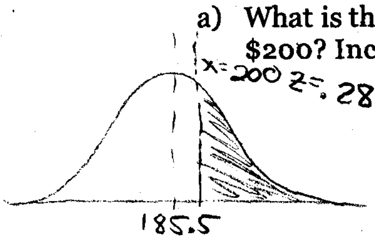
MIDTERM EXAM II
(75 points)

Name

Read each question carefully. Write your work and answers on this test paper or, if you prefer, you may attach printed sheets to this test paper. If you need more space to write, you may continue your answer on the reverse but please indicate that you have done so. You may use resources including SPSS, applets, and your books and notes.

1) **(20 points)** The average outstanding bill for delinquent customer accounts for a national department store chain is \$185.50 with a standard deviation of \$51.50. Assume the outstanding bill amounts are normally distributed.

a) What is the probability that a customer chosen at random has an outstanding bill over \$200? Include a sketch.



p=.3891 or 38.91%

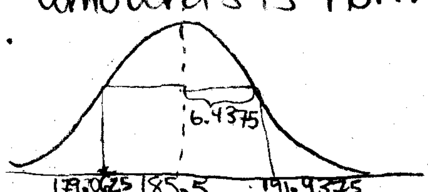
b) A simple random sample of 64 delinquent accounts is taken. What is the approximate sampling distribution of the mean outstanding bill for these accounts? (Note: Be sure to state the shape, mean, and standard deviation.) Include a sketch.

$\mu = 185.50$ $\sigma = 51.50$

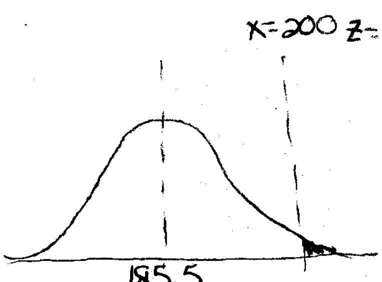
$\frac{\sigma}{\sqrt{n}} = \frac{51.5}{\sqrt{64}} = 6.4375$ ← new s.d.

$n=64$

Shape = Normal because the original graph of outstanding bill amounts is normal



c) What is the probability that the sample mean of the 64 delinquent accounts' outstanding bill is over \$200? Include a sketch.



p=.0121 or 1.21%

d) It is actually very unlikely that outstanding bill amounts are normally distributed. Given this fact, which, if any, of your previous answers, a), b), or c), is likely to be incorrect? Briefly explain why.

A would be incorrect because the calculations were done assuming a normal shape. With a different distribution the probability of being above 200 in a would appear differently. The others would still be valid because we have a large sample and because of the central limit theorem can assume a normal dist. for the sample.

2) (15 points) A Bureau of Labor Statistics report gives a 90% confidence interval for the average income of full-time female workers in 2001 as $28,944 to $29,486. This result was calculated by advanced methods from the Current Population Survey, a multistage random sample of about 50,000 households.

a) Would a 95% confidence interval be wider or narrower? Explain your answer.

A 95% confidence interval would be wider. To have a higher confidence, we must have a larger margin of error to include more possible values of μ, even if they are very unlikely.

b) Would the null hypothesis that the 2001 average income of all full-time female workers was $28,000 be rejected at the 10% significance level in favor of the two-sided alternative? What about the null hypothesis that the average was $29,000? Explain briefly how you can answer without doing any calculations.

The null hypothesis that mean income was $28,000 would be rejected at the 10% significance level. A level 0.1 significance test rejects $H_0: \mu = \mu_0$ when μ_0 falls outside a level 1-0.1 confidence interval for μ. $28,000 falls outside the 90% confidence interval (28,944, 29,486).

However, the H_0 that the mean 2001 income was $29,000 would not be rejected at the 10% significance level. $29,000 does fall inside the 90% confidence interval (28,944, 29,486).

c) If instead of a sample of 50,000 households, only 12,500 households were included in the study, find the 90% confidence interval. (Note: Assume that changing the sample size has the same effect here as it would if the sample were an SRS.)

$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ $\quad \bar{x} = ? \quad \sigma = ? \quad n = 12{,}500 \quad z^* = 1.645$

$\bar{x} - m = 28{,}944$

$\bar{x} + m = 29{,}486$

$\bar{x} = 29{,}486 - m$

$29{,}486 - m - m = 28{,}944$

$-2m = -542$

$m = 271$

$29{,}486 - 271$

$= 29{,}215 = \bar{x}$

$= 28{,}944 + 271$

$\boxed{\bar{x} = 29{,}215}$

$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 271$

$271 \times \sqrt{50{,}000} = z^* \sigma$

$= 60597.4$

For $n = 12{,}500$:

$\bar{x} \pm \frac{60597.4}{\sqrt{12500}} = \bar{x} \pm 542$

$= 29{,}215 \pm 542$

$= \boxed{28{,}673 \text{ to } 29{,}757}$

3) (20 points) An author of a new book claims that adults following her suggested diet program will lose an average of 5.8 pounds per week. A researcher believes that the true figure will be lower and carries out a test involving a random sample of 100 adults. The average weight loss of the 100 people under the program was 5.1 pounds per week. Assume the standard deviation is 4.5 pounds per week.

Are the results of this experiment good evidence that the true average weight loss differs from what is claimed by the author? Is the evidence significant at the $\alpha = 0.05$ level? Use the 4-step method to present your solution and be sure to interpret the *p*-value in context.

State: Do the data suggest good evidence against the hypothesis that the average weight loss per week (μ) is 5.8lb? Is this evidence significant at the $\alpha = 0.05$ level?

Formulate:

$H_0: \mu = 5.8$ (The average weight loss per week of all adults on this diet is 5.8 pounds)

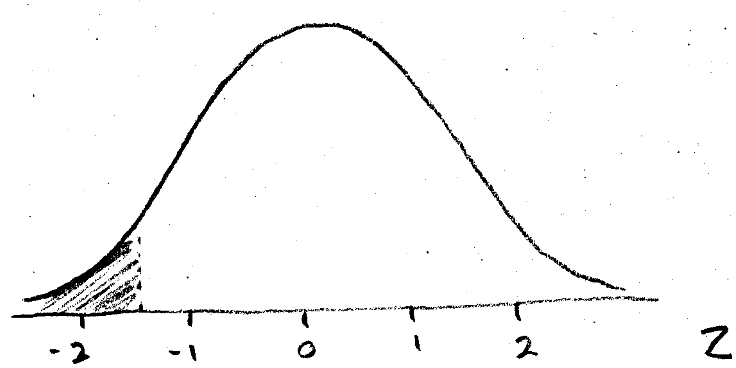
$H_a: \mu < 5.8$ (The average weight loss per week of all adults on this diet is less than 5.8 pounds)

Solve: The simple conditions have been met – it is an SRS with a Normal distribution and a known σ.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{5.1 - 5.8}{4.5/\sqrt{100}} = \frac{-0.7}{.45} = -1.56$$

$P(Z \leq -1.56) = 0.0594$

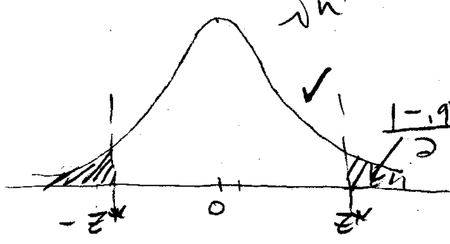
Conclude: The data give slight evidence against the null hypothesis. The probability that a sample would produce the result of an average weight loss of 5.1 pounds per week if the true mean were 5.8 pounds per week is 0.0594. This evidence is significant at the $\alpha = 0.10$ level, but not at the $\alpha = 0.05$ level.



4) **(20 points)** To answer this problem, use the information provided in Question 3).

a) Calculate a 97% confidence interval for the population average weight loss based on the sample. Provide a sketch showing how you obtained the value for z^* needed for your interval.

$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$ $\bar{x} = 5.1$ $\sigma = 4.5$ $n = 100$ $z^* = 2.17$



$5.1 \pm 2.17\left(\frac{4.5}{\sqrt{100}}\right) = 5.1 \pm 0.9765$
$= 4.12$ to 6.08

$\boxed{(4.12,\ 6.08)}$

b) How large a sample would be required to estimate the mean weight loss within ±0.5 pounds with 97% confidence?

$n = \left(\frac{z^* \sigma}{m}\right)^2$ $n = \left(\frac{2.17(4.5)}{0.5}\right)^2 = 381.4209$

A sample of 382 would be needed to estimate mean weight loss w/in ±0.5 pounds.

c) Upon further investigation, it is found that the sample of 100 people used for this study was drawn from the membership list of a national health club chain. Does this fact cause either the significance test you carried out in the previous problem or the confidence interval in a) to be invalid? Explain briefly why or why not.

This fact causes both the significance test from the previous problem and the confidence interval in a) to be invalid. If the sample was drawn from a membership list of a national health club chain, it is certainly not an SRS of all the people on the author's diet program, and will not be representative of the entire population. Our significance tests and confidence intervals do not compensate for flaws in collecting data; z-procedures are not appropriate for non-probability methods.