Statistics 13V
Fall 2003
Final Exam Key

Name: ____________________________________

Last 6 digits of Student ID#: __________________

Open book and notes. If you need more space use the back of the page or a separate sheet of paper. You should have 5 questions on 4 pages, make sure you have them all.

1. (5 points each) Answer each of the following questions:
   a. Suppose A and B are two events each with probability greater than 0. Under what circumstances does P(A|B) = P(A)? If it is impossible, state that.

   This is true when A and B are independent events.

   b. Suppose A and B are two events each with probability greater than 0. Under what circumstances does P(A|B) = 0? If it is impossible, state that.

   This is true when A and B are mutually exclusive events.

   c. An ABC news poll asked visitors to its website to respond to a question about the economy and there were 17,251 responses. Explain why the results of this poll should not be generalized to the American population.

   The answer should address some aspect of the fact that the sample would not be representative of the American population. The most obvious reason is that only visitors to that website would even see the poll, thus excluding anyone who doesn't use the internet, who isn't likely to visit news sites even if they do use the internet, etc. Even those who do visit the site may not be interested in a question about the economy and thus would not respond.

   d. Is it possible for a statistically significant relationship between two variables to have little practical significance? Explain.

   Yes. This is most likely to happen if the sample size is large and the actual population relationship is weak.

   e. Suppose a relationship between two variables exists in a population and is of practical importance. Is it possible that in a sample from the population the two variables will not have a statistically significant relationship? Explain.

   Yes. This is most likely to happen with a small sample, in which case even if the relationship in the population is moderately large the test will have low power. Thus, the data may not exhibit a strong enough relationship to find it statistically significant.

2. (2 pts each) In each case, is the given percent a statistic or a parameter? Circle your answer.
   a. Of 10 students sampled from a high school, 8 (80%) said they would like the school library to have longer hours.

      Statistic   Parameter

   b. 75% of all students at UCD are in favor of having more bicycle cops on campus.

      Statistic   ***Parameter***

   c. A customs inspector sampled 5 passengers from among those landing on a large flight from Europe to the United States. He found that one of the five (20%) had an illegal food item.

      Statistic   Parameter

   d. Based on the 2000 Census, 39.5% of the California population over 5 years old speaks languages other than English at home.

      Statistic   ***Parameter***

   e. In a national poll it was found that 46% of respondents favored raising the minimum wage.

      Statistic   Parameter

3. (2 pts each) In each case, specify whether the statement applies to a statistic, a parameter or both. Circle your answer.
   a. It has a sampling distribution.

      Statistic   Parameter   Both

   b. It doesn't change from one sample to the next from the same population.

      Statistic   ***Parameter***   Both

   c. Examples of notation used for it are $\bar{x}$ and $\hat{p}$.

      Statistic   Parameter   Both

   d. It is a random variable.

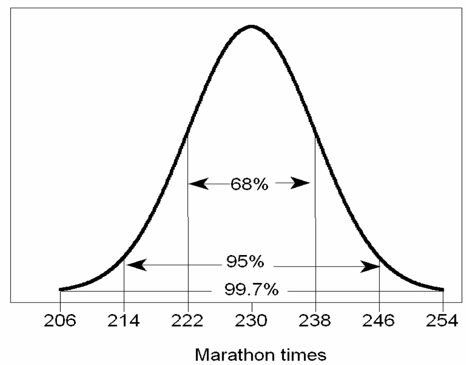
      Statistic   Parameter   Both

   e. The mean of the sampling distribution for a sample proportion is an example of one of these.

      Statistic   ***Parameter***   Both

4. (5 pts each) In order to qualify for the Boston Marathon, a woman in the 18-34 year old age group must run an official qualifying marathon in 3 hours and 40 minutes (220 minutes) or less. Samantha is 25 years old. Suppose the times at which she can run a marathon are normally distributed with mean of 230 minutes and standard deviation of 8 minutes.

   a. Draw a picture of the population of marathon times for Samantha. Show the appropriate ranges into which 68%, 95% and 99.7% of her times would fall.



   b. If Samantha runs one marathon, what is the probability that her time will qualify her for the Boston marathon? Show your work.

To qualify her time must be $\leq$ *220 minutes. P(Time* $\leq$ *220) = P(Z* $\leq \frac{220-230}{8}$*) = P(Z*$\leq$*−1.25) = .1056.*

   c. Samantha plans to run 4 marathons in the next year. Suppose her finishing times are independent from one marathon to the next. What is the probability that she has at least one time that qualifies her for the Boston Marathon? Show your work. (Note: You will need to use your answer from part b. If you couldn't solve part b, call the result you would have found *p* and use it here, showing the formula you would use.)

First find the probability that none of her times will qualify. For each race, the probability that she will not qualify is (1 − .1056) = .8944. Therefore, the probability that she will not qualify based on any of the four races is $(.8944)^4$ *= .6399. The complement of that event is that she will qualify based on at least one of the races, so the probability is 1 − .6399 = .3601.*

   d. Refer to part c. Describe the sampling distribution for the mean of her four times. Be sure to include the shape, mean and standard deviation.

The sampling distribution is approximately normal with mean = her mean time = 230 minutes. The standard deviation is $\sigma/\sqrt{n}$ *= 8/2 = 4 minutes.*

   e. If the Boston Marathon were to change its rules so that people could qualify using the *mean* time from four races, would Samantha's probability of qualifying increase or decrease from what it is based on one race (from part b)? Assume the mean would still need to be 220 minutes or less. Show your work.

$P(\bar{x} \leq 220) = P(Z \leq \frac{220-230}{4}) = P(Z \leq -2.5) = .0062.$ *So, her probability of qualifying would decrease.*

5. A survey done in early February asked a random sample of 1200 adults whether or not they had made a New Year's resolution on January 1st. Of the 1200 asked, 756 said they had.

   a. (10 pts) Find a 95% confidence interval for the proportion of all adults who made a New Year's resolution that year. Show your work.

A 95% confidence interval for a proportion is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. *In this case,* $\hat{p} = \frac{756}{1200} = .63$

So a 95% confidence interval is $.63 \pm 1.96\sqrt{\frac{.63(1-.63)}{1200}}$ *or .63 ± (1.96)(.0139) or .63 ± .027*

So the confidence interval is .603 to .657.

   b. (20 pts) The 756 respondents who said they had made a resolution were then asked whether or not they were still keeping it, now that it was a month later. Of the 756 in this sample, 401 said they were still keeping it. Is this sufficient evidence to conclude that a majority (more than 50%) of those who made a New Year's resolution were still keeping it in early February? Use $\alpha = 0.05$. Show your work.

The parameter of interest is p = the proportion of the population who had made a New Year's resolution, who were still keeping it in early February. The sample proportion is $\hat{p} = \frac{401}{756} = .53$.

Step 1 (5 pts): *Specify the hypotheses.*

The hypotheses are:
H_0: $p = .50$
H_a: $p > .50$

Step 2 (5 pts): *Verify necessary conditions and compute the test statistic.*

The conditions require that the sample size be large enough, which clearly holds.

The test statistic is $z = \frac{\hat{p} - .50}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.53 - .50}{\sqrt{\frac{.5(1-.5)}{756}}} = \frac{.03}{.0182} = 1.65$

Step 3 (5 pts): *Find the p-value.*
The test is one-sided, so find the area above 1.65 on a standard normal curve. From Table A.1, the area above 1.65 is .0495, so p-value = .0495.

Step 4 (2 pts): *Make a conclusion about statistical significance.*
The p-value is less than 0.05, so the result is statistically significant. Reject the null hypothesis.

Step 5 (3 pts): *Make a conclusion in context.*
There is sufficient evidence to conclude that a majority of those who make a New Year's resolution are still keeping it in early February.