My signature on this assessment confirms I have used no outside resources and adhered to all assessment protocols assigned to this daily grade/quiz/test/exam.

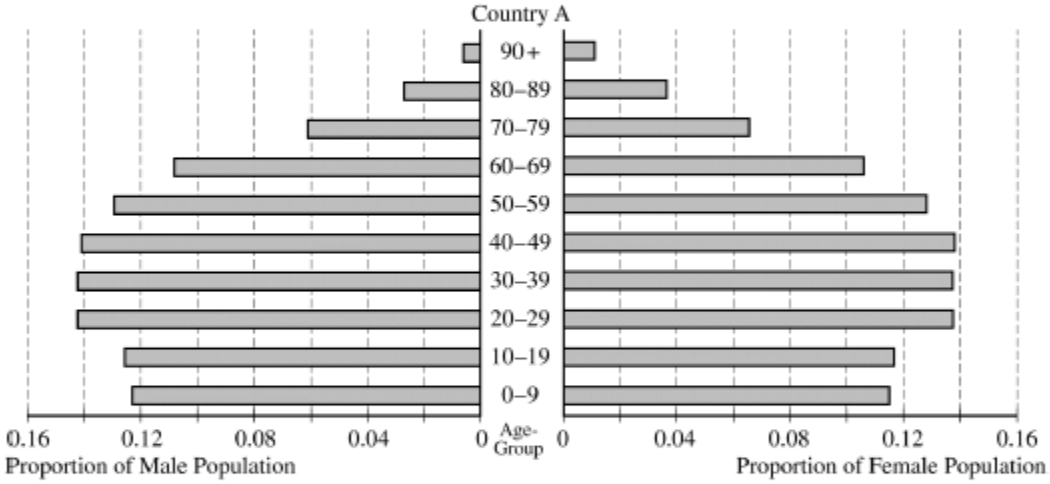
_____

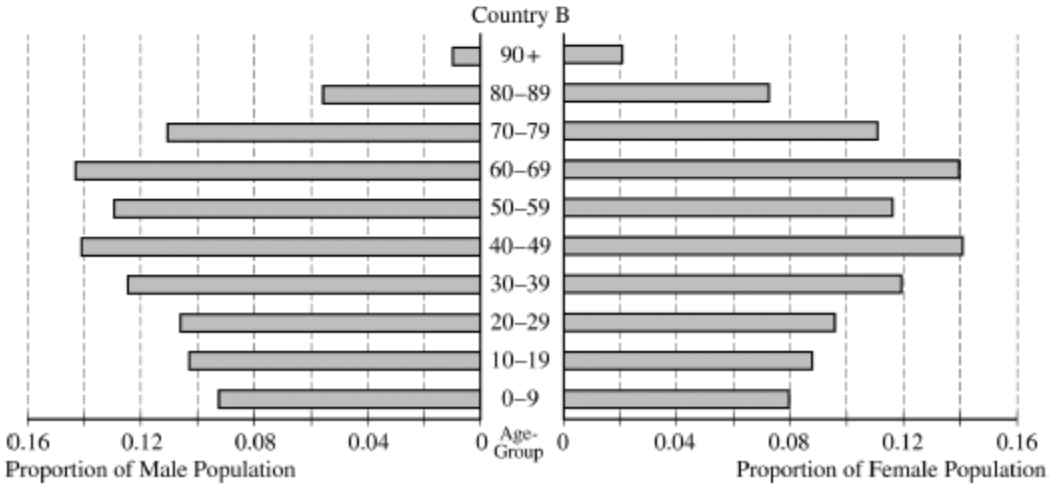
AP STAT PRACTICE EXAM #1

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

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

1. Population pyramids are a type of bar chart that show the distribution of ages of a country's population. The distributions of ages of men and women for two countries, A and B, for the year 2015 are shown in the population pyramids below. The age-groups, in years, are listed in the center columns, and the proportions are shown on the horizontal axes. Each bar represents the proportion of the age-group in the population for that sex.





(a) Is the proportion of the female population age 60 or older in Country A greater than the proportion of the female population age 60 or older in Country B? Justify your answer.

(b) One of the two countries experienced an increase in the birth rate in the years 1946 to 1955 and another increase in the birth rate about 20 years later. Based on the graphs, which country experienced the described increases in birth rate? Justify your answer using information from the graphs.

(c) For Country A, in which age-group is the median age of the male population? Justify your answer.

2. The ability to visually search, such as when reading an x-ray or interpreting a satellite image, is an important skill in many jobs. Researchers conducted a study to investigate whether playing video games could improve a person's ability to visually search. Three video games were used in the study: one was a driving game, one was a sports game, and one was a puzzle game. The participants consisted of 60 volunteers who had no experience playing video games before the study. Each participant was randomly assigned to one of the three games so that there were 20 participants per game.

   (a) Describe an appropriate method for randomly assigning 60 participants to three groups so that each group has 20 participants.

   The time to complete a visual search task was recorded for each participant before the assigned game was played. The time to complete a visual search task was again recorded for each participant after the assigned game was played. For each game, the mean improvement time (time before minus time after) was calculated.

   (b) One researcher expressed an interest in investigating whether playing a driving game would lead to a different mean improvement time to complete a visual search task than would playing a sports game. Assuming the appropriate population values are approximately normally distributed, state the name of a test with the appropriate null and alternative hypotheses that could be used to investigate the researcher's interest.

(c) When the appropriate test was performed, no significant difference was found in mean improvement time between the driving game and the sports game. Name one change the researchers could make in a future study to increase the power of the test. Explain why such a change would increase the power.

3. In women's tennis, a player must win 2 out of 3 sets to win a match. If a player wins the first 2 sets, she wins the match and the third set is not played. Player V and Player M will compete in a match.

   (a) Let V represent the event that Player V wins a set, and let M represent the event that Player M wins a set.

      (i) List all possible sequences of events V and M by set played that will result in Player V winning the match.

      (ii) List all possible sequences of events V and M by set played that will result in Player M winning the match.

   Player V and Player M have competed against each other many times. Historical data show that each player is equally likely to win the first set. If Player V wins the first set, the probability that she will win the second set is 0.60. If Player V loses the first set, the probability that she will lose the second set is 0.70. If Player V wins exactly one of the first two sets, the probability that she will win the third set is 0.45.

   (b) What is the probability that Player V will win a match against Player M?

   (c) What is the probability that a match between Player V and Player M will consist of 3 sets given that Player V wins the match?

4. A sociologist was investigating the ages of grandparents of high school students. From a random sample of 10 high school students, the sociologist collected data on the current ages, in years, of the students' maternal grandparents. The data are shown in the table below.

	Student											
	A	B	C	D	E	F	G	H	I	J	Mean	Standard Deviation
Age of grandmother	75	70	55	75	55	70	80	70	74	74	69.8	8.38
Age of grandfather	74	75	65	67	60	78	83	74	70	70	71.6	6.65
Difference	1	–5	–10	8	–5	–8	–3	–4	4	4	–1.8	5.81

(a) Construct and interpret a 95 percent confidence interval for the population mean difference in age (age of grandmother minus age of grandfather) of the maternal grandparents of high school students.

(b) One of the sociologist's research questions was about the mean difference in age without regard to which grandparent is older. The interval constructed in part (a) does not address such a question. Based on the sample of high school students, give the value of the point estimate for the mean difference in age that could be used to address the sociologist's question.

5. An automobile manufacturer sold 30,000 new cars, one to each of 30,000 customers, in a certain year. The manufacturer was interested in investigating the proportion of the new cars that experienced a mechanical problem within the first 5,000 miles driven.

   (a) A list of the names and addresses of all customers who bought the new cars is available. Describe a sampling plan that could be used to obtain a simple random sample of 1,000 customers from the list.

Each customer from a simple random sample of 1,000 customers who bought one of the new cars was asked whether they experienced any mechanical problems within the first 5,000 miles driven. Forty customers from the sample reported a problem. Of the 40 customers who reported a problem, 13 customers, or 32.5%, reported a problem specifically with the power door locks.

   (b) Explain why 0.325 should not be used to estimate the population proportion of the 30,000 new cars sold that experienced a problem with the power door locks within the first 5,000 miles driven.

   (c) Based on the results of the sample, give a point estimate of the number of new cars sold that experienced a problem with the power door locks within the first 5,000 miles driven.

STATISTICS

SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

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

6. Emily walks every day, and she keeps a record of the number of miles she walks each day. The histogram and five-number summary below were created from the recorded miles for a random sample of 25 of the days Emily walked.

Minimum	Q1	Median	Q3	Maximum
1.4	2.6	3.25	3.8	7.5

On one of the 25 days in the sample, Emily walked 7.5 miles. From the histogram, it appears that the value 7.5 might be an outlier relative to the other values. Two methods are proposed for identifying an outlier in a set of data.

(a) One method for identifying an outlier is to use the interquartile range (IQR). An outlier is any number that is greater than the upper quartile by at least 1.5 times the IQR or less than the lower quartile by at least 1.5 times the IQR. Does such a method identify the value of 7.5 miles as an outlier for Emily's set of data? Justify your answer.

Another method of identifying an outlier is to investigate whether there is evidence that a value might have come from a population with a mean different from the mean of the population of the other values.

Let X and Y represent random variables. X is distributed normally with mean μ_x and standard deviation σ, and Y is distributed normally with mean μ_y and standard deviation σ. Consider 1 randomly selected value of Y and $n-1$ randomly selected values of X.

(b) Consider the difference $Y - \overline{X}$.

(i) In terms of μ_y and μ_x, what is the mean of the difference $Y - \overline{X}$?

(ii) In terms of n and σ, what is the standard deviation of the difference $Y - \overline{X}$?

Suppose that of the $n = 25$ recorded values from Emily's sample, the value of 7.5 comes from the distribution of Y and the remaining 24 values come from the distribution of X. The summary statistics for the 24 values that come from the distribution of X are given below.

$n-1=24$
$\overline{x}=3.171$
$s=0.821$

(c) Use the value of the potential outlier and the summary statistics of the remaining 24 values to estimate the mean and standard deviation of the difference $Y - \overline{X}$.

The estimated mean of $Y - \overline{X}$:

The estimated standard deviation of $Y - \overline{X}$:

Recall that a method for identifying an outlier is to investigate whether there is evidence that a value might have come from a population with a mean different from the mean of the population of the other values. The following hypotheses can be used for such an investigation.

$$H_0 : \mu_y = \mu_x$$
$$H_a : \mu_y \neq \mu_x$$

(d) Calculate the value of the test statistic used for evaluating the hypotheses.

(e) The p-value for the hypothesis test described above is less than 0.0001. What conclusion can be made about the population means, and what conclusion can be made about identifying 7.5 as an outlier? Justify your answers.