2022

AP CollegeBoard

AP® Statistics

Free-Response Questions

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

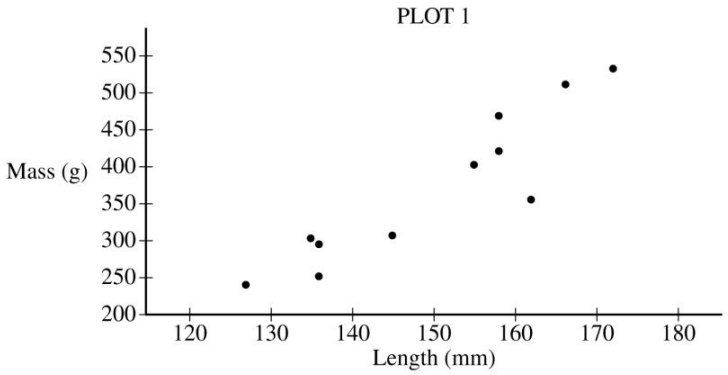
Part A

Suggested Time—1 hour and 5 minutes

5 Questions

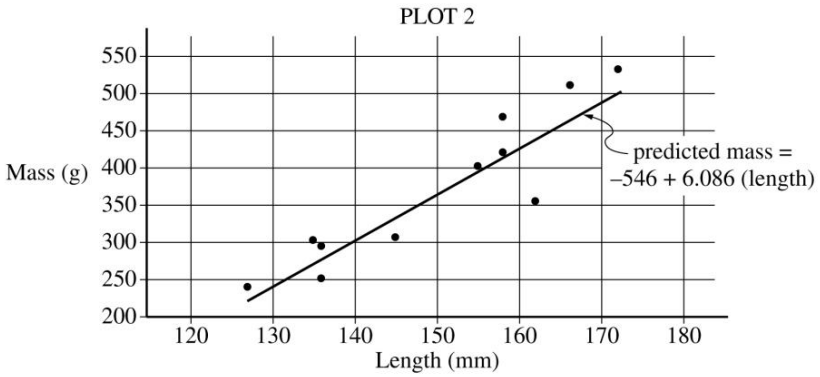
Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. A biologist gathered data on the length, in millimeters (mm), and the mass, in grams (g), for 11 bullfrogs. The data are shown in Plot 1.



(a) Based on the scatterplot, describe the relationship between mass and length, in context.

From the data, the biologist calculated the least-squares regression line for predicting mass from length. The least-squares regression line is shown in Plot 2.



(b) Identify and interpret the slope of the least-squares regression line in context.

(c) Interpret the coefficient of determination of the least-squares regression line, $r^2 \approx 0.819$, in context.

(d) From Plot 2, consider the residuals of the 11 bullfrogs.

(i) Based on the plot, approximately what is the length and mass of the bullfrog with the largest absolute value residual?

(ii) Does the least-squares regression line overestimate or underestimate the mass of the bullfrog identified in part (d-i)? Explain your answer.

2. A dermatologist will conduct an experiment to investigate the effectiveness of a new drug to treat acne. The dermatologist has recruited 36 pairs of identical twins. Each person in the experiment has acne and each person in the experiment will receive either the new drug or a placebo. After each person in the experiment uses either the new drug or the placebo for 2 weeks, the dermatologist will evaluate the improvement in acne severity for each person on a scale from 0 (no improvement) to 100 (complete cure).

   (a) Identify the treatments, experimental units, and response variable of the experiment.

   - Treatments:
   - Experimental units:
   - Response variable:

   Each twin in the experiment has a severity of acne similar to that of the other twin. However, the severity of acne differs from one twin pair to another.

   (b) For the dermatologist's experiment, describe a statistical advantage of using a matched-pairs design where twins are paired rather than using a completely randomized design.

   (c) For the dermatologist's experiment, describe how the treatments can be randomly assigned to people using a matched-pairs design in which twins are paired.

3. A machine at a manufacturing company is programmed to fill shampoo bottles such that the amount of shampoo in each bottle is normally distributed with mean 0.60 liter and standard deviation 0.04 liter. Let the random variable A represent the amount of shampoo, in liters, that is inserted into a bottle by the filling machine.

   (a) A bottle is considered to be underfilled if it has less than 0.50 liter of shampoo. Determine the probability that a randomly selected bottle of shampoo will be underfilled. Show your work.

   After the bottles are filled, they are placed in boxes of 10 bottles per box. After the bottles are placed in the boxes, several boxes are placed in a crate for shipping to a beauty supply warehouse. The manufacturing company's contract with the beauty supply warehouse states that one box will be randomly selected from a crate. If 2 or more bottles in the selected box are underfilled, the entire crate will be rejected and sent back to the manufacturing company.

   (b) The beauty supply warehouse manager is interested in the probability that a crate shipped to the warehouse will be rejected. Assume that the amounts of shampoo in the bottles are independent of each other.

   (i) Define the random variable of interest for the warehouse manager and state how the random variable is distributed.

   (ii) Determine the probability that a crate will be rejected by the warehouse manager. Show your work.

   To reduce the number of crates rejected by the beauty supply warehouse manager, the manufacturing company is considering adjusting the programming of the filling machine so that the amount of shampoo in each bottle is normally distributed with mean 0.56 liter and standard deviation 0.03 liter.

   (c) Would you recommend that the manufacturing company use the original programming of the filling machine or the adjusted programming of the filling machine? Provide a statistical justification for your choice.

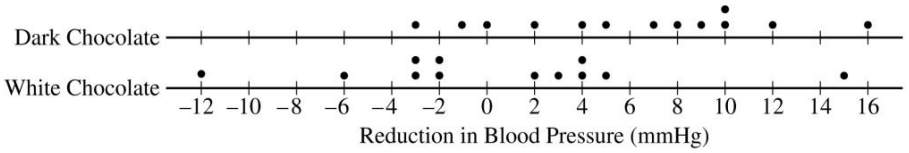
4. A survey conducted by a national research center asked a random sample of 920 teenagers in the United States how often they use a video streaming service. From the sample, 59% answered that they use a video streaming service every day.

   (a) Construct and interpret a 95% confidence interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.

(b) Based on the confidence interval in part (a), do the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would respond that they use a video streaming service every day is not 0.5? Justify your answer.

5. Studies have shown that foods rich in compounds known as flavonoids help lower blood pressure. Researchers conducted a study to investigate whether there was a greater reduction in blood pressure for people who consumed dark chocolate, which contains flavonoids, than people who consumed white chocolate, which does not contain flavonoids. Twenty-five healthy adults agreed to participate in the study and add 3.5 ounces of chocolate to their daily diets. Of the 25 participants, 13 were randomly assigned to the dark chocolate group and the rest were assigned to the white chocolate group. All participants had their blood pressure recorded, in millimeters of mercury (mmHg), before adding chocolate to their daily diets and again 30 days after adding chocolate to their daily diets.

   The reduction in blood pressure (before minus after) for each of the participants in the two groups is shown in the dotplots below.

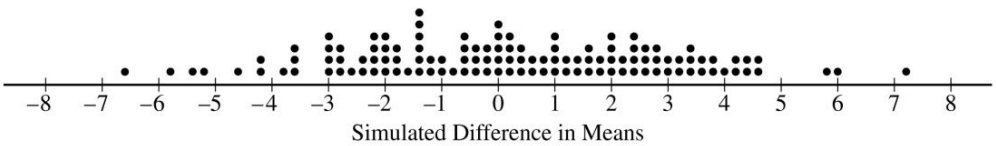


   (a) Determine and compare the medians of the reduction in blood pressure for the two groups.

   The researchers found the mean reduction in blood pressure for those who consumed dark chocolate is $\bar{x}_{dark} = 6.08$ mmHg and the mean reduction in blood pressure for those who consumed white chocolate is $\bar{x}_{white} = 0.42$ mmHg.

   (b) One researcher indicated that because the difference in sample means of 5.66 mmHg is greater than 0 there is convincing statistical evidence to conclude that the population mean blood pressure reduction for those who consume dark chocolate is greater than for those who consume white chocolate. Why might the researcher's conclusion, based only on the difference in sample means of 5.66 mmHg, not necessarily be true?

A simulation was conducted to investigate whether there is a greater reduction of blood pressure for those who consume dark chocolate than for those who consume white chocolate. The simulation was conducted under the assumption that no difference exists. The results of 120 trials of the simulation are shown in the following dotplot.



   (c) Use the results of the simulation to determine whether the results from the 25 participants in the study provide convincing statistical evidence, at a 5 percent level of significance, that adding dark chocolate to a daily diet will result in a greater reduction in blood pressure, on average, than adding white chocolate to a daily diet. Justify your answer.

STATISTICS

SECTION II, Part B

Suggested Time—25 minutes

1 Question

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. To compare success rates for treating allergies at two clinics that specialize in treating allergy sufferers, researchers selected random samples of patient records from the two clinics. The following table summarizes the data.

	Clinic A	Clinic B	**Total**
Unsuccessful treatment	51	33	84
Successful treatment	88	35	123
Total	139	68	207

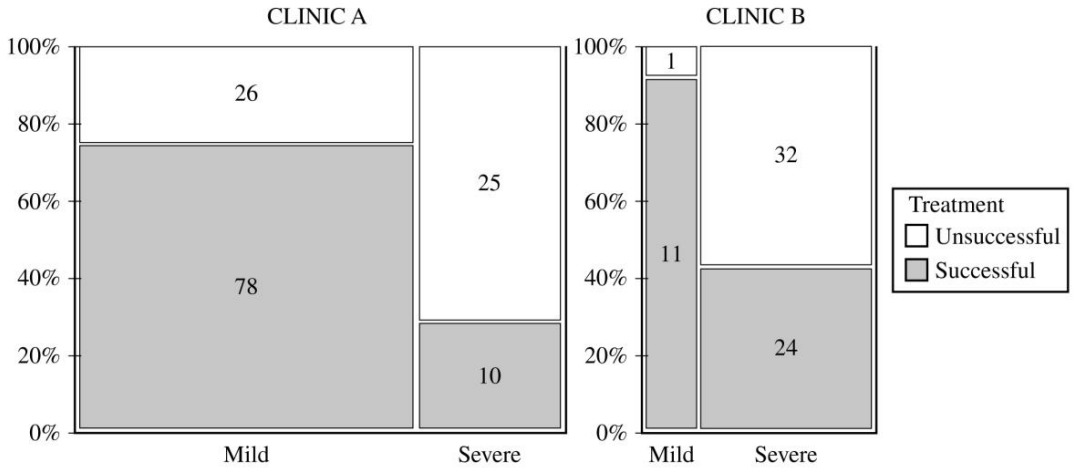
(a) (i) Complete the following table by recording the relative frequencies of successful and unsuccessful treatments at each clinic.

	Clinic A	Clinic B
Unsuccessful treatment		
Successful treatment		

(ii) Based on the relative frequency table in part (a-i), which clinic is more successful in treating allergy sufferers? Justify your answer.

(b) Based on the design of the study, would a statistically significant result allow the researchers to conclude that receiving treatments at the clinic you selected in part (a-ii) causes a higher percentage of successful treatments than at the other clinic? Explain your answer.

A physician who worked at both clinics believed that it was important to separate the patients in the study by severity of the patient's allergy (severe or mild). The physician constructed the following mosaic plot. The values in the mosaic plot represent the number of patients who were either successfully treated or unsuccessfully treated in each allergy severity group within each clinic. For example, the value 78 represents the number of patients successfully treated in the mild group within Clinic A.



Based on the mosaic plot, the physician concluded the following:

For mild allergy sufferers, **Clinic B** was more successful in treating allergies.

For severe allergy sufferers, **Clinic B** was more successful in treating allergies.

(c) (i) For each clinic, which allergy severity is treated more successfully? Justify your answer.

- Clinic A:

- Clinic B:

(ii) For each clinic, which allergy severity is more likely to be treated? Justify your answer.

- Clinic A:

- Clinic B:

(d) Using your answers from part (c), give a reasonable explanation of why the more successful clinic identified in part (a-ii) is the same as or different from the physician's conclusion that Clinic B is more successful in treating both severe and mild allergies.

STOP

END OF EXAM