From the 2019 Administration

AP® CollegeBoard

AP® Statistics

Practice Exam

Contents

Note: This publication shows the page numbers that appeared in the ***2018–19 AP Exam Instructions*** book and in the actual exam. This publication was not repaginated to begin with page 1.

AP Statistics Exam

Regularly Scheduled Exam Date: Thursday afternoon, May 16, 2019

Late-Testing Exam Date: Wednesday morning, May 22, 2019

Section I	**Total Time:** 1 hour and 30 minutes Graphing calculator expected **Number of Questions:** 40 *(The number of questions may vary slightly depending on the form of the exam.)* **Percent of Total Score:** 50% **Writing Instrument:** Pencil required
Section II	**Total Time:** 1 hour and 30 minutes Graphing calculator expected **Number of Questions:** 6 **Percent of Total Score:** 50% **Writing Instrument:** Either pencil or pen with black or dark blue ink

Before Distributing Exams: Check that the title on all exam covers is ***Statistics***. If there are any exam booklets with a different title, contact the AP coordinator immediately.

What Proctors Need to Bring to This Exam

- ☐ Exam packets
- ☐ Answer sheets
- ☐ AP Student Packs
- ☐ *2018-19 AP Coordinator's Manual*
- ☐ This book—*2018-19 AP Exam Instructions*
- ☐ AP Exam Seating Chart template
- ☐ School Code and Homeschool/Self-Study Codes
- ☐ Extra graphing calculators
- ☐ Pencil sharpener
- ☐ Container for students' electronic devices (if needed)
- ☐ Extra No. 2 pencils with erasers
- ☐ Extra pens with black or dark blue ink
- ☐ Extra paper
- ☐ Stapler
- ☐ Watch
- ☐ Signs for the door to the testing room
  - "Exam in Progress"
  - "Phones of any kind are prohibited during the test administration, including breaks"

Students are expected to bring graphing calculators with statistical capabilities to the AP Statistics Exam. Nongraphing scientific calculators are permitted as long as they have the required computational capabilities. Before starting the exam administration, make sure each student has a graphing calculator from the approved list on page 53 of the *2018-19 AP Coordinator's Manual* or a scientific calculator. It is up to the student to determine if a nongraphing scientific calculator has the required computational capabilities. If a student does not have a graphing calculator from the approved list or an appropriate scientific calculator, you may provide one from your supply. See pages 50–53 of the *AP Coordinator's Manual* for more information. If the student does not want to use the calculator you provide, or does not want to use a calculator at all, he or she must hand copy, date, and sign the release statement on page 52 of the *AP Coordinator's Manual*.

Students may have no more than two calculators on their desks. Calculators may not be shared. Calculator memories do not need to be cleared before or after the exam. Students with Hewlett-Packard 48–50 Series and Casio FX-9860 graphing calculators may use cards designed for use with these calculators. Proctors should make sure infrared ports (Hewlett-Packard) are not facing each other. **Since graphing calculators can be used to store data, including text, proctors should monitor that students are using their calculators appropriately. Attempts by students to use the calculator to remove exam questions and/or answers from the room may result in the cancellation of AP Exam scores.**

SECTION I: Multiple Choice

› **Do not begin the exam instructions below until you have completed the appropriate General Instructions for your group.**

Make sure you begin the exam at the designated time. Remember, you must complete a seating chart for this exam. See pages 295–296 for a seating chart template and instructions. See the *2018-19 AP Coordinator's Manual* for exam seating requirements (pages 56–59).

If you are giving the regularly scheduled exam, say:
It is Thursday afternoon, May 16, and you will be taking the AP Statistics Exam.

If you are giving the alternate exam for late testing, say:
It is Wednesday morning, May 22, and you will be taking the AP Statistics Exam.

Look at your exam packet and confirm that the exam title is "AP Statistics." Raise your hand if your exam packet contains any title other than "AP Statistics," and I will help you.

Once you confirm that all students have the correct exams, say:

In a moment, you will open the exam packet. By opening this packet, you agree to all of the AP Program's policies and procedures outlined in the *2018-19 Bulletin for AP Students and Parents*.

You may now remove the shrinkwrap from the outside only of your exam packet. Do not open the Section I booklet; do not remove the shrinkwrap from the Section II materials. Put the white seals and the shrinkwrapped Section II booklet aside. . . .

Carefully remove the AP Exam label found near the top left of your exam booklet cover. Place it on page 1 of your answer sheet on the light blue box near the top right corner that reads "AP Exam Label." . . .

If students accidentally place the exam label in the space for the number label or vice versa, advise them to leave the labels in place. They should not try to remove the label; their exam can still be processed correctly.

Listen carefully to all my instructions. I will give you time to complete each step. Please look up after completing each step. Raise your hand if you have any questions.

Give students enough time to complete each step. Don't move on until all students are ready.

Read the statements on the front cover of the Section I booklet. . . .

Sign your name, and write today's date. . . .

Now print your full legal name where indicated. . . .

Turn to the back cover of your exam booklet and read it completely. . . .

Give students a few minutes to read the entire cover.

Are there any questions? . . .

You will now take the multiple-choice portion of the exam. You should have in front of you the multiple-choice booklet and your answer sheet. You may never discuss the multiple-choice exam content at any time in any form with anyone, including your teacher and other students. If you disclose the multiple-choice exam content through any means, your AP Exam score will be canceled.

Open your answer sheet to page 2. You must complete the answer sheet using a No. 2 pencil only. Mark all of your responses beginning on page 2 of your answer sheet, one response per question. Completely fill in the circles. If you need to erase, do so carefully and completely. No credit will be given for anything written in the exam booklet. Scratch paper is not allowed, but you may use the margins or any blank space in the exam booklet for scratch work. Calculators may be used for both sections of this exam. You may place your calculators on your desk. Are there any questions? . . .

You have 1 hour and 30 minutes for this section. Open your Section I booklet and begin.

Note Start Time _____. Note Stop Time _____.

Check that students are marking their answers in pencil on their answer sheets and that they have not opened their shrinkwrapped Section II booklets. Proctors should walk around and make sure Hewlett-Packard calculators' infrared ports are not facing each other and that students are not sharing calculators.

After 1 hour and 20 minutes, say:

There are 10 minutes remaining.

After 10 minutes, say:

Stop working. Close your booklet and put your answer sheet on your desk, faceup. Make sure you have your AP number label and an AP Exam label on page 1 of your answer sheet. Sit quietly while I collect your answer sheets.

Collect an answer sheet from each student. Check that each answer sheet has an AP number label and an AP Exam label.

After all answer sheets have been collected, say:

Now you must seal your exam booklet using the white seals you set aside earlier. Remove the white seals from the backing and press one on each area of your exam booklet cover marked "PLACE SEAL HERE." Fold each seal over the

back cover. When you have finished, place the booklet on your desk, faceup. I will now collect your Section I booklet. . . .

Collect a Section I booklet from each student. Check that each student has signed the front cover of the sealed Section I booklet.

There is a 10-minute break between Sections I and II.

When all Section I materials have been collected and accounted for and you are ready for the break, say:

Please listen carefully to these instructions before we take a 10-minute break. All items you placed under your chair at the beginning of this exam, including your Student Pack, must stay there, and you are not permitted to open or access them in any way. Leave your shrinkwrapped Section II packet on your desk during the break. You are not allowed to consult teachers, other students, notes, textbooks, or any other resources during the break. You may not make phone calls, send text messages, use your calculators, check email, use a social networking site, or access any electronic or communication device. You may not leave the designated break area. Remember, you may never discuss the multiple-choice exam content with anyone, and if you disclose the content through any means, your AP Exam score will be canceled. Are there any questions? . . .

You may begin your break. Testing will resume at _____ .

SECTION II: Free Response

After the break, say:

May I have everyone's attention? Place your Student Pack on your desk. . . .

You may now remove the shrinkwrap from the Section II packet, but do not open the exam booklet until you are told to do so. . . .

Read the bulleted statements on the front cover of the exam booklet. Look up when you have finished. . . .

Now take an AP number label from your Student Pack and place it on the shaded box. If you don't have any AP number labels, write your AP number in the box. Look up when you have finished. . . .

Read the last statement. . . .

Using your pen, print the first, middle, and last initials of your legal name in the boxes and print today's date where indicated. This constitutes your signature and your agreement to the statements on the front cover. . . .

Now turn to the back cover. Using your pen, complete Items 1 through 3 under "Important Identification Information." . . .

Read Item 4. . . .

Are there any questions? . . .

If this is your last AP Exam, you may keep your Student Pack. Place it under your chair for now. Otherwise if you are taking any other AP Exams this year, leave your Student Pack on your desk and I will collect it now. . . .

Read the information on the back cover of the exam booklet. Do not open the booklet until you are told to do so. Look up when you have finished. . . .

Collect the Student Packs from students who are taking any other AP Exams this year.

Then say:

Are there any questions? . . .

Section II has two parts. You have 1 hour and 30 minutes to complete all of Section II. You are responsible for pacing yourself and may proceed freely from one part to the next. You must write your answers in the exam booklet using a pen with black or dark blue ink or a No. 2 pencil. If you use a pencil, be sure that your writing is dark enough to be easily read. If you need more paper to complete your responses, raise your hand. At the top of each extra sheet of paper you use, write only:

- **your AP number,**
- **the exam title, and**
- **the question number you are working on.**

Do not write your name. Are there any questions? . . .

You may begin Section II.

Note Start Time _____. Note Stop Time _____.

You should also make sure that Hewlett-Packard calculators' infrared ports are not facing each other and that students are not sharing calculators.

After 1 hour and 5 minutes, say:

There are 25 minutes remaining and you may want to move on to Part B, if you have not already started answering that question.

After 15 minutes, say:

There are 10 minutes remaining.

After 10 minutes, say:

Stop working and close your exam booklet. Place it on your desk, faceup. . . .

If any students used extra paper for a question in the free-response section, have those students staple the extra sheet(s) to the first page corresponding to that question in their free-response exam booklets. Complete an Incident Report after the exam and return these free-response booklets with the extra sheets attached in the Incident Report return envelope (see page 68 of the *2018-19 AP Coordinator's Manual* for complete details).

Then say:

Remain in your seat, without talking, while the exam materials are collected. . . .

Collect a Section II booklet from each student. Check for the following:

- Exam booklet front cover: The student placed an AP number label on the shaded box and printed their initials and today's date.
- Exam booklet back cover: The student completed the "Important Identification Information" area.

When all exam materials have been collected and accounted for, return to students any electronic devices you may have collected before the start of the exam.

If you are giving the regularly scheduled exam, say:

You may not discuss or share the free-response exam content with anyone unless it is released on the College Board website in about two days. Your AP Exam score results will be available online in July.

If you are giving the alternate exam for late testing, say:

None of the content in this exam may ever be discussed or shared in any way at any time. Your AP Exam score results will be available online in July.

If any students completed the AP number card at the beginning of this exam, say:

Please remember to take your AP number card with you. You will need the information on this card to view your scores and order AP score reporting services online.

Then say:

You are now dismissed.

After-Exam Tasks

Be sure to give the completed seating chart to the AP coordinator. Schools must retain seating charts for at least six months (unless the state or district requires that they be retained for a longer period of time). Schools should not return any seating charts in their exam shipments unless they are required as part of an Incident Report.

NOTE: If you administered exams to students with accommodations, review the *2018-19 AP Coordinator's Manual* and the *2018-19 AP SSD Guidelines* for information about completing the Nonstandard Administration Report (NAR) form, and returning these exams.

The exam proctor should complete the following tasks if asked to do so by the AP coordinator. Otherwise, the AP coordinator must complete these tasks:

- Complete an Incident Report for any students who used extra paper for the free-response section. (Incident Report forms are provided in the coordinator packets sent with the exam shipments.) **These forms must be completed with a No. 2 pencil.** It is best to complete a single Incident Report for multiple students per exam subject, per administration (regular or late testing), as long as all required information is provided. Include all exam booklets with extra sheets of paper in an Incident Report return envelope (see page 68 of the *2018-19 AP Coordinator's Manual* for complete details).
- Return all exam materials to secure storage until they are shipped back to the AP Program. (See page 27 of the *2018-19 AP Coordinator's Manual* for more information about secure storage.) Before storing materials, check the "School Use Only" section on page 1 of the answer sheet and:
  - Fill in the appropriate section number circle in order to access a separate AP Instructional Planning Report (for regularly scheduled exams only) or subject score roster at the class section or teacher level. See "Post-Exam Activities" in the *2018-19 AP Coordinator's Manual.*
  - Check your list of students who are eligible for fee reductions and fill in the appropriate circle on their registration answer sheets.

Name:_____

Answer Sheet for AP Statistics

Practice Exam, Section I

No.	Answer
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

No.	Answer
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	

AP® Statistics Exam

SECTION I: Multiple Choice

2019

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

At a Glance

Total Time
1 hour and 30 minutes
Number of Questions
40
Percent of Total Score
50%
Writing Instrument
Pencil required
Electronic Device
Graphing calculator expected

Instructions

Section I of this exam contains 40 multiple-choice questions. Fill in only the circles for numbers 1 through 40 on your answer sheet.

Indicate all of your answers to the multiple-choice questions on the answer sheet. No credit will be given for anything written in this exam booklet, but you may use the booklet for notes or scratch work. After you have decided which of the suggested answers is best, completely fill in the corresponding circle on the answer sheet. Give only one answer to each question. If you change an answer, be sure that the previous mark is erased completely. Here is a sample question and answer.

Sample Question	Sample Answer
Chicago is a (A) state (B) city (C) country (D) continent (E) village	Ⓐ ● Ⓒ Ⓓ Ⓔ

Use your time effectively, working as quickly as you can without losing accuracy. Do not spend too much time on any one question. Go on to other questions and come back to the ones you have not answered if you have time. It is not expected that everyone will know the answers to all of the multiple-choice questions.

Your total score on the multiple-choice section is based only on the number of questions answered correctly. Points are not deducted for incorrect answers or unanswered questions.

Form I
Form Code 4PBP4-S

90

Formulas

(I) Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$s_p = \sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{\left(n_1 - 1\right) + \left(n_2 - 1\right)}}$$

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}}{\sqrt{\sum\left(x_i - \bar{x}\right)^2}}$$

(II) Probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$E(X) = \mu_x = \sum x_i p_i$$

$$\text{Var}(X) = \sigma_x^2 = \sum \left(x_i - \mu_x\right)^2 p_i$$

If X has a binomial distribution with parameters n and p, then:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mu_x = np$$

$$\sigma_x = \sqrt{np(1-p)}$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If $\bar{x}$ is the mean of a random sample of size n from an infinite population with mean μ and standard deviation σ, then:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(III) Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: statistic $\pm$ (critical value) $\bullet$ (standard deviation of statistic)

Single-Sample

Statistic	Standard Deviation of Statistic
Sample Mean	$\dfrac{\sigma}{\sqrt{n}}$
Sample Proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

Two-Sample

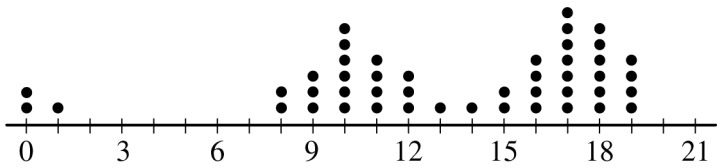
Statistic	Standard Deviation
Difference of sample means	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

Chi-square test statistic $= \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

STATISTICS
SECTION I
Time—1 hour and 30 minutes
Number of questions—40
Percent of total score—50

Directions: Solve each of the following problems, using the available space for scratch work. Decide which is the best of the choices given and fill in the corresponding circle on the answer sheet. No credit will be given for anything written in the test book. Do not spend too much time on any one problem.

1. A group of students played a game in which they earned points for answering questions correctly. The following dotplot shows the total number of points earned by each student.



Which of the following is the best description of the distribution of points earned?

(A) Approximately normal

(B) Bimodal without a gap

(C) Bimodal with a gap

(D) Skewed to the right without a gap

(E) Skewed to the right with a gap

GO ON TO THE NEXT PAGE.

2. The table shows the responses from 103 people when asked if they support a proposal to expand the public library.

	Under the Age of 55	Age 55 or Older	Total
Yes	17	8	25
No	42	36	78
Total	59	44	103

One person from those who responded will be selected at random. Which of the following is closest to the probability that the person selected will be someone who responded no, given that the person selected is age 55 or older?

(A) 0.350

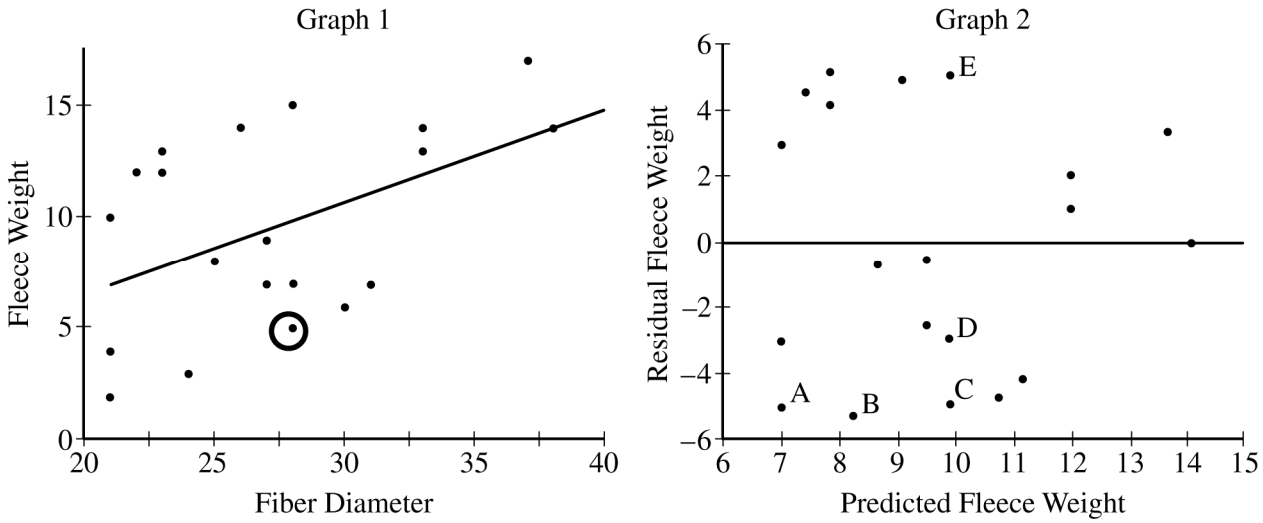
(B) 0.427

(C) 0.462

(D) 0.757

(E) 0.818

3. Data were collected on the fiber diameter and the fleece weight of wool taken from a sample of 20 sheep. The data are shown in the following graphs. Graph 1 is a scatterplot of fleece weight versus fiber diameter with the respective least-squares regression line shown. Graph 2 is the associated plot of the residuals versus the predicted values.



One point is circled on graph 1. Five points labeled A, B, C, D, and E are identified on graph 2. Which point on graph 2 represents the residual for the circled point on graph 1 ?

(A) A

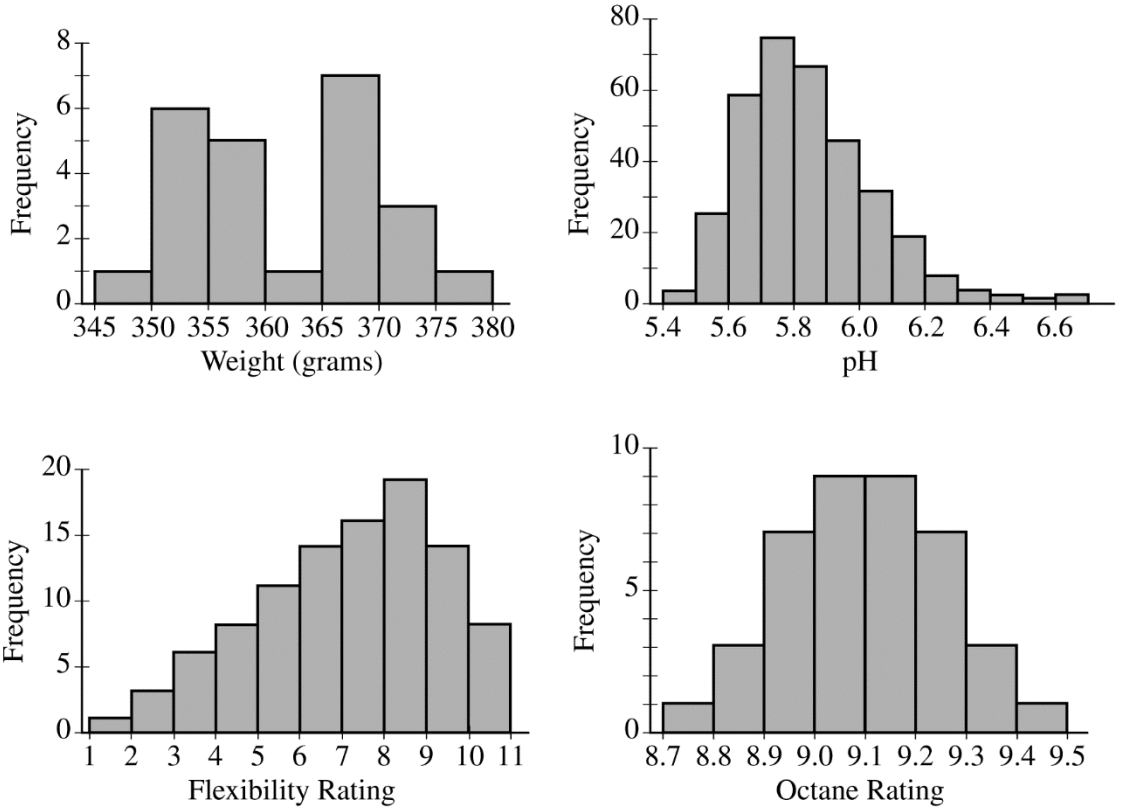
(B) B

(C) C

(D) D

(E) E

4. The distributions of four variables are shown in the following histograms.



Which of the following shapes is NOT represented by one of the four distributions?

(A) Uniform

(B) Bimodal

(C) Skewed to the left

(D) Skewed to the right

(E) Symmetric and unimodal

5. A random sample of 1,092 people were asked whether color was a consideration in buying a new car. They were also asked to identify one additional feature that was important. The responses are shown in the table.

	Color Consideration			
	Yes	No	Maybe	Total
Comfort	40	96	12	148
Cost	108	68	8	184
Performance	62	62	12	136
Reliability	128	116	4	248
Safety	152	192	32	376
Total	490	534	68	1,092

Which of the following is closest to the proportion of people who responded no to color consideration and who identified safety as the additional feature that was important?

(A) 0.18

(B) 0.34

(C) 0.36

(D) 0.49

(E) 0.51

6. Scientists estimate that the distribution of the life span of the Galápagos Islands giant tortoise is approximately normal with mean 100 years and standard deviation 15 years. Based on the estimate, which of the following is closest to the age of a Galápagos Islands giant tortoise at the 90th percentile of the distribution?

   (A) 80 years

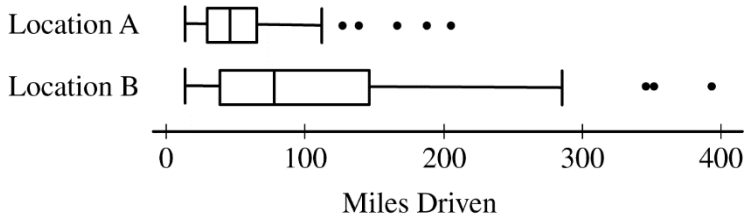
   (B) 115 years

   (C) 120 years

   (D) 125 years

   (E) 130 years

7. A car rental agency has two locations in a city. The boxplots below summarize the miles driven for one day of single-day car rentals at each location.



Based on the boxplots, which statement provides the best comparison of the two locations?

(A) The number of single-day rentals is greater for location A than for location B.

(B) The number of single-day rentals is less for location A than for location B.

(C) Compared with location A, the miles driven for location B display more variability, and the median is greater.

(D) Compared with location A, the miles driven for location B display less variability, and the median is greater.

(E) Compared with location A, the miles driven for location B display less variability, and the median is about the same.

8. For the purpose of determining the value of its end-of-year inventory, a clothing store creates a list at the end of the year of every item currently in stock along with each item's wholesale price. Which of the following is the best description of the end-of-the-year activity?

   (A) An experiment, because the items are treatments and wholesale prices are responses.

   (B) An experiment, because the store does not know the total wholesale price of all the items.

   (C) A sample survey, because the store wants to estimate the value of all items for the entire year.

   (D) A sample survey, because the items currently in the store at the end of the year are a random sample of all items in the store for the entire year.

   (E) A census, because the wholesale prices of all items are listed.

9. A grocery store receives deliveries of corn from two farms, one in Iowa and the other in Ohio. Both farms produce ears of corn with mean weight 1.26 pounds. The standard deviation of the weights of the ears of corn from the farm in Ohio is 0.01 pound greater than that from the farm in Iowa. A randomly selected ear of corn from the farm in Iowa weighed 1.39 pounds, which has a standardized score of 1.645 for the distribution of weights for the Iowa corn. If an ear of corn from the farm in Ohio weighs 1.39 pounds, how many standard deviations from the mean is the weight with respect to the Ohio distribution?

(A) 1.46 standard deviations below the mean

(B) 1.46 standard deviations above the mean

(C) 1.65 standard deviations above the mean

(D) 1.88 standard deviations below the mean

(E) 1.88 standard deviations above the mean

---

10. The distribution of number of hours worked by volunteers last year at a large hospital is approximately normal with mean 80 and standard deviation 7. Volunteers in the top 20 percent of hours worked will receive a certificate of merit. If a volunteer from last year is selected at random, which of the following is closest to the probability that the volunteer selected will receive a certificate of merit given that the number of hours the volunteer worked is less than 90 ?

(A) 0.077

(B) 0.123

(C) 0.134

(D) 0.618

(E) 0.923

GO ON TO THE NEXT PAGE.

11. Resting heart rates, in beats per minute, were recorded for two samples of people. One sample was from people in the age-group of 20 years to 30 years, and the other sample was from people in the age-group of 40 years to 50 years. The five-number summaries are shown in the table.

Age-Group (years)	Minimum	Q1	Median	Q3	Maximum
20 to 30	60	71	72	75	84
40 to 50	60	70	73	76	85

The values of 60, 62, and 84 were common to both samples. The three values are identified as outliers with respect to the age-group 20 years to 30 years because they are either 1.5 times the interquartile range (IQR) greater than the upper quartile or 1.5 times the IQR less than the lower quartile.

Using the same method for identifying outliers, which of the three values are identified as outliers for the age-group 40 years to 50 years?

(A) None of the three values is identified as an outlier.

(B) Only 60 is identified as an outlier.

(C) Only 60 and 62 are identified as outliers.

(D) Only 60 and 84 are identified as outliers.

(E) The three values are all identified as outliers.

12. As part of a demographic study, a college administrator needed to survey a sample of students from the college. From each major offered at the college, the administrator randomly selected 5 percent of the students with that major to participate in the survey. Which of the following is the best description of the type of sample selected by the administrator?

(A) Cluster sample

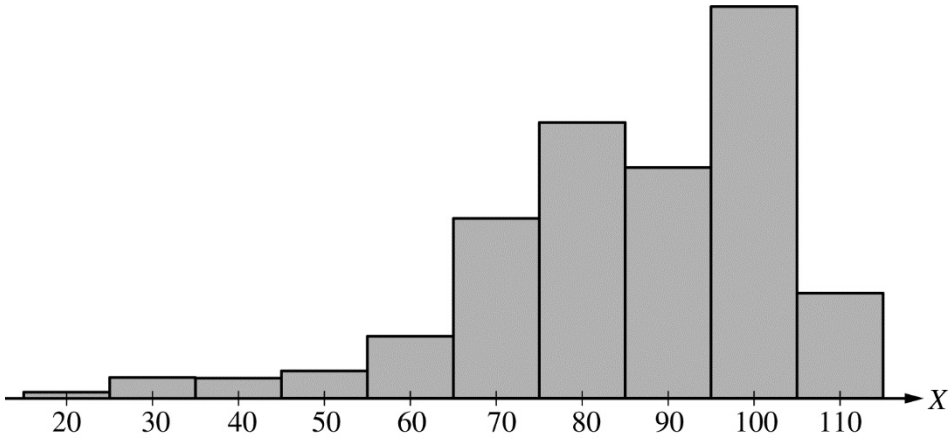
(B) Convenience sample

(C) Simple random sample

(D) Stratified random sample

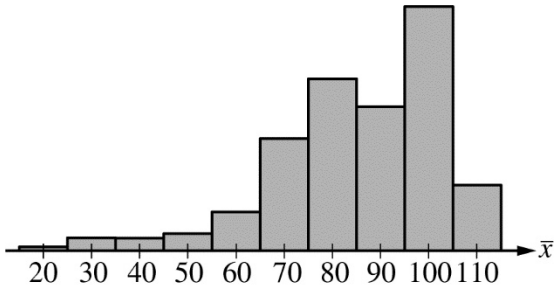
(E) Systematic random sample

13. The graph shows the population distribution of random variable X with mean 85 and standard deviation 18.

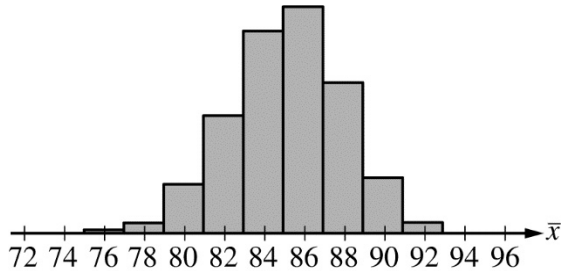


Which of the following graphs is a sampling distribution of the sample mean $\bar{x}$ for samples of size 40 taken from the population?

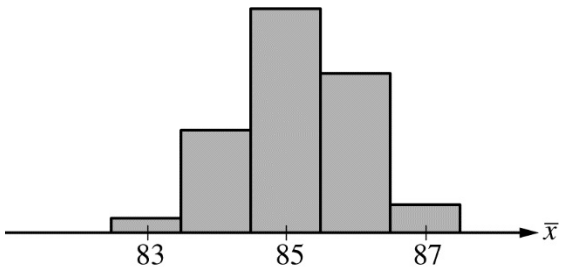
(A)



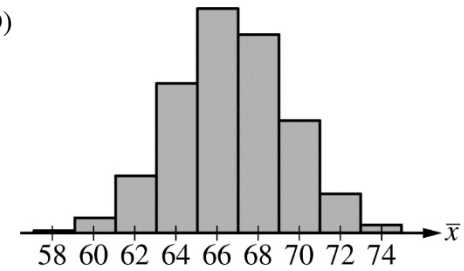
(B)



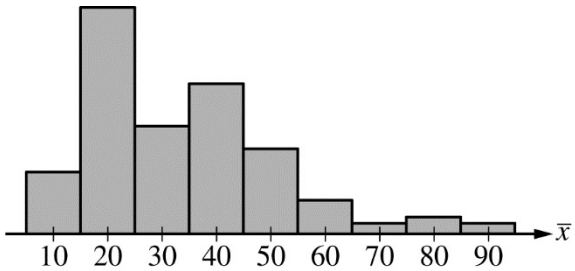
(C)



(D)



(E)



GO ON TO THE NEXT PAGE.

14. A biologist studying trees constructed the confidence interval $(0.14, 0.20)$ to estimate the proportion of trees in a large forest that are dead but still standing. Using the same confidence level, the interval was later revised because the sample proportion had been miscalculated. The correct sample proportion was 0.27. Which of the following statements about the revised interval based on the correct sample proportion is true?

(A) The revised interval is narrower than the original interval because the correct sample proportion is farther from 0.5 than the miscalculated proportion is.

(B) The revised interval is narrower than the original interval because the correct sample proportion is closer to 0.5 than the miscalculated proportion is.

(C) The revised interval is wider than the original interval because the correct sample proportion is farther from 0.5 than the miscalculated proportion is.

(D) The revised interval is wider than the original interval because the correct sample proportion is closer to 0.5 than the miscalculated proportion is.

(E) The revised interval has the same width as the original interval.

15. A research organization reported that 41 percent of adults who were asked to describe their day responded that they were having a good day rather than a typical day or a bad day. To investigate whether the percent would be different for high school students, 600 high school students were randomly selected. When asked to describe their day, 245 students reported that they were having a good day rather than a typical day or a bad day. Do the data provide convincing statistical evidence that the proportion of all high school students who would respond that they were having a good day is different from 0.41 ?

(A) No, because the *p*-value is less than any reasonable significance level.

(B) No, because the *p*-value is greater than any reasonable significance level.

(C) Yes, because the *p*-value is less than any reasonable significance level.

(D) Yes, because the *p*-value is greater than any reasonable significance level.

(E) Yes, because the expected value of the number of students who will report having a good day is 246, not 245.

16. The director of a community recreation center conducted a six-week study to examine the effects of four types of exercise—strength training, flexibility training, aerobics, and jogging—on maximal oxygen consumption. From the 40 members who participated, the director randomly assigned 10 members to each exercise type. Maximal oxygen consumption was measured for each member at the beginning of the study and again at the end of the six weeks. The director examined the change in maximal oxygen consumption for each member. Which of the following statements is a correct description of a feature of the study?

(A) The study has replication because there are four types of exercise.

(B) The study has replication because it was conducted over a six-week period.

(C) The response variable is the type of exercise with the greatest change in maximal oxygen consumption.

(D) The treatments in the study are strength training, flexibility training, aerobics, and jogging.

(E) The experimental units are the four different types of exercise.

17. Sean and Evan are college roommates who have part-time jobs as servers in restaurants. The distribution of Sean's weekly income is approximately normal with mean $225 and standard deviation $25. The distribution of Evan's weekly income is approximately normal with mean $240 and standard deviation $15. Assuming their weekly incomes are independent of each other, which of the following is closest to the probability that Sean will have a greater income than Evan in a randomly selected week?

(A) 0.067

(B) 0.159

(C) 0.227

(D) 0.303

(E) 0.354

---

18. According to data from the United States Elections Project, only 36 percent of eligible voters voted in the 2014 elections. For random samples of size 40, which of the following best describes the sampling distribution of $\hat{p}$, the sample proportion of people who voted in the 2014 elections?

(A) The sampling distribution is skewed to the left, with mean 0.36 and standard deviation 0.076.

(B) The sampling distribution is skewed to the right, with mean 0.64 and standard deviation 0.006.

(C) The sampling distribution is approximately normal, with mean 0.36 and standard deviation 0.076.

(D) The sampling distribution is approximately normal, with mean 0.36 and standard deviation 0.006.

(E) The sampling distribution is approximately normal, with mean 0.64 and standard deviation 0.076.

GO ON TO THE NEXT PAGE.

19. To investigate whether the consumption of beetroot juice enhances exercise performance, a researcher selected a random sample of 50 student athletes from all the student athletes at a college. The athletes in the sample were randomly assigned to one of two groups. In one group, 25 athletes were given a daily dose of beetroot juice, and in the other group, the remaining athletes were given a daily dose of a placebo. At the end of six weeks of exercise training, the researcher compared the performances of the two groups. Based on the design of the investigation, which of the following is the largest population to which the results can be generalized?

(A) The 25 student athletes assigned to the beetroot juice group

(B) The 50 student athletes in the sample

(C) All student athletes at the college

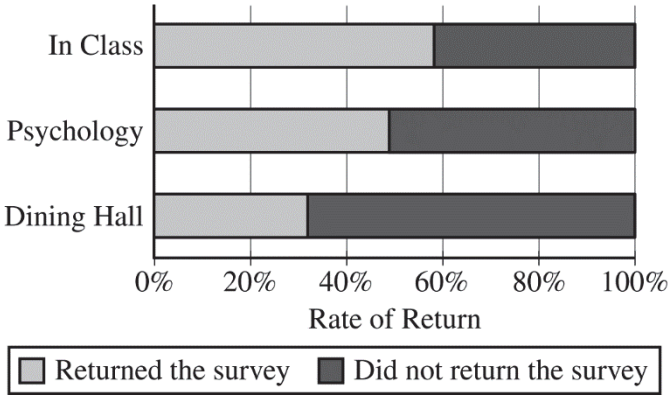
(D) All students at the college

(E) All people who exercise

20. College researchers wanted to know under what conditions people are more likely to complete and return a survey. As part of a study, the researchers prepared three sets of identical surveys and used three methods of delivering and returning the surveys. The methods are described as follows.

- In Class: The surveys were given to students in a class, and students were asked to return completed surveys to their instructor.
- Psychology: The surveys were given to students participating in a psychology experiment, and students were asked to return completed surveys to a collection box in the hallway of the psychology building.
- Dining Hall: The surveys were given to students in the dining hall, and students were asked to return completed surveys to a collection box outside the dining hall.

The graph shows the percent of surveys returned and not returned for each delivery method.



Which statement about delivery method and rate of survey return is supported by the graph?

(A) There is a positive association between delivery method and rate of return.

(B) There is a negative association between delivery method and rate of return.

(C) The number of surveys given using the Dining Hall delivery method was less than the number given using either of the other delivery methods.

(D) The Psychology delivery method displays the most symmetric results; the other delivery methods display skewed results.

(E) The In Class delivery method had the greatest rate of return, and the Dining Hall delivery method had the least rate of return.

21. The height and age of each child in a random sample of children was recorded. The value of the correlation coefficient between height and age for the children in the sample was 0.8. Based on the least-squares regression line created from the data to predict the height of a child based on age, which of the following is a correct statement?

(A) On average, the height of a child is 80% of the age of the child.

(B) The least-squares regression line of height versus age will have a slope of 0.8.

(C) The proportion of the variation in height that is explained by a regression on age is 0.64.

(D) The least-squares regression line will correctly predict height based on age 80% of the time.

(E) The least-squares regression line will correctly predict height based on age 64% of the time.

---

22. At a certain restaurant, the distribution of wait times between ordering a meal and receiving the meal has mean 11.4 minutes and standard deviation 2.6 minutes. The restaurant manager wants to find the probability that the mean wait time will be greater than 12.0 minutes for a random sample of 84 customers. Assuming the wait times among customers are independent, which of the following describes the sampling distribution of the sample mean wait time for random samples of size 84 ?

(A) Approximately normal with mean 11.4 minutes and standard deviation 2.6 minutes

(B) Approximately normal with mean 11.4 minutes and standard deviation $\frac{2.6}{\sqrt{84}}$ minute

(C) Approximately normal with mean 12.0 minutes and standard deviation 2.6 minutes

(D) Binomial with mean $84(0.41)$ minutes and standard deviation $\sqrt{84(0.41)(0.59)}$ minutes

(E) Binomial with mean $84(0.5)$ minutes and standard deviation $\sqrt{84(0.5)(0.5)}$ minutes

GO ON TO THE NEXT PAGE.

23. Two high schools have a similar number of students and parking lots of similar size. The safety officers at both schools want to investigate whether there is an average difference in the number of cars parked per day in the student parking lots for the school year. A random sample of 15 school days will be selected. For each selected day, the number of cars parked in the student parking lots will be counted at both schools and the difference will be recorded. Assuming all conditions for inference are met, which of the following is the appropriate test for the investigation?

(A) A two-sample z-test for a difference between proportions

(B) A two-sample t-test for a difference between means

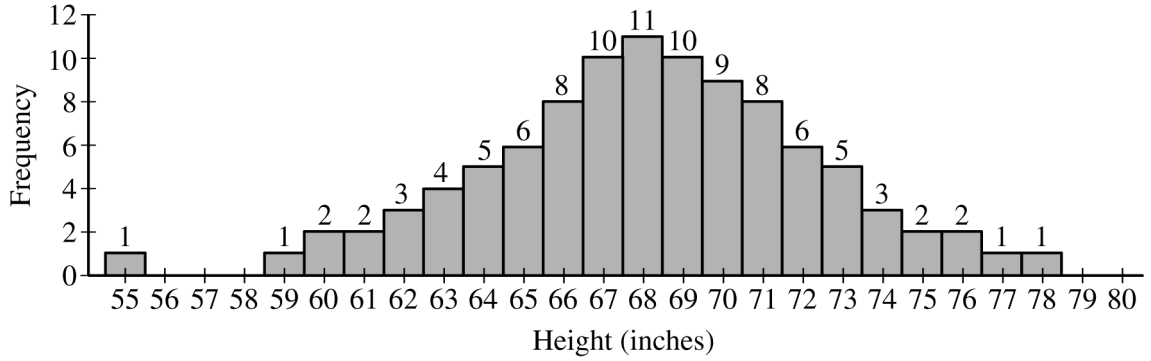
(C) A matched-pairs t-test for a mean difference

(D) A chi-square test of homogeneity

(E) A chi-square test of independence

GO ON TO THE NEXT PAGE.

24. The histogram shows the distribution of heights, in inches, of 100 adult men.



Based on the histogram, which of the following is closest to the interquartile range, in inches, of the distribution?

(A) 2

(B) 5

(C) 9

(D) 12

(E) 15

25. A state study on labor reported that one-third of full-time teachers in the state also worked part time at another job. For those teachers, the average number of hours worked per week at the part-time job was 13. After an increase in state teacher salaries, a random sample of 400 teachers who worked part time at another job was selected. The average number of hours worked per week at the part-time job for the teachers in the sample was 12.5 with standard deviation 6.5 hours. Is there convincing statistical evidence, at the level of $\alpha = 0.05$, that the average number of hours worked per week at part-time jobs decreased after the salary increase?

(A) No. The p-value of the appropriate test is greater than 0.05.

(B) No. The p-value of the appropriate test is less than 0.05.

(C) Yes. The p-value of the appropriate test is greater than 0.05.

(D) Yes. The p-value of the appropriate test is less than 0.05.

(E) Not enough information is given to determine whether there is convincing statistical evidence.

26. A city department of transportation studied traffic congestion on a certain highway. To encourage carpooling, the department will recommend a carpool lane if the average number of people in passenger cars on the highway is less than 2. The probability distribution of the number of people in passenger cars on the highway is shown in the table.

Number of people	1	2	3	4	5
Probability	0.56	0.28	0.08	0.06	0.02

Based on the probability distribution, what is the mean number of people in passengers cars on the highway?

(A) 0.28

(B) 0.56

(C) 1.7

(D) 2

(E) 3

GO ON TO THE NEXT PAGE.

27. To compare the effectiveness of two treatments, researchers conducted a well-designed experiment using a randomized block design in which the subjects were blocked by age-group (under 40 years and 40 years or older). Which of the following must be true about the randomized block design of the experiment?

(A) The number of subjects in each block is different.

(B) Treatments are randomly assigned to subjects within each block.

(C) The design cannot have a control group because subjects are blocked by age-group.

(D) The experiment uses a matched-pairs design, where subjects from one block are paired with subjects from the other block.

(E) The subjects in one block receive one treatment, and the subjects in the other block receive the other treatment.

GO ON TO THE NEXT PAGE.

28. A die used in a certain board game has eight faces, of which 3 are red, 3 are yellow, and 2 are blue. Each face is equally likely to land faceup when the die is tossed. In the game, a player tosses the die until blue lands faceup, and the number of tosses before blue lands faceup is counted. For example, a player who tosses the sequence shown in the following table has tossed the die 3 times before blue lands faceup.

Toss number	1	2	3	4
Face color	yellow	yellow	red	blue

What is the probability that a player will toss the die <u>at least</u> 2 times before blue lands faceup?

(A) 0.1406

(B) 0.4219

(C) 0.4375

(D) 0.5625

(E) 0.5781

29. A statistical test involves the following null and alternative hypotheses.

$$H_0: \mu = 64$$
$$H_a: \mu > 64$$

Which of the following describes a Type II error?

(A) Failing to reject the null hypothesis when the population mean is 64

(B) Failing to reject the null hypothesis when the population mean is greater than 64

(C) Rejecting the null hypothesis when the population mean is 64

(D) Rejecting the null hypothesis when the population mean is greater than 64

(E) Failing to reject the null hypothesis when the *p*-value is less than the significance level

30. The marketing director for an ice cream company investigated whether there was a difference in preference for two new ice cream flavors—cotton candy and mango. Each participant from a large group of people was randomly assigned to taste one of the two flavors. After tasting, each person rated the flavor on a numerical scale from 1 to 5, where 1 represented strongly dislike and 5 represented strongly like. A two-sample *t*-interval for a difference between means (cotton candy minus mango) was constructed. Based on the interval, there was convincing statistical evidence of a difference in population mean flavor ratings, with mango having the greater sample mean rating. Which of the following could be the constructed interval?

(A) $(-20, -15)$

(B) $(-2.1, -1.3)$

(C) $(-1.4, 2.6)$

(D) $(1.5, 2.7)$

(E) $(15, 20)$

---

31. The director of a marketing department wants to estimate the proportion of people who purchase a certain product online. The director originally planned to obtain a random sample of 2,500 people who purchased the product. However, because of budget concerns, the sample size will be reduced to 1,500 people. Which of the following describes the effect of reducing the number of people in the sample?

(A) The variance of the sample will increase.

(B) The variance of the population will decrease.

(C) The variance of the sampling distribution of the estimator will increase.

(D) The variance of the sampling distribution of the estimator will decrease.

(E) The variance of the sampling distribution of the estimator will remain the same.

GO ON TO THE NEXT PAGE.

32. From a random sample of 1,005 adults in the United States, it was found that 32 percent own an e-reader. Which of the following is the appropriate 90 percent confidence interval to estimate the proportion of all adults in the United States who own an e-reader?

(A) $0.32 \pm 1.960\left(\frac{(0.32)(0.68)}{\sqrt{1{,}005}}\right)$

(B) $0.32 \pm 1.645\left(\frac{(0.32)(0.68)}{\sqrt{1{,}005}}\right)$

(C) $0.32 \pm 2.575\sqrt{\frac{(0.32)(0.68)}{1{,}005}}$

(D) $0.32 \pm 1.960\sqrt{\frac{(0.32)(0.68)}{1{,}005}}$

(E) $0.32 \pm 1.645\sqrt{\frac{(0.32)(0.68)}{1{,}005}}$

GO ON TO THE NEXT PAGE.

33. Measuring the height of a tree is usually more difficult than measuring the diameter of the tree. Therefore, many researchers use regression models to predict the height of a tree from its diameter measured at 4 feet 6 inches from the ground. The following computer output shows the results of a linear regression based on the heights, in feet, and the diameters, in inches, recorded from 31 felled trees.

	Estimate	Std Error	t value	Pr(> $\|t\|$)
Intercept	62.031	4.383	14.15	0.0000
Diameter	1.054	0.322	3.27	0.0028

Which of the following is a 95 percent confidence interval for the slope of the population regression line?

(A) $(0.001, 2.107)$

(B) $(0.396, 1.712)$

(C) $(0.423, 1.685)$

(D) $(0.732, 1.376)$

(E) $(53.07, 70.99)$

34. A group of 80 people who had been diagnosed as prediabetic because of high blood glucose levels volunteered to participate in a study designed to investigate the use of cinnamon to reduce blood glucose to a normal level. Of the 80 people, 40 were randomly assigned to take a cinnamon tablet each day and the other 40 were assigned to take a placebo each day. The people did not know which tablet they were taking. Their blood glucose levels were measured at the end of one month. The results showed that 14 people in the cinnamon group and 10 people in the placebo group had normal blood glucose levels. For people similar to those in the study, do the data provide convincing statistical evidence that the proportion who would be classified as normal after one month of taking cinnamon is greater than the proportion who would be classified as normal after one month of not taking cinnamon?

(A) No conclusion can be made about the use of cinnamon because the people in the study were volunteers.

(B) There is convincing statistical evidence at the level of 0.01.

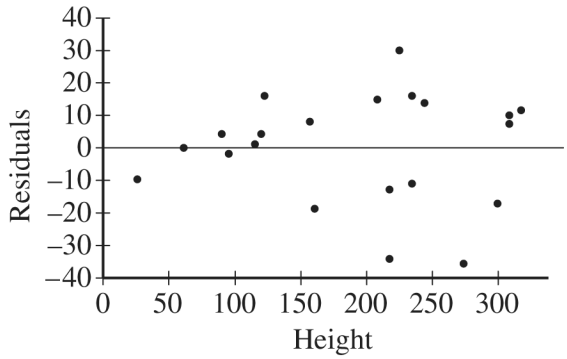
(C) There is convincing statistical evidence at the level of 0.05 but not at the level of 0.01.

(D) There is convincing statistical evidence at the level of 0.10 but not at the level of 0.05.

(E) There is not convincing statistical evidence at any reasonable significance level.

GO ON TO THE NEXT PAGE.

35. A student working on a physics project investigated the relationship between the speed and the height of roller coasters. The student collected data on the maximum speed, in miles per hour, and the maximum height, in feet, for a random sample of 21 roller coasters, with the intent of testing the slope of the linear relationship between maximum speed and maximum height. However, based on the residual plot shown, the conditions for such a test might not be met.



Based on the residual plot, which condition appears to have been violated?

(A) The errors are independent.

(B) The sum of the residuals is 0.

(C) The expected value of the errors is 0.

(D) There is a linear relationship between the response variable and the explanatory variable.

(E) The variance of the response variable is constant for all values of the explanatory variable.

GO ON TO THE NEXT PAGE.

36. A researcher conducted a *t*-test of the hypotheses $H_0 : \mu = 38$ versus $H_a : \mu \neq 38$. The sample mean was 35 and the *p*-value for the test was 0.0627. What would the *p*-value have been if the researcher had used $H_a : \mu < 38$ as the alternative hypothesis?

(A) $1 - 0.0627$

(B) $1 - 2(0.0627)$

(C) $1 - \left(\frac{1}{2}\right)(0.0627)$

(D) $2(0.0627)$

(E) $\frac{1}{2}(0.0627)$

37. High school students from track teams in the state participated in a training program to improve running times. Before the training, the mean running time for the students to run a mile was 402 seconds with standard deviation 40 seconds. After completing the program, the mean running time for the students to run a mile was 368 seconds with standard deviation 30 seconds. Let X represent the running time of a randomly selected student before training, and let Y represent the running time of the same student after training. Which of the following is true about the distribution of $X - Y$?

(A) The variables X and Y are independent; therefore, the mean is 34 seconds and the standard deviation is 10 seconds.

(B) The variables X and Y are independent; therefore, the mean is 34 seconds and the standard deviation is 50 seconds.

(C) The variables X and Y are not independent; therefore, the standard deviation is 50 seconds and the mean cannot be determined with the information given.

(D) The variables X and Y are not independent; therefore, the mean is 34 seconds and the standard deviation cannot be determined with the information given.

(E) The variables X and Y are not independent; therefore, neither the mean nor the standard deviation can be determined with the information given.

GO ON TO THE NEXT PAGE.

38. A polling organization surveyed 2,002 randomly selected adults who are not scientists and 3,748 randomly selected adults who are scientists. Each adult was asked the question "Do you think that genetically modified foods are safe to eat?" Of those who are not scientists, 37 percent responded yes, and of those who are scientists, 88 percent responded yes. Which of the following is the standard error used to construct a confidence interval for the difference between the proportions of all adults who are not scientists and all adults who are scientists who would answer yes to the question?

(A) $\sqrt{\frac{(0.37)(0.63)}{2{,}002} + \frac{(0.88)(0.12)}{3{,}748}}$

(B) $\sqrt{\frac{(0.37)(0.63)}{2{,}002} - \frac{(0.88)(0.12)}{3{,}748}}$

(C) $\sqrt{\frac{(0.37)(0.63)}{2{,}002}} + \sqrt{\frac{(0.88)(0.12)}{3{,}748}}$

(D) $\sqrt{\frac{(0.70)(0.30)}{2{,}002}} + \sqrt{\frac{(0.70)(0.30)}{3{,}748}}$

(E) $\frac{(0.37)(0.63)}{\sqrt{2{,}002}} + \frac{(0.88)(0.12)}{\sqrt{3{,}748}}$

GO ON TO THE NEXT PAGE.

39. A polling agency conducted a survey about social media in which each person in random samples of 1,000 men and 1,000 women was asked what factor he or she considers to be the most important when deciding whether to connect on social media with another person. The responses are shown in the table.

	Factor				
	Personal Friend	Stay in Touch	Mutual Friends	Business Networking	Other
Men	600	210	105	45	40
Women	650	224	65	15	46

What is the contribution to the chi-square test statistic for men who selected business networking as the most important factor?

(A) 0.5

(B) 5

(C) 7.5

(D) 30

(E) 45

40. A national survey asked 1,501 randomly selected employed adults how many hours they work per week. Based on the collected data, a 95 percent confidence interval for the mean number of hours worked per week for all employed adults was given as $(41.18, 42.63)$. Which of the following statements is a correct interpretation of the interval?

(A) Ninety-five percent of all employed adults work between 41.18 hours and 42.63 hours per week.

(B) The probability is 0.95 that a sample of size 1,501 will produce a mean between 41.18 hours and 42.63 hours.

(C) Of all samples of size 1,501 taken from the population, 95% of the samples will have a mean between 41.18 hours and 42.63 hours.

(D) We are 95% confident that the mean number of hours worked per week for employed adults in the sample is between 41.18 hours and 42.63 hours.

(E) We are 95% confident that the mean number of hours worked per week for all employed adults is between 41.18 hours and 42.63 hours.

END OF SECTION I

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION.

DO NOT GO ON TO SECTION II UNTIL YOU ARE TOLD TO DO SO.

MAKE SURE YOU HAVE DONE THE FOLLOWING.

- **PLACED YOUR AP NUMBER LABEL ON YOUR ANSWER SHEET**
- **WRITTEN AND GRIDDED YOUR AP NUMBER CORRECTLY ON YOUR ANSWER SHEET**
- **TAKEN THE AP EXAM LABEL FROM THE FRONT OF THIS BOOKLET AND PLACED IT ON YOUR ANSWER SHEET**

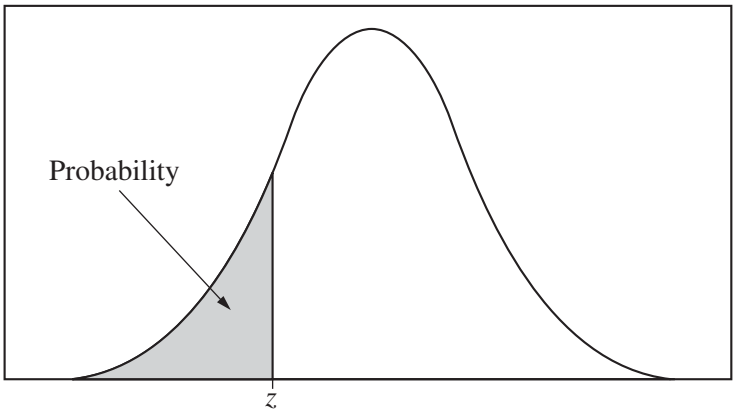


Table entry for z is the probability lying below z.

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

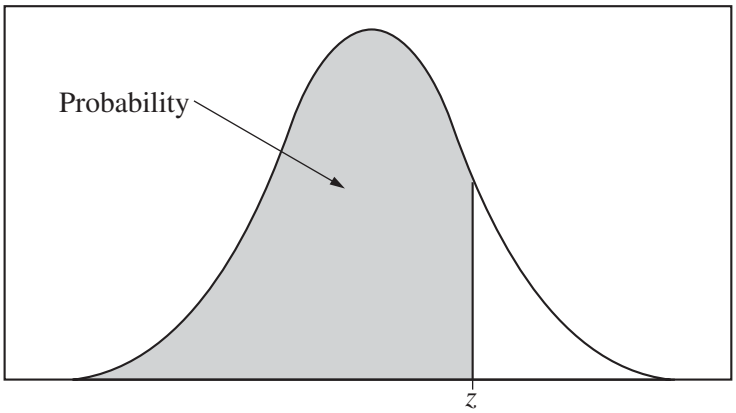


Table entry for z is the probability lying below z.

Table A (***Continued***)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.

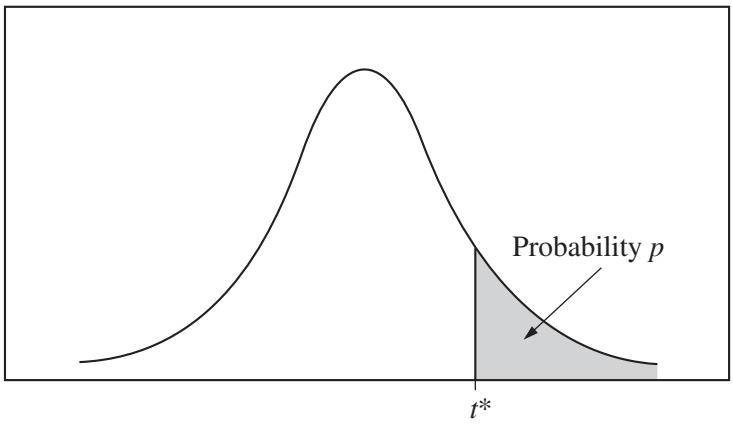


Table B t **distribution critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Table entry for p is the point (χ^2) with probability p lying above it.

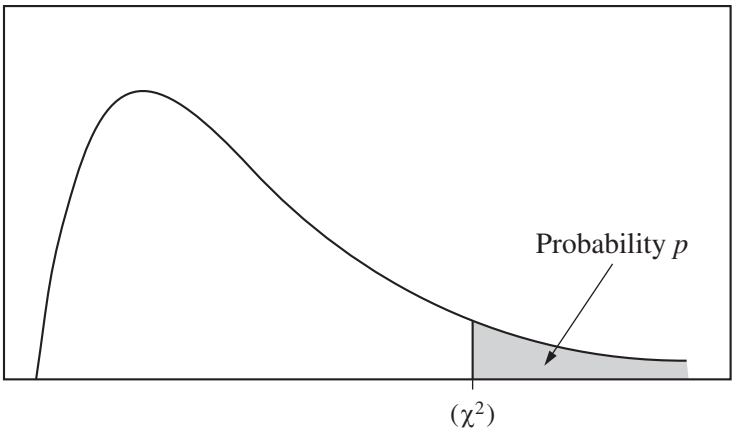


Table C χ^2 critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

AP® Statistics Exam

SECTION II: Free Response

2019

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

At a Glance

Total Time
1 hour and 30 minutes
Number of Questions
6
Percent of Total Score
50%
Writing Instrument
Either pencil or pen with black or dark blue ink
Electronic Device
Graphing calculator expected

Part A

Number of Questions
5
Suggested Time
1 hour and 5 minutes
Percent of Section II Score
75%

Part B

Number of Questions
1
Suggested Time
25 minutes
Percent of Section II Score
25%

IMPORTANT Identification Information

PLEASE PRINT WITH PEN:

1. First two letters of your last name ☐☐
   First letter of your first name ☐
2. Date of birth
   ☐☐ Month ☐☐ Day ☐☐☐☐ Year
3. Six-digit school code
   ☐☐☐☐☐☐
4. Unless I check the box below, I grant the College Board the unlimited right to use, reproduce, and publish my free-response materials, both written and oral, for educational research and instructional purposes. My name and the name of my school will not be used in any way in connection with my free-response materials. I understand that I am free to mark "No" with no effect on my score or its reporting.

   No, I do not grant the College Board these rights. ☐

Instructions

The questions for both Part A and Part B are printed in this booklet. You may use any blank space in the booklet to organize your answers and for scratch work, but you must write your answers in the spaces provided for each answer. Pages containing statistical tables and useful formulas are printed in this booklet.

You may wish to look over the questions before starting to work on them. It is not expected that everyone will be able to complete all parts of all questions. Show all your work. Indicate clearly the methods you use because you will be scored on the correctness of your methods as well as the accuracy and completeness of your results and explanations. Correct answers without supporting work may not receive credit. Write your solution to each part of each question in the space provided for that part. Write clearly and legibly. Cross out any errors you make; erased or crossed-out work will not be scored.

Manage your time carefully. The proctor will announce the suggested time for Part A and Part B, but you may proceed freely from one question to the next. You may review your responses if you finish before the end of the exam is announced.

Form I
Form Code 4PBP4-S

90

Formulas

(I) Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$s_p = \sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{\left(n_1 - 1\right) + \left(n_2 - 1\right)}}$$

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}}{\sqrt{\sum\left(x_i - \bar{x}\right)^2}}$$

(II) Probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$E(X) = \mu_x = \sum x_i p_i$$

$$\text{Var}(X) = \sigma_x^2 = \sum \left(x_i - \mu_x\right)^2 p_i$$

If X has a binomial distribution with parameters n and p, then:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mu_x = np$$

$$\sigma_x = \sqrt{np(1-p)}$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If $\bar{x}$ is the mean of a random sample of size n from an infinite population with mean μ and standard deviation σ, then:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(III) Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: statistic $\pm$ (critical value) $\bullet$ (standard deviation of statistic)

Single-Sample

Statistic	Standard Deviation of Statistic
Sample Mean	$\dfrac{\sigma}{\sqrt{n}}$
Sample Proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

Two-Sample

Statistic	Standard Deviation
Difference of sample means	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

Chi-square test statistic $= \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

STATISTICS
SECTION II
Part A
Questions 1-5

Spend about 1 hour and 5 minutes on this part of the exam.

Percent of Section II score—75

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

1. Thirty-four college students were asked how much money they spent on textbooks for the current semester. Their responses are shown in the following stemplot.

1	2 3 3 4 5 5 6 7 8
2	1 2 3 4 5 6 8 8 9 9 9
3	1 2 2 7 8 9
4	1 4 5 7
5	1 3
6	2
7	
8	1

Key: 1 | 2 = $120

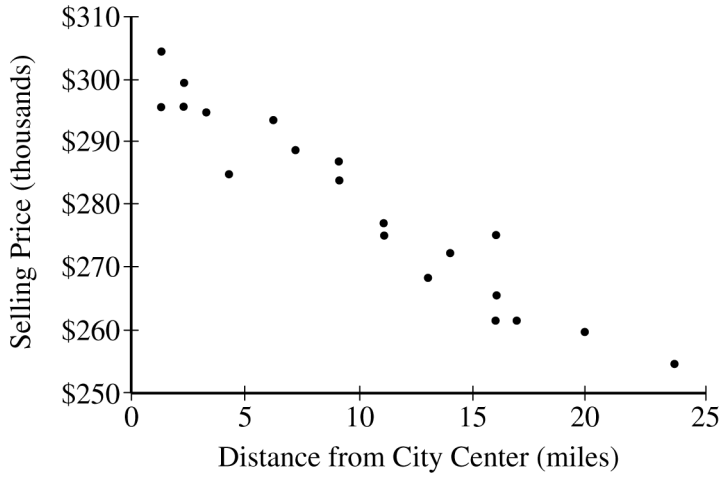
(a) Describe a procedure for identifying potential outliers, and use the procedure to decide whether there are outliers among the responses for the money spent on textbooks.

GO ON TO THE NEXT PAGE.

(b) Based on the stemplot, write a few sentences describing the distribution of money spent on textbooks for the 34 students.

GO ON TO THE NEXT PAGE.

2. A real estate agent working in a large city believes that, for three-bedroom houses, the selling price of the house decreases by approximately $2,000 for every mile increase in the distance of the house from the city center. To investigate the belief, the agent obtained a random sample of 20 three-bedroom houses that sold in the last year. The selling price, in thousands of dollars, and the distance from the city center, in miles, for each of the 20 houses are shown in the scatterplot. The table shows computer output from a regression analysis of the data.



Predictor	Coef	SE Coef	T	P
Constant	301.7	1.80	167.17	0.000
Distance	–2.158	0.149	–14.45	0.000
	S = 4.4336		R-sq = 92.1%	

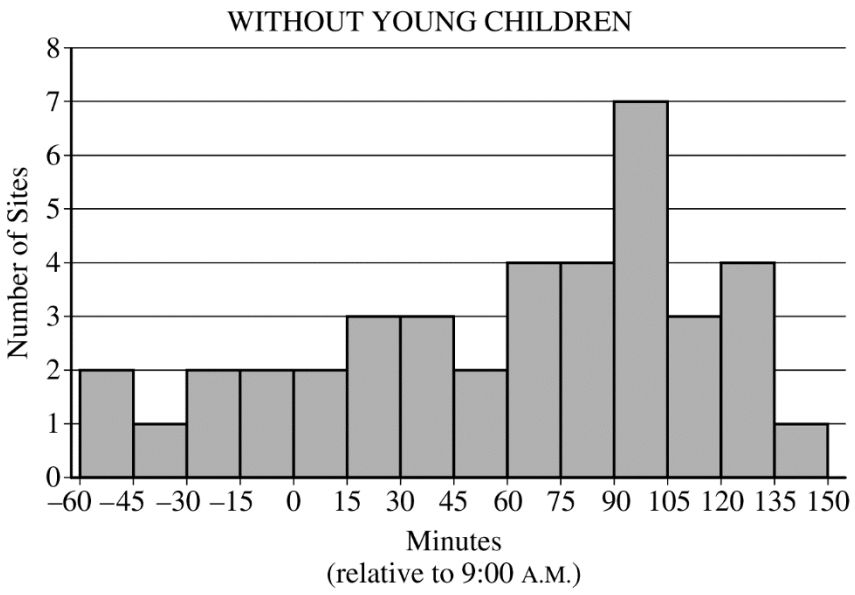
(a) Assume all conditions for inference are met. Construct and interpret a 95 percent confidence interval for the slope of the least-squares regression line.

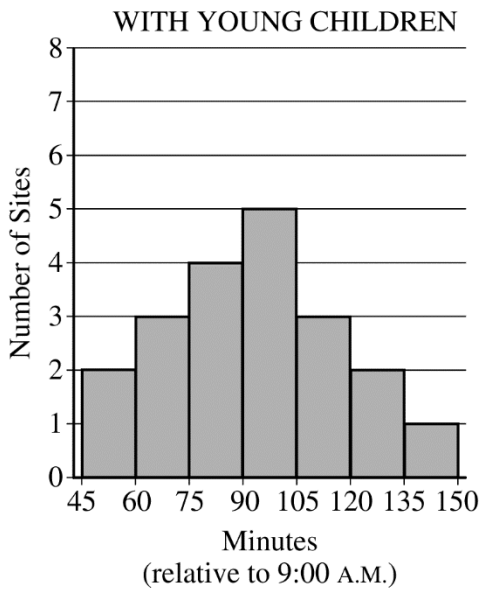
GO ON TO THE NEXT PAGE.

(b) Does the confidence interval contradict the agent's belief about the relationship between selling price and distance from the city center? Justify your answer.

3. River Run campground has sites for people to use for camping. The sites can be reserved for a certain number of days. To help with cleaning and maintenance, the campground requests an exit time (the time at which campers leave the site) of 9 A.M. on the last day of the reservation.

   To estimate the typical exit time, the manager of River Run selected a random sample of 60 sites. Of the selected sites, 40 were reserved by people without young children, and 20 were reserved by people with young children. The following histograms summarize the exit times, recorded as minutes relative to 9 A.M. For example, an exit time of 9:30 A.M. is 30 minutes relative to an exit time of 9 A.M. Each interval contains possible values from the left endpoint up to but not including the right endpoint.





(a) Consider the two histograms.

(i) How many of the 60 sites had an exit time before 8:30 A.M.?

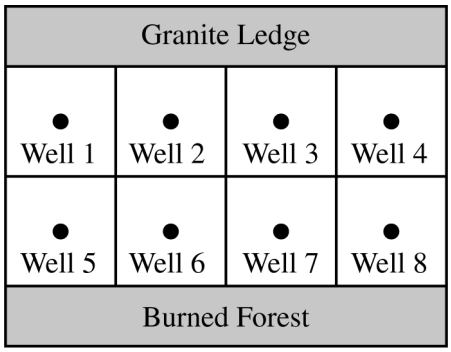
(ii) How many of the 60 sites had an exit time of 11:00 A.M. or later?

(b) Compare the distributions of the exit times for those without young children and those with young children.

(c) Based on the histograms, what is a reasonable estimate of the median exit time for the random sample of 60 sites? Explain your reasoning.

4. Arsenic is a naturally occurring chemical that can enter groundwater through eroding granite or from a burned forest. A health organization recommends drinking water should contain no more than 10 parts per billion (ppb) of arsenic. A company produces filters to clean arsenic from private wells that could be affected by the groundwater.

   The company wants to investigate the effectiveness of a new filter compared to that of an older filter. They will test the filters on a field that is bordered on one side by a granite ledge and on the other side by a burned forest. The field is divided into 8 square plots of equal size, and a well to collect groundwater is drilled in the center of each plot. One filter will be used in each well. The following diagram shows the placement of the wells in the field.



   The company will use four of each type of filter to conduct the investigation. A randomized block design will be used.

   (a) Assuming there is a difference in the effectiveness of the two filters, under what conditions will a randomized block design be better for detecting the difference than a completely randomized design?

GO ON TO THE NEXT PAGE.

(b) Identify the wells, by number, that will be included in each block.

(c) Describe how to assign filters to wells to create a randomized block design.

GO ON TO THE NEXT PAGE.

5. For each day that Sasha travels to work, the probability that she will experience a delay due to traffic is 0.2. Each day can be considered independent of the other days.

   (a) For the next 21 days that Sasha travels to work, what is the probability that Sasha will experience a delay due to traffic on at least 3 of the days?

   (b) What is the probability that Sasha's first delay due to traffic will occur after the fifth day of travel to work?

(c) Consider a random sample of 21 days that Sasha will travel to work. For the proportion of those days that she will experience a delay due to traffic, is the sampling distribution of the sample proportion approximately normal? Justify your answer.

STATISTICS

SECTION II

Part B

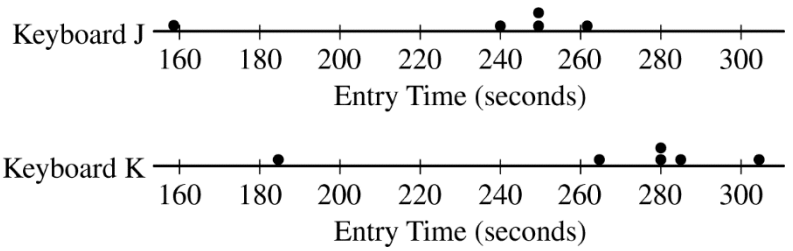
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

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

6. Researchers are studying two different designs of computer keyboards, J and K, to investigate the effectiveness of the design on the speed of data entry. The researchers believe there is a tendency for people entering data with keyboard J to have faster entry times compared with people entering data with keyboard K. Using the same set of data for entry, the researchers randomly assigned 5 people to keyboard J and 6 people to keyboard K and recorded the number of seconds each person took to enter the data. The following dotplots show the observed entry times for the two keyboards.



(a) Explain why it is <u>not</u> appropriate to conduct a two-sample *t*-test for the difference in population means.

(b) Based on the dotplots, explain why it might be more appropriate to compare population medians instead of population means.

One test used to compare population medians is the Wilcoxon Rank Sum Test. Under the assumption that the shape and variability of the distributions are the same, the test uses the rankings of the combined observed values. To conduct the test, the entry times for keyboards J and K are combined into one group and then ranked from 1 to n_T, the total number of observed values in the combined group. The observed entry times, in seconds, for both keyboard types are shown in the following table.

	Observed Entry Times					
J	158	240	248	251	261	
K	184	267	279	280	284	305

(c) Consider the observed entry times for keyboards J and K.

(i) Complete the following table to assign ranks to the observed entry times for keyboards J and K combined.

Rank	1	2	3	4	5	6	7	8	9	10	11
Keyboard	J	K	J							K	K
Time	158	184	240							284	305

(ii) Use the completed table in (i) to calculate the sum of the ranks assigned to each keyboard.

Sum of ranks for J (SR_J):

Sum of ranks for K (SR_K):

The hypotheses for the Wilcoxon Rank Sum Test are as follows.

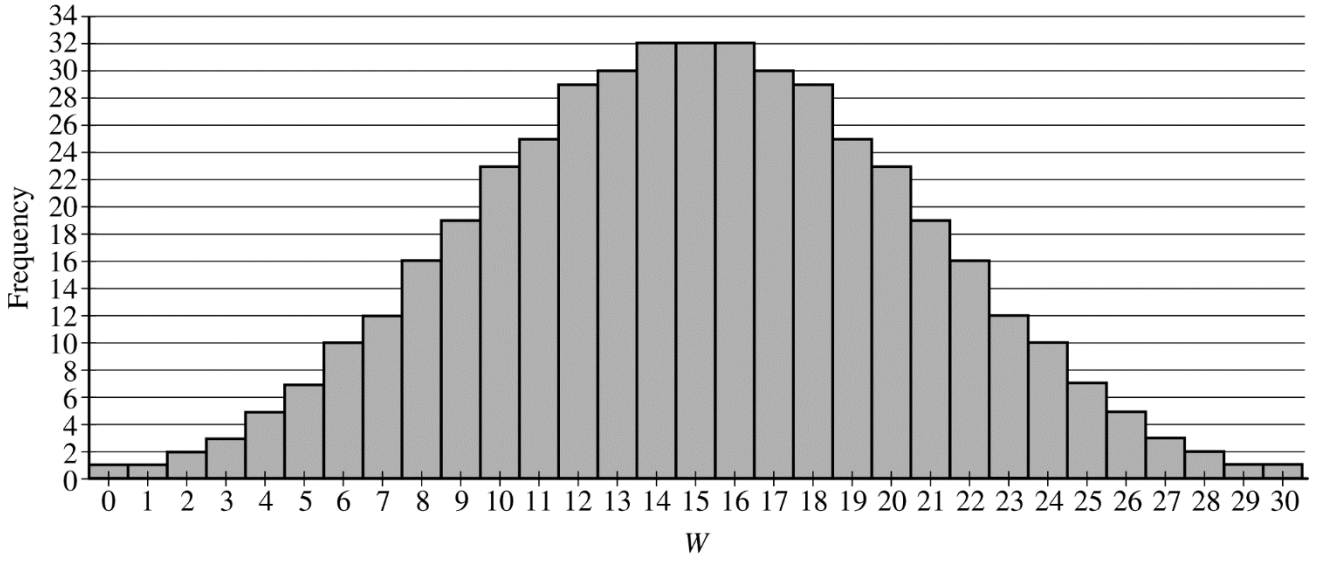
H_0 : The median of the distribution of entry times for all users of keyboard J and the median of the distribution of entry times for all users of keyboard K are the same.

H_a : The median of the distribution of entry times for all users of keyboard J is less than the median of the distribution of entry times for all users of keyboard K.

The test statistic W for the test is $W = SR_J - \frac{n_J(n_J + 1)}{2}$, where n_J is the number of observations for keyboard J.

(d) Use the formula to calculate the test statistic W for the keyboard data.

There are 462 possible assignments of 11 ranks to 5 Js and 6 Ks. If the null hypothesis is true, the 462 assignments are equally likely. The following graph shows the sampling distribution of all possible values of W resulting from the 462 assignments.



(e) The least possible value of W in the sampling distribution is 0.

(i) Find the value of SR_J for $W = 0$.

(ii) Assign 5 ranks to keyboard J and 6 ranks to keyboard K so that $W = 0$. Show your assignments by completing the following table.

Rank	1	2	3	4	5	6	7	8	9	10	11
Keyboard											

(f) Use the test statistic from part (d) and the graph of the sampling distribution to decide whether there is convincing statistical evidence, at the level of $\alpha = 0.05$, that the median of the distribution for all users of keyboard J is less than the median of the distribution for all users of keyboard K. Support your answer.

STOP

END OF EXAM

THE FOLLOWING INSTRUCTIONS APPLY TO THE COVERS OF THE SECTION II BOOKLET.

- **MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE FRONT AND BACK COVERS OF THE SECTION II BOOKLET.**
- **CHECK TO SEE THAT YOUR AP NUMBER LABEL APPEARS IN THE BOX ON THE FRONT COVER.**
- **MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON ALL AP EXAMS YOU HAVE TAKEN THIS YEAR.**

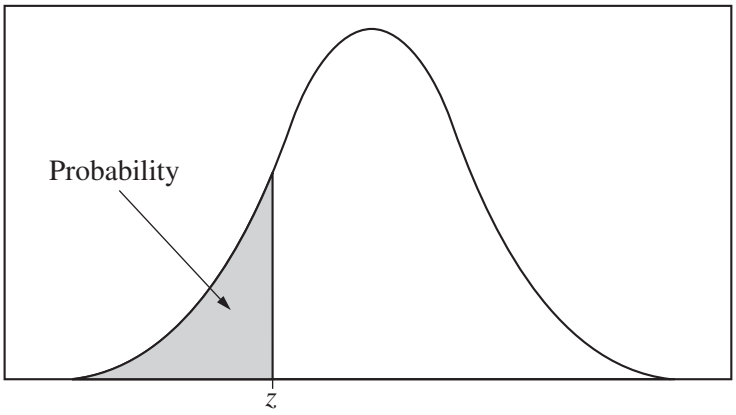


Table entry for z is the probability lying below z.

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table entry for z is the probability lying below z.

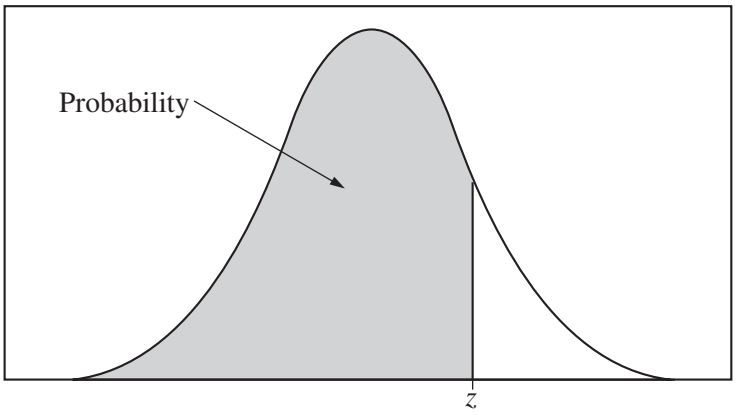


Table A (***Continued***)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.

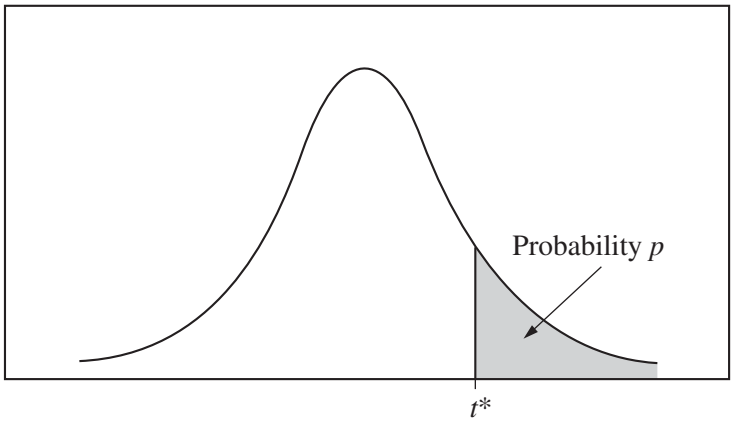


Table B t **distribution critical values**

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Table entry for p is the point (χ^2) with probability p lying above it.

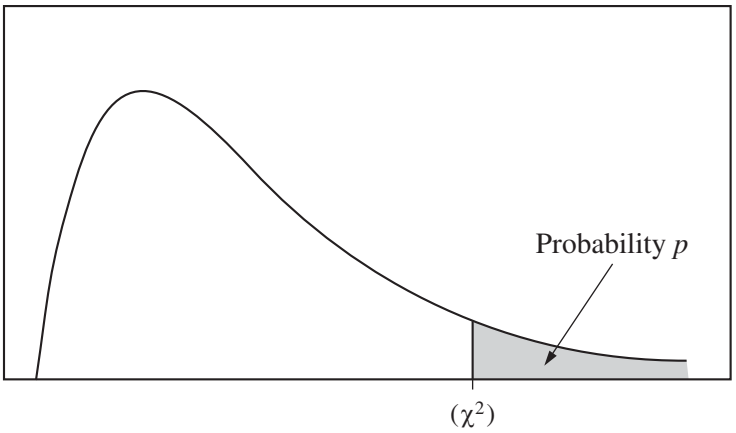


Table C χ^2 critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

Answer Key for AP Statistics
Practice Exam, Section I

Question 1: C
Question 2: E
Question 3: C
Question 4: A
Question 5: A
Question 6: C
Question 7: C
Question 8: E
Question 9: B
Question 10: C
Question 11: B
Question 12: D
Question 13: B
Question 14: D
Question 15: B
Question 16: D
Question 17: D
Question 18: C
Question 19: C
Question 20: E
Question 21: C
Question 22: B
Question 23: C
Question 24: B
Question 25: A
Question 26: C
Question 27: B
Question 28: D
Question 29: B
Question 30: B
Question 31: C
Question 32: E
Question 33: B
Question 34: E
Question 35: E
Question 36: E
Question 37: D
Question 38: A
Question 39: C
Question 40: E

Multiple-Choice Section for Statistics
2019 Course Framework Alignment and Rationales

Question 1

Skill	Learning Objective	Topic
2.A	UNC-1.H	Describing the Distribution of a Quantitative Variable

(A)	Incorrect. The distribution is <u>not</u> approximately normal, since the distribution is neither mound shaped nor symmetric.
(B)	Incorrect. It is true that the distribution is bimodal. However, there are no observed data values between 1 and 8, so there is a gap displayed in the distribution.
(C)	**Correct**. The distribution is bimodal, with one mode at 10 and another mode at 17. Also, there are no observed data values between 1 and 8, so there is a gap displayed in the distribution.
(D)	Incorrect. The distribution is <u>not</u> skewed to the right. A distribution is skewed to the right when the right tail is longer than the left. However, there are no observed data values between 1 and 8, so there is a gap displayed in the distribution.
(E)	Incorrect. It is correct that there is a gap in the distribution. However, the distribution is <u>not</u> skewed to the right. A distribution is skewed to the right when the right tail is longer than the left tail.

Question 2

Skill		Learning Objective	Topic
3.A		VAR-4.D	Conditional Probability
(A)	Incorrect. This is the probability that the person selected is age 55 or older and responded no; it is not the probability that the person selected will be someone who responded no, given that the person selected is age 55 or older.		
(B)	Incorrect. This is the probability that the person selected is age 55 or older; it is not the probability that the person selected will be someone who responded no, given that the person selected is age 55 or older.		
(C)	Incorrect. This is the probability that the person selected was age 55 or older given that the person selected is someone who responded no; it is not the probability that the person selected will be someone who responded no, given that the person selected is age 55 or older.		
(D)	Incorrect. This is the probability that the person selected answered no; it is not the probability that the person selected will be someone who responded no, given that the person selected is age 55 or older.		
(E)	**Correct**. The condition given specifies that the person selected is age 55 or older, and this condition restricts the sample space to 44 people. Of those 44 people, 36 responded no, so the probability is found by $\frac{36}{44} \approx 0.818$.		

Question 3

Skill	Learning Objective	Topic
2.A	DAT-1.F	Residuals

(A)	Incorrect. Point A in Graph 2 has a predicted fleece weight of approximately 7, not a predicted fleece weight of approximately 10, so it cannot be the residual for the circled point in Graph 1.
(B)	Incorrect. Point B in Graph 2 has a predicted fleece weight of approximately 8, not a predicted fleece weight of approximately 10, so it cannot be the residual for the circled point in Graph 1.
(C)	**Correct**. The circled point in Graph 1 corresponds to the sample value that has a fiber diameter of approximately 26 and a predicted fleece weight of approximately 10. For that point, the value of the residual fleece weight can be found using values for the observed fleece weight and predicted fleece weight from Graph 1. The value of the residual is given by $\text{residual} = \text{observed} - \text{predicted} \approx 5 - 10 \approx -5$. Point C is the point on Graph 2 that has a predicted fleece weight of approximately 10 and that has a residual fleece weight that is approximately -5.
(D)	Incorrect. Point D in Graph 2 has a predicted fleece weight of approximately 10, but a residual value of approximately -3, not -5, so it cannot be the residual for the circled point in Graph 1.
(E)	Incorrect. Point E in Graph 2 has a predicted fleece weight of approximately 10, but a residual value of approximately 5, not -5, so it cannot be the residual for the circled point in Graph 1.

Question 4

Skill		Learning Objective	Topic
2.A		UNC-1.H	Describing the Distribution of a Quantitative Variable
(A)	**Correct**. The only shape listed that is not represented by one of the distributions is a uniform shape. The shape of the weight distribution is bimodal. The shape of the pH distribution is skewed to the right. The shape of the flexibility rating distribution is skewed to the left. The shape of the octane rating distribution is symmetric and unimodal.		
(B)	Incorrect. The shape of the weight distribution is bimodal.		
(C)	Incorrect. The shape of the flexibility rating distribution is skewed to the left.		
(D)	Incorrect. The shape of the pH distribution is skewed to the right.		
(E)	Incorrect. The shape of the octane rating distribution is symmetric and unimodal.		

Question 5

Skill		Learning Objective	Topic
2.C		UNC-1.Q	Statistics for Two Categorical Variables
(A)	**Correct**. Of the $1{,}092$ people who responded, 192 responded no to color consideration and also identified safety as the additional feature that is important. The proportion of people who responded no to color consideration and who identified safety as the additional feature that was important is $\frac{192}{1{,}092} \approx 0.18$.		
(B)	Incorrect. This is the proportion of the $1{,}092$ people who responded that safety was the additional feature that was important.		
(C)	Incorrect. This is the proportion of the 534 people who responded no to color consideration who also identified safety as the additional feature that was important.		
(D)	Incorrect. This is the proportion of the $1{,}092$ people who responded no to color consideration.		
(E)	Incorrect. This is the proportion of the $1{,}092$ people who did <u>not</u> respond no to color consideration.		

Question 6

Skill		Learning Objective	Topic
3.A		VAR-2.B	The Normal Distribution
(A)	Incorrect. This is an age that is close to the age of a tortoise at the 10th percentile, not the 90th percentile, of the distribution.		
(B)	Incorrect. This is an age that is close to the age of a tortoise at the 85th percentile, not the 90th percentile, of the distribution.		
(C)	**Correct**. The value of approximately 119.22, found using technology, is the value that has 90 percent of the area to the left of it in the normal distribution with mean 100 and standard deviation 15. Of the values listed, 120 is the tortoise age that is closest to 119.22.		
(D)	Incorrect. This is an age that is close to the age of a tortoise at the 95th percentile, not the 90th percentile, of the distribution.		
(E)	Incorrect. This is an age that is close to the age of a tortoise at the 98th percentile, not the 90th percentile, of the distribution.		

Question 7

Skill		Learning Objective	Topic
2.D		UNC-1.N	Comparing Distributions of a Quantitative Variable
(A)	Incorrect. Boxplots provide information on the proportion of values between certain measures in a distribution, but they give no information about the number of rentals for the locations.		
(B)	Incorrect. Boxplots provide information on the proportion of values between certain measures in a distribution, but they give no information about the number of rentals for the locations.		
(C)	**Correct**. There is more variability in the miles driven for location B than for location A since the interquartile range is greater for B than for A $(120 > 50)$ and the range of values for B is greater than the range of values for A. Also, the median number of miles driven is greater for location B than for location A $(80 > 50)$.		
(D)	Incorrect. It is true that the median is greater for B than for A. However, the miles driven for location B display more variability, not less variability.		
(E)	Incorrect. The miles driven for location B display more variability, not less variability, and the median is <u>not</u> about the same for B as it is for A.		

Question 8

Skill	Learning Objective	Topic
1.C	DAT-2.C	Random Sampling and Data Collection

(A)	Incorrect. No experiment was conducted; the items and prices were observed and recorded.
(B)	Incorrect. No experiment was conducted; the items and prices were observed and recorded.
(C)	Incorrect. The end-of-year activity was not a sample survey, since no sample was selected; every item in stock was used.
(D)	Incorrect. The end-of-year activity was not a sample survey, since no sample was selected; every item in stock was used.
(E)	**Correct**. The end-of-year activity described is a census, since a list is made of every item in stock along with its corresponding wholesale price.

Question 9

Skill	Learning Objective	Topic
3.A	VAR-2.B	The Normal Distribution

(A)	Incorrect. The z-score for the Ohio weight should be positive, so the number of standard deviations should be above the mean, not below the mean.
(B)	**Correct**. The number of standard deviations from the mean is given by $z = \frac{x - \mu}{\sigma}$. For the farm in Iowa, the z-score is 1.645, the value of x is 1.39, the value of μ is 1.26, and the value of σ is unknown. Thus, $1.645 = \frac{1.39 - 1.26}{\sigma}$, and solving for σ yields approximately 0.079. For the farm in Ohio, the value of σ is 0.01 greater than the value of σ for Iowa, so $\sigma = 0.079 + 0.01 = 0.089$. The z-score for Ohio is equal to $z = \frac{1.39 - 1.26}{0.089} \approx 1.46$, so the weight with respect to the Ohio distribution is 1.46 standard deviations above the mean.
(C)	Incorrect. The z-score for the Ohio weight was incorrectly calculated by using a standard deviation of 0.079; 0.089 should have been used.
(D)	Incorrect. The z-score for the Ohio weight was incorrectly calculated by using a standard deviation of 0.069; 0.089 should have been used. Also, the number of standard deviations should be above the mean, not below the mean.
(E)	Incorrect. The z-score for the Ohio weight was incorrectly calculated by using a standard deviation of 0.069; 0.089 should have been used.

Question 10

Skill	Learning Objective	Topic
3.A	VAR-6.B	The Normal Distribution, Revisited

(A)	Incorrect. This is the probability that the number of hours worked by a volunteer selected at random is greater than 90 in a normal distribution with mean 80 and standard deviation 7, not the probability that the volunteer selected will receive the certificate of merit given that the number of hours the volunteer worked is less than 90.
(B)	Incorrect. This is the probability that a volunteer selected at random will have worked between 85.89 hours and 90 hours, not the probability that the volunteer selected will receive the certificate of merit given that the number of hours the volunteer worked is less than 90.
(C)	**Correct**. If X represents the number of hours worked, then the value of X for which 20 percent of the hours worked are greater than X in a normal distribution with mean 80 and standard deviation 7 can be found using technology to be approximately 85.89. Then the probability that the volunteer selected will receive a certificate of merit given that the number of hours the volunteer worked is less than 90 is given by $P(X > 85.89 \mid X < 90) = \frac{P(85.89 < X < 90)}{P(X < 90)}$. Technology can be used to find that $P(85.89 < X < 90) \approx 0.1235$ and that $P(X < 90) \approx 0.924$ in a normal distribution with mean 80 and standard deviation 7, so $P(X > 85.89 \mid X < 90) = \frac{P(85.89 < X < 90)}{P(X < 90)} \approx \frac{0.1235}{0.9234} \approx 0.134.$
(D)	Incorrect. This is approximately equal to dividing the probability that a volunteer selected at random will have worked greater than 90 hours by the probability that a volunteer selected at random will have worked between 85.89 hours and 90 hours.
(E)	Incorrect. This is the probability that a volunteer selected at random will have worked less than 90 hours, not the probability that the volunteer selected will receive the certificate of merit given that the number of hours the volunteer worked is less than 90.

Question 11

Skill		Learning Objective	Topic
2.A		UNC-1.H	Describing the Distribution of a Quantitative Variable
(A)	Incorrect. One of the three values (60) is an outlier.		
(B)	**Correct**. The interquartile range is $76 - 70 = 6$ for the age-group 40 to 50, and 1.5 times the interquartile range is $(1.5)(6) = 9$. Then $Q1 - 9 = 70 - 9 = 61$, and $Q3 + 9 = 76 + 9 = 85$. Of the numbers 60, 62, and 84, only 60 is less than 61 or greater than 85, so 60 is the only outlier.		
(C)	Incorrect. It is true that the value 60 is an outlier. However, the value 62 is not an outlier because 62 is not less than $Q1 - 1.5(\text{IQR})$ or greater than $Q3 + 1.5(\text{IQR})$.		
(D)	Incorrect. It is true that the value 60 is an outlier. However, the value 84 is not an outlier because 84 is not less than $Q1 - 1.5(\text{IQR})$ or greater than $Q3 + 1.5(\text{IQR})$.		
(E)	Incorrect. Only one of the three values (60) is an outlier.		

Question 12

Skill		Learning Objective	Topic
1.C		DAT-2.C	Random Sampling and Data Collection
(A)	Incorrect. A cluster sample involves dividing a population into smaller subgroups. However, the college administrator did not select a simple random sample of all subgroups (majors), and there is no indication that there is heterogeneity within each subgroup (major).		
(B)	Incorrect. A convenience sample was not selected, because a single easily available group of students was not selected to serve as the sample.		
(C)	Incorrect. A simple random sample was not selected, because students were not selected at random from the entire population of students.		
(D)	**Correct**. The administrator selected a stratified random sample, because all of the students at the college were separated into strata (the majors) and a random sample was selected from each of the strata.		
(E)	Incorrect. A systematic random sample was not selected, because the students were selected at random from the majors; it was not the case that every kth student was selected to be in the sample for some integer k.		

Question 13

Skill	Learning Objective	Topic
4.B	UNC-3.Q	Sampling Distributions for Sample Means

(A)	Incorrect. This sampling distribution has the same shape as the population distribution (left-skewed). Because the sample size is sufficiently large, the sampling distribution of the sample mean should be approximately normal.
(B)	**Correct**. For samples of size 40, the sampling distribution of the sample mean should be approximately normal, with a mean equal to $\mu_{\bar{x}} = \mu = 85$ and standard deviation equal to $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{40}} \approx 2.85$. This graph appears to be approximately normal, centered at 85, and with a standard deviation of approximately 2.85.
(C)	Incorrect. It is correct that the sampling distribution of the sample mean should be approximately normal with a mean of 85. However, the standard deviation of the sampling distribution of the sample mean should be equal to $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{40}} \approx 2.85$, and the standard deviation in this graph appears to be much less than 2.85.
(D)	Incorrect. It is correct that the sampling distribution of the sample mean should be approximately normal. However, the sampling distribution of the sample mean should be centered at the population mean of 85, not centered at 66.
(E)	Incorrect. Because the sample size is sufficiently large, the sampling distribution of the sample mean should be approximately normal, not right-skewed. Also, the sampling distribution of the sample mean should be centered at the population mean of 85, not at approximately 35.

Question 14

Skill	Learning Objective	Topic
4.A	UNC-4.H	Justifying a Claim Based on a Confidence Interval for a Population Proportion

(A)	Incorrect. It is true that the interval will be narrower when the sample proportion is farther from 0.5, but in this instance the sample proportion is closer to 0.5.
(B)	Incorrect. The revised interval will be wider, not narrower, for sample proportion values closer to 0.5.
(C)	Incorrect. It is true that the revised interval will be wider than the original interval, but the reason is not because the sample proportion is farther from 0.5 than the miscalculated proportion is.
(D)	**Correct**. The confidence interval is given by the formula $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. When the interval is revised, the value of z^* remains the same since the same confidence level is used, and the value of n remains the same. The original value of $\hat{p}$ was the midpoint of the confidence interval (0.17), but it has now changed to 0.27. The greatest value of $\hat{p}(1-\hat{p})$ will occur when $\hat{p} = 0.5$, and the value will decrease for values closer to 0 or 1. Since $\hat{p} = 0.27$ is closer to 0.5 than $\hat{p} = 0.17$, the revised confidence interval will be wider than the original interval since z^* and n remain the same but $\hat{p}(1-\hat{p})$ will increase.
(E)	Incorrect. The original and revised intervals would have the same width only if the values of $\hat{p}$ were the same, but they are different.

Question 15

Skill		Learning Objective	Topic
4.E		DAT-3.B	Concluding a Test for a Population Proportion
(A)	Incorrect. It is true that the data do not provide convincing statistical evidence, but the p-value is very large, so it is not less than any reasonable significance level.		
(B)	**Correct**. The test statistic for testing the hypotheses $H_0: p = 0.41$ and $H_a: p \neq 0.41$ can be found using $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ or technology. The test statistic has the value -0.083, with the corresponding p-value of approximately 0.934 found using technology. This p-value is greater than any reasonable value for the significance level, so the null hypothesis would not be rejected, and the data do not provide convincing statistical evidence that the proportion of all high school students who would respond they are having a good day is different from 0.41.		
(C)	Incorrect. The data do <u>not</u> provide convincing statistical evidence, and the p-value is very large, so it is not less than any reasonable significance level.		
(D)	Incorrect. It is true that the p-value is greater than any reasonable significance level, but it is not true that the data provide convincing statistical evidence.		
(E)	Incorrect. By itself, the expected value of the number of students who will report having a good day does not provide convincing statistical evidence.		

Question 16

Skill	Learning Objective	Topic
1.C	VAR-3.A	Introduction to Experimental Design

(A)	Incorrect. Replication exists because there were 10 members assigned to each exercise type, not because there are four types of exercise.
(B)	Incorrect. Replication exists because there were 10 members assigned to each exercise type, not because the experiment was conducted over a six-week period.
(C)	Incorrect. The response variable is the change in maximal oxygen consumption measured, not the type of exercise.
(D)	**Correct**. The values for the explanatory variable (exercise) are the treatments, and these values are strength training, flexibility training, aerobics, and jogging.
(E)	Incorrect. An experimental unit is the smallest unit to which a treatment is applied. Each of the 40 members who participated is an experimental unit, not the four different types of exercise.

Question 17

Skill	Learning Objective	Topic
3.B	VAR-5.E	Combining Random Variables

(A)	Incorrect. This value was calculated by using a normal distribution with a correct mean of -15 but using a standard deviation that was incorrectly calculated by subtracting the standard deviations of Sean and Evan.
(B)	Incorrect. This value was calculated by using a normal distribution with a correct mean of -15 but by incorrectly using Evan's standard deviation.
(C)	Incorrect. This value was calculated by using a normal distribution with a correct mean of -15 but using a standard deviation that was incorrectly calculated as $\sqrt{25^2 - 15^2} = 20$.
(D)	**Correct**. Let S and E represent Sean's weekly income and Evan's weekly income, respectively. Because S and E are both approximately normal and independent, the distribution of $S - E$ will be approximately normal with mean $\bar{x}_S - \bar{x}_E = 225 - 240 = -15$ and standard deviation $\sqrt{\sigma_{S-E}^2} = \sqrt{\sigma_S^2 + \sigma_E^2} = \sqrt{25^2 + 15^2} = \sqrt{850}$. The probability that Sean's income is greater than Evan's income is $P(S - E > 0)$ in a normal distribution with mean -15 and standard deviation $\sqrt{850}$, which can be found using technology to be approximately 0.303.
(E)	Incorrect. This value was calculated by using a normal distribution with a correct mean of -15 but using a standard deviation that was incorrectly calculated as $25 + 15 = 40$.

Question 18

Skill	Learning Objective	Topic
3.C	UNC-3.L|UNC-3.K	Sampling Distributions for Sample Proportions

(A)	Incorrect. It is correct that the mean is 0.36 and the standard deviation is 0.076. However, the sampling distribution of the sample proportion is approximately normal because the sample size is large enough.
(B)	Incorrect. The sampling distribution of the sample proportion is approximately normal because the sample size is large enough. Also, the mean and standard deviation are not correct. The mean of the sampling distribution of the sample proportion is given by $\mu_{\hat{p}} = p$, and the standard deviation is given by $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
(C)	**Correct**. The sampling distribution of the sample proportion is approximately normal because the sample size is large enough ($np = 40(0.36) = 14.4$ and $n(1-p) = 40(1-0.36) = 25.6$, each of which is greater than 10). The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p = 0.36$, and the standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.36)(0.64)}{40}} \approx 0.076$.
(D)	Incorrect. It is correct that the sampling distribution is approximately normal and the mean is 0.36. However, the standard deviation is incorrect. The standard deviation of the sampling distribution of the sample proportion is given by $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
(E)	Incorrect. It is correct that the sampling distribution is approximately normal and the standard deviation is 0.076. However, the mean is incorrect. The mean of the sampling distribution of the sample proportion is given by $\mu_{\hat{p}} = p$.

Question 19

Skill		Learning Objective	Topic
4.B		VAR-3.E	Inference and Experiments
(A)	Incorrect. The 25 student athletes who received the beetroot juice are the athletes in the treatment group, but the results of the study can be generalized to the population from which the sample was selected.		
(B)	Incorrect. The 50 student athletes in the sample are the athletes used in the experiment, but the results of the study can be generalized to the population from which the sample was selected.		
(C)	**Correct**. The largest population to which the results can be generalized is the population from which the sample was selected, which is all student athletes at the college.		
(D)	Incorrect. The results of the study can only be generalized to the population from which the sample was selected, which only includes student athletes at the college, not other students at the college who are not athletes.		
(E)	Incorrect. The results of the study can only be generalized to the population from which the sample was selected, which only includes student athletes at the college and does not include other people who exercise but are not from the college.		

Question 20

Skill		Learning Objective	Topic
2.D		UNC-1.P	Representing Two Categorical Variables
(A)	Incorrect. Association cannot be determined from the bar graph.		
(B)	Incorrect. Association cannot be determined from the bar graph.		
(C)	Incorrect. The graph shows the percents of returned surveys, but the numbers cannot be determined unless the total number of surveys is known.		
(D)	Incorrect. Symmetric and skewed results have no meaning in the context of the bar graph.		
(E)	**Correct**. According to the graph, the rate of return for the Dining Hall delivery method was approximately 33 percent, for the Psychology delivery method was approximately 48 percent, and for the In Class delivery method was approximately 58 percent. The In Class delivery method had the greatest rate of return, and the Dining Hall delivery method had the least rate of return.		

Question 21

Skill	Learning Objective	Topic
2.A	DAT-1.G	Least Squares Regression

(A)	Incorrect. This incorrectly describes the meaning of the correlation coefficient r; the correlation coefficient is a measure of the strength of the linear association between age and height and does not give the relationship between an individual age and height.
(B)	Incorrect. The correlation coefficient r is not equal to the slope of the regression line; the correlation coefficient is a measure of the strength of the linear association between age and height.
(C)	**Correct**. The coefficient of determination, r^2, is the proportion of the variation in height that is explained by the least-squares regression line. The value of the coefficient of determination is $r^2 = (0.8)^2 = 0.64$, so the proportion of the variation in height that is explained by a regression on age is 0.64.
(D)	Incorrect. The correlation coefficient r does not give a probability of predicting the height; the correlation coefficient is a measure of the strength of the linear association between age and height.
(E)	Incorrect. The square of the correlation coefficient, r^2, does not give a probability of predicting the height; the coefficient of determination $\left(r^2\right)$ is the proportion of the variation in the response variable explained by the least-squares regression line.

Question 22

Skill	Learning Objective	Topic
3.C	UNC-3.R|UNC-3.Q	Sampling Distributions for Sample Means

(A)	Incorrect. It is correct that the sampling distribution of the sample mean is approximately normal and that the mean is 11.4. However, the standard deviation is incorrect. The standard deviation is given by the formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
(B)	**Correct**. The distribution of wait times is approximately normal because the sample size of 84 is greater than 30. The mean of the sampling distribution of the sample mean is $\mu_{\bar{x}} = \mu = 11.4$, and the standard deviation of the sampling distribution of the sample mean is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.6}{\sqrt{84}}$.
(C)	Incorrect. It is correct that the sampling distribution of the sample mean is approximately normal. However, the value of 12.0 is the value for which the manager wishes to calculate a probability; it is not equal to the mean of the sampling distribution. The mean of the sampling distribution of the sample mean is $\mu_{\bar{x}} = \mu$. Also, the value of 2.6 is not the correct standard deviation for the sampling distribution of the sample mean. The standard deviation of the sampling distribution is given by $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
(D)	Incorrect. The distribution of the sample mean is not binomial. The mean and standard deviation are not correct since they are calculated using formulas for the binomial distribution.
(E)	Incorrect. The distribution of the sample mean is not binomial. The mean and standard deviation are not correct since they are calculated using formulas for the binomial distribution.

Question 23

Skill	Learning Objective	Topic
1.E	VAR-7.B	Setting Up a Test for a Population Mean

(A)	Incorrect. The safety officers want to investigate whether there is a mean difference in the number of cars, not a difference between proportions.
(B)	Incorrect. A two-sample z-test for a difference between means is not appropriate because the days on which the number of cars were recorded are not independent. The numbers were recorded on the same days for each school.
(C)	**Correct**. The cars in the investigation are matched by day; the number of cars were recorded for the same day at each school. Because the measurements taken at each school were matched by day and the safety officers want to investigate whether there is an average difference for the 15 differences calculated from the matched pairs, the appropriate test is a matched-pairs t-test for a mean difference.
(D)	Incorrect. A chi-square test is not appropriate because the data is quantitative, not qualitative.
(E)	Incorrect. A chi-square test is not appropriate because the data is quantitative, not qualitative.

Question 24

Skill	Learning Objective	Topic
2.C	UNC-1.J	Summary Statistics for a Quantitative Variable

(A)	Incorrect. The interquartile range represents the middle 50 percent of the data. There is no interval of width 2 that contains 50 percent of the data values.
(B)	**Correct**. The first quartile, Q1, is the value that has 25 percent of the data values at or below it, so Q1 = 66. The third quartile, Q3, is the value that has 25 percent of the data values at or above it, so Q3 = 71. The interquartile range is Q3 − Q1 = 71 − 66 = 5.
(C)	Incorrect. The interquartile range represents the middle 50 percent of the data. There is no interval of length 9 such that 25 percent of the data values are less than the left endpoint and 25 percent of the data values are greater than the right endpoint.
(D)	Incorrect. The interquartile range represents the middle 50 percent of the data. There is no interval of length 12 such that 25 percent of the data values are less than the left endpoint and 25 percent of the data values are greater than the right endpoint.
(E)	Incorrect. The interquartile range represents the middle 50 percent of the data. There is no interval of length 15 such that 25 percent of the data values are less than the left endpoint and 25 percent of the data values are greater than the right endpoint.

Question 25

Skill	Learning Objective	Topic
4.E	DAT-3.F	Carrying Out a Test for a Population Mean

(A)	**Correct**. The hypotheses tested are H_0: $\mu = 13$ versus H_a: $\mu < 13$. The test statistic is equal to $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{12.5 - 13}{\frac{6.5}{\sqrt{40}}} \approx -1.54$, with the number of degrees of freedom equal to $n - 1 = 400 - 1 = 399$. The p-value is 0.0624, found using technology. Since the p-value is greater than the value of alpha $(0.0624 > 0.05)$, the null hypothesis is not rejected and there is not convincing statistical evidence to conclude that the average number of hours worked per week at part-time jobs decreased after the salary increase.
(B)	Incorrect. It is correct that there is not convincing statistical evidence to conclude that the average number of hours worked per week at part-time jobs decreased after the salary increase. However, the p-value of the appropriate test is not less than 0.05.
(C)	Incorrect. It is incorrect that there is convincing statistical evidence, but it is correct that the p-value of the appropriate test is greater than 0.05.
(D)	Incorrect. It is incorrect that there is convincing statistical evidence, and it is also incorrect that the p-value of the appropriate test is less than 0.05.
(E)	Incorrect. There is enough information to conduct the appropriate hypothesis test and to make a conclusion.

Question 26

Skill		Learning Objective	Topic
3.B		VAR-5.C	Mean and Standard Deviation of Random Variables
(A)	Incorrect. This is the probability that there are 2 people in a passenger car.		
(B)	Incorrect. This is the probability that there is 1 person in a passenger car.		
(C)	**Correct**. The mean number of people in passenger cars is $1(0.56) + 2(0.28) + 3(0.08) + 4(0.06) + 5(0.02) = 1.7$.		
(D)	Incorrect. The department will base their recommendation on this number of people.		
(E)	Incorrect. This is the mean of the numbers of people, $\frac{1+2+3+4+5}{5} = 3$.		

Question 27

Skill		Learning Objective	Topic
1.B		VAR-3.B	Introduction to Experimental Design
(A)	Incorrect. It is not a requirement that the number of subjects in each block in a randomized block design be different. The number of subjects in each block can be equal or different.		
(B)	**Correct**. A feature of a well-designed experiment is randomization, which reduces the chance of bias in experimental groups. Randomization can be achieved in an experiment by randomly assigning treatments to subjects within each block.		
(C)	Incorrect. Blocking by age-group does not mean that there cannot be a control group.		
(D)	Incorrect. There is no matching between groups in this experiment. The subjects in one group and the subjects in the other group are different and not paired in any way.		
(E)	Incorrect. In a randomized block design, subjects within each of the blocks are randomly assigned to the two treatments.		

Question 28

Skill	Learning Objective	Topic
3.A	UNC-3.E	The Geometric Distribution

(A)	Incorrect. The value 0.1406 represents the probability that a color other than blue lands faceup on the first toss, followed by a color other than blue on the second toss, followed by a blue on the third toss, which is not equal to the probability that the player will toss the die at least 2 times before blue lands faceup.
(B)	Incorrect. The value 0.4219 represents the probability that a color other than blue lands faceup 3 times when the die is tossed 3 times, which is not equal to the probability that the player will toss the die at least 2 times before blue lands faceup.
(C)	Incorrect. The value 0.4375 represents the probability that a player will toss the die fewer than 2 times before blue lands faceup, which is not equal to the probability that the player will toss the die at least 2 times before blue lands faceup.
(D)	**Correct**. Let B represent the number of tosses until a blue lands faceup. The random variable B follows a geometric distribution with $p = 0.25$. The probability that a player will toss the die at least 2 times before blue lands faceup is $P(B \geq 3) = 1 - P(B < 3) = 1 - [P(B = 2) + P(B = 1)] = 1 - [0.25 + (0.25)(0.75)]$.
(E)	Incorrect. The value 0.5781 represents the probability that a player will toss the die fewer than 3 times before blue lands faceup, which is not equal to the probability that the player will toss the die at least 2 times before blue lands faceup.

Question 29

Skill	Learning Objective	Topic
1.B	UNC-5.A	Potential Errors When Performing Tests

(A)	Incorrect. Failing to reject the null hypothesis is a correct decision, not an error, when the null hypothesis is true.
(B)	**Correct**. A Type II error occurs when the null hypothesis is not rejected but it should have been rejected. Not rejecting the null hypothesis means that a conclusion is reached where there is not enough statistical evidence to conclude that the population mean is greater than 64, but in fact the population mean is greater than 64.
(C)	Incorrect. Rejecting the null hypothesis when the null hypothesis is true is a Type I error, not a Type II error.
(D)	Incorrect. Rejecting the null hypothesis when the population mean is greater than 64 is a correct decision, not an error.
(E)	Incorrect. Failing to reject the null hypothesis when the p-value is less than the signficance level is an incorrect decision, but it is neither a Type I nor Type II error.

Question 30

Skill	Learning Objective	Topic
4.B	UNC-4.AA	Justifying a Claim About the Difference of Two Means Based on a Confidence Interval

(A)	Incorrect. The values in the interval are all negative, which is necessary if mango has the greater sample mean rating, but the difference in means must be between -4 and 0, and these values do not meet that condition.
(B)	**Correct**. If there was a statistically significant difference in mean flavor rating, with mango having the greater sample mean rating, then the difference in means (cotton candy minus mango) must be negative. Also, the difference in means must be between -4 and 0 because the ratings for each flavor were between 1 and 5 and mango had the greater sample mean rating. Of the intervals listed, only $(-2.1, -1.3)$ has values that are all negative between -4 and 0.
(C)	Incorrect. The interval represents the set of plausible values for the difference in population means. Because the interval contains negative values, 0, and positive values, it is plausible that the cotton candy mean is greater than the mango mean. It is also plausible that there is no difference in population means (as indicated by 0 in the interval).
(D)	Incorrect. The interval represents the set of plausible values for the difference in population means. This interval provides evidence that the cotton candy mean is greater than the mango mean because all values in the interval are positive.
(E)	Incorrect. The interval represents the set of plausible values for the difference in population means. This interval provides evidence that the cotton candy mean is greater than the mango mean because all values in the interval are positive. Also, it is not possible to construct this interval from the data because the flavors are rated on a scale of 1 to 5.

Question 31

Skill	Learning Objective	Topic
3.B	UNC-3.K	Sampling Distributions for Sample Proportions

(A)	Incorrect. There is no variance associated with a single sample proportion.
(B)	Incorrect. There is no variance associated with a single population proportion.
(C)	**Correct**. The variance of the sampling distribution of the sample proportion is given by $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$. If the value of n is decreased, the value of the fraction will increase. Therefore, the variance of the sampling distribution of the sample proportion will increase.
(D)	Incorrect. As sample size decreases, the variance of the sampling distribution of the sample proportion will increase, not decrease.
(E)	Incorrect. The variance of the sampling distribution of the sample proportion will change as the value of n changes in the formula $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$.

Question 32

Skill	Learning Objective	Topic
3.D	UNC-4.C	Constructing a Confidence Interval for a Population Proportion

(A)	Incorrect. The z^* value used in the confidence interval formula is for a 95 percent confidence interval, not a 90 percent confidence interval. Also, the square root should contain the entire fraction, not just the denominator of the fraction. The correct confidence interval formula is given by $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
(B)	Incorrect. The square root should contain the entire fraction, not just the denominator of the fraction. The correct confidence interval formula is given by $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
(C)	Incorrect. The z^* value used in the confidence interval formula is for a 99 percent confidence interval, not a 90 percent confidence interval.
(D)	Incorrect. The z^* value used in the confidence interval formula is for a 95 percent confidence interval, not a 90 percent confidence interval.
(E)	**Correct**. The formula $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ gives a confidence interval for one-sample proportion. Technology can be used to find critical value z^* for a 90 percent confidence interval. Substituting the values $\hat{p} = 0.32,\ z^* = 1.645,$ and $n = 1{,}005$ into the confidence interval formulas yields $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.32 \pm 1.645\sqrt{\frac{(0.32)(1-0.32)}{1{,}005}} = 0.32 \pm 1.645\sqrt{\frac{(0.32)(0.68)}{1{,}005}}$.

Question 33

Skill	Learning Objective	Topic
4.B	UNC-4.AF	Confidence Intervals for the Slope of a Regression Model

(A)	Incorrect. This interval was obtained by incorrectly using the t-value in the computer output (3.27) in the formula for the confidence interval, which is not the t-value for the confidence interval. The t-value for a confidence interval for the slope is found in a t-table, or using technology for the t-distribution with 29 degrees of freedom.
(B)	**Correct**. The interval estimate for the slope of a regression model is given by the formula $b \pm t^*(\text{SE}_b)$, where b is the slope of the line of best fit, and SE_b is the standard error for the slope of the regression line. The value of b is the estimate of the diameter in the computer output (1.054), and the value of SE_b is the standard error of the diameter in the computer output (0.322). The value of t^* for a 95 percent confidence interval is found using technology to be 2.045, with $n - 2 = 31 - 2 = 29$ degrees of freedom. The confidence interval is thus $1.054 \pm 2.045(0.322)$, which yields the confidence interval $(0.396, 1.712)$.
(C)	Incorrect. This confidence interval was calculated using correct values for b and SE_b in the confidence interval formula $b \pm t^*(\text{SE}_b)$, but incorrectly used the z^* value for a 95 percent interval, not a t^* value with $n - 2 = 31 - 2 = 29$ degrees of freedom.
(D)	Incorrect. This confidence interval used the incorrect formula $b \pm (\text{SE}_b)$, which omits the required t^* value. The correct formula is $b \pm t^*(\text{SE}_b)$.
(E)	Incorrect. This confidence interval used the values for the estimate and standard error for the intercept in the formula but should have used the values of the estimate and standard error for the diameter in the formula. The correct value of t^* was used in the formula.

Question 34

Skill	Learning Objective	Topic
4.E	DAT-3.D	Carrying Out a Test for the Difference of Two Population Proportions

(A)	Incorrect. Randomization was used in the study to randomly assign treatments to the volunteer subjects, so a conclusion can be made.
(B)	Incorrect. The p-value of 0.1645 for the hypothesis test is greater than 0.01, so the null hypothesis is not rejected and there is insufficient statistical evidence to conclude that the proportion of people who would be classified as normal after taking cinnamon is greater than the proportion who would be classified as normal after not taking cinnamon.
(C)	Incorrect. The p-value of 0.1645 for the hypothesis test is greater than either 0.01 or 0.05, so the null hypothesis is not rejected and there is insufficient statistical evidence to conclude that the proportion of people who would be classified as normal after taking cinnamon is greater than the proportion who would be classified as normal after not taking cinnamon.
(D)	Incorrect. The p-value of 0.1645 for the hypothesis test is greater than either 0.10 or 0.05, so the null hypothesis is not rejected and there is insufficient statistical evidence to conclude that the proportion of people who would be classified as normal after taking cinnamon is greater than the proportion who would be classified as normal after not taking cinnamon.
(E)	**Correct**. A two-sample z-test for a difference in population proportions can be conducted to test the hypothesis $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 > 0$, where the subscript 1 represents the cinnamon group and the subscript 2 represents the placebo group. The combined (or pooled) proportion needed for the test is given by $\hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{40\left(\frac{14}{40}\right) + 40\left(\frac{10}{40}\right)}{40 + 40} = 0.3$. The test statistic is equal to $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_c(1-\hat{p}_c)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\frac{14}{40} - \frac{10}{40}}{\sqrt{\frac{3}{10}\left(1 - \frac{3}{10}\right)}\sqrt{\frac{1}{40} + \frac{1}{40}}} \approx 0.976$. The corresponding p-value, found using technology, is approximately 0.1645, which is very large, so there is not convincing statistical evidence at any reasonable significance level.

Question 35

Skill	Learning Objective	Topic
4.C	VAR-7.L	Setting Up a Test for the Slope of a Regression Model

(A)	Incorrect. A residual plot does not indicate if the errors from a sample are independent. To check for independence, data should be collected using a random sample or a randomized experiment, and when a sample is selected without replacement, the sample size must be less than or equal to 10 percent of the population size.
(B)	Incorrect. It is true that the sum of the residuals is 0, but this is not a condition for the test which must be checked.
(C)	Incorrect. It is true that the expected value of the errors is 0, but this is not a condition for the test which must be checked.
(D)	Incorrect. This is a condition for the test to be checked. However, the residual plot is not the most appropriate display to check this condition. A scatterplot of the explanatory variable and response variable is more appropriate to check this condition.
(E)	**Correct**. To test the claim that the maximum height and the maximum speed are linearly related, one of the conditions that must be satisfied is that the residuals must have constant error variance. The displayed residuals are not evenly spread around the horizontal line at 0 since the residual points are closer to the line for heights below 125 and further from the line for heights greater than 125. Thus the requirement of constant error variance for all values of the explanatory variable has not been satisfied.

Question 36

Skill	Learning Objective	Topic
4.B	DAT-3.A	Interpreting P-Values

(A)	Incorrect. This is the probability of obtaining a sample statistic that is <u>not</u> as extreme as the one observed under the assumption that the null hypothesis in the original set of hypotheses is true. However, it cannot be a p -value, since a p -value is the probability of obtaining a test statistic that <u>is</u> as extreme or more extreme than the test statistic observed under the assumption that the null hypothesis is true.
(B)	Incorrect. The value $2(0.0627)$ is the area in the tails of a two-tailed test corresponding to an alternative hypothesis containing a hypothesized value different from 38. Therefore, the value $1-2(0.0627)$ is not equal to the p -value.
(C)	Incorrect. The new test is left tailed, and the value $\frac{1}{2}(0.0627)$ is the area in the left tail. The value $1-\frac{1}{2}(0.0627)$ is the probability of obtaining a sample statistic that is <u>not</u> as extreme as the one observed under the assumption that the null hypothesis in the original set of hypotheses is true, so does not meet the definition of a p -value.
(D)	Incorrect. The new alternative hypothesis corresponds to a left-tailed test, so the area in the left tail should be half of what the area in the two tails was, not twice that area.
(E)	**Correct**. A p -value is the probability of obtaining a test statistic as extreme or more extreme than the test statistic observed under the assumption that the null hypothesis is true. The original set of hypotheses indicates that a two-tailed test is to be conducted, which means that the p -value comprises the sum of the area in the right tail and the area in the left tail. Also, the areas in the tails are equal. If the alternative hypothesis is changed so that the test is left tailed, then the p -value is halved to find the area in only the left tail. Thus the p value would have been $\frac{1}{2}(0.0627)$.

Question 37

Skill	Learning Objective	Topic
3.B	VAR-5.E	Combining Random Variables

(A)	Incorrect. It is true that the mean is 34 seconds. It is not true, however, that the variables are independent, since X and Y represent the running times before and after training for the same student, and it is not true that the standard deviation is 10 seconds.
(B)	Incorrect. It is true that the mean is 34 seconds. It is not true, however, that the variables are independent, since X and Y represent the running times before and after training for the same student, and it is not true that the standard deviation is 50 seconds.
(C)	Incorrect. It is true that the variables X and Y are not independent, since X and Y represent the running times before and after training for the same student. There is, however, enough information to calculate the mean, but there is not enough information provided to calculate the standard deviation.
(D)	**Correct**. The random variables X and Y represent the running times before and after training for the same student, so the variables are dependent, not independent. The mean of $X - Y$ is $\mu_{X-Y} = \mu_X - \mu_Y = 402 - 368 = 34$ seconds. If X and Y are independent, the variance σ^2_{X-Y} of $X - Y$ is equal to $\sigma^2_X + \sigma^2_Y$. Since X and Y are not independent, the variance and hence the standard deviation cannot be determined with the given information.
(E)	Incorrect. It is true that the variables X and Y are not independent, since X and Y represent the running times before and after training for the same student, and it is true that there is not enough information to calculate the standard deviation. There is, however, enough information to calculate the mean.

Question 38

Skill	Learning Objective	Topic
3.D	UNC-4.K	Confidence Intervals for the Difference of Two Proportions

(A)	**Correct**. Let the subscript 1 denote adults who are not scientists, and the subscript 2 denote adults who are scientists. Then $n_1 = 2{,}002$, $n_2 = 3{,}748$, $p_1 = 0.37$, and $p_2 = 0.88$. The standard error is equal to $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{(0.37)(0.63)}{2{,}002} + \frac{(0.88)(0.12)}{3{,}748}}$.
(B)	Incorrect. In this response, the fractions are subtracted, instead of added, in the formula for standard deviation. The standard error is given by $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.
(C)	Incorrect. The two fractions should be added under one square root, not added after the square root is applied to each fraction. The standard error is given by $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.
(D)	Incorrect. The two fractions should be added under one square root, not added after the square root is applied to each fraction. Also, the pooled proportion is incorrectly used for $\hat{p}_1$ and $\hat{p}_2$. The standard error is given by $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.
(E)	Incorrect. The correct values for the sample sizes and proportions were used. A square root of the sum of two fractions should have been used, however, but the square root was applied only to the denominators. The standard error is given by $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.

Question 39

Skill	Learning Objective	Topic
3.E	VAR-8.L	Carrying Out a Chi-Square Test for Homogeneity or Independence

(A)	Incorrect. This incorrectly used $\text{observed count} - \text{expected count}$ in the calculation, but should have used $(\text{observed count} - \text{expected count})^2$. The correct formula is $\frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$.
(B)	Incorrect. This divided by the observed count in the calculation, but should have divided by the expected count. The correct formula is $\frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$.
(C)	**Correct**. The chi-square test statistic is calculated by summing the values $\frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$. The expected count is found by $\frac{(\text{row total})(\text{column total})}{\text{table total}} = \frac{(1{,}000)(60)}{2{,}000} = 30$. The contribution to the test statistic is equal to $\frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} = \frac{(45 - 30)^2}{30} = 7.5$.
(D)	Incorrect. This is the expected count, not the contribution to the chi-square test statistic.
(E)	Incorrect. This is the count of men who considered business networking important, but it is not the contribution to the chi-square test statistic.

Question 40

Skill	Learning Objective	Topic
4.B	UNC-4.S	Justifying a Claim About a Population Mean Based on a Confidence Interval

(A)	Incorrect. The percent is how much confidence exists that the interval has captured the population mean; it is not about the percentage of individual observations in the population that fall within the interval.
(B)	Incorrect. Once the interval is constructed, the interpretation of the confidence interval should not be a statement about probability. Once the sample has been selected and the interval constructed, the unknown population mean was either captured by the interval (probability equal to 1) or not (probability equal to 0).
(C)	Incorrect. Different samples can yield different results. The interval is a statement about how confident we are that we have captured the population parameter, not any possible sample proportion.
(D)	Incorrect. The interval is used to estimate the unknown population mean, not the sample mean. The sample mean is not estimated. It is used to create the interval and will always be at the midpoint of the interval.
(E)	**Correct**. The percent is how much confidence exists that the interval has captured the population mean.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe a procedure for identifying a potential outlier, (2) apply the outlier identification procedure to data presented in a stemplot and (3) describe the distribution of data presented in a stemplot.

Solution

Part (a):

Outliers are identified as any observation outside of the interval bounded by $Q1 - 1.5(IQR)$ and $Q3 + 1.5(IQR)$. The IQR is $Q3 - Q1 = 390 - 180 = \$210$. The interval lower bound is $180 - 1.5(210) = -\$135$ and the interval upper bound is $390 + 1.5(210) = \$705$. Since \$810 is outside of this interval, it is an outlier. No other observation is outside the interval.

Part (b):

The distribution of the amount of money students spent on textbooks is unimodal and skewed to the right. The sample median is between \$280 and \$290. Money spent on textbooks ranges from about \$120 to \$810 for this sample of students, with fifty percent of the amounts between about \$180 and \$390. The largest observation is a possible outlier.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 1 (continued)

Scoring

Parts (a) and (b) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response includes the following two components:

1. Describes a reasonable procedure for identifying potential outliers.
2. Correctly applies the described procedure for detecting potential outliers to the data presented in the stemplot *AND* clearly identifies potential outliers.

Partially correct (P) if the response includes only one of the two components.

Incorrect (I) if the response does not meet the requirements for E or P.

Notes:

- Responses that that satisfy component 1 include, but are not limited to, checking for observations that are more than $(1.5)(\text{IQR})$ above the upper quartile or more than $(1.5)(\text{IQR})$ below the lower quartile, checking for observations that are more than two (or three) standard deviations away from the center of the data (sample mean or sample median), looking for one or more large gaps between any extreme observation and the rest of the data.
- Using the values in the stemplot, the sample mean is \$308.82 and the sample standard deviation is \$155.71. The sample median is between \$280 and \$290. There are no observations more than two standard deviations below the sample mean or sample median. The largest observation is a potential outlier because it is more than two (and three) standard deviations above the sample mean (or sample median).
- If the response does not have a clearly described procedure but shows correct work for the IQR or standard deviation approach, credit is given for component 2.
- Responses that interpret values in the stemplot as purchase amounts that are rounded down, should not be penalized. For example, the second largest observation may be interpreted as a purchase amount between \$620 and \$629.
- If the values in the stemplot are interpreted as being rounded down, the largest observation is an outlier because it is more than two (and three) standard deviations above the sample mean (or sample median). The second largest observation is a potential outlier based on the two standard deviation criterion because $(\text{sample mean}) + 2(\text{standard deviations}) = \$620.24$ and the second largest observation is interpreted as a purchase amount between \$620 and \$629.
- If the procedure checks for a gap, it must identify the largest observation as a potential outlier and it must not identify any other observation as a potential outlier.

Question 1 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response includes reasonable comments on the following four components.

1. The shape of the distribution is skewed to the right.
2. The center of the distribution is around $300. (Any value between $280 and $320 is acceptable.)
3. The spread of the distribution.
4. The response includes context.

Partially correct (P) if the response includes only three of the four components.

Incorrect (I) if the response does not meet the requirements for E or P.

Notes:

- Because part (a) addresses possible outliers, the response to part (b) is not required to address potential outliers.
- Responses that satisfy component 3 include, but are not limited to, referring to the range ($690), referring to the standard deviation (any value between $150 and $160 is acceptable), referring to the interquartile range (any value between $200 and $240 is acceptable), or stating that most or all purchase amounts are between two reasonable amounts (simply stating the values without linking them as endpoints of an interval would not receive credit).
- Responses that interpret values in the stemplot as purchase amounts that are rounded down, should not be penalized; for example, stating that the mean amount of money spent on textbooks is between $308.82 and $317.82.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 1 (continued)

4 **Complete Response**

Both parts essentially correct

3 **Substantial Response**

One part essentially correct and one part partially correct

2 **Developing Response**

One part essentially correct and one part incorrect

OR

Both parts partially correct

1 **Minimal Response**

One part partially correct and one part incorrect

AP® STATISTICS
2019 SCORING GUIDELINES

Question 2

Intent of Question

The primary goal of this question were to assess a student's ability to (1) construct and interpret a confidence interval for the slope of a regression line; and (2) determine if the confidence interval contradicts or supports a prior belief/claim.

Solution:

(a) The 95 percent confidence interval for the slope is computed as follows:

$$-2.158 \pm t_{18,0.975} \times (0.149) = -2.158 \pm (2.101) \times (0.149)$$
$$= (-2.471, -1.845)$$

We are 95% confident that the slope of the true regression line is between -2.471 and -1.845 thousands of dollars per mile. This implies that for each additional mile that a three-bedroom house is away from the city center, the selling price of the house is expected to decline between \$1,845 and \$2,471.

(b) Because the confidence interval contains -2, corresponding to a \$2,000 decrease, it is a plausible value for the slope of the regression line. Consequently, the data do not contradict the agent's belief that the selling prices of three bedroom houses decrease about \$2,000 for every one-mile increase in the distance of the house from the city center.

Scoring

This question is scored in three sections. Section 1 consists of computing the 95% confidence interval for the slope of the true regression line in part (a). Section 2 consists of the interpretation of the confidence interval in part (a). Section 3 consists of the response to part (b). Sections 1, 2, and 3 are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the confidence interval displays the form

(estimated slope) $\pm$ critical value $\times$ *SE*(estimated slope)

AND

correct values are used for the estimated slope (-2.158), the standard error for the estimated slope (0.149), and the 0.975 percentile of the *t*-distribution with 18 degrees of freedom (2.101)

AND

correct values are reported for the endpoints of the confidence interval.

Partially correct (P) if the response displays the form (estimated slope) $\pm$ critical value $\times$ *SE*(estimated slope), but uses an incorrect value for one of the following: the t-percentile, the estimate of the slope, or the standard error of the slope.

OR

if the response gives the correct endpoints of the confidence interval but does not provide sufficient work.

Incorrect if the response does not meet the criteria for E or P.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 2 (continued)

Section 2 is scored as follows:

Essentially correct (E) if the response satisfies the following four components:

1. Uses a 95% level of confidence
2. Provides a correct statement about potential values for the population or expected slope of the regression line.
3. Uses the end points of the confidence interval.
4. Presents the interpretation in context.

Partially correct (P) if the response satisfies only three of the four components.

Incorrect (I) if the response does not meet the criteria for E or P.

Section 3 is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

1. Provides a correct conclusion about whether the confidence interval contradicts the agent's belief based on the results from part (a). If the correct confidence interval is reported in part (a), then the correct conclusion is that the confidence interval does not contradict (or supports) the agents belief, but it should not indicate that it proves the agent's belief.
2. Links the conclusion to the confidence interval based on where $-2{,}000$, a decrease of 2,000, or -2, is relative to the interval.

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) if the response does not meet the criteria for E or P.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 2 (continued)

4 **Complete Response**

Three sections essentially correct

3 **Substantial Response**

Two sections essentially correct and one section partially correct

2 **Developing Response**

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 **Minimal Response**

One section essentially correct

OR

No section essentially correct and two sections partially correct

AP® STATISTICS
2019 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) extract information about counts from a histogram; (2) use histograms to compare distributions; (3) estimate a median from the combined information in two histograms.

Solution

Part (a):

(i) Three sites were vacated before 8:30 A.M. This is the sum of the counts represented by the two leftmost bars of the histogram of exit times for campsites without young children. No campers with small children vacated their campsites before 8:30 A.M.

(ii) Eight sites were vacated at 11:00 A.M. or later. This is the sum of the counts represented by the two rightmost bars on each histogram.

Part (b):

The distribution of exit times for campers without young children is skewed to the left while the distribution for campers with young children is roughly symmetric. Both exit time distributions appear to be unimodal. The distribution of exit times is more spread out for campers without young children; the largest possible range is around 210 minutes compared to a largest possible range of around 105 minutes for campers with young children. Campers without young children tend to leave the campground earlier than campers with young children; the median exit time for campers without young children is between 60 and 75 minutes after 9:00 A.M. which is less than the median exit time for campers with young children, which is between 90 and 105 minutes after 9:00 A.M.

Part (c)

There are a total of 60 exit times, so the median falls between the 30^{th} and 31^{st} exit times. Since the 30^{th} and 31^{st} exit times fall between 10:15 and 10:30, any time between 10:15 A.M. and 10:30 A.M. provides a reasonable estimate of the median exit time.

Scoring

Parts (a), (b) and (c) are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:
1. Correctly states that 3 sites were vacated before 8:30 A.M. in part (i).
2. Correctly states that 8 sites were vacated at 11:00 A.M. or later in part (ii)

Partially correct (P) if the response contains only one of the two components.
OR
Correctly estimates the counts at each site separately, but does not combine them for a total count.

Incorrect (I) if the response does not meet the criteria for E or P.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 3 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following four components:

1. Correct comparison of the centers of the two distributions
2. Correct comparison of the spread of the two distributions
3. Correct comparison of the shapes of the two distributions
4. Includes context

Partially correct (P) if the response satisfies only two or three of the four components.

Incorrect (I) if the response does not meet the criteria for E or P.

Part (c) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

1. Reports a specific time or range of times between 10:15 A.M. and 10:30 A.M. (between 75 and 90 minutes after 9:00 A.M)
2. Gives a reasonable justification.

Partially correct (P) if the response satisfies only one of the two components
OR
if the response correctly estimates median exit times for both sets of sites (between 10:00 A.M and 10:15 A.M., or 60-75 minutes, for campers without young children, and between 10:30 A.M and 10:45 A.M., or 90-105 minutes, for campers with young children)

Incorrect if the response does not meet the criteria for E or P.

Notes: A reasonable justification can be:

- Markings on the histogram(s) that illustrate a histogram of the combined exit times.
- A combination of the medians that properly weights the medians in a 2:1 ratio.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 3 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

AP® STATISTICS
2019 SCORING GUIDELINES

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to (1) identify conditions under which a randomized block design would be better than a completely randomized design for detecting a difference in mean responses for two treatments, and (2) describe how an experiment can be conducted as a randomized block design.

Solution

Part (a):

A randomized block experiment will be better for detecting a difference between the abilities of the new and old filters to reduce arsenic concentrations when (1) there is at least one feature of the wells that enable the researchers to create blocks such that each block consists of wells with similar arsenic concentrations and (2) the average arsenic concentration varies among blocks. This would occur, for example, when well water arsenic concentrations at all four wells near the burned forest are higher than the arsenic concentration at any well near the granite ledge.

Part (b):

To take advantage of the differences among wells, well water arsenic concentrations must be similar for wells with each block, but average arsenic concentrations should vary among blocks. A randomized block experiment with two blocks could be conducted by including wells 1, 2, 3, and 4, on the granite ledge side of the field in one block and including wells 5, 6, 7, and 8, on the burned forest side of the field in the other block.

Part (c)

Within each block, the new filter should be randomly assigned to the same number of wells as the old filter. For the blocks identified in part (b), this can be accomplished by writing well numbers 1, 2, 3, and 4 on four slips of paper, placing the four slips into a hat, and mixing them. New filters would be used for the wells on the first two slips drawn from the hat and old filters would be used for the other two wells. Similarly, well numbers 5, 6, 7, and 8 can be written on four slips of paper that are put into a hat and mixed. New filters would be used for the wells on the first two slips drawn from the hat and old filters would be used for the other two wells.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 4 (continued)

Scoring

Parts (a), (b) and (c) are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:
1. Well water arsenic concentrations are similar for all wells within the same block.
2. Average well water arsenic concentrations vary substantially from block to block.
3. Statements are presented in the context of arsenic levels and wells.

Partially correct (P) if the response includes only two of the three components.

Incorrect if the response does not meet the criteria for E or P.

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following two components. Blocks are formed so that
1. Each block contains wells in locations with similar features, consistent with the scenario identified in the response,
2. Each block contains the same number of wells (either two or four).

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) it the response does not meet the criteria for E or P.

Notes:
- Suppose the response to part (a) indicates that arsenic concentrations might be similar for wells on the granite ledge side of the field, and arsenic concentrations might be similar for wells of the burned forest side of the field, but arsenic concentrations might be quite different on those two sides of the field. Then, the two components required for E are satisfied by including wells 1, 2, 3, 4, in one block and including wells 5, 6, 7 and 8 in the other block. The two components required for E are also satisfied with four blocks with wells 1 and 2 in one block, wells 3 and 4 in a second block, wells 5 and 6 in a third block, and wells 7 and 8 in a fourth block. For this response to part (a), component 1 is <u>not</u> satisfied with four blocks consisting of wells 1 and 5, 2 and 6, 3 and 7, 4 and 8.
- If the response to part (a) uses a scenario in which arsenic levels decrease as one moves from the left toward the right side of the field, for example, then components 1 and 2 are satisfied with four blocks consisting of wells 1 and 5, 2 and 6, 3 and 7, 4 and 8. For this response to part (a), however, component 1 is <u>not</u> satisfied with four blocks consisting of wells 1 and 2, 3 and 4, 5 and 6, 7 and 8.
- If no conditions are identified in the response to part (a) and no conditions are given in the response to part (b), use the granite ledge and burned forest sides of the field as the conditions for creating blocks in scoring the response to part (b).
- If additional or alternative conditions are identified in part (b), use those conditions in scoring the response to part (b).

AP® STATISTICS
2019 SCORING GUIDELINES

Question 4 (continued)

Part (c) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. An indication that types of filters are randomly assigned to wells.
2. Within each block, each type of filter is assigned to half of the wells.
3. Describes how to correctly implement the random assignment process.

Partially correct (P) if the response satisfies only two of the three components.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:

- Sufficient detail must be provided to enable a knowledgeable statistics user to implement the randomization method. Some additional acceptable methods are:
  - Using a random number generator to select half of the wells in a block for which the newer filters will be used.
  - For blocks consisting of just two wells, tossing a coin.
- Each well must have a 50% chance of using a newer filter and a 50% chance of using an older filter, and half of the wells within the block must be assigned to each filter.
- Responses to part (c) for designs that do not have an even number of wells in each block are scored no higher than P.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 4 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

AP® STATISTICS
2019 SCORING GUIDELINES

Question 5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) recognize a situation in which a binomial probability distribution may be applied; (2) discuss conditions for accurately using a normal approximation to a binomial probability; (3) compute appropriate binomial probabilities.

Solution

Part (a):

The probability that Sasha will experience at least three days with traffic delays in the next 21 days that she travels to work is

$$\begin{aligned} P(\text{Number of delays} \geq 3) &= 1 - [P(\text{no delays}) + P(\text{exactly one delay}) + P(\text{exactly 2 delays})] \\ &= 1 - \left[\binom{21}{0}(0.2)^0(0.8)^{21} + \binom{21}{1}(0.2)^1(0.8)^{20} + \binom{21}{2}(0.2)^2(0.8)^{19} \right] \\ &\approx 1 - [0.00922 + 0.04842 + 0.12106] \\ &\approx 0.82130. \end{aligned}$$

Part (b):

The probability that Sasha's first traffic delay will occur **after** the fifth day is the probability that no traffic delay occurs on any of the first five days:

$$P(\text{No delay during first 5 days}) = (0.8)^5 = 0.32768.$$

Part (c):

No. The sampling distribution of the sample proportion is right skewed. It may not be well approximated by a normal distribution because the expected number of traffic delays in 21 days, $np = (21)(0.20) = 4.2$, is less than 10.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 5 (continued)

Scoring

Parts (a), (b) and (c) are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. Clearly indicates a binomial distribution with $n = 21$ and $p = 0.2$.
2. Indicates the correct boundary value and direction of the event.
3. Reports the correct probability.

Partially correct (P) if the response satisfies component 1 and it does not satisfy one or both of the other two components

OR

the response does not satisfy component 1 and both of the other two components are satisfied.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:

- The response B(21, 0.2) satisfies component 1.
- Component 1 and 2 are satisfied by displaying the correct formula for computing the binomial probability using the correct values for n and p, e.g.
  $$1 - \left[\binom{21}{0}(0.2)^0(0.8)^{21} + \binom{21}{1}(0.2)^1(0.8)^{20} + \binom{21}{2}(0.2)^2(0.8)^{19}\right]$$
  Only component 1 is satisfied if the correct binomial distribution is used in an incorrect probability formula, e.g.
  $$\binom{21}{3}(0.2)^3(0.8)^{18}.$$
- For component 2, the boundary value and direction may be described in words, e.g. *P*(at least three delays in the next 21 days).
- Component 2 may be satisfied by displaying a bar graph of a binomial distribution with the appropriate bars shaded.
- The response of $1 - binomcdf\,(n = 21, p = 0.2, \text{upper bound} = 2) \approx 0.8213$ is scored E since *n, p* and the boundary value are clearly identified.
  The response of $1 - binomcdf\,(n = 21, p = 0.2, 2) \approx 0.8213$ is scored P since *n, p* are clearly identified and the boundary value is not identified.
  The response of $1 - binomcdf\,(21, 0.2, 2) \approx 0.8213$ is scored I.

Question 5 (continued)

- A normal approximation to the binomial is not appropriate since $np = 21 \times 0.2 = 4.2 < 5$.

  A response using the normal approximation can score at most P. To score P, the response must include all of the following:
  - An indication that the probability calculated is a normal approximation for the binomial probability
  - A correct mean and standard deviation based on the binomial parameters
  - Clear indication of boundary and direction with a z-score or diagram
  - The probability computed correctly

  An example of a response, which meets these four criteria, is $P\left(Z \geq \frac{3 - np}{\sqrt{np(1-p)}}\right) = P\left(Z \geq \frac{3 - (21)(0.2)}{\sqrt{(21)(0.2)(0.8)}}\right) \approx 0.744$ and the binomial distribution is mentioned.

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:
1. The correct probability
2. An appropriate justification.

Partially correct (P) if the response satisfies only one of the two components
OR
if the response correctly computes the probability that the first traffic delay is on day 6, $(0.8)(0.8)(0.8)(0.8)(0.8)(0.2) = 0.065536,$ with supporting work.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:
- Component 2 is satisfied if the response uses a binomial distribution with $n = 5$ and $p = 0.2$ to compute the probability of no successes in $n = 5$ trials; *or u*ses direct probability rules to obtain $(0.8)(0.8)(0.8)(0.8)(0.8) = 0.32768.$
- Responses that find the probability of the first delay occurring on day 6 or more, with an incorrect upper bound for the sum, should be scored P.

  For example, $0.8^5(0.2) + 0.8^6(0.2) + ... + 0.8^{20}(0.2).$
- Responses that find the probability of traffic delays on each of the first five days, $(0.2)(0.2)(0.2)(0.2)(0.2) = 0.00032,$ should be scored P.
- Responses using a normal approximation to the binomial distribution should be scored I.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 5 (continued)

Part (c) is scored as follows:

Essentially correct (E) if the response contains the following three components:

1. Calculates $np = (21)(0.2) = 4.2$
2. Indicates that np is less than 10 (or less than 5).
3. Reaches the conclusion that the condition is not met.

Partially correct (P) if the response satisfies only two of the three components;
OR
if the response satisfies component 3 and indicates that the sample size is too small with respect to a commonly used standard (for example, $n < 30$), but does not establish a link to the expected count condition.

Incorrect (I) if the response does not meet the criteria for E or P.

Note: A response that simply indicates that the sample size is too small for the distribution of the sample proportion to be approximately normal is scored I.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 5 (continued)

4 **Complete Response**

Three parts essentially correct

3 **Substantial Response**

Two parts essentially correct and one part partially correct

2 **Developing Response**

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 **Minimal Response**

One part essentially correct

OR

No parts essentially correct and two parts partially correct

AP® STATISTICS
2019 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) analyze results from a completely randomized experiment; (2) identify a situation in which it may be better to compare medians instead of means; (3) develop a test procedure for comparing medians; and (4) interpret results of the test.

Solution

Part (a):

Because the samples are small and the dot plots suggest potential outliers for both keyboards, it is not appropriate to perform a two-sample *t*-test for the difference between the population means.

Part (b):

It is more appropriate to compare population medians because the dot plots indicate that the distributions of data entry times have outliers to the left for both keyboards. When outliers are present, the median provides a better indicator of the typical values than the mean. For each keyboard, the mean of the population of data entry times will be pulled away from the typical values toward the extreme values in the left tail of the distribution.

Part (c):

(i) The completed table rankings for the combined set of 11 data entry times is shown below

Rank	1	2	3	4	5	6	7	8	9	10	11
Keyboard	J	K	J	J	J	J	K	K	K	K	K
Time	158	184	240	248	251	261	267	279	280	284	305

(ii) Sum of ranks for keyboard J: $SR_J = 1 + 3 + 4 + 5 + 6 = 19$

Sum of ranks for keyboard K: $SR_K = 2 + 7 + 8 + 9 + 10 + 11 = 47$

Part (d):

$$W = SR_{\text{J}} - \frac{n_{\text{J}}(n_{\text{J}} + 1)}{2} = 19 - \frac{(5)(5 + 1)}{2} = 4$$

AP® STATISTICS
2019 SCORING GUIDELINES

Question 6 (continued)

Part (e):

(i) To obtain $W = 0$, we must have $0 = SR_J - \frac{n_J(n_J + 1)}{2}$ which implies that $SR_J = \frac{n_J(n_J + 1)}{2} = \frac{(5)(5 + 1)}{2} = 15$.

(ii) All five ranks for keyboard J must be smaller than any of the ranks for keyboard K.

Rank	1	2	3	4	5	6	7	8	9	10	11
Keyboard	J	J	J	J	J	K	K	K	K	K	K

Part (f):

For the one-sided alternative that the median of the distribution of data entry times for keyboard J is smaller than the median of the distribution of data entry times for keyboard K, the *p*-value is the proportion of possible arrangements with W values of 4 or less. From the graph, the *p*-value is $\frac{12}{462} = 0.026$. Because the *p*-value is smaller than the $\alpha = 0.05$ significance level, there is sufficient evidence to reject the null hypothesis in favor of the alternative that the median of the distribution of data entry times for keyboard J is smaller than the median of the distribution of data entry times for keyboard K.

Scoring

This question is scored in four sections. Section 1 consists of parts (a) and (b), section 2 consists of parts (c) and (d), section 3 consists of part (e), and section 4 consists of part (f). Sections 1, 2, 3 and 4 are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response includes the following two components:

1. The response to part (a) correctly argues that a two-sample *t*-test is not appropriate because the sample sizes are small and the dot plots show potential outliers (or left skewed distributions).
2. The response to part (b) uses the outliers (skewness) in the entry time distributions to argue that the medians are better indicators of typical entry times than the means.

Partially correct (P) if the response includes only one of the two components.

Incorrect (I) if the response does not meet the criteria for E or P.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 6 (continued)

Section 2 is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. The table of rankings in part (c-i) is completed correctly.
2. The sum of ranks in part (c-ii) is computed correctly for both keyboards, using the ranks reported in the table in part (c-i).
3. The value of W is calculated correctly in part (d), based on the value of SR_J in part (c-ii) and the correct value of n_J.

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response does not meet the criteria for E or P.

Section 3 is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. The response to part (e-i) correctly finds the value of SR_J when $W = 0$, using the value of n_J from the response to part (d).
2. The correct assignment of ranks is displayed in the response to part (e-ii) to give $W = 0$ or the value of SR_J shown in part (e-i).

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) if the response does not meet the criteria for E or P.

Section 4 is scored as follows:

Essentially correct if the response in part (f) includes the following four components:
1. The critical region is identified as the possible values of W that are less than or equal to the value of W reported in part (d). This can be done by marking on the graph shown in the stem prior to part (e).
2. A correct *p*-value is reported for the indicated critical region.
3. A correct conclusion is reached based on a reasonable reported *p*-value.
4. The conclusion is presented in the context of comparing the medians of the data entry time distributions for keyboards J and K.

Partially correct if the response satisfies only two or three of the four components.

Incorrect if the response does not meet the criteria for E or P.

AP® STATISTICS
2019 SCORING GUIDELINES

Question 6 (continued)

Each essentially correct (E) section counts as 1 point, and each partially correct (P) section counts as ½ point.

4 **Complete Response**

3 **Substantial Response**

2 **Developing Response**

1 **Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.

2019 AP Statistics Scoring Worksheet

Section I: Multiple Choice

_____	×	1.2500	=	_____
Number Correct (out of 40)				Weighted Section I Score (Do not round)

Section II: Free Response

Question 1	_____ (out of 4)	×	1.8750	=	_____ (Do not round)
Question 2	_____ (out of 4)	×	1.8750	=	_____ (Do not round)
Question 3	_____ (out of 4)	×	1.8750	=	_____ (Do not round)
Question 4	_____ (out of 4)	×	1.8750	=	_____ (Do not round)
Question 5	_____ (out of 4)	×	1.8750	=	_____ (Do not round)
Question 6	_____ (out of 4)	×	3.1250	=	_____ (Do not round)

Sum = _____
Weighted Section II Score (Do not round)

Composite Score

_____	+	_____	=	_____
Weighted Section I Score		Weighted Section II Score		Composite Score (Round to nearest whole number)

AP Score Conversion Chart
Statistics

Composite Score Range	AP Score
73-100	5
59-72	4
44-58	3
32-43	2
0-31	1

2019 AP Statistics
Question Descriptors and Performance Data

Multiple-Choice Questions

Question	Skill	Learning Objective	Topic	Key	% Correct
1	2.A	UNC-1.H	Describing the Distribution of a Quantitative Variable	C	61
2	3.A	VAR-4.D	Conditional Probability	E	72
3	2.A	DAT-1.F	Residuals	C	63
4	2.A	UNC-1.H	Describing the Distribution of a Quantitative Variable	A	79
5	2.C	UNC-1.Q	Statistics for Two Categorical Variables	A	78
6	3.A	VAR-2.B	The Normal Distribution	C	65
7	2.D	UNC-1.N	Comparing Distributions of a Quantitative Variable	C	92
8	1.C	DAT-2.C	Random Sampling and Data Collection	E	73
9	3.A	VAR-2.B	The Normal Distribution	B	62
10	3.A	VAR-6.B	The Normal Distribution, Revisited	C	21
11	2.A	UNC-1.H	Describing the Distribution of a Quantitative Variable	B	60
12	1.C	DAT-2.C	Random Sampling and Data Collection	D	76
13	4.B	UNC-3.Q	Sampling Distributions for Sample Means	B	58
14	4.A	UNC-4.H	Justifying a Claim Based on a Confidence Interval for a Population Proportion	D	35
15	4.E	DAT-3.B	Concluding a Test for a Population Proportion	B	62
16	1.C	VAR-3.A	Introduction to Experimental Design	D	76
17	3.B	VAR-5.E	Combining Random Variables	D	33
18	3.C	UNC-3.L|UNC-3.K	Sampling Distributions for Sample Proportions	C	68
19	4.B	VAR-3.E	Inference and Experiments	C	76
20	2.D	UNC-1.P	Representing Two Categorical Variables	E	86
21	2.A	DAT-1.G	Least Squares Regression	C	42
22	3.C	UNC-3.R|UNC-3.Q	Sampling Distributions for Sample Means	B	74
23	1.E	VAR-7.B	Setting Up a Test for a Population Mean	C	31
24	2.C	UNC-1.J	Summary Statistics for a Quantitative Variable	B	69
25	4.E	DAT-3.F	Carrying Out a Test for a Population Mean	A	56
26	3.B	VAR-5.C	Mean and Standard Deviation of Random Variables	C	79
27	1.B	VAR-3.B	Introduction to Experimental Design	B	77
28	3.A	UNC-3.E	The Geometric Distribution	D	45
29	1.B	UNC-5.A	Potential Errors When Performing Tests	B	67
30	4.B	UNC-4.AA	Justifying a Claim About the Difference of Two Means Based on a Confidence Interval	B	52
31	3.B	UNC-3.K	Sampling Distributions for Sample Proportions	C	41

2019 AP Statistics
Question Descriptors and Performance Data

Question	Skill	Learning Objective	Topic	Key	% Correct
32	3.D	UNC-4.C	Constructing a Confidence Interval for a Population Proportion	E	74
33	4.B	UNC-4.AF	Confidence Intervals for the Slope of a Regression Model	B	32
34	4.E	DAT-3.D	Carrying Out a Test for the Difference of Two Population Proportions	E	43
35	4.C	VAR-7.L	Setting Up a Test for the Slope of a Regression Model	E	25
36	4.B	DAT-3.A	Interpreting P-Values	E	55
37	3.B	VAR-5.E	Combining Random Variables	D	28
38	3.D	UNC-4.K	Confidence Intervals for the Difference of Two Proportions	A	77
39	3.E	VAR-8.L	Carrying Out a Chi-Square Test for Homogeneity or Independence	C	40
40	4.B	UNC-4.S	Justifying a Claim About a Population Mean Based on a Confidence Interval	E	66

Free-Response Questions

Question	Skill	Learning Objective	Topic	Mean Score
1	2.A|4.B	UNC-1.K|UNC-1.H|UNC-1.M	1.7|1.6|1.8	2.25
2	1.D|3.D|4.B|4.D	UNC-4.AC|UNC-4.AF|UNC-4.AG|UNC-4.AH	9.2|9.3	1.2
3	2.A|2.D	UNC-1.N|UNC-1.M	1.9|1.8	2.19
4	1.B|1.C	VAR-3.D|VAR-3.B	3.6|3.5	1.47
5	3.A|3.C	UNC-3.B|VAR-4.E|UNC-3.L	4.10|4.6|5.5	0.97
6	2.A|2.B|4.B|4.C	VAR-7.H|UNC-1.M	7.8|1.8	1.93