AP Statistics- Practice Exam (2018)

SECTION 2:
Free Response Questions 1-5 spend approximately 65 minutes

Free Response Question 6 spend approximately 25 minutes

Name: ______________________________

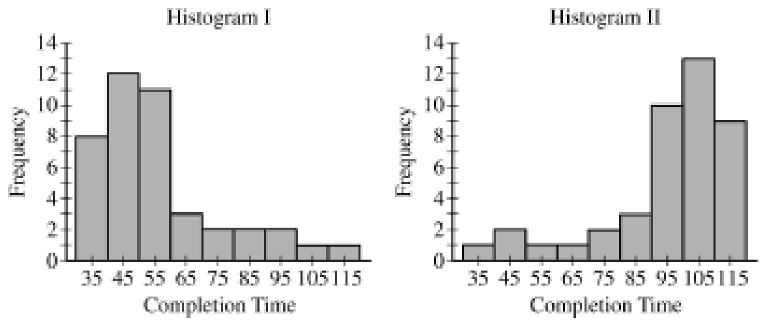
Date: ___________

Period: ______

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

---

1. The students enrolled in honors biology at a high school were given the task of using a spreadsheet program to investigate a topic in genetics. All students in the class had similar background knowledge of the topic. Some students in the class had no spreadsheet experience, Group R, and needed time to learn the program to complete the task. The rest of the students, Group S, had previous spreadsheet experience and typically took less time to complete the task. Each of the histograms below show the distribution of completion times, in minutes for one of the two groups.



a) Of the two histograms, I and II, which is more likely to be the distribution of completion times for the students in Group R? Justify your answer.

b) Describe the shape of a histogram created from the data of the two groups of students combined.

c) Consider the population of all students in honors biology classes in the high school's state who are given the task of using the spreadsheet program to investigate the topic in genetics. The distribution of the completion times has a shape similar to the combined histograms of the students at the high school, with mean 70 minutes and standard deviation 26.5 minutes. For random samples of 50 students taken from the population, describe the sampling distribution of the sample mean completion time.

2. Researchers are designing an experiment to compare two different types of running shoes, jA and B, to investigate which type is better for minimizing running time for a one-mile run. The experiment will consist of distributing the shoes to runners who are classified as either professional or recreational. A randomized block design is planned, with blocking by classification of runner. Random samples of 50 professional runners and 50 recreational runners will be selected. Each runner within each classification will be randomly assigned to wear either the type A shoe or the type B shoe, and their running time will be recorded for a one-mile run.

a) What is a statistical advantage of blocking by the classification of runner?

b) Why is it important to randomize the type of shoe the runner will wear instead of allowing the runner to choose the shoe?

c) Explain how the design of the experiment will address replication. What is the benefit of the replication?

3. A large university offers STEM (science, technology, engineering, and mathematics) internships to women in STEM majors at the university. A woman must be 20 years or older to meet the age requirement for the internships. The table show the probability distribution of the ages of the women in STEM majors at the university.

Age (years)	17	18	19	20	21	22	23 or older
Probability	0.005	0.107	0.111	0.252	0.249	0.213	0.063

a) Suppose one woman is selected at random from the women in STEM majors at the university. What is the probability that the woman selected will **not** meet the age requirement for the internships?

The university will select a sample of 100 women in STEM majors to participate in a focus group about the internships.

b) Suppose a simple random sampling process is used to select the sample of 100 women. What is the probability that at least 30 percent of the women in the sample will **not** meet the age requirement for the internships?

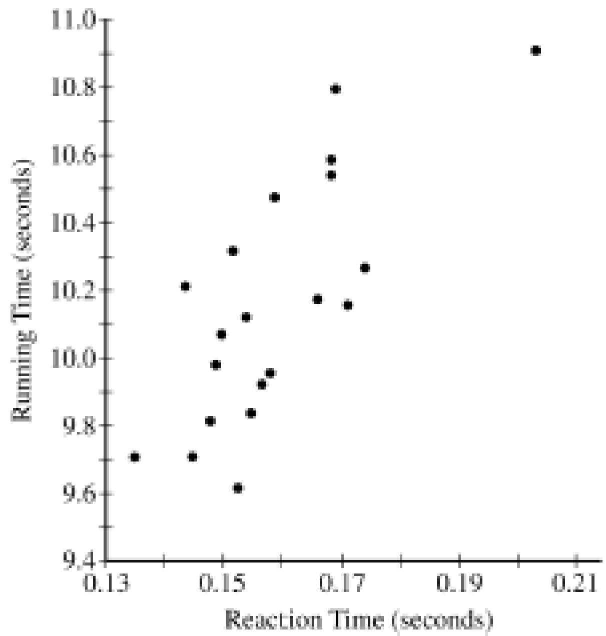
c) Suppose a stratified random sampling design is used to select a sample of 30 women who do **not** meet the age requirement and a sample of 70 women who do meet the age requirement. Based on the probability distribution is a woman who does **not** meet the age requirement more likely, less likely, or equally likely to be selected with a stratified random sample than with a simple random sample? Justify your answer.

4. Activity trackers are electronic devices that people wear to record physical activity. Researchers wanted to estimate the mean number of steps taken on a typical workday f0r people working in New York City who wear such trackers. A random sample of 61 people working in New York City who wear an activity tracker was selected. The number of steps taken on a typical workday for each person in the sample was recorded. The mean was 9,797 steps and the standard deviation was 2,313 steps.

a) Construct and interpret a 99 percent confidence interval for the mean number of steps taken on a typical workday for all people working in New York City who wear an activity tracker.

b) A wellness director at a company in New York City wants to investigate whether it is unusual for one person working in the city who wears an activity tracker to record approximately 8,500 steps on a typical workday. Is it appropriate to use the confidence interval found in part (a) to conduct the investigation? Explain your answer.

5. The total race time for a 100-meter dash can be considered the sum of two variables: the reaction time to the starting signal and the running time for 100 meters. The scatter plot shows reaction times and running times for 20 runners in a certain race. The winner was the runner with the least total race time.



a) Circle the point on the graph that represents the runner who won the race and approximate the total race time for that runner.

b) Based on the graph, is it reasonable to assume that reaction time and running timer are independent? Justify your answer.

c) Based on the least-squares regression model created from the data, explain why the use of extrapolation to predict the running time for a runner whose reaction time is 0.30 second might not be appropriate.

6. A large company produces an equal number of brand-name light bulbs and generic light bulbs. The director of quality control sets guidelines that production will be stopped if there is evidence that the proportion of **all** light bulbs that are defective is greater than 0.10. The director also believes that the proportion of brand-name light bulbs that are defective is not equal to the proportion of generic light bulbs that are defective. Therefore, the director wants to estimate the average off the two proportions.

To estimate the proportion of brand-name light bulbs that are defective a simple random sample of 400 brand-name light bulbs is taken and 44 are found to be defective. Let X represent the number of brand-name light bulbs that are defective in a sample of 400, and let p_X represent the proportion of all brand-name light bulbs that are defective. It is reasonable to assume that X is a binomial random variable.

a) One condition for obtaining an interval estimate for p_X is that the distribution of $\hat{p}_X$ is approximately normal. Is it reasonable to assume that the condition is met? Justify your answer.

b) The standard error of $\hat{p}_X$ is approximately 0.0156. Show how the value of the standard error is calculated.

c) How many standard errors is the observed value of $\hat{p}_X$ from 0.10?

To estimate the proportion of generic light bulbs that are defective, a simple random sample of 400 generic light bulbs is taken and 104 are found to be defective. Let Y represent the number of generic light bulbs that are defective in a sample of 400. It is reasonable to assume that Y is a binomial random variable and the distribution of $\hat{p}_Y$ is approximately normal, with an approximate standard error of 0.0219. It is also reasonable to assume that X and Y are independent.

The parameter of interest for the manager of quality control is D, the average proportion of defective light bulbs for the brand-name and the generic light bulbs. D is defined as $D = \frac{p_X + p_Y}{2}$.

d) Consider $\hat{D}$, the point estimate of D.

(i) Calculate $\hat{D}$ using data from the sample of brand-name light bulbs and the sample of generic light bulbs.

(ii) Calculate $s_{\hat{D}}$, the standard error of $\hat{D}$.

Consider the following hypotheses.

H_o: $The\ average\ proportion\ of\ all\ light\ bulbs\ that\ are\ defective\ is\ 0.10\ (D = 0.10)$

H_a: $The\ average\ proportion\ of\ all\ light\ bulbs\ that\ are\ defective\ is\ greater\ than\ 0.10\ (D > 0.10)$

A reasonable test statistic for the hypothesis is *W*, defined as $W = \frac{\hat{D} - 0.10}{s_{\hat{D}}}$.

e) Calculate W using your answer to part (d).

f) Chebyshev's inequality states that the proportion of any distribution that lies within *k* standard errors of the mean is at least…

$$1 - \frac{1}{k^2}$$

Use Chebyshev's inequality and the value of W to decide whether there is statistical evidence, at the significance level of $\alpha = 0.05$, that *D* , the average proportion of all light bulbs that are defective, is greater than 0.10.

STOP