



# AP<sup>®</sup> Statistics

## Free-Response Questions

### STATISTICS

### SECTION II

Total Time—1 hour and 30 minutes

6 Questions

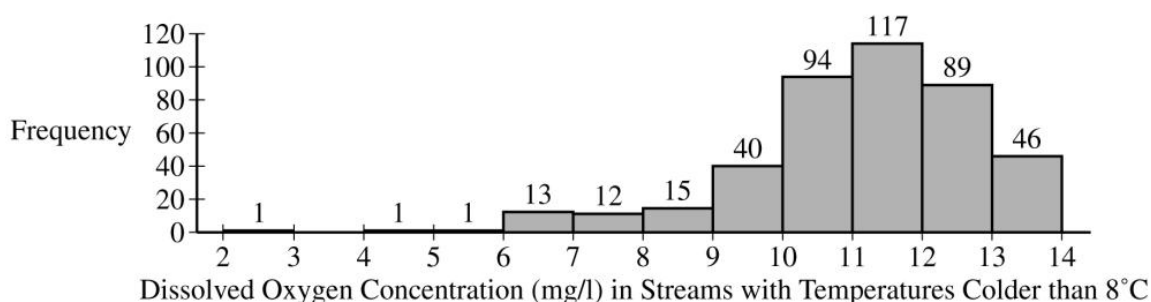
#### Part A

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. As part of a study on the chemistry of Alaskan streams, researchers took water samples from many streams with temperatures colder than  $8^{\circ}\text{C}$  and from many streams with temperatures warmer than  $8^{\circ}\text{C}$ . For each sample, the researchers measured the dissolved oxygen concentration, in milligrams per liter (mg/l).

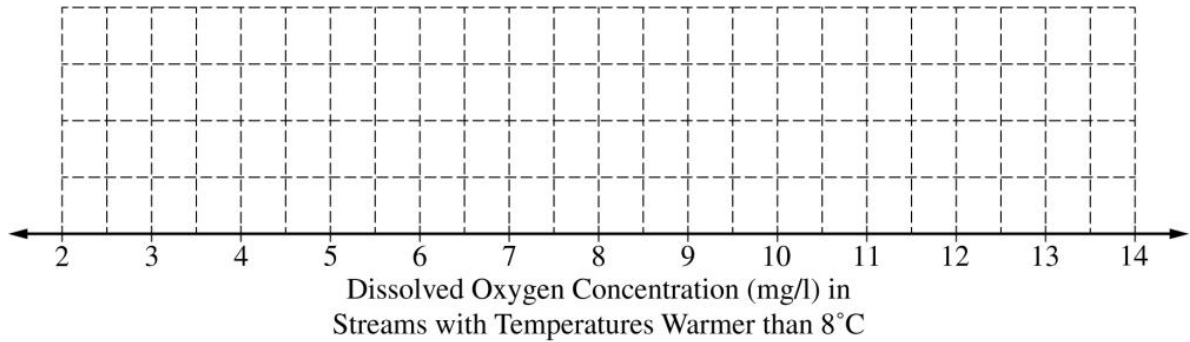


- (a) The researchers constructed the histogram shown for the dissolved oxygen concentration in streams from the sample with water temperatures colder than  $8^{\circ}\text{C}$ . Based on the histogram, describe the distribution of dissolved oxygen concentration in streams with water temperatures colder than  $8^{\circ}\text{C}$ .

Continue your response to **QUESTION 1** on this page.

Min	Q1	Median	Q3	Max	Mean	Std. Dev.
2.10	4.39	5.43	6.12	13.45	5.54	1.64

- (b) The researchers computed the summary statistics shown in the table for the dissolved oxygen concentration in streams from the sample with water temperatures warmer than  $8^{\circ}\text{C}$ . Use the summary statistics to construct a box plot for the dissolved oxygen concentration in streams with water temperatures warmer than  $8^{\circ}\text{C}$ . Do not indicate outliers.



- (c) The researchers believe that streams with higher dissolved oxygen concentration are generally healthier for wildlife. Which streams are generally healthier for wildlife, those with water temperature colder than  $8^{\circ}\text{C}$  or those with water temperature warmer than  $8^{\circ}\text{C}$ ? Using characteristics of the distribution of dissolved oxygen concentration for temperatures colder than  $8^{\circ}\text{C}$  and characteristics of the distribution of dissolved oxygen concentration for temperatures warmer than  $8^{\circ}\text{C}$ , justify your answer.

2. A developer wants to know whether adding fibers to concrete used in paving driveways will reduce the severity of cracking, because any driveway with severe cracks will have to be repaired by the developer. The developer conducts a completely randomized experiment with 60 new homes that need driveways. Thirty of the driveways will be randomly assigned to receive concrete that contains fibers, and the other 30 driveways will receive concrete that does not contain fibers. After one year, the developer will record the severity of cracks in each driveway on a scale of 0 to 10, with 0 representing not cracked at all and 10 representing severely cracked.

(a) Based on the information provided about the developer's experiment, identify each of the following.

- Experimental units
  
- Treatments
  
- Response variable

(b) Describe an appropriate method the developer could use to randomly assign concrete that contains fibers and concrete that does not contain fibers to the 60 driveways.

Suppose the developer finds that there is a statistically significant reduction in the mean severity of cracks in driveways using the concrete that contains fibers compared to the driveways using concrete that does not contain fibers.

(c) In terms of the developer's conclusion, what is the benefit of randomly assigning the driveways to either the concrete that contains fibers or the concrete that does not contain fibers?

3. Bath fizzies are mineral tablets that dissolve and create bubbles when added to bathwater. In order to increase sales, the Fizzy Bath Company has produced a new line of bath fizzies that have a cash prize in every bath fizzy. Let the random variable,  $X$ , represent the dollar value of the cash prize in a bath fizzy. The probability distribution of  $X$  is shown in the table.

Cash prize, $x$	\$1	\$5	\$10	\$20	\$50	\$100
Probability of cash prize, $P(X = x)$	$P(X = \$1)$	0.2	0.05	0.05	0.01	0.01

(a) Based on the probability distribution of  $X$ , answer the following. Show your work.

(i) Calculate the proportion of bath fizzies that contain \$1.

(ii) Calculate the proportion of bath fizzies that contain at least \$10.

(b) Based on the probability distribution of  $X$ , calculate the probability that a randomly selected bath fizzy contains \$100, given that it contains at least \$10. Show your work.

(c) Based on the probability distribution of  $X$ , calculate and interpret the expected value of the distribution of the cash prize in the bath fizzies. Show your work.

(d) The Fizzy Bath Company would like to sell the bath fizzies in France, where the currency is euros. Suppose the conversion rate for dollars to euros is 1 dollar = 0.89 euros. Using your expected value from part (c), calculate the expected value, in euros, of the distribution of the cash prize in the bath fizzies. Show your work.



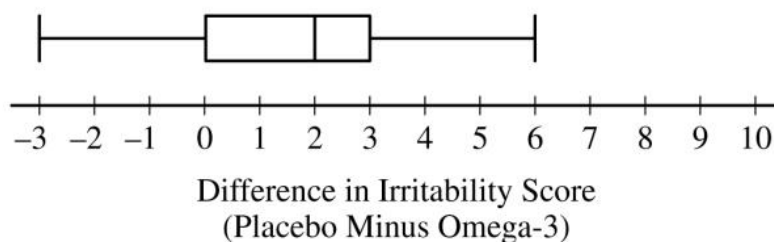
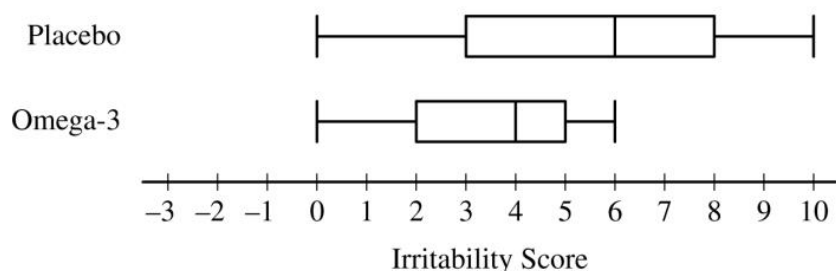
4. A medical researcher completed a study comparing an omega-3 fatty acids supplement to a placebo in the treatment of irritability in patients with a certain medical condition. Nineteen patients with the medical condition volunteered to participate in the study. The study was conducted using the following weekly schedule.

- Week 1: Each patient took a randomly assigned treatment, omega-3 supplement or placebo.
- Week 2: The patients did not take either the omega-3 supplement or the placebo. This was necessary to reduce the possibility of any carryover effect from the assigned treatment taken during week 1.
- Week 3: Each patient took the treatment, omega-3 supplement or placebo, that they did not take during week 1.

At the end of week 1 and week 3, each patient's irritability was given a score on a scale of 0 to 10, with 0 representing no irritability and 10 representing the highest level of irritability.

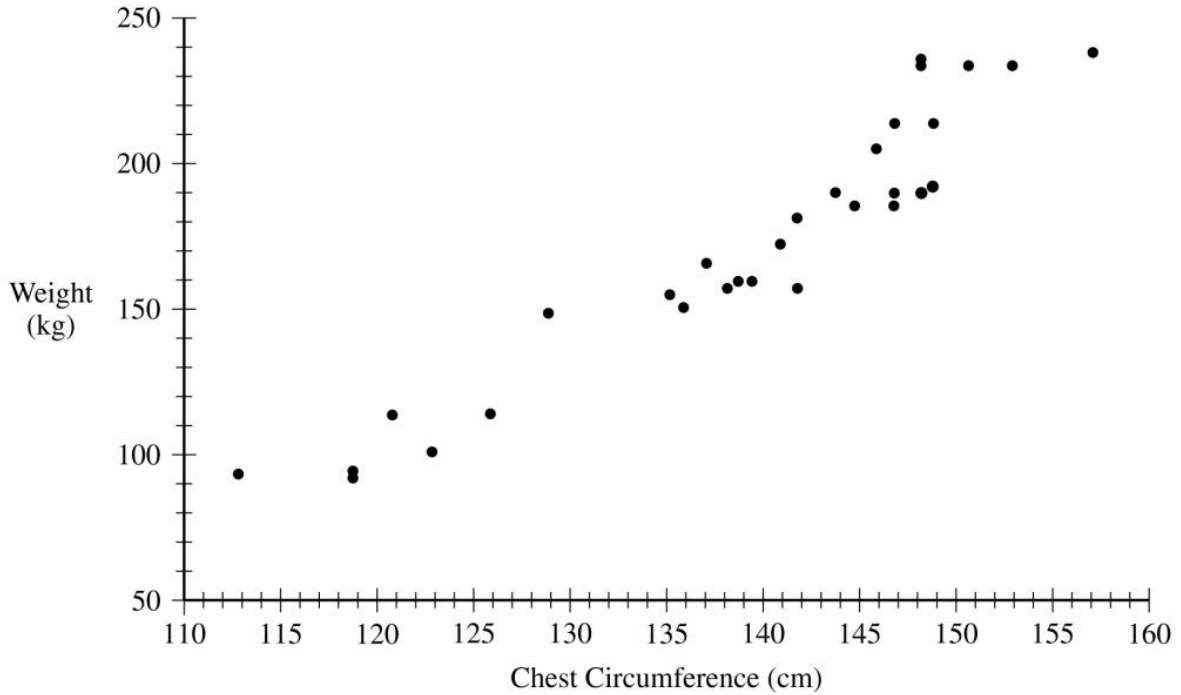
For each patient, the two irritability scores and the difference in their scores (placebo minus omega-3) were recorded. The results are summarized in the table and boxplots.

	<i>n</i>	Mean	Standard Deviation
Placebo	19	5.421	2.987
Omega-3	19	3.632	1.739
Difference (placebo minus omega-3)	19	1.789	2.485



The researcher claims the omega-3 supplement will decrease the mean irritability score of all patients with the medical condition similar to the volunteers who participated in the study. Is there convincing statistical evidence to support the researcher's claim at a significance level of  $\alpha = 0.05$ ? Complete the appropriate inference procedure to support your answer.

5. Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.



(a) Describe the relationship between chest circumference and weight of male tule elk in context.

Following is the equation of the least-squares regression line relating chest circumference and weight for male tule elk.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

(b) The weight of one male tule elk with a chest circumference of 145.9 cm is 204.3 kg.

(i) Using the equation of the least-squares regression line, calculate the predicted weight for this male tule elk. Show your work.

(ii) Calculate the residual for this male tule elk. Show your work.



The equation of the least-squares regression line relating chest circumference and weight for male tule elk is repeated here.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

(c) Interpret the slope of the least-squares regression line in context.

(d) The sambar, another species of deer, is similar in size to the tule elk. The slope of the population regression line relating chest circumference and weight for all male sambars is 4.5 kilograms per centimeter. A wildlife biologist wants to determine whether the slope of the population regression line for male tule elk is different than that for male sambars. Let  $\beta$  represent the slope of the population regression line for male tule elk. The wildlife biologist conducted a test of the following hypotheses using the sample of 30 tule elk.

$$H_0: \beta = 4.5$$

$$H_a: \beta \neq 4.5$$

The test statistic was calculated to be 3.408. Assume all conditions for inference were met.

(i) Determine the  $p$ -value of the test.

(ii) At a significance level of  $\alpha = 0.05$ , what conclusion should the wildlife biologist make regarding the slope of the population regression line for male tule elk? Justify your response.

## 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. A jewelry company uses a machine to apply a coating of gold on a certain style of necklace. The amount of gold applied to a necklace is approximately normally distributed. When the machine is working properly, the amount of gold applied to a necklace has a mean of 300 milligrams (mg) and standard deviation of 5 mg.
- (a) A necklace is randomly selected from the necklaces produced by the machine. Assuming that the machine is working properly, calculate the probability that the amount of gold applied to the necklace is between 296 mg and 304 mg.

The jewelry company wants to make sure the machine is working properly. Each day, Cleo, a statistician at the jewelry company, will take a random sample of the necklaces produced that day. Each selected necklace will be melted down and the amount of the gold applied to that necklace will be determined. Because a necklace must be destroyed to determine the amount of gold that was applied, Cleo will use random samples of size  $n = 2$  necklaces.

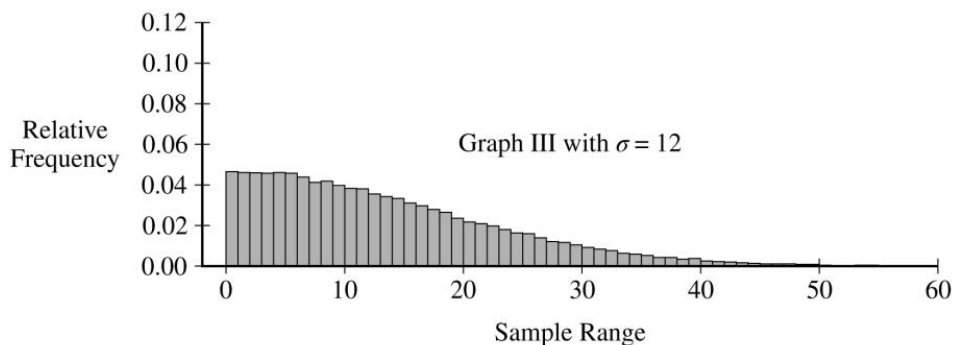
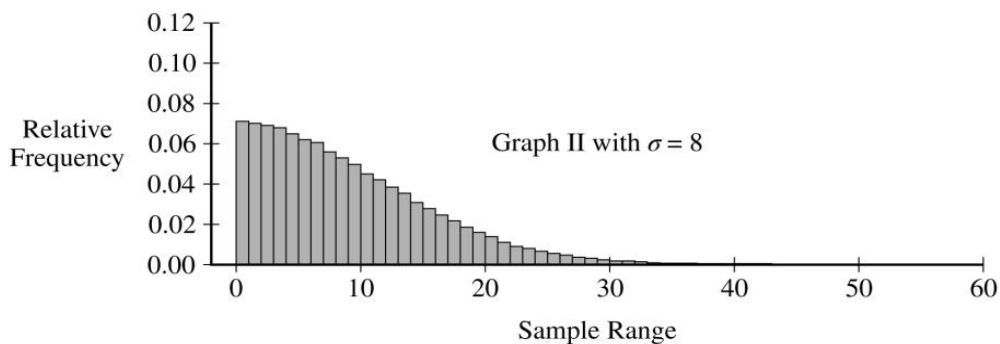
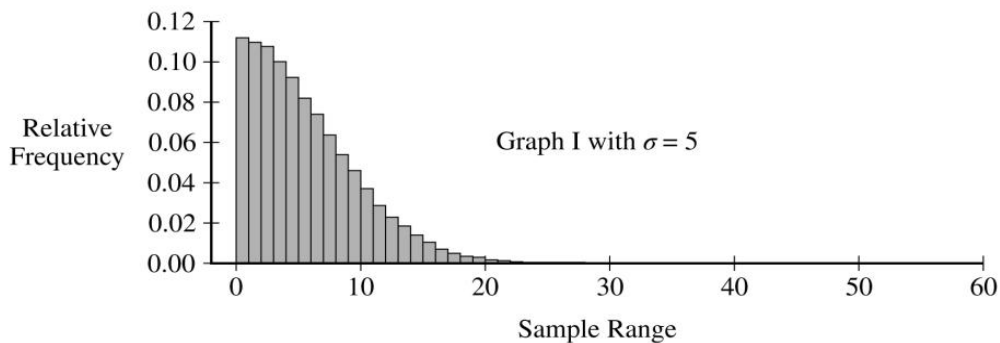
Cleo starts by considering the mean amount of gold being applied to the necklaces. After Cleo takes a random sample of  $n = 2$  necklaces, she computes the sample mean amount of gold applied to the two necklaces.

- (b) Suppose the machine is working properly with a population mean amount of gold being applied of 300 mg and a population standard deviation of 5 mg.
- (i) Calculate the probability that the sample mean amount of gold applied to a random sample of  $n = 2$  necklaces will be greater than 303 mg.

- (ii) Suppose Cleo took a random sample of  $n = 2$  necklaces that resulted in a sample mean amount of gold applied of 303 mg. Would that result indicate that the population mean amount of gold being applied by the machine is different from 300 mg? Justify your answer without performing an inference procedure.

Now, Cleo will consider the variation in the amount of gold the machine applies to the necklaces. Because of the small sample size,  $n = 2$ , Cleo will use the sample range of the data for the two randomly selected necklaces, rather than the sample standard deviation.

Cleo will investigate the behavior of the range for samples of size  $n = 2$ . She will simulate the sampling distribution of the range of the amount of gold applied to two randomly sampled necklaces. Cleo generates 100,000 random samples of size  $n = 2$  independent values from a normal distribution with mean  $\mu = 300$  and standard deviation  $\sigma = 5$ . The range is calculated for the two observations in each sample. The simulated sampling distribution of the range is shown in Graph I. This process is repeated using  $\sigma = 8$ , as shown in Graph II, and again using  $\sigma = 12$ , as shown in Graph III.



(c) Use the information in the graphs to complete the following.

(i) Describe the sampling distribution of the sample range for random samples of size  $n = 2$  from a normal distribution with standard deviation  $\sigma = 5$ , as shown in Graph I.

(ii) Describe how the sampling distribution of the sample range for samples of size  $n = 2$  changes as the value of the population standard deviation  $\sigma$  increases.

Recall that Cleo needs to consider both the mean and standard deviation of the amount of gold applied to necklaces to determine whether the machine is working properly. Suppose that one month later, Cleo is again checking the machine to make sure it is working properly. Cleo takes a random sample of 2 necklaces and calculates the sample mean amount of gold applied as 303mg and the sample range as 10mg.

(d) Recall that the machine is working properly if the amount of gold applied to the necklaces has a mean of 300mg and standard deviation of 5mg.

(i) Consider Cleo's range of 10mg from the sample of size  $n = 2$ . If the machine is working properly with a standard deviation of 5mg, is a sample range of 10mg unusual? Justify your answer.

(ii) Do Cleo's sample mean of 303mg and range of 10mg indicate that the machine is not working properly? Explain your answer.

**STOP**

**END OF EXAM**