Name: __________________________ Time of Section: ________

All of your work on this exam must be entirely your own. You may use a calculator and an 8.5×11 sheet of notes. Remember to show your work.

1a) [5 pts.] Suppose 10 people in California are chosen at random, and their ages are recorded. These 10 ages are then plotted in a histogram. In view of the Central Limit Theorem, would you expect the histogram to look like a symmetric bell-shaped curve? Explain your answer in a sentence or two.

Solution: There is no reason to expect a bell-shaped curve. The Central Limit Theorem applies to averages, not individual ages, and even if the ages were normally distributed, we would not necessarily see a bell-shaped curve from just 10 observations.

b) [10 pts.] Suppose instead 1000 people in California are chosen at random, and their ages are recorded and plotted in a histogram. This time would you expect the histogram to look like a symmetric bell-shaped curve? Explain your answer in a sentence or two.

Solution: Even with 1000 people, we do not expect a bell-shaped curve because the Central Limit Theorem applies to sums and averages, not individual observations. Most likely the distribution of the ages of Californians is not normally distributed.

2a) [15 pts.] Suppose 1100 people in the United States are selected at random and asked about their views on abortion. Suppose that 45% describe themselves as pro-life. Find a 95% confidence interval for the proportion of all people in the United States who are pro-life.

Solution: The sample proportion is $\hat{p} = .45$, and for a 95 percent confidence interval the critical value is $z^* = 1.96$. The sample size is $n = 1100$, so the margin of error is

$$ME = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (1.96)\sqrt{\frac{(.45)(.55)}{1100}} \approx .0294.$$

Therefore, a 95% confidence interval is $(.45 - .0294, .45 + .0294) \approx (.431, .479)$.

b) [5 pts.] About how many people would we need to survey if we wanted to cut in half the width of the confidence interval computed in part a)?

Solution: Cutting the width of the interval in half requires multiplying the sample size by 4, so we would need to sample 4400 people.

3. Suppose the amount of time (in hours) you will have to wait for the next bus to arrive is a random variable with density

$$f(x) = \begin{cases} 8x & \text{if } 0 \le x \le 1/2 \\ 0 & \text{otherwise} \end{cases}$$

a) [10 pts.] What is the probability that you will have to wait at least 15 minutes for the next bus.

Solution: Letting X be the time to wait for the next bus, this is

$$P(X > 1/4) = \int_{1/4}^{1/2} 8x \, dx = 4x^2 \Big|_{x=1/4}^{x=1/2} = 1 - 1/4 = 3/4.$$

b) [10 pts.] What is the expected value of the amount of time you will have to wait for the next bus.

Solution: We have

$$E[X] = \int_0^{1/2} x \cdot 8x \, dx = \int_0^{1/2} 8x^2 \, dx = \frac{8x^3}{3} \Big|_{x=0}^{x=1/2} = \frac{1}{3} - 0 = \frac{1}{3}.$$

4. Suppose a company promoting a weight-loss program wants to establish that its program leads to more weight loss on average than the weight loss of 5 pounds achieved by a competitor's program. Suppose the new weight-loss program is tested on 46 people, who lose an average of 6.7 pounds with a standard deviation of 5.3 pounds.

a) [5 pts] State your null and alternative hypotheses.

Solution: Letting μ be the mean weight loss resulting from the new program, we test $H_0 : \mu = 5$ against $H_A : \mu > 5$.

b) [5 pts] Find a test statistic. Show the formula that you use, not just the numerical answer.

Solution: Letting $\bar{X}$ and s be the sample mean and sample standard deviation of the amount of weight the 46 people lost and letting n be the number of people in the study, our test statistic is

$$T = \frac{\bar{X} - 5}{s/\sqrt{n}} = \frac{6.7 - 5}{5.3/\sqrt{46}} \approx 2.175.$$

c) [5 pts] Find the p-value for your test (or find an interval containing the p-value, if that is the best you can do with your tables).

Solution: If H_0 is true, then T has a t-distribution with $n - 1 = 45$ degrees of freedom, so from the table we see that the p-value is between .01 and .025 (using the one-tailed probabilities). Using a calculator, we can find the exact p-value of .0174.

d) [5 pts] Write a sentence explaining the meaning of the p-value you calculated in part c, in the context of the problem. (If you got stuck before part c, pretend the p-value is .03 and use this number for parts d, e, and f.)

Solution: The p-value means that if there were in fact no difference between this weight loss program and the competitor's, then the 46 people in the sample would have lost as much weight as they did (or more) only about 1.7 percent of the time.

e) [5 pts] Are you able to reject the null hypothesis (use significance level .05)?

Solution: Because the p-value is less than .05, we reject H_0.

f) [5 pts] Write a sentence explaining your conclusion in the context of the problem.

Solution: Our data provide strong evidence that the new weight loss program is more effective than the competitor's.

5. [15 pts.] Suppose a company's annual costs (in millions of dollars) are normally distributed with a mean of 36 and a standard deviation of 3, and its annual revenues (in millions of dollars) are normally distributed with a mean of 46 and a standard deviation of 4. Assume that revenues and costs are independent. Find the probability that the company's revenues in the next year exceed its costs.

Solution: Let X be the company's costs, and let Y be the revenues, both in millions of dollars. We want to find $P(Y - X > 0)$. Note that $E[Y - X] = E[Y] - E[X] = 46 - 36 = 10$ and because X and Y are independent, $\text{Var}(Y - X) = \text{Var}(Y) + \text{Var}(X) = 3^2 + 4^2 = 25$, so the standard deviation of $SD(Y - X) = \sqrt{25} = 5$. Since X and Y are independent and normally distributed, $Y - X$ is normally distributed. Therefore, $Z = (Y - X - 10)/5$ has approximately a standard normal distribution. Thus, the probability that revenues exceed costs is

$$P(Y - X > 0) = P\left(\frac{Y - X - 10}{5} > \frac{0 - 10}{5}\right) = P(Z > -2) \approx .9772,$$

where the last number comes from tables.