

Cumulative AP Practice Test 1

Section I: Multiple Choice Choose the best answer for Questions AP1.1 to AP1.14.

AP1.1. You look at real estate ads for houses in Sarasota, Florida. Many houses range from \$200,000 to \$400,000 in price. The few houses on the water, however, have prices up to \$15 million. Which of the following statements best describes the distribution of home prices in Sarasota?

- (a) The distribution is most likely skewed to the left, and the mean is greater than the median.
- (b) The distribution is most likely skewed to the left, and the mean is less than the median.
- (c) The distribution is roughly symmetric with a few high outliers, and the mean is approximately equal to the median.
- (d) The distribution is most likely skewed to the right, and the mean is greater than the median.
- (e) The distribution is most likely skewed to the right, and the mean is less than the median.

AP1.2. A child is 40 inches tall, which places her at the 90th percentile of all children of similar age. The heights for children of this age form an approximately Normal distribution with a mean of 38 inches. Based on this information, what is the standard deviation of the heights of all children of this age?

- (a) 0.20 inches (c) 0.65 inches (e) 1.56 inches
- (b) 0.31 inches (d) 1.21 inches

AP1.3. A large set of test scores has mean 60 and standard deviation 18. If each score is doubled, and then 5 is subtracted from the result, the mean and standard deviation of the new scores are

- (a) mean 115; std. dev. 31. (d) mean 120; std. dev. 31.
- (b) mean 115; std. dev. 36. (e) mean 120; std. dev. 36.
- (c) mean 120; std. dev. 6.

AP1.4. For a certain experiment, the available experimental units are eight rats, of which four are female (F1, F2, F3, F4) and four are male (M1, M2, M3, M4). There are to be four treatment groups, A, B, C, and D. If a randomized block design is used, with the experimental units blocked by gender, which of the following assignments of treatments is impossible?

- (a) A → (F1, M1), B → (F2, M2),
C → (F3, M3), D → (F4, M4)
- (b) A → (F1, M2), B → (F2, M3),
C → (F3, M4), D → (F4, M1)
- (c) A → (F1, M2), B → (F3, F2),
C → (F4, M1), D → (M3, M4)
- (d) A → (F4, M1), B → (F2, M3),
C → (F3, M2), D → (F1, M4)
- (e) A → (F4, M1), B → (F1, M4),
C → (F3, M2), D → (F2, M3)

AP1.5. For a biology project, you measure the weight in grams (g) and the tail length in millimeters (mm) of a group of mice. The equation of the least-squares line for predicting tail length from weight is

$$\text{predicted tail length} = 20 + 3 \times \text{weight}$$

Which of the following is not correct?

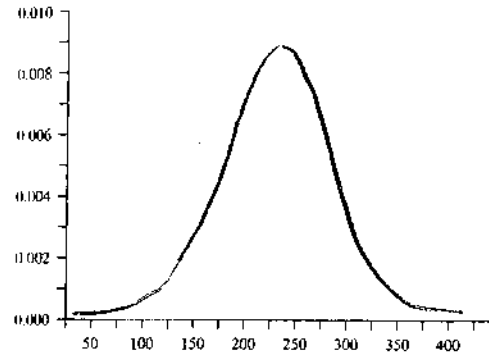
- (a) The slope is 3, which indicates that a mouse's weight should increase by about 3 grams for each additional millimeter of tail length.
- (b) The predicted tail length of a mouse that weighs 38 grams is 134 millimeters.

(c) By looking at the equation of the least-squares line, you can see that the correlation between weight and tail length is positive.

(d) If you had measured the tail length in centimeters instead of millimeters, the slope of the regression line would have been $3/10 = 0.3$.

(e) One mouse weighed 29 grams and had a tail length of 100 millimeters. The residual for this mouse is -7 .

AP1.6. The figure below shows a Normal density curve. Which of the following gives the best estimates for the mean and standard deviation of this Normal distribution?



- (a) $\mu = 200, \sigma = 50$ (d) $\mu = 225, \sigma = 25$
- (b) $\mu = 200, \sigma = 25$ (e) $\mu = 225, \sigma = 275$
- (c) $\mu = 225, \sigma = 50$

AP1.7. The owner of a chain of supermarkets notices that there is a positive correlation between the sales of beer and the sales of ice cream over the course of the previous year. During seasons when sales of beer were above average, sales of ice cream also tended to be above average. Likewise, during seasons when sales of beer were below average, sales of ice cream also tended to be below average. Which of the following would be a valid conclusion from these facts?

- (a) Sales records must be in error. There should be no association between beer and ice cream sales.
- (b) Evidently, for a significant proportion of customers of these supermarkets, drinking beer causes a desire for ice cream or eating ice cream causes a thirst for beer.
- (c) A scatterplot of monthly ice cream sales versus monthly beer sales would show that a straight line describes the pattern in the plot, but it would have to be a horizontal line.
- (d) There is a clear negative association between beer sales and ice cream sales.
- (e) The positive correlation is most likely a result of the lurking variable temperature; that is, as temperatures increase, so do both beer sales and ice cream sales.

AP1.8. Here are the IQ scores of 10 randomly chosen fifth-grade students:

145 139 126 122 125 130 96 110 118 118

Which of the following statements about this data set is *not* true?

- (a) The student with an IQ of 96 is considered an outlier by the $1.5 \times IQR$ rule.
- (b) The five-number summary of the 10 IQ scores is 96, 118, 123.5, 130, 145.
- (c) If the value 96 were removed from the data set, the mean of the remaining 9 IQ scores would be higher than the mean of all 10 IQ scores.
- (d) If the value 96 were removed from the data set, the standard deviation of the remaining 9 IQ scores would be lower than the standard deviation of all 10 IQ scores.
- (e) If the value 96 were removed from the data set, the IQR of the remaining 9 IQ scores would be lower than the IQR of all 10 IQ scores.

AP1.9. Before he goes to bed each night, Mr. Kleen pours dishwasher powder into his dishwasher and turns it on. Each morning, Mrs. Kleen weighs the box of dishwasher powder. From an examination of the data, she concludes that Mr. Kleen dispenses a rather consistent amount of powder each night. Which of the following statements is true?

- I. There is a high positive correlation between the number of days that have passed since the box of dishwasher powder was opened and the amount of powder left in the box.
 - II. A scatterplot with days since purchase as the explanatory variable and amount of dishwasher powder used as the response variable would display a strong positive association.
 - III. The correlation between the amount of powder left in the box and the amount of powder used should be -1 .
- (a) I only (c) III only (e) I, II, and III
 (b) II only (d) II and III only

AP1.10. The General Social Survey (GSS), conducted by the National Opinion Research Center at the University of Chicago, is a major source of data on social attitudes in the United States. Once each year, 1500 adults are interviewed in their homes all across the country. The subjects are asked their opinions about sex and marriage, attitudes toward women, welfare, foreign policy, and many other issues. The GSS begins by selecting a sample of counties from the 3000 counties in the country. The counties are divided into urban, rural, and suburban; a separate sample is chosen at random from each group. This is a

- (a) simple random sample.
- (b) systematic random sample.
- (c) cluster sample.
- (d) stratified random sample.
- (e) voluntary response sample.

AP1.11. You are planning an experiment to determine the effect of the brand of gasoline and the weight of a car on gas mileage measured in miles per gallon. You will use a single test car, adding weights so that its total weight is 3000, 3500, or 4000 pounds. The car will drive on a test track at each weight using each of Amoco, Marathon, and Speedway gasoline. Which is the best way to organize the study?

- (a) Start with 3000 pounds and Amoco and run the car on the test track. Then do 3500 and 4000 pounds. Change to Marathon and go through the three weights in order. Then change to Speedway and do the three weights in order once more.
- (b) Start with 3000 pounds and Amoco and run the car on the test track. Then change to Marathon and then to Speedway without changing the weight. Then add weights to get 3500 pounds and go through the three gasolines in the same order. Then change to 4000 pounds and do the three gasolines in order again.
- (c) Choose a gasoline at random, and run the car with this gasoline at 3000, 3500, and 4000 pounds in order. Choose one of the two remaining gasolines at random and again run the car at 3000, then 3500, then 4000 pounds. Do the same with the last gasoline.
- (d) There are nine combinations of weight and gasoline. Run the car several times using each of these combinations. Make all these runs in random order.
- (e) Randomly select an amount of weight and a brand of gasoline, and run the car on the test track. Repeat this process a total of 30 times.

AP1.12. A linear regression was performed using the five following data points: A(2, 22), B(10, 4), C(6, 14), D(14, 2), E(18, -4). The residual for which of the five points has the largest absolute value?

- (a) A (b) B (c) C (d) D (e) E

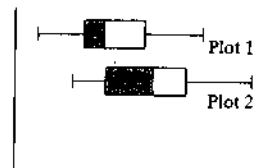
AP1.13. The frequency table below summarizes the times in the last month that patients at the emergency room of a small-city hospital waited to receive medical attention.

Waiting time	Frequency
Less than 10 minutes	5
At least 10 but less than 20 minutes	24
At least 20 but less than 30 minutes	45
At least 30 but less than 40 minutes	38
At least 40 but less than 50 minutes	19
At least 50 but less than 60 minutes	7
At least 60 but less than 70 minutes	2

Which of the following represents possible values for the median and mean waiting times for the emergency room last month?

- (a) median = 27 minutes and mean = 24 minutes
- (b) median = 28 minutes and mean = 30 minutes
- (c) median = 31 minutes and mean = 35 minutes
- (d) median = 35 minutes and mean = 39 minutes
- (e) median = 45 minutes and mean = 46 minutes

AP1.14. Boxplots of two data sets are shown.



Based on the boxplots, which statement below is true?

- (a) The spread of both plots is about the same.
- (b) The means of both plots are approximately equal.
- (c) Plot 2 contains more data points than Plot 1.
- (d) The medians are approximately equal.
- (e) Plot 1 is more symmetric than Plot 2.

Section II: Free Response 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.

AP1.15. The manufacturer of exercise machines for fitness centers has designed two new elliptical machines that are meant to increase cardiovascular fitness. The two machines are being tested on 30 volunteers at a fitness center near the company's headquarters. The volunteers are randomly assigned to one of the machines and use it daily for two months. A measure of cardiovascular fitness is administered at the start of the experiment and again at the end. The following table contains the differences in the two scores (After - Before) for the two machines. Note that higher scores indicate larger gains in fitness.

Machine A		Machine B
	5	3, 5, 9
6, 1	4	2, 5, 7
9, 7, 4, 1, 1	3	2, 4, 8, 9
8, 7, 6, 3, 2, 0	2	1, 5, 9
5, 4	1	0
	0	2

- Write a few sentences comparing the distributions of cardiovascular fitness gains from the two elliptical machines.
- Which machine should be chosen if the company wants to advertise it as achieving the highest overall gain in cardiovascular fitness? Explain your reasoning.
- Which machine should be chosen if the company wants to advertise it as achieving the most consistent gain in cardiovascular fitness? Explain your reasoning.
- Give one reason why the advertising claims of the company (the scope of inference) for this experiment would be limited. Explain how the company could broaden that scope of inference.

AP1.16. Those who advocate for monetary incentives in a work environment claim that this type of incentive has the greatest appeal because it allows the winners to do what they want with their winnings. Those in favor of tangible incentives argue that money lacks the emotional appeal of, say, a weekend for two at a romantic country inn or elegant hotel or a weeklong trip to Europe.

A few years ago a national tire company, in an effort to improve sales of a new line of tires, decided to test which method—offering cash incentives or offering non-cash prizes such as vacations—was more successful in increasing sales. The company had 60 retail sales districts of various sizes across the country and data on the previous sales volume for each district.

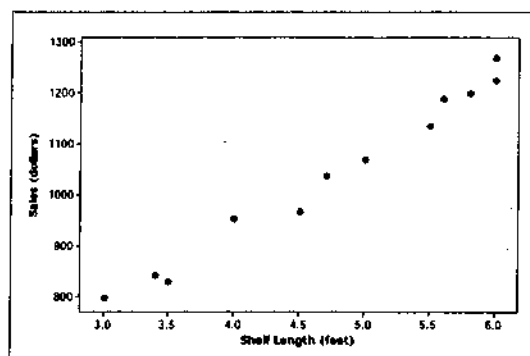
- Describe a completely randomized design using the 60 retail sales districts that would help answer this question.
- Explain how you would use the table of random digits below to do the randomization that your design requires. Then use your method to assign treatments to the first 3 experimental units. Show your work clearly.

07511 88915 41267 16853 84569 79367 32337 03316
81486 69487 60513 09297 00412 71238 27649 39950

- One of the company's officers suggested that it would be better to use a matched pairs design instead of a completely randomized design. Explain how you would change your design to accomplish this.

AP1.17. In retail stores, there is a lot of competition for shelf space. Not only are there national brands for most products, but many stores have their own in-house brands. Since shelf space is not infinite, the question is how many linear feet to allocate to each product and which shelf (top, bottom, or somewhere in the middle) to put it on. The middle shelf is the most popular and lucrative, since many shoppers, if undecided, will simply pick the product that is at eye level.

A local store that sells many upscale goods is trying to determine how much shelf space to allocate to its own brand of men's personal-grooming products. The middle shelf space is randomly varied between three and six linear feet over the next 12 weeks, and weekly sales revenue (in dollars) from the store's brand of personal-grooming products for men is recorded. Below is some computer output from the study, along with a scatterplot.



Predictor	Coef	SE Coef	T	P
Constant	317.94	31.32	10.15	0.000
Shelf length	152.680	6.445	23.69	0.000

S = 22.9212 R-Sq = 98.2% R-Sq(adj) = 98.1%

- Does it appear that the weekly sales revenue is related to the shelf length allocated to the house brand? Justify your answer.
- Write the equation of the least-squares regression line. Be sure to define any variables you use.
- If the store manager were to decide to allocate five linear feet of shelf space to the store's brand of men's grooming products, what is the best estimate of the weekly sales revenue?
- Interpret the value of s .
- Identify and interpret the coefficient of determination.
- The store manager questions the intercept of the regression line: "Am I supposed to believe that this analysis tells me that I can sell these products with no shelf space?" How do you answer her?