AP Statistics Fall Semester Review Packet

Chapters 1 – 7 (Units 1 – 4 from CED)

# Chapter 1 AP® Statistics Practice Test

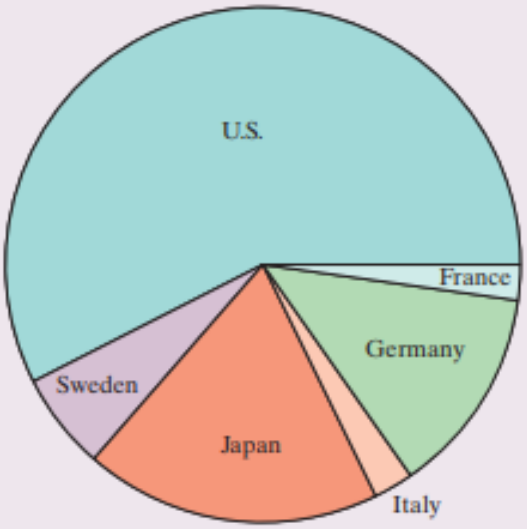
**Section I: Multiple Choice** *Select the best answer for each question.*

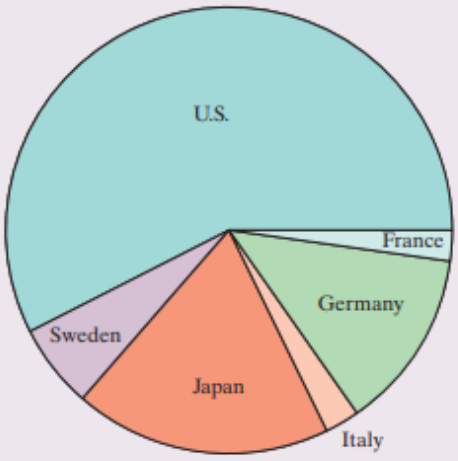
**T1.1** You record the age, marital status, and earned income of a sample of 1463 women. The number and type of variables you have recorded is

**(a)** 3 quantitative, 0 categorical.
**(b)** 4 quantitative, 0 categorical.
**(c)** 3 quantitative, 1 categorical.
**(d)** 2 quantitative, 1 categorical.
**(e)** 2 quantitative, 2 categorical.

**T1.2** Consumers Union measured the gas mileage in miles per gallon of 38 vehicles from the same model year on a special test track. The pie chart provides information about the country of manufacture of the model cars tested by Consumers Union. Based on the pie chart, we conclude that

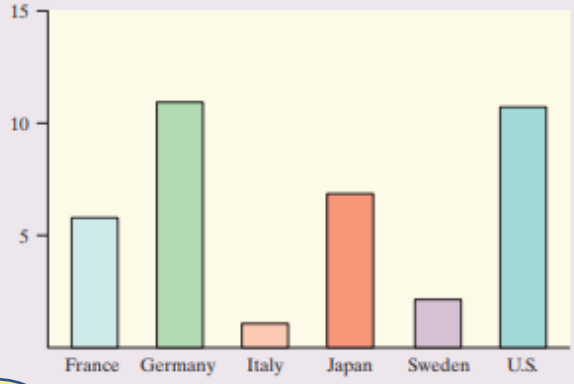
(a) Japanese cars get significantly lower gas mileage than cars from other countries.
**(b)** U.S. cars get significantly higher gas mileage than cars from other countries.
**(c)** Swedish cars get gas mileages that are between those of Japanese and U.S. cars.
**(d)** cars from France have the lowest gas mileage.
**(e)** more than half of the cars in the study were from the United States.



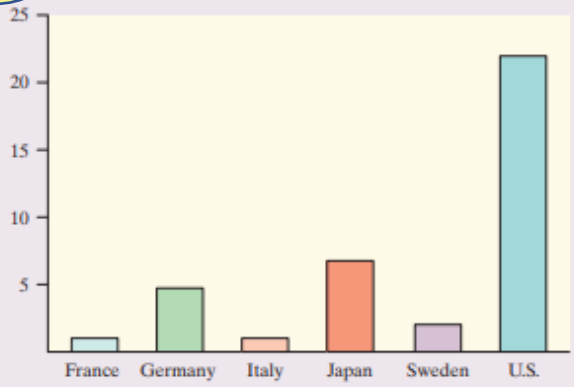



**T1.3** Which of the following bar graphs is equivalent to the pie chart in Question T1.2?

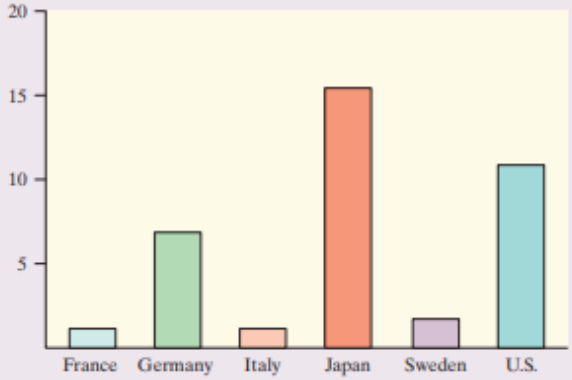
(a)



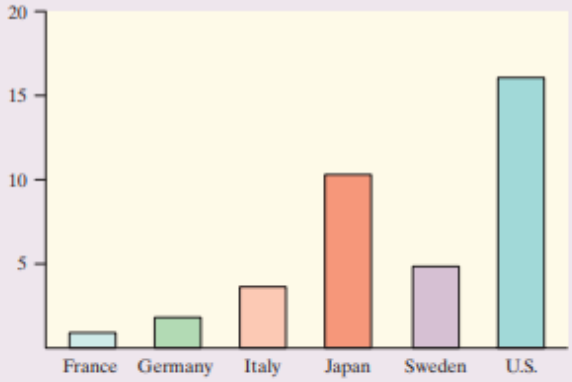

(b)




(c)




(d)




(e) None of these.

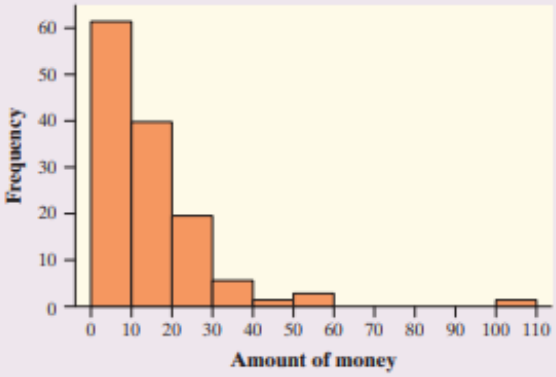
**T1.4** Earthquake intensities are measured using a device called a seismograph, which is designed to be most sensitive to earthquakes with intensities between 4.0 and 9.0 on the Richter scale. Measurements of nine earthquakes gave the following readings:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4.5 | L | 5.5 | H | 8.7 | 8.9 | 6.0 | H | 5.2 |

where L indicates that the earthquake had an intensity below 4.0 and an H indicates that the earthquake had an intensity above 9.0. The median earthquake intensity of the sample is

(a) 5.75.
(b) 6.00.
(c) 6.47.
(d) 8.70.
(e) Cannot be determined.

*Questions T1.5 and T1.6 refer to the following setting.* In a statistics class with 136 students, the professor records how much money (in dollars) each student has in his or her possession during the first class of the semester. The histogram shows the data that were collected.




**T1.5** The percentage of students with less than \$10 in their possession is closest to

(a) 30%. (b) 35%. (c) 45%. (d) 60%. (e) 70%.

**T1.6** Which of the following statements about this distribution is *not* correct?

(a) The histogram is right-skewed.

(b) The median is less than \$20.

(c) The *IQR* is \$35.

(d) The mean is greater than the median.

(e) The histogram is unimodal.

**T1.7** Forty students took a statistics examination having a maximum of 50 points. The score distribution is given in the following stem-and-leaf plot:

```
0 | 28
1 | 2245
2 | 01333358889
3 | 001356679
4 | 22444466788
5 | 000
```

The third quartile of the score distribution is equal to

(a) 45. (b) 44. (c) 43. (d) 32. (e) 23.

**T1.8** The mean salary of all female workers is \$35,000. The mean salary of all male workers is \$41,000. What must be true about the mean salary of all workers?

(a) It must be \$38,000.

(b) It must be larger than the median salary.

(c) It could be any number between \$35,000 and \$41,000.

(d) It must be larger than \$38,000.

(e) It cannot be larger than \$40,000.

*Questions T1.9 and T1.10 refer to the following setting.* A survey was designed to study how business operations vary according to their size. Companies were classified as small, medium, or large. Questionnaires were sent to 200 randomly selected businesses of each size. Because not all questionnaires in a survey of this type are returned, researchers decided to investigate the relationship between the response rate and the size of the business. The data are given in the following two-way table:

| | Business size | | |
|---|---|---|---|
| **Response?** | **Small** | **Medium** | **Large** |
| Yes | 125 | 81 | 40 |
| No | 75 | 119 | 160 |

**T1.9** What percent of all small companies receiving questionnaires responded?

(a) 12.5% (c) 33.3% (e) 62.5%

(b) 20.8% (d) 50.8%

**T1.10** Which of the following conclusions seems to be supported by the data?

(a) There are more small companies than large companies in the survey.

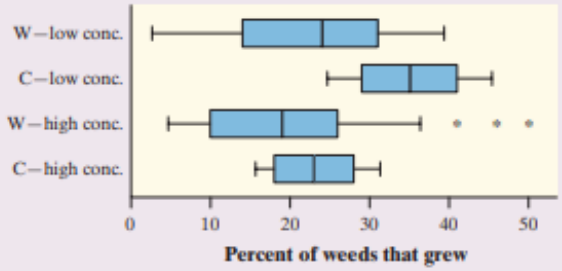
(b) Small companies appear to have a higher response rate than medium or big companies.

(c) Exactly the same number of companies responded as didn't respond.

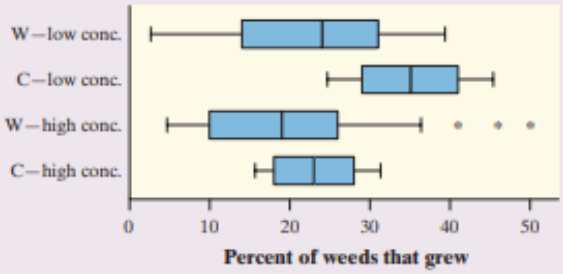
(d) Overall, more than half of companies responded to the survey.

(e) If we combined the medium and large companies, then their response rate would be equal to that of the small companies.

**T1.11** An experiment was conducted to investigate the effect of a new weed killer to prevent weed growth in onion crops. Two chemicals were used: the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations on a total of 50 test plots. The percent of weeds that grew in each plot was recorded. Here are some boxplots of the results. Which of the following is *not* a correct statement about the results of this experiment?



**T1.11** An experiment was conducted to investigate the effect of a new weed killer to prevent weed growth in onion crops. Two chemicals were used: the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations on a total of 50 test plots. The percent of weeds that grew in each plot was recorded. Here are some boxplots of the results. Which of the following is *not* a correct statement about the results of this experiment?




(a) At both high and low concentrations, the new chemical (W) gives better weed control than the standard weed killer (C).

(b) Fewer weeds grew at higher concentrations of both chemicals.

(c) The results for the standard weed killer (C) are less variable than those for the new chemical (W).

(d) High and low concentrations of either chemical have approximately the same effects on weed growth.

(e) Some of the results for the low concentration of weed killer W show fewer weeds growing than some of the results for the high concentration of W.

FRQ (or FRAPPY)

**T1.14** The back-to-back stemplot shows the lifetimes of several Brand X and Brand Y batteries.

| Brand X | | Brand Y |
|---|---|---|
| | 1 | |
| | 1 | 7 |
| | 2 | 2 |
| | 2 | 6 |
| 2110 | 3 | |
| 99775 | 3 | |
| 3221 | 4 | 223334 |
| | 4 | 56889 |
| 4 | 5 | 0 |
| 5 | 5 | |

Key: 4|2 represents 420–429 hours.

(a) What is the longest that any battery lasted?

(b) Give a reason someone might prefer a Brand X battery.

(c) Give a reason someone might prefer a Brand Y battery.

a) The longest that any battery lasted was between 550 to 559 hours.
b) You might prefer Brand X battery because it is more consistent (less variability) and has a longer minimum life as compared to Brand Y
c) You might prefer Brand Y battery because it has a greater median lifetime than that of Brand X

**T2.1** Many professional schools require applicants to take a standardized test. Suppose that 1000 students take such a test. Several weeks after the test, Pete receives his score report: he got a 63, which placed him at the 73rd percentile. This means that

(a) Pete's score was below the median.

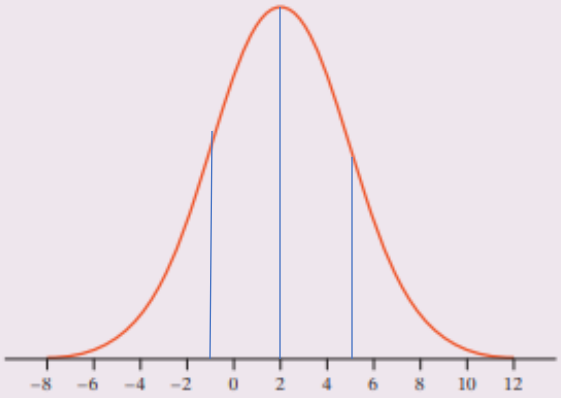
(b) Pete did worse than about 63% of the test takers.

(c) Pete did worse than about 73% of the test takers.

(d) Pete did better than about 63% of the test takers.

(e) Pete did better than about 73% of the test takers.

**T2.2** For the Normal distribution shown, the standard deviation is closest to



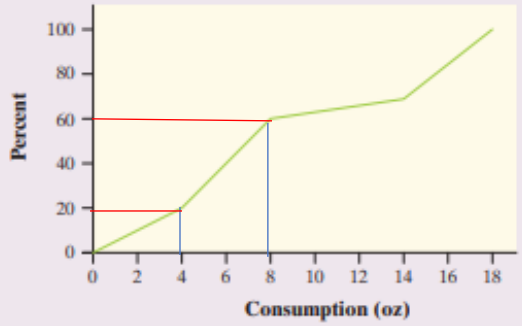

(a) 0 (b) 1 (c) 2 (d) 3 (e) 5

**T2.3** Rainwater was collected in water collectors at 30 different sites near an industrial complex, and the amount of acidity (pH level) was measured. The mean and standard deviation of the values are 4.60 and 1.10, respectively. When the pH meter was recalibrated back at the laboratory, it was found to be in error. The error can be corrected by adding 0.1 pH units to all of the values and then multiplying the result by 1.2. The mean and standard deviation of the corrected pH measurements are

(a) 5.64, 1.44 (c) 5.40, 1.44 (e) 5.64, 1.20

(b) 5.64, 1.32 (d) 5.40, 1.32

**T2.4** The figure shows a cumulative relative frequency graph of the number of ounces of alcohol consumed per week in a sample of 150 adults who report drinking alcohol occasionally. About what percent of these adults consume between 4 and 8 ounces per week?




(a) 20% (b) 40% (c) 50% (d) 60% (e) 80%

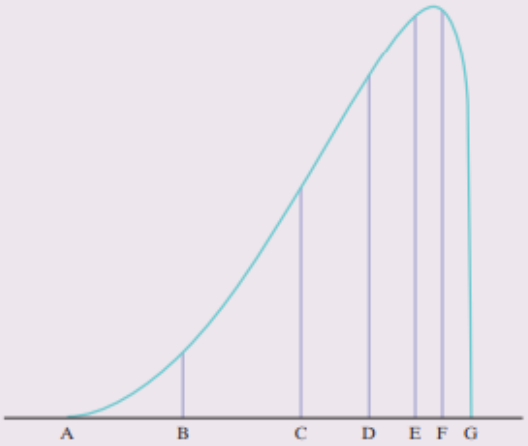
**T2.5** The average yearly snowfall in Chillyville is Normally distributed with a mean of 55 inches. If the snowfall in Chillyville exceeds 60 inches in 15% of the years, what is the standard deviation?

(a) 4.83 inches (d) 8.93 inches

(b) 5.18 inches (e) The standard deviation cannot be computed from the given information.

(c) 6.04 inches

**T2.6** The figure shown is the density curve of a distribution. Seven values are marked on the density curve. Which of the following statements is true?




(a) The mean of the distribution is E.

(b) The area between B and F is 0.50.

(c) The median of the distribution is C.

(d) The 3rd quartile of the distribution is D.

(e) The area between A and G is 1.

**T2.7** If the heights of a population of men follow a Normal distribution, and 99.7% have heights between 5′0″ and 7′0″, what is your estimate of the standard deviation of the heights in this population?

(a) 1″ (b) 3″ (c) 4″ (d) 6″ (e) 12″

**T2.8** Which of the following is *not* correct about a standard Normal distribution?

(a) The proportion of scores that satisfy $0 < z < 1.5$ is 0.4332.

(b) The proportion of scores that satisfy $z < -1.0$ is 0.1587.

(c) The proportion of scores that satisfy $z > 2.0$ is 0.0228.

(d) The proportion of scores that satisfy $z < 1.5$ is 0.9332.

(e) The proportion of scores that satisfy $z > -3.0$ is 0.9938.

# 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.

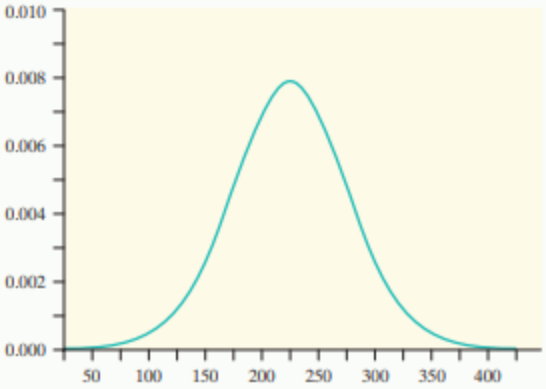
(b) The predicted tail length of a mouse that weighs 38 grams is 134 millimeters.

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

**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 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 greater 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 less 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 less 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
(b) II only
(c) III only
(d) II and III only
(e) I, II, and III

**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.

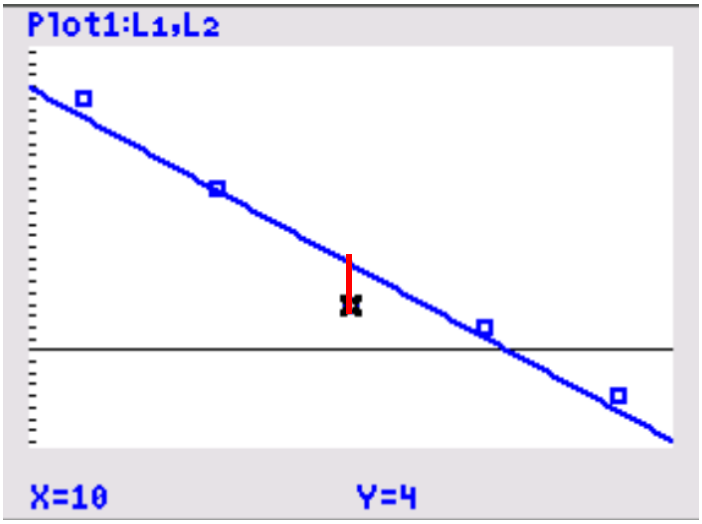
(c) Choose a gasoline at random, and run the car with this gasoline at 3000, 3500, and 4000 pounds in order. Choose one of the two remaining gasolines at random and again run the car at 3000, then 3500, then 4000 pounds. Do the same with the last gasoline.

(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

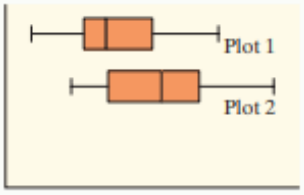
(b) median = 28 minutes and mean = 30 minutes

(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 range 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.

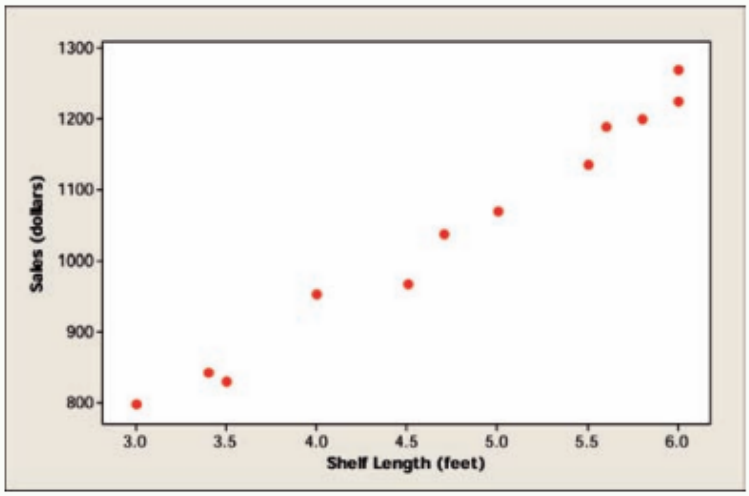
(d) The medians are approximately equal.

(e) Plot 1 is more symmetric than Plot 2.

FRQ (or FRAPPY)

**AP1.17** In retail stores, there is a lot of competition for shelf space. There are national brands for most products, and many stores carry their own line of in-house brands, too. 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, because 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 3 and 6 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%

(a) Describe the relationship between shelf length and sales.

(b) Write the equation of the least-squares regression line. Be sure to define any variables you use.

(c) If the store manager were to decide to allocate 5 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?

(d) Interpret the value of $s$.

(e) Identify and interpret the coefficient of determination.

(f) 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?

(a) There is a strong, positive, linear correlation between shelf length and sales.

(b) LSRL: $\hat{y} = 317.94 + 152.68x$ ,
where $\hat{y} = predicted\ weekly\ sales$ (\$) and
$x = shelf\ length\ (ft.)$

(c) $\hat{y} = 317.94 + 152.68x$
$\hat{y} = 317.94 + 152.68(5)$ , $\hat{y} = \$1081.34$

(d) $S = s_e = 22.9212$, ∴ When comparing shelf space to predict weekly sales, the typical deviation from the LSRL model will be about \$23.

(e) $r^2 = 98.2\%$, so The LSRL comparing shelf space to predict weekly sales accounts for 98.2% of the variation of weekly sales prices.

(f) Extrapolation is problematic! It is non-sensical to consider any weekly sales if a product has no (zero) shelf space.

# Cumulative AP® Practice Test 2

**Section I: Multiple Choice** *Choose the best answer for each question.*

**AP2.1** The five-number summary for a data set is given by min = 5, $Q_1$ = 18, median = 20, $Q_3$ = 40, max = 75. If you wanted to construct a boxplot for the data set (that is, one that would show outliers, if any existed), what would be the maximum possible length of the right-side "whisker"?

(a) 33 (b) 35 (c) 45 (d) 53 (e) 55

**AP2.2** The probability distribution for the number of heads in four tosses of a coin is given by

| Number of heads: | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Probability: | 0.0625 | 0.2500 | 0.3750 | 0.2500 | 0.0625 |

The probability of getting at least one *tail* in four tosses of a coin is

(a) 0.2500. (c) 0.6875. (e) 0.0625.
(b) 0.3125. (d) 0.9375.

**AP2.3** In a certain large population of adults, the distribution of IQ scores is strongly left-skewed with a mean of 122 and a standard deviation of 5. Suppose 200 adults are randomly selected from this population for a market research study. The distribution of the sample mean of IQ scores is

(a) left-skewed with mean 122 and standard deviation 0.35.
(b) exactly Normal with mean 122 and standard deviation 5.
(c) exactly Normal with mean 122 and standard deviation 0.35.
(d) approximately Normal with mean 122 and standard deviation 5.
(e) approximately Normal with mean 122 and standard deviation 0.35.

**AP2.4** A 10-question multiple-choice exam offers 5 choices for each question. Jason just guesses the answers, so he has probability 1/5 of getting any one answer correct. You want to perform a simulation to determine the number of correct answers that Jason gets. One correct way to use a table of random digits to do this is the following:

(a) One digit from the random digit table simulates one answer, with 5 = right and all other digits = wrong. Ten digits from the table simulate 10 answers.
(b) One digit from the random digit table simulates one answer, with 0 or 1 = right and all other digits = wrong. Ten digits from the table simulate 10 answers.
(c) One digit from the random digit table simulates one answer, with odd = right and even = wrong. Ten digits from the table simulate 10 answers.
(d) One digit from the random digit table simulates one answer, with 0 or 1 = right and all other digits = wrong, ignoring repeats. Ten digits from the table simulate 10 answers.
(e) Two digits from the random digit table simulate one answer, with 00 to 20 = right and 21 to 99 = wrong. Ten pairs of digits from the table simulate 10 answers.

**AP2.5** Suppose we roll a fair die four times. The probability that a 6 occurs on exactly one of the rolls is

(a) $4\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^1$ (c) $4\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3$ (e) $6\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3$
(b) $\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^1$ (d) $\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3$

**AP2.6** You want to take an SRS of 50 of the 816 students who live in a dormitory on a college campus. You label the students 001 to 816 in alphabetical order. In the table of random digits, you read the entries

95592 94007 69769 33547 72450 16632 81194 14873

The first three students in your sample have labels

(a) 955, 929, 400. (d) 929, 400, 769.
(b) 400, 769, 769. (e) 400, 769, 335.
(c) 559, 294, 007.

**AP2.7** The number of unbroken charcoal briquets in a 20-pound bag filled at the factory follows a Normal distribution with a mean of 450 briquets and a standard deviation of 20 briquets. The company expects that a certain number of the bags will be underfilled, so the company will replace for free the 5% of bags that have too few briquets. What is the minimum number of unbroken briquets the bag would have to contain for the company to avoid having to replace the bag for free?

(a) 404 (b) 411 (c) 418 (d) 425 (e) 448

**AP2.8** You work for an advertising agency that is preparing a new television commercial to appeal to women. You have been asked to design an experiment to compare the effectiveness of three versions of the commercial. Each subject will be shown one of the three versions and then asked about her attitude toward the product. You think there may be large differences between women who are employed and those who are not. Because of these differences, you should use

(a) a block design, but not a matched pairs design.
(b) a completely randomized design.
(c) a matched pairs design.
(d) a simple random sample.
(e) a stratified random sample.

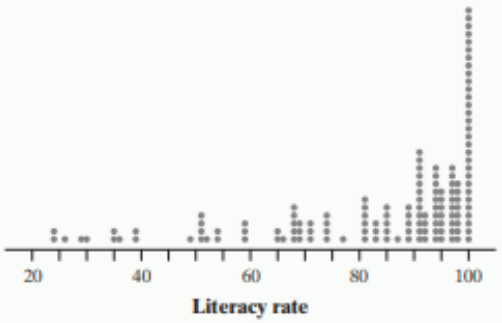
**AP2.9** Suppose that you have torn a tendon and are facing surgery to repair it. The orthopedic surgeon explains the risks to you. Infection occurs in 3% of such operations, the repair fails in 14%, and both infection and failure occur together 1% of the time. What is the probability that the operation is successful for someone who has an operation that is free from infection?

(a) 0.8342 (c) 0.8600 (e) 0.9900
(b) 0.8400 (d) 0.8660

**AP2.10** Social scientists are interested in the association between high school graduation rate (HSGR, measured as a percent) and the percent of U.S. families living in poverty (POV). Data were collected from all 50 states and the District of Columbia, and a regression analysis was conducted. The resulting least-squares regression line is given by $\widehat{POV} = 59.2 - 0.620(HSGR)$ with $r^2 = 0.802$. Based on the information, which of the following is the best interpretation for the slope of the least-squares regression line?

(a) For each 1% increase in the graduation rate, the percent of families living in poverty is predicted to decrease by approximately 0.896.
(b) For each 1% increase in the graduation rate, the percent of families living in poverty is predicted to decrease by approximately 0.802.
(c) For each 1% increase in the graduation rate, the percent of families living in poverty is predicted to decrease by approximately 0.620.
(d) For each 1% increase in the percent of families living in poverty, the graduation rate is predicted to increase by approximately 0.802.
(e) For each 1% increase in the percent of families living in poverty, the graduation rate is predicted to decrease by approximately 0.620.

Here is a dotplot of the adult literacy rates in 177 countries in a recent year, according to the United Nations. For example, the lowest literacy rate was 23.6%, in the African country of Burkina Faso. Mali had the next lowest literacy rate at 24.0%. Use the graph to answer Questions AP2.11 to AP2.13.




**AP2.11** The overall shape of this distribution is

(a) clearly skewed to the right.
(b) clearly skewed to the left.
(c) roughly symmetric.
(d) uniform.
(e) There is no clear shape.

**AP2.12** The mean of this distribution (*don't* try to find it) will be

(a) very close to the median.
(b) greater than the median.
(c) less than the median.
(d) You can't say, because distribution isn't symmetric.
(e) You can't say, because the distribution isn't Normal.

**AP2.13** Based on the shape of this distribution, what measures of center and spread would be most appropriate to report?

(a) The mean and standard deviation
(b) The mean and the interquartile range
(c) The median and the standard deviation
(d) The median and the interquartile range
(e) The mean and the range

**AP2.14** The correlation between the age and height of children under the age of 12 is found to be $r = 0.60$. Suppose we use the age $x$ of a child to predict the height $y$ of the child. What can we conclude?

(a) The height is generally 60% of a child's weight.
(b) About 60% of the time, age will accurately predict height.
(c) Thirty-six percent of the variation in height is accounted for by the linear model relating height to age.
(d) For every 1 year older a child is, the regression line predicts an increase of 0.6 feet in height.
(e) Thirty-six percent of the time, the least-squares regression line accurately predicts height from age.

**AP2.15** An agronomist wants to test three different types of fertilizer (A, B, and C) on the yield of a new variety of wheat. The yield will be measured in bushels per acre. Six 1-acre plots of land were randomly assigned to each of the three fertilizers. The treatment, experimental unit, and response variable are, respectively,

(a) a specific fertilizer, bushels per acre, a plot of land.
(b) a plot of land, bushels per acre, a specific fertilizer.
(c) random assignment, a plot of land, wheat yield.
(d) a specific fertilizer, a plot of land, wheat yield.
(e) a specific fertilizer, the agronomist, wheat yield.

**AP2.16** According to the U.S. Census, the proportion of adults in a certain county who owned their own home was 0.71. An SRS of 100 adults in a certain

**AP2.16** According to the U.S. Census, the proportion of adults in a certain county who owned their own home was 0.71. An SRS of 100 adults in a certain section of the county found that 65 owned their home. Which one of the following represents the approximate probability of obtaining a sample of 100 adults in which fewer than 65 own their home, assuming that this section of the county has the same overall proportion of adults who own their home as does the entire county?

(a) $\binom{100}{65}(0.71)^{65}(0.29)^{35}$

(b) $\binom{100}{65}(0.29)^{65}(0.71)^{35}$

(c) $P\left(Z < \dfrac{0.65 - 0.71}{\sqrt{\dfrac{(0.71)(0.29)}{100}}}\right)$

(d) $P\left(Z < \dfrac{0.65 - 0.71}{\sqrt{\dfrac{(0.65)(0.35)}{100}}}\right)$

(e) $P\left(Z < \dfrac{0.65 - 0.71}{\dfrac{(0.71)(0.29)}{\sqrt{100}}}\right)$

**AP2.17** Which one of the following would be a correct interpretation if you have a $z$-score of +2.0 on an exam?

(a) It means that you missed two questions on the exam.

(b) It means that you got twice as many questions correct as the average student.

(c) It means that your grade was 2 points higher than the mean grade on this exam.

(d) It means that your grade was in the upper 2% of all grades on this exam.

(e) It means that your grade is 2 standard deviations above the mean for this exam.

**AP2.18** Records from a random sample of dairy farms yielded the information below on the number of male and female calves born at various times of the day.

| | Day | Evening | Night | Total |
|---|---|---|---|---|
| Males | 129 | 15 | 117 | **261** |
| Females | 118 | 18 | 116 | **252** |
| **Total** | **247** | **33** | **233** | **513** |

What is the probability that a randomly selected calf was born in the night or was a female?

(a) $\frac{369}{513}$ (b) $\frac{485}{513}$ (c) $\frac{116}{513}$ (d) $\frac{116}{252}$ (e) $\frac{116}{233}$

**AP2.19** When people order books from a popular online source, they are shipped in standard-sized boxes. Suppose that the mean weight of the boxes is 1.5 pounds with a standard deviation of 0.3 pounds, the mean weight of the packing material is 0.5 pounds with a standard deviation of 0.1 pounds, and the mean weight of the books shipped is 12 pounds with a standard deviation of 3 pounds. Assuming that the weights are independent, what is the standard deviation of the total weight of the boxes that are shipped from this source?

(a) 1.84 (c) 3.02 (e) 9.10

(b) 2.60 (d) 3.40

**AP2.20** A grocery chain runs a prize game by giving each customer a ticket that may win a prize when the box is scratched off. Printed on the ticket is a dollar value ($500, $100, $25) or the statement "This ticket is not a winner." Monetary prizes can be redeemed for groceries at the store. Here is the probability distribution of the amount won on a randomly selected ticket:

| **Amount won:** | $500 | $100 | $25 | $0 |
|---|---|---|---|---|
| **Probability:** | 0.01 | 0.05 | 0.20 | 0.74 |

Which of the following are the mean and standard deviation, respectively, of the winnings?

(a) $15.00, $2900.00

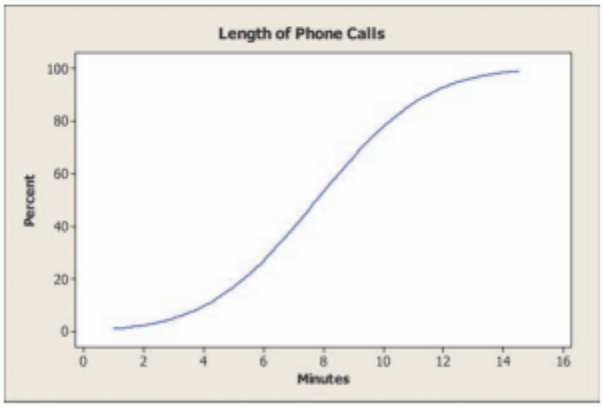
(b) $15.00, $53.85

(c) $15.00, $26.93

(d) $156.25, $53.85

(e) $156.25, $26.93

**AP2.21** A large company is interested in improving the efficiency of its customer service and decides to examine the length of the business phone calls made to clients by its sales staff. A cumulative relative frequency graph is shown below from data collected over the past year. According to the graph, the shortest 80% of calls will take how long to complete?




(a) Less than 10 minutes

(b) At least 10 minutes

(c) Exactly 10 minutes

(d) At least 5.5 minutes

(e) Less than 5.5 minutes

*These exercises are designed to help you review the important ideas and methods of the chapter.*

**R5.1** **Rainy days** The TV weatherman says, "There's a 30% chance of rain tomorrow." Explain what this statement means.

**R5.2** **Click it or else** From police records, it has been determined that 15% of drivers stopped for routine license checks are not wearing seat belts. If a police officer stops 10 vehicles, how likely is it that two consecutive drivers won't be wearing their seat belts?

(a) Describe the design of a simulation to estimate this probability. Explain clearly how you will use the partial table of random digits below to carry out your simulation.

(b) Carry out three repetitions of the simulation. Copy the random digits below onto your paper. Then mark on or directly above the table to show your results.

| | | | | | |
|---|---|---|---|---|---|
| 29077 | 14863 | 61683 | 47052 | 62224 | 51025 |
| 95052 | 90908 | 73592 | 75186 | 87136 | 95761 |
| 27102 | 56027 | 55892 | 33063 | 41842 | 81868 |
| 43367 | 49497 | 72719 | 96758 | 27611 | 91596 |

**R5.3** **Weird dice** Nonstandard dice can produce interesting distributions of outcomes. Suppose you have two balanced, six-sided dice. Die A has faces with 2, 2, 2, 2, 6, and 6 spots. Die B has faces with 1, 1, 1, 5, 5, and 5 spots. Imagine that you roll both dice at the same time.

(a) Find a probability model for the difference (Die A − Die B) in the total number of spots on the up-faces.

(b) Which die is more likely to roll a higher number? Justify your answer.

**R5.4** **Race and ethnicity** The Census Bureau allows each person to choose from a long list of races. That is, in the eyes of the Census Bureau, you belong to whatever race you say you belong to. Hispanic (also called Latino) is a separate category. Hispanics may be of any race. If we choose a resident of the United States at random, the Census Bureau gives these probabilities:[25]

| | Hispanic | Not Hispanic |
|---|---|---|
| Asian | 0.001 | 0.044 |
| Black | 0.006 | 0.124 |
| White | 0.139 | 0.674 |
| Other | 0.003 | 0.009 |

(a) Verify that this is a legitimate assignment of probabilities.

(b) What is the probability that a randomly chosen American is Hispanic?

(c) Non-Hispanic whites are the historical majority in the United States. What is the probability that a randomly chosen American is *not* a member of this group?

(d) Explain why $P(\text{white or Hispanic}) \neq P(\text{white}) + P(\text{Hispanic})$. Then find $P(\text{white or Hispanic})$.

**R5.5** In 2012, fans at Arizona Diamondbacks home games would win 3 free tacos from Taco Bell if the Diamondbacks scored 6 or more runs. In the 2012 season, the Diamondbacks won 41 of their 81 home games and gave away free tacos in 30 of their 81 home games. In 26 of the games, the Diamondbacks won and gave away free tacos. Choose a Diamondbacks home game at random.

(a) Make a Venn diagram to model this chance process.

(b) What is the probability that the Diamondbacks lost and did not give away free tacos?

(c) What is the probability that the Diamondbacks won the game or fans got free tacos?

**R5.6** **Steroids** A company has developed a drug test to detect steroid use by athletes. The test is accurate 95% of the time when an athlete has taken steroids. It is 97% accurate when an athlete hasn't taken steroids. Suppose that the drug test will be used in a population of athletes in which 10% have actually taken steroids. Let's choose an athlete at random and administer the drug test.

(a) Make a tree diagram showing the sample space of this chance process.

(b) What's the probability that the randomly selected athlete tests positive? Show your work.

(c) Suppose that the chosen athlete tests positive. What's the probability that he or she actually used steroids? Show your work.

**Section I: Multiple Choice** *Select the best answer for each question.*

**T5.1** Dr. Stats plans to toss a fair coin 10,000 times in the hope that it will lead him to a deeper understanding of the laws of probability. Which of the following statements is true?

(a) It is unlikely that Dr. Stats will get more than 5000 heads.
(b) Whenever Dr. Stats gets a string of 15 tails in a row, it becomes more likely that the next toss will be a head.
(c) The fraction of tosses resulting in heads should be exactly 1/2.
(d) The chance that the 100th toss will be a head depends somewhat on the results of the first 99 tosses.
(e) It is likely that Dr. Stats will get about 50% heads.

**T5.2** China has 1.2 billion people. Marketers want to know which international brands they have heard of. A large study showed that 62% of all Chinese adults have heard of Coca-Cola. You want to simulate choosing a Chinese at random and asking if he or she has heard of Coca-Cola. One correct way to assign random digits to simulate the answer is:

(a) One digit simulates one person's answer; odd means "Yes" and even means "No."
(b) One digit simulates one person's answer; 0 to 6 mean "Yes" and 7 to 9 mean "No. "
(c) One digit simulates the result; 0 to 9 tells how many in the sample said "Yes."
(d) Two digits simulate one person's answer; 00 to 61 mean "Yes" and 62 to 99 mean "No. "
(e) Two digits simulate one person's answer; 00 to 62 mean "Yes" and 63 to 99 mean "No. "

**T5.3** Choose an American household at random and record the number of vehicles they own. Here is the probability model if we ignore the few households that own more than 5 cars:

| Number of cars: | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Probability: | 0.09 | 0.36 | 0.35 | 0.13 | 0.05 | 0.02 |

A housing company builds houses with two-car garages. What percent of households have more cars than the garage can hold?

(a) 7% (b) 13% (c) 20% (d) 45% (e) 55%

**T5.4** Computer voice recognition software is getting better. Some companies claim that their software correctly recognizes 98% of all words spoken by a trained user. To simulate recognizing a single word when the probability of being correct is 0.98, let two digits simulate one word; 00 to 97 mean "correct." The program recognizes words (or not) independently. To simulate the program's performance on 10 words, use these random digits:

60970 70024 17868 29843 61790 90656 87964

The number of words recognized correctly out of the 10 is

(a) 10 (b) 9 (c) 8 (d) 7 (e) 6

*Questions T5.5 to T5.7 refer to the following setting. One thousand students at a city high school were classified according to both GPA and whether or not they consistently skipped classes. The two-way table below summarizes the data. Suppose that we choose a student from the school at random.*

| | GPA | | |
|---|---|---|---|
| **Skipped Classes** | **<2.0** | **2.0–3.0** | **>3.0** |
| Many | 80 | 25 | 5 |
| Few | 175 | 450 | 265 |

**T5.5** What is the probability that a student has a GPA under 2.0?

(a) 0.227 (b) 0.255 (c) 0.450 (d) 0.475 (e) 0.506

**T5.6** What is the probability that a student has a GPA under 2.0 or has skipped many classes?

(a) 0.080 (b) 0.281 (c) 0.285 (d) 0.365 (e) 0.727

**T5.7** What is the probability that a student has a GPA under 2.0 given that he or she has skipped many classes?

(a) 0.080 (b) 0.281 (c) 0.285 (d) 0.314 (e) 0.727

**T5.8** For events $A$ and $B$ related to the same chance process, which of the following statements is true?

(a) If $A$ and $B$ are mutually exclusive, then they must be independent.
(b) If $A$ and $B$ are independent, then they must be mutually exclusive.
(c) If $A$ and $B$ are not mutually exclusive, then they must be independent.
(d) If $A$ and $B$ are not independent, then they must be mutually exclusive.
(e) If $A$ and $B$ are independent, then they cannot be mutually exclusive.

**T5.9** Choose an American adult at random. The probability that you choose a woman is 0.52. The probability that the person you choose has never married is 0.25. The probability that you choose a woman who has never married is 0.11. The probability that the person you choose is either a woman or has never been married (or both) is therefore about

(a) 0.77. (b) 0.66. (c) 0.44. (d) 0.38. (e) 0.13.

**T5.10** A deck of playing cards has 52 cards, of which 12 are face cards. If you shuffle the deck well and turn over the top 3 cards, one after the other, what's the probability that all 3 are face cards?

(a) 0.001 (c) 0.010 (e) 0.02
(b) 0.005 (d) 0.012

**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.*

**T5.11** Your teacher has invented a "fair" dice game to play. Here's how it works. Your teacher will roll one fair *eight-sided* die, and you will roll a fair *six-sided* die. Each player rolls once, and the winner is the person with the higher number. In case of a tie, neither player wins. The table shows the sample space of this chance process.

| | Teacher Rolls | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **You Roll** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | | | | | | | | |
| 2 | | | | | | | | |
| 3 | | | | | | | | |
| 4 | | | | | | | | |
| 5 | | | | | | | | |
| 6 | | | | | | | | |

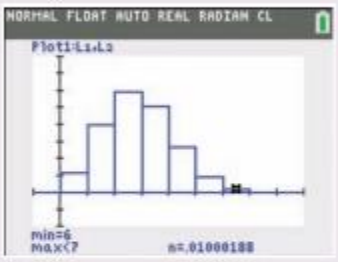
(a) Let $A$ be the event "your teacher wins." Find $P(A)$.
(b) Let $B$ be the event "you get a 3 on your first roll." Find $P(A \cup B)$.
(c) Are events $A$ and $B$ independent? Justify your answer.

**T5.13** Researchers are interested in the relationship between cigarette smoking and lung cancer. Suppose an adult male is randomly selected from a particular population. The following table shows the probabilities of some events related to this chance process:

| Event | Probability |
|---|---|
| Smokes | 0.25 |
| Smokes and gets cancer | 0.08 |
| Does not smoke and does not get cancer | 0.71 |

(a) Find the probability that the individual gets cancer given that he is a smoker. Show your work.
(b) Find the probability that the individual smokes or gets cancer. Show your work.
(c) Two adult males are selected at random. Find the probability that at least one of the two gets cancer. Show your work.

**T6.9** The figure shows the probability distribution of a discrete random variable X. Note that the cursor is on the histogram bar representing a value of 6. Which of the following best describes this random variable?




(a) Binomial with $n = 8, p = 0.1$
(b) Binomial with $n = 8, p = 0.3$
(c) Binomial with $n = 8, p = 0.8$
(d) Geometric with $p = 0.1$
(e) Geometric with $p = 0.2$

**T6.10** A test for extrasensory perception (ESP) involves asking a person to tell which of 5 shapes—a circle, star, triangle, diamond, or heart—appears on a hidden computer screen. On each trial, the computer is equally likely to select any of the 5 shapes. Suppose researchers are testing a person who does not have ESP and so is just guessing on each trial. What is the probability that the person guesses the first 4 shapes incorrectly but gets the fifth correct?

(a) 1/5
(b) $\left(\frac{4}{5}\right)^4$
(c) $\left(\frac{4}{5}\right)^4 \cdot \left(\frac{1}{5}\right)$
(d) $\binom{5}{1} \cdot \left(\frac{4}{5}\right)^4 \cdot \left(\frac{1}{5}\right)$
(e) 4/5

**Section I: Multiple Choice** *Select the best answer for each question.*

*Questions T6.1 to T6.3 refer to the following setting. A psychologist studied the number of puzzles that subjects were able to solve in a five-minute period while listening to soothing music. Let X be the number of puzzles completed successfully by a randomly chosen subject. The psychologist found that X had the following probability distribution:*

| Value: | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Probability: | 0.2 | 0.4 | 0.3 | 0.1 |

**T6.1** What is the probability that a randomly chosen subject completes more than the expected number of puzzles in the five-minute period while listening to soothing music?

(a) 0.1
(b) 0.4
(c) 0.8
(d) 1
(e) Cannot be determined

**T6.2** The standard deviation of X is 0.9. Which of the following is the best interpretation of this value?

(a) About 90% of subjects solved 3 or fewer puzzles.
(b) About 68% of subjects solved between 0.9 puzzles less and 0.9 puzzles more than the mean.
(c) The typical subject solved an average of 0.9 puzzles.
(d) The number of puzzles solved by subjects typically differed from the mean by about 0.9 puzzles.
(e) The number of puzzles solved by subjects typically differed from one another by about 0.9 puzzles.

**T6.3** Let $D$ be the difference in the number of puzzles solved by two randomly selected subjects in a five-minute period. What is the standard deviation of $D$?

(a) 0 (b) 0.81 (c) 0.9 (d) 1.27 (e) 1.8

**T6.4** Suppose a student is randomly selected from your school. Which of the following pairs of random variables are most likely independent?

(a) X = student's height; Y = student's weight
(b) X = student's IQ; Y = student's GPA
(c) X = student's PSAT Math score; Y = student's PSAT Verbal score
(d) X = average amount of homework the student does per night; Y = student's GPA
(e) X = average amount of homework the student does per night; Y = student's height

**T6.5** A certain vending machine offers 20-ounce bottles of soda for $1.50. The number of bottles X bought from the machine on any day is a random variable with mean 50 and standard deviation 15. Let the random variable Y equal the total revenue from this machine on a given day. Assume that the machine works properly and that no sodas are stolen from the machine. What are the mean and standard deviation of Y?

(a) $\mu_Y = \$1.50$, $\sigma_Y = \$22.50$
(b) $\mu_Y = \$1.50$, $\sigma_Y = \$33.75$
(c) $\mu_Y = \$75$, $\sigma_Y = \$18.37$
(d) $\mu_Y = \$75$, $\sigma_Y = \$22.50$
(e) $\mu_Y = \$75$, $\sigma_Y = \$33.75$

**T6.6** The weight of tomatoes chosen at random from a bin at the farmer's market follows a Normal distribution with mean $\mu = 10$ ounces and standard deviation $\sigma = 1$ ounce. Suppose we pick four tomatoes at random from the bin and find their total weight $T$. The random variable $T$ is

(a) Normal, with mean 10 ounces and standard deviation 1 ounce.
(b) Normal, with mean 40 ounces and standard deviation 2 ounces.
(c) Normal, with mean 40 ounces and standard deviation 4 ounces.
(d) binomial, with mean 40 ounces and standard deviation 2 ounces.
(e) binomial, with mean 40 ounces and standard deviation 4 ounces.

**T6.7** Which of the following random variables is geometric?

(a) The number of times I have to roll a die to get two 6s.
(b) The number of cards I deal from a well-shuffled deck of 52 cards until I get a heart.
(c) The number of digits I read in a randomly selected row of the random digits table until I find a 7.
(d) The number of 7s in a row of 40 random digits.
(e) The number of 6s I get if I roll a die 10 times.

**T6.8** Seventeen people have been exposed to a particular disease. Each one independently has a 40% chance of contracting the disease. A hospital has the capacity to handle 10 cases of the disease. What is the probability that the hospital's capacity will be exceeded?

(a) 0.011
(b) 0.035
(c) 0.092
(d) 0.965
(e) 0.989

FRQ (or FRAPPY)

**T6.11** Let Y denote the number of broken eggs in a randomly selected carton of one dozen "store brand" eggs at a local supermarket. Suppose that the probability distribution of Y is as follows.

| **Value $y_i$:** | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| **Probability $p_i$:** | 0.78 | 0.11 | 0.07 | 0.03 | 0.01 |

(a) What is the probability that at least 10 eggs in a randomly selected carton are *unbroken*?

(b) Calculate and interpret $\mu_Y$.

(c) Calculate and interpret $\sigma_Y$. Show your work.

(d) A quality control inspector at the store keeps looking at randomly selected cartons of eggs until he finds one with at least 2 broken eggs. Find the probability that this happens in one of the first three cartons he inspects.

Experimental Design Practice

89. To investigate whether standing up while studying affects performance in an algebra class, a teacher assigns half of the 30 students in his class to stand up while studying and assigns the other half to not stand up while studying. To determine who receives which treatment, the teacher identifies the two students who did best on the last exam and randomly assigns one to stand and one to not stand. The teacher does the same for the next two highest-scoring students and continues in this manner until each student is assigned a treatment. Which of the following best describes this plan?

(a) This is an observational study.
(b) This is an experiment with blocking.
(c) This is a completely randomized experiment.
(d) This is a stratified random sample.
(e) This is a cluster sample.

90. A gardener wants to try different combinations of fertilizer (none, 1 cup, 2 cups) and mulch (none, wood chips, pine needles, plastic) to determine which combination produces the highest yield for a variety of green beans. He has 60 green-bean plants to use in the experiment. If he wants an equal number of plants to be assigned to each treatment, how many plants will be assigned to each treatment?

(a) 1 (b) 3 (c) 4 (d) 5 (e) 12

91. Corn variety 1 yielded 140 bushels per acre last year at a research farm. This year, corn variety 2, planted in the same location, yielded only 110 bushels per acre. Based on these results, is it reasonable to conclude that corn variety 1 is more productive than corn variety 2?

(a) Yes, because 140 bushels per acre is greater than 110 bushels per acre.
(b) Yes, because the study was done at a research farm.
(c) No, because there may be other differences between the two years besides the corn variety.
(d) No, because there was no use of a placebo in the experiment.
(e) No, because the experiment wasn't double-blind.

92. A report in a medical journal notes that the risk of developing Alzheimer's disease among subjects who regularly opted to take the drug ibuprofen was about half the risk among those who did not. Is this good evidence that ibuprofen is effective in preventing Alzheimer's disease?

(a) Yes, because the study was a randomized, comparative experiment.
(b) No, because the effect of ibuprofen is confounded with the placebo effect.
(c) Yes, because the results were published in a reputable professional journal.
(d) No, because this is an observational study. An experiment would be needed to confirm (or not confirm) the observed effect.
(e) Yes, because a 50% reduction can't happen just by chance.

93. A farmer is conducting an experiment to determine which variety of apple tree, Fuji or Gala, will produce more fruit in his orchard. The orchard is divided into 20 equally sized square plots. He has 10 trees of each variety and randomly assigns each tree to a separate plot in the orchard. What are the experimental unit(s) in this study?

(a) The trees (c) The apples (e) The orchard
(b) The plots (d) The farmer

94. Two essential features of all statistically designed experiments are

(a) compare several treatments; use the double-blind method.
(b) compare several treatments; use chance to assign subjects to treatments.
(c) always have a placebo group; use the double-blind method.
(d) use a block design; use chance to assign subjects to treatments.
(e) use enough subjects; always have a control group.

95. **Seed weights** (2.2) Biological measurements on the same species often follow a Normal distribution quite closely. The weights of seeds of a variety of winged bean are approximately Normal with mean 525 milligrams (mg) and standard deviation 110 mg.

(a) What percent of seeds weigh more than 500 mg? Show your method.
(b) If we discard the lightest 10% of these seeds, what is the smallest weight among the remaining seeds? Show your method.

96. **Twins** (1.3, 3.1) A researcher studied a group of identical twins who had been separated and adopted at birth. In each case, one twin (Twin A) was adopted by a low-income family and the other (Twin B) by a high-income family. Both twins were given an IQ test as adults. Here are their scores:[48]

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Twin A:** | 120 | 99 | 99 | 94 | 111 | 97 | 99 | 94 | 104 | 114 | 113 | 100 |
| **Twin B:** | 128 | 104 | 108 | 100 | 116 | 105 | 100 | 100 | 103 | 124 | 114 | 112 |

(a) How well does one twin's IQ predict the other's? Give appropriate evidence to support your answer.
(b) Do identical twins living in low-income homes tend to have lower IQs later in life than their twins who live in high-income homes? Give appropriate evidence to support your answer.

**Section I: Multiple Choice** *Select the best answer for each question.*

**T4.1** When we take a census, we attempt to collect data from

(a) a stratified random sample.

(b) every individual chosen in a simple random sample.

(c) every individual in the population.

(d) a voluntary response sample.

(e) a convenience sample.

**T4.2** You want to take a simple random sample (SRS) of 50 of the 816 students who live in a dormitory on

**T4.2** You want to take a simple random sample (SRS) of 50 of the 816 students who live in a dormitory on campus. You label the students 001 to 816 in alphabetical order. In the table of random digits, you read the entries

95592 94007 69769 33547 72450 16632 81194 14873

The first three students in your sample have labels

(a) 955, 929, 400.
(b) 400, 769, 769.
(c) 559, 294, 007.
(d) 929, 400, 769.
(e) 400, 769, 335.

**T4.3** A study of treatments for angina (pain due to low blood supply to the heart) compared bypass surgery, angioplasty, and use of drugs. The study looked at the medical records of thousands of angina patients whose doctors had chosen one of these treatments. It found that the average survival time of patients given drugs was the highest. What do you conclude?

(a) This study proves that drugs prolong life and should be the treatment of choice.
(b) We can conclude that drugs prolong life because the study was a comparative experiment.
(c) We can't conclude that drugs prolong life because the patients were volunteers.
(d) We can't conclude that drugs prolong life because this was an observational study.
(e) We can't conclude that drugs prolong life because no placebo was used.

**T4.4** Tonya wanted to estimate the average amount of time that students at her school spend on Facebook each day. She gets an alphabetical roster of students in the school from the registrar's office and numbers the students from 1 to 1137. Then Tonya uses a random number generator to pick 30 distinct labels from 1 to 1137. She surveys those 30 students about their Facebook use. Tonya's sample is a simple random sample because

(a) it was selected using a chance process.
(b) it gave every individual the same chance to be selected.
(c) it gave every possible sample of the same size an equal chance to be selected.
(d) it doesn't involve strata or clusters.
(e) it is guaranteed to be representative of the population.

**T4.5** Consider an experiment to investigate the effectiveness of different insecticides in controlling pests and their impact on the productivity of tomato plants. What is the best reason for randomly assigning treatment levels (spraying or not spraying) to the experimental units (farms)?

(a) Random assignment allows researchers to generalize conclusions about the effectiveness of the insecticides to all farms.
(b) Random assignment will tend to average out all other uncontrolled factors such as soil fertility so that they are not confounded with the treatment effects.
(c) Random assignment eliminates the effects of other variables, like soil fertility.
(d) Random assignment eliminates chance variation in the responses.
(e) Random assignment helps avoid bias due to the placebo effect.

**T4.6** The most important advantage of experiments over observational studies is that

(a) experiments are usually easier to carry out.
(b) experiments can give better evidence of causation.
(c) confounding cannot happen in experiments.
(d) an observational study cannot have a response variable.
(e) observational studies cannot use random samples.

**T4.7** A TV station wishes to obtain information on the TV viewing habits in its market area. The market area contains one city of population 170,000, another city of 70,000, and four towns of about 5000 inhabitants each. The station suspects that the viewing habits may be different in larger and smaller cities and in the rural areas. Which of the following sampling designs would give the type of information that the station requires?

(a) A cluster sample using the cities and towns as clusters
(b) A convenience sample from the market area
(c) A simple random sample from the market area
(d) A stratified sample from the cities and towns in the market area
(e) An online poll that invites all people from the cities and towns in the market area to participate

**T4.8** *Bias* in a sampling method is

(a) any difference between the sample result and the truth about the population.
(b) the difference between the sample result and the truth about the population due to using chance to select a sample.
(c) any difference between the sample result and the truth about the population due to practical difficulties such as contacting the subjects selected.
(d) any difference between the sample result and the truth about the population that tends to occur in the same direction whenever you use this sampling method.
(e) racism or sexism on the part of those who take the sample.

**T4.9** You wonder if TV ads are more effective when they are longer or repeated more often or both. So you design an experiment. You prepare 30-second and 60-second ads for a camera. Your subjects all watch the same TV program, but you assign them at random to four groups. One group sees the 30-second ad once during the program; another sees it three times; the third group sees the 60-second ad once; and the last group sees the 60-second ad three times. You ask all subjects how likely they are to buy the camera.