

AP[®] Statistics Exam

SECTION I: Multiple Choice

2021

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) ● (C) (D) (E)
(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 begin on page 3.
Questions begin on page 6.
Tables begin on page 42.

Formulas

(I) Descriptive Statistics

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

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

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

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

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

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

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

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

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

(II) Probability

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

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

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

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

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:

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

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

Formulas

(I) Descriptive Statistics

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

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

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

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

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

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

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

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

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

(II) Probability

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

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

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

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

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: $\frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: $\text{statistic} \pm (\text{critical value}) \cdot (\text{standard deviation of statistic})$

Single-Sample

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

Two-Sample

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

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

(III) Inferential Statistics

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

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

Single-Sample

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

Two-Sample

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

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

Statistics 2021Intl MCQ

1.

0	5
1	45668
2	3334
3	2449
4	2456
5	0 4 49
6	456
7	47

Key: 7|4 = 74%

In 2011, the Gallup polling organization selected random samples of residents from each of the 27 countries in the European Union and found the percentage of residents in each country who consider themselves to be thriving. The 27 percentages are shown in the stemplot.

What is the interquartile range for this data set?

- (A) 27
- (B) 31
- (C) 39
- (D) 41
- (E) 72

2. The final scores for 309 participants in a national golf tournament are summarized in the following table.

N	Mean	SE Mean	St Dev	Min	Q1	Median	Q3	Max
309	149.35	0.37	6.42	131	145	150	153.5	168

Which of the following is closest to the z-score of the minimum golf score?

- (A) -2.16
- (B) -2.86
- (C) -2.96
- (D) -18.40
- (E) -278.10

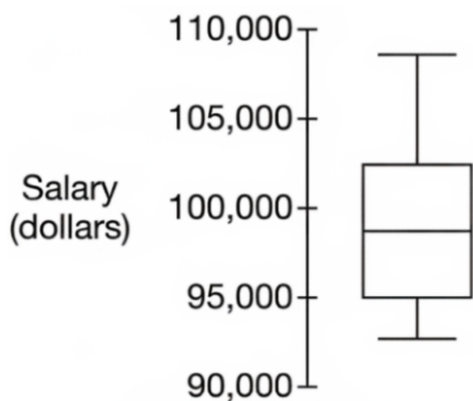
Statistics 2021 Intl MCQ

3. A random sample of middle school students from District 1 and a random sample of middle school students from District 2 were selected with an even number of students in each sample. The number of minutes each student spent playing a certain mobile game application in a typical day was recorded. The data are summarized in the following table.

	Min	Q1	Med	Q3	Max
District 1	112	120	130	137	145
District 2	112	128	130	145	152

If no two students within each district spent the same amount of time playing the mobile game application, which statement must be true?

- (A) The number of students in the sample from District 1 was less than the number of students in the sample from District 2.
- (B) The percentage of students in both samples who spent less than 130 minutes playing the game is equal.
- (C) No students in either district spent 115 minutes playing the game.
- (D) The range of the number of minutes spent playing the game was greater for the District 1 sample than for the District 2 sample.
- (E) The interquartile range of the number of minutes spent playing the game was less for the District 1 sample than for the District 2 sample.
4. The boxplot shown summarizes the 100 highest salaries for municipal employees in a large city.

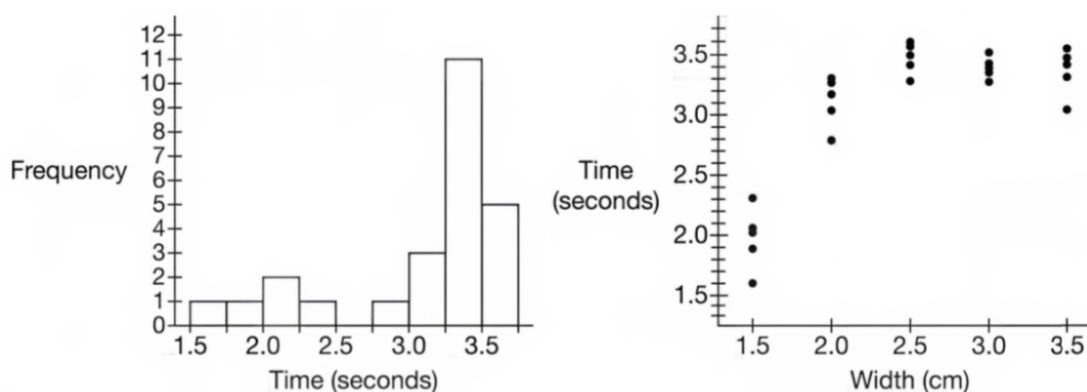


A new municipal employee was hired and was told that the salary would be within the middle 50 percent of the current 100 highest salaries.

Based on the boxplot, which of the following is closest to the width of the interval in which the new employee's salary will fall?

Statistics 2021 Intl MCQ

- (A) \$1,000
(B) \$3,000
(C) \$7,000
(D) \$10,000
(E) \$15,000
5. Students in a physics class designed an experiment to investigate the effect of blade width on flight time for paper helicopters. A histogram of the flight times, in seconds, and a scatterplot of flight time versus blade width, in centimeters (cm), are shown for 25 helicopters.



Which of the following best describes the clusters and the gap in the histogram?

- (A) The left cluster displays flight times for blade widths from 1.5 cm to 2.5 cm, and the right cluster displays flight times for the remaining blade widths.
- (B) The left cluster displays flight times for blade widths of 1.5 cm, and the right cluster displays flight times for the remaining blade widths.
- (C) The left cluster displays minimum flight times for each of the blade widths, and the right cluster displays the remaining flight times.
- (D) The clusters are not related to the blade widths because flight time is measured in seconds and blade widths are measured in centimeters.
- (E) The clusters are not related to the blade widths because the relationship between blade width and flight time is not linear.

Statistics 2021 Intl MCQ

6. A company conducted a charity drive and recorded the donations made by the employees of two departments, A and B. The back-to-back stemplot below shows the amounts given by the employees from department A and the amounts given by the employees from department B.

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

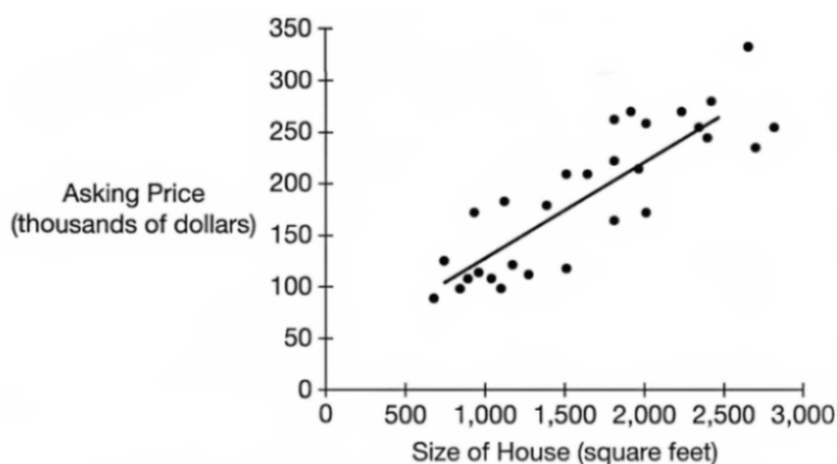
Key: 7|5 represents \$75

Based on the back-to-back stemplot, which of the following is correct?

- (A) The variability of the donations for the employees from department A is less than the variability of the donations for the employees from department B.
- (B) The median donation for the employees from department A is greater than the median donation for the employees from department B.
- (C) The shape of the distribution of donations for each department is skewed left.
- (D) The number of employees from department A who donated at most \$35 is greater than the number of employees from department B who donated at most \$35.
- (E) The number of employees from department A who donated \$50 is greater than the number of employees from department B who donated \$50.

Statistics 2021Intl MCQ

7. A realtor collected data from a random sample of 29 houses for sale in a certain town. The scatterplot shows the size of each house, in square feet, and the asking price, in thousands of dollars. The least-squares regression line is also shown.



Based on the recorded data, the realtor created the following regression equation to predict asking price, in thousands of dollars, based on the size, in square feet, of the house.

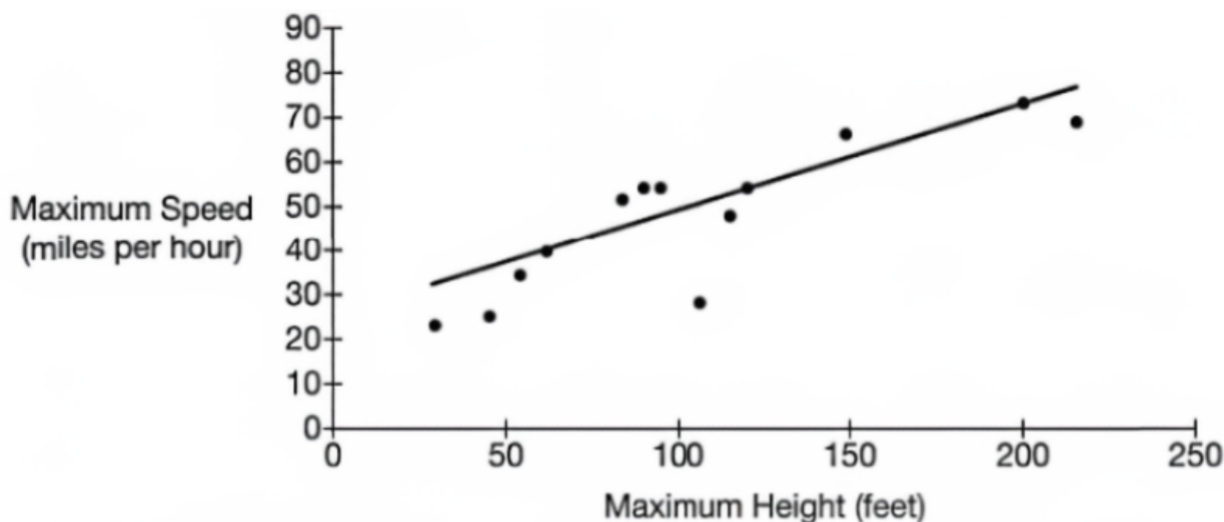
$$\text{predicted asking price} = 35.050 + 0.096(\text{size of house})$$

Which of the following provides the best interpretation of the slope of the regression line?

- (A) For each additional square foot, the predicted asking price of a house in this town increases by \$96.
- (B) For each additional square foot, the predicted asking price of a house in this town increases by \$35.
- (C) For each additional square foot, the predicted asking price of a house in this town increases by \$0.096.
- (D) For each additional \$1,000 of asking price, the predicted size of a house increases by 35 square feet.
- (E) For each additional \$1,000 of asking price, the predicted size of a house increases by 0.096 square feet.

Statistics 2021 Intl MCQ

8.



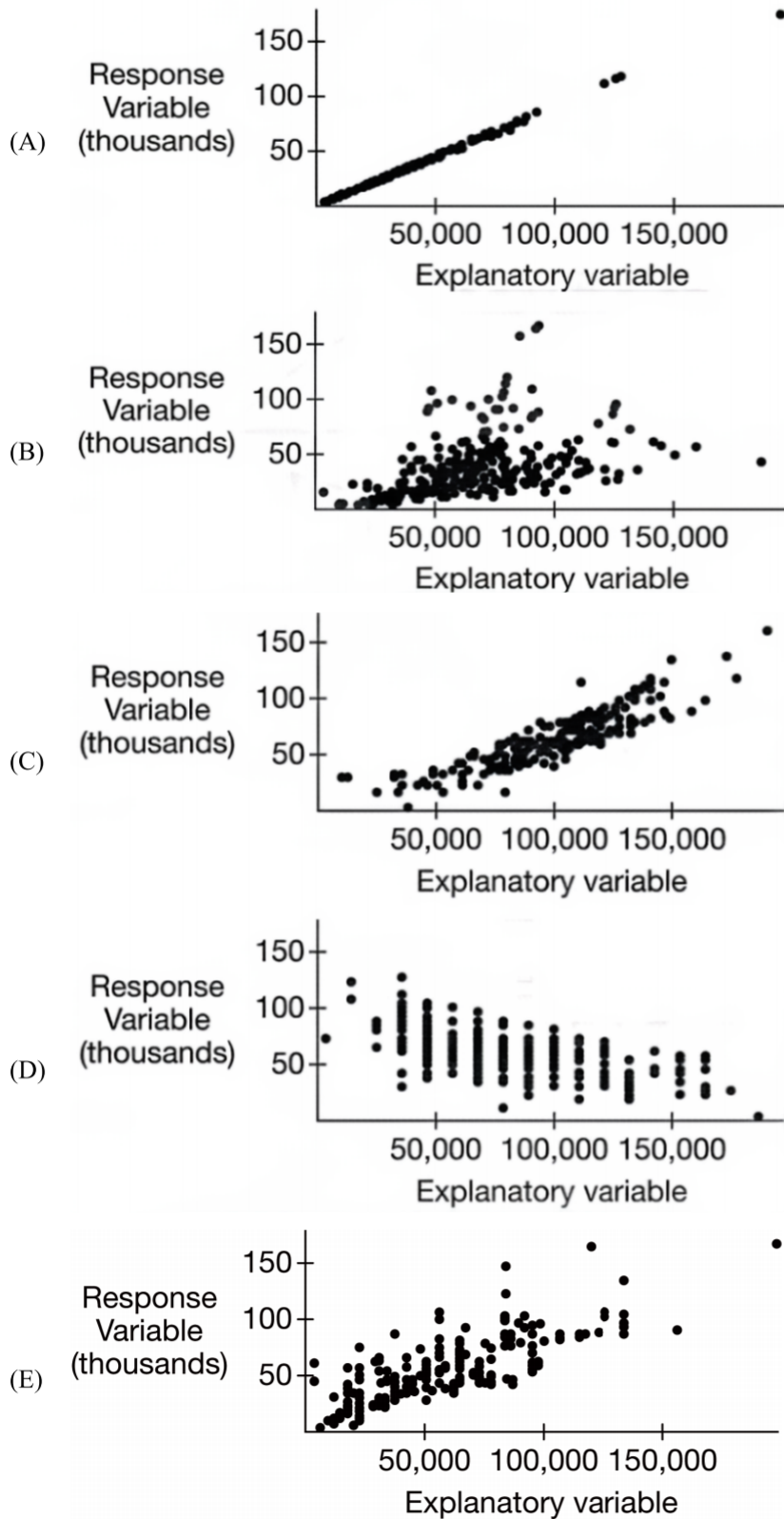
An engineering student collected data on the maximum height, in feet, and maximum speed, in miles per hour, of thirteen roller coasters. A scatterplot of the data and the least-squares regression line are shown. The maximum heights of five of the roller coasters are 60, 105, 150, 200, and 215 feet.

If the least-squares regression line is used to predict the maximum speed for the five roller coasters, for which maximum height, in feet, would the absolute value of the residual be largest?

- (A) 60
 - (B) 105
 - (C) 150
 - (D) 200
 - (E) 215
9. One condition for inference for slope of a regression line is that the standard deviation of the response variable does not vary with the explanatory variable.

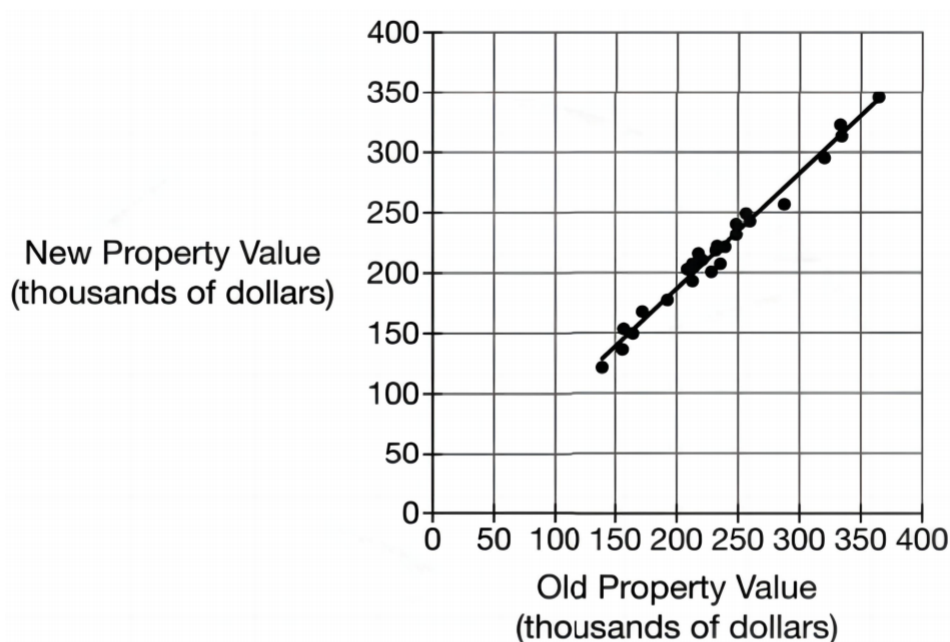
Which of the following scatterplots most likely exhibits a violation of that condition?

Statistics 2021Intl MCQ



Statistics 2021 Intl MCQ

10. A certain state law requires residential properties to receive an appraisal for tax purposes every six years. A random sample of 25 appraised properties was selected. The following scatterplot shows the value of the properties, in thousands of dollars, before and after the appraisal. Also shown is the least-squares regression line and corresponding computer output.



Term	Coef	SE Coef	T-Value	P-Value
Constant	-3.817	1.23	-0.60	0.5562
Old Property Value	0.95597	0.02634	36.30	< 0.0001
S = 7.53867		R-Sq = 98.28%		R-Sq(adj) = 98.21%

Which of the following is not an appropriate description of the data in the sample?

- (A) New property values tend to be greater than old property values.
- (B) As old property values increase, new property values tend to increase.
- (C) The relationship between old and new property value is strong and positive.
- (D) For each additional \$1,000 in old property value, the new property value increases by \$955.97, on average.
- (E) The observed new property values typically deviate from the predicted new property values by about \$7,539.

Statistics 2021 Intl MCQ

11. A medical researcher conducted a study to investigate the effectiveness of two drugs, L and M, designed to reduce the number of seizures that people with epilepsy might have. Each person in a group of 120 adult volunteers who had epilepsy was randomly assigned to one of two groups. Those assigned to one group took drug L for one year, and those assigned to the other group took drug M for one year. Each participant recorded the number of seizures that occurred during the year.

To which group is it most reasonable to generalize the results of this study?

- (A) All adults
 - (B) All adults with epilepsy
 - (C) All adults with epilepsy who take drugs to treat the disease
 - (D) All adults with epilepsy who are similar to those who volunteered for the study
 - (E) All adults who experience a reduction in the number of seizures
12. Researchers conducted a study to investigate whether watching an outdoor scene while exercising improves health more than exercise alone. Fourteen fifth-grade students participated in two trials of the study. For one trial, the students rode a stationary bicycle for 15 minutes while watching a video of someone cycling through a forest. For the other trial, the students rode a stationary bicycle for 15 minutes without watching a video. Each student completed both trials in a randomized order. The researchers recorded the students' heart rates after each trial.

Which of the following is the best description of the study?

- (A) A randomized matched-pairs experiment
 - (B) A randomized placebo-controlled experiment
 - (C) A completely randomized experiment
 - (D) A sample survey
 - (E) An observational study
13. A new laser-guided surgical procedure for removing an inflamed appendix is believed to have a shorter recovery time for patients than an alternative surgical procedure that is not guided by laser has. In an experiment to evaluate recovery time, patients needing the surgery will be randomly assigned to one of two surgical procedures. A surgeon will perform the surgery for one group using lasers and for the other group without the use of lasers. Both types of surgeries are performed with similarly sized incisions and patients are unaware of which procedure is performed.

What type of blinding can be accomplished in the experiment?

- (A) Both the patient and the surgeon can be blind to the treatment.
- (B) Neither the patient nor the surgeon can be blind to the treatment.
- (C) The patient can be blind to the treatment, but not the surgeon.
- (D) The surgeon can be blind to the treatment, but not the patient.
- (E) Blinding is not possible because the design has no control group.

Statistics 2021 Intl MCQ

14. A medical researcher is interested in assessing the effectiveness of a new medication for pain relief. Three hundred volunteers will be randomly assigned to groups. One group will receive the new medication at a low dose, one group will receive the new medication at a high dose, and one group will receive a placebo that looks like the new medication but does not contain any active ingredients. Each person will be asked to rate his or her level of pain at the beginning of the experiment and again two hours after taking the medication or placebo.

What is the number of treatments for this experiment?

- (A) 2
 - (B) 3
 - (C) 6
 - (D) 100
 - (E) 300
15. The height requirement for a certain ride at an amusement park is 48 inches or taller. The distribution of height of six-year-old children is approximately normal with mean 45.2 inches and standard deviation 1.75 inches.

Which of the following is closest to the probability that a randomly selected six-year-old child will not meet the height requirement for the ride?

- (A) 0.0401
 - (B) 0.0548
 - (C) 0.9452
 - (D) 0.9599
 - (E) 0.9974
16. A manufacturer of a certain item knows that 1 percent of the items produced on a certain assembly line are defective. Suppose that 100 items produced on that assembly line will be selected at random.

Which of the following expressions represents the probability that 3 of the 100 items selected will be defective?

- (A) $\binom{100}{3}(0.01)^3(0.99)^{97}$
- (B) $\binom{100}{97}(0.01)^{97}(0.99)^3$
- (C) $\binom{97}{3}(0.01)^3(0.99)^{97}$
- (D) $\binom{97}{3}(0.01)^{97}(0.99)^3$
- (E) $\binom{100}{3}(0.01)^3(0.99)^3$

Statistics 2021 Intl MCQ

17. The distribution of sale price for a certain model of car is approximately normal with mean \$22,500 and standard deviation \$600.

Based on the distribution, which statement must be true?

- (A) Approximately 99% of the cars have a sale price less than \$23,700.
- (B) Approximately 84% of the cars have a sale price greater than \$21,900.
- (C) Approximately 95% of the cars have a sale price between \$21,900 and \$23,100.
- (D) The interquartile range of the sale prices is approximately \$600.
- (E) The minimum sale price is \$20,700.

18.

Number of cookies sold	50	100	200
Probability	0.60	0.35	0.05

The probability distribution of the number of cookies sold by a store each day is shown in the table. The store earns a profit of \$0.50 per cookie sold.

What is the store's expected daily profit for the sales of the cookies?

- (A) \$37.50
 - (B) \$75.00
 - (C) \$116.67
 - (D) \$150.00
 - (E) \$175.00
19. An elevator has a posted weight limit of 4,200 pounds and an occupancy limit of 21 people. If the weights of adult males are approximately normally distributed with $\mu = 190$ pounds and $\sigma = 32$ pounds, which of the following best approximates the probability that a random sample of 21 adult males will exceed the posted weight limit of the elevator?
- (A) 0.001
 - (B) 0.076
 - (C) 0.155
 - (D) 0.378
 - (E) 0.500

Statistics 2021 Intl MCQ

20. Clark drives through two traffic lights on his way to work each day. From experience, he estimates that the probability that the first light will be red when he reaches it is 0.3. He also estimates that the probability that the second light will be red when he reaches it is 0.45. The lights are far enough apart that they can be considered to operate independently of each other.

Based on Clark's estimates, what is the probability that at least one of the two lights will be red on a randomly selected drive to work?

- (A) 0.135
 - (B) 0.165
 - (C) 0.385
 - (D) 0.615
 - (E) 0.750
21. The driving route between a student's home and the student's high school has two traffic lights. The distribution of total time spent driving has mean 5 minutes and standard deviation 1 minute. The distribution of time spent waiting at a light has mean 0.75 minute and standard deviation 0.25 minute. The total time spent driving and the time spent waiting at a light are independent. The total travel time from the student's home to the student's high school consists of the total time spent driving and the time spent waiting at two lights.

What are the mean and the standard deviation of the total travel time?

- (A) The mean is 5.75 minutes and the standard deviation is about 1.03 minutes.
 - (B) The mean is 5.75 minutes and the standard deviation is about 1.06 minutes.
 - (C) The mean is 6.5 minutes and the standard deviation is about 1.06 minutes.
 - (D) The mean is 6.5 minutes and the standard deviation is about 1.12 minutes.
 - (E) The mean is 6.5 minutes and the standard deviation is about 1.5 minutes.
22. A six-week training program was developed to help student athletes improve performance on a certain task. To study the effectiveness of the program, researchers randomly selected 50 student athletes. The students were tested before participating in the program. Then, half the athletes were randomly assigned to participate in the program and the other half were allowed to train on their own. At the end of the six weeks, all athletes were tested again. The researchers reported that the athletes who participated in the program increased their scores, on average, by 200 points, whereas those who did not participate raised their scores, on average, by 80 points.

After reading the results of the study, Jasmine, a student athlete, decided to enroll in the program. At the end of six weeks, her score had increased by 80 points. Which of the following is an explanation for the difference between Jasmine's score and the scores that the study reported for student athletes who participated in the program?

- (A) The results of the study were false; there must have been a bias in the sample that was not reported.
- (B) Jasmine probably did not follow the program correctly; if she had, she would have seen a 200-point increase in her score.
- (C) The increase of 200 points reported was an average; some athletes could have increased their scores by more than 200 points, and some by less than 200 points.
- (D) The sample sizes used in the study were too small for any conclusions about the training to be valid.
- (E) The study was not a well-designed experiment because causation cannot be attributed to the program.

Statistics 2021 Intl MCQ

23. Which of the following statements illustrates the Law of Large Numbers?

- I. For n flips of a fair coin with one side heads, the proportion of flips that land heads up approaches 0.5 as n increases.
- II. For n tosses of a fair six-sided die with one face labeled 1, the proportion of tosses that land with the face of 1 up approaches $\frac{1}{6}$ as n increases.
- III. If each of 10 consecutive tosses of a fair coin resulted in a head, the probability that the next toss of the coin will result in a tail is greater than 0.5.

- (A) I only
- (B) II only
- (C) III only
- (D) I and II only
- (E) I, II, and III

24. A 95 percent confidence interval for the mean weight, in thousands of pounds, of African bush elephants is determined to be (12.2, 14.6).

Which of the following is a correct interpretation of the interval?

- (A) We are 95% confident that 95% of African bush elephants weigh between 12.2 thousand pounds and 14.6 thousand pounds.
- (B) The probability is 0.95 that the population mean weight of African bush elephants is between 12.2 thousand pounds and 14.6 thousand pounds.
- (C) We are 95% confident that the population mean weight of African bush elephants is between 12.2 thousand pounds and 14.6 thousand pounds.
- (D) We are 95% confident that the population mean weight of African bush elephants is 13.4 thousand pounds, which is the midpoint of the interval.
- (E) We are 95% confident that the mean weight calculated from a different sample of African bush elephants would be between 12.2 thousand pounds and 14.6 thousand pounds.

25. Two brands of golf cart batteries, R and S, were tested to investigate whether there is a difference between the proportions of fully charged batteries that could power a golf cart for at least thirty miles. A random sample of 50 brand R batteries was selected and a random sample of 50 brand S batteries was selected. Batteries were randomly assigned to one of 100 different golf carts. Each cart was driven until the battery failed. Of the carts with brand R batteries, 40 traveled over thirty miles, and of the carts with brand S batteries, 35 traveled over thirty miles.

Which of the following is a 95 percent confidence interval for the difference in the proportions of fully charged batteries that can power a golf cart for at least thirty miles for the two brands?

Statistics 2021 Intl MCQ

- (A) $(0.8 - 0.7) \pm 1.645 \sqrt{\frac{(0.8)(0.2)}{50} + \frac{(0.7)(0.3)}{50}}$
- (B) $(0.8 - 0.7) \pm 1.645 \sqrt{(0.75)(0.25) \left(\frac{1}{50} + \frac{1}{50} \right)}$
- (C) $(0.8 - 0.7) \pm 1.645 \sqrt{\frac{(0.8)(0.2)}{50} - \frac{(0.7)(0.3)}{50}}$
- (D) $(0.8 - 0.7) \pm 1.960 \sqrt{\frac{(0.8)(0.2)}{50} + \frac{(0.7)(0.3)}{50}}$
- (E) $(0.8 - 0.7) \pm 1.960 \sqrt{(0.75)(0.25) \left(\frac{1}{50} + \frac{1}{50} \right)}$

26. An agriculturalist wants to estimate the average time from germination to ripe fruit for a new type of seedless strawberry. From a random sample of 25 seedless strawberries of the new type, the mean time from germination to ripe fruit is 42.1 days with standard deviation 3.6 days.

Assume all conditions for inference were met. Which of the following is a 90 percent confidence interval for the mean time from germination to ripe fruit for the new type of seedless strawberry?

- (A) 42 ± 6.16
- (B) 42.1 ± 1.49
- (C) 42.1 ± 1.23
- (D) 42.1 ± 1.18
- (E) 42.1 ± 0.25
27. Researchers selected a random sample of 100 students from a certain college to investigate whether students from the college who are sleep deprived will have lower reading-comprehension scores, on average, than students from the college who are not sleep deprived. Half the students from the sample were randomly assigned to a treatment group that received sleep restrictions, and the other half were assigned to a control group that received no sleep restrictions. The students were given a reading-comprehension test before the study began and at the end of two days. The mean score of the test for the treatment group was similar to that of the control group before the study. A 95 percent confidence interval constructed from the difference in mean scores (treatment minus control) at the end of two days was $(-7.3, -1.5)$.

Based on the confidence interval, which of the following is an appropriate conclusion?

- (A) The data provide no evidence that the population of sleep-deprived students from the college has a lower mean reading-comprehension score than the population of students from the college who are not sleep deprived.
- (B) The data provide convincing statistical evidence that the population of sleep-deprived students from the college has a lower mean reading-comprehension score than the population of students from the college who are not sleep deprived.
- (C) No conclusion can be made because a confidence interval cannot be used to assess statistical evidence.
- (D) The probability that a student from the college who is sleep deprived will have a lower reading-comprehension score than a student from the college who is not sleep deprived is 0.95.
- (E) More than half the students from the college do worse on a reading-comprehension test after being sleep deprived.

Statistics 2021 Intl MCQ

28. An engineer wanted to investigate the relationship between total stopping distance required (reaction distance and actual breaking distance) and speed of a car. The engineer completed an experiment measuring the total stopping distance of the car at 5 different speeds, and the resulting regression output is shown below.

Predictor	Coef	SECoef	T	P
Constant	-284.29	63.032	-4.51	0.0204
Speed	11.905	1.220	9.76	0.0023
S = 35.56		R-sq = 96.9%		R-sq (adj) = 95.9%

Assume the conditions for inference are met. Which of the following represents a 95 percent confidence interval for the slope of the regression line?

- (A) $11.905 \pm 3.182(1.220)$
- (B) $11.905 \pm 3.182\left(\frac{1.220}{\sqrt{5}}\right)$
- (C) $11.905 \pm 3.182\left(\frac{1.220}{\sqrt{4}}\right)$
- (D) $11.905 \pm 2.776(1.220)$
- (E) $11.905 \pm 2.776\left(\frac{1.220}{\sqrt{5}}\right)$
29. To estimate the proportion of Europeans who walk for exercise at least 10 minutes a week, data were collected from a random sample of 27,919 people from Europe. The 95 percent confidence interval created from the data was (0.856, 0.864).

Which of the following is an appropriate interpretation of the confidence interval?

- (A) The probability that a randomly selected person from Europe walks for exercise at least 10 minutes a week is 0.95.
- (B) The proportion of Europeans who walk between 0.856 and 0.864 minute a week for exercise is 0.95.
- (C) There is a 95% chance that the sample proportion of Europeans who walk for exercise at least 10 minutes a week is between 0.856 and 0.864.
- (D) Between 85.6% and 86.4% of all Europeans walk for exercise at least 10 minutes a week.
- (E) We can be 95% confident that the proportion of all Europeans who walk for exercise at least 10 minutes a week is between 0.856 and 0.864.

Statistics 2021 Intl MCQ

30. A study was conducted to investigate whether there is a difference in the mean life span of a new brand of battery compared with an older brand of battery. Independent random samples of 9 new-brand batteries and 8 older-brand batteries were selected, and the life span (in hours) of each battery was recorded. Summary statistics obtained from the data are shown in the table below. Assume that the conditions for inference are met.

	New Brand	Older Brand
Mean	426.8	355.2
Standard Deviation	49.5	96.7

Which of the following best approximates the p -value for a test of $H_0 : \mu_{\text{new}} - \mu_{\text{older}} = 0$ versus $H_a : \mu_{\text{new}} - \mu_{\text{older}} \neq 0$, where μ_{new} is the population mean life span of the new brand of battery and μ_{older} is the population mean life span of the older brand of battery?

- (A) $P\left(t > \frac{426.8 - 355.2}{\sqrt{\frac{49.5}{9} + \frac{96.7}{8}}}\right)$
- (B) $P\left(t > \frac{426.8 - 355.2}{\sqrt{\frac{49.5^2}{9} - \frac{96.7^2}{8}}}\right)$
- (C) $P\left(t > \frac{426.8 - 355.2}{\sqrt{\frac{49.5^2}{9} + \frac{96.7^2}{8}}}\right)$
- (D) $2 \cdot P\left(t > \frac{426.8 - 355.2}{\sqrt{\frac{49.5}{9} + \frac{96.7}{8}}}\right)$
- (E) $2 \cdot P\left(t > \frac{426.8 - 355.2}{\sqrt{\frac{49.5^2}{9} + \frac{96.7^2}{8}}}\right)$

Statistics 2021 Intl MCQ

31. The campaign manager for a candidate for governor wants to estimate the difference between the percentage of voters who viewed an advertisement and favor the candidate and the percentage of voters who had not viewed the advertisement and favor the candidate. Two independent random samples of 500 voters were selected. One random sample consisted of voters who viewed the advertisement and the other random sample consisted of voters who had not viewed the advertisement. The table shows the frequencies of responses for those surveyed.

	Have Not Seen Advertisement	Have Seen Advertisement
In favor	224	283
Not in favor	276	217

Which of the following is the most appropriate method for the campaign manager to use to estimate the difference between the percentages of voters who favor the candidate?

- (A) A one-sample z -interval for estimating a population proportion
 - (B) A t -interval for estimating the slope of the least-squares regression line
 - (C) A matched-pairs t -interval for estimating a mean difference
 - (D) A two-sample z -interval for estimating a difference between sample proportions
 - (E) A two-sample z -interval for estimating a difference between population proportions
32. A national newspaper reported that 40 percent of high school students engage in community service activities. A school administrator wondered whether the proportion of high school students at the administrator's school who engage in community service activities is less than 40 percent. The administrator selected a random sample of 100 students, and 36 responded that they do engage in community service activities. The conditions for inference were met, and an appropriate hypothesis test resulted in a p -value of 0.207.

Which of the following is a correct interpretation of this p -value?

- (A) The probability that a randomly selected student at this high school engages in community service activities is 0.207.
- (B) The probability is 0.40 that less than 20.7 percent of students at this high school engage in community service activities.
- (C) The probability is 0.207 that less than 40 percent of students at this high school engage in community service activities.
- (D) If the proportion of all students at the school who engage in community service activities is 0.4, the probability of observing 36 or fewer in a random sample of 100 students who engage in community service activities is 0.207.
- (E) If the proportion of all students at the school who engage in community service activities is 0.36, the probability of obtaining 40 or fewer in a random sample of 100 who engage in community service activities is 0.207.

Statistics 2021 Intl MCQ

33. The editor of a local newspaper wants to know whether readers would be willing to pay an additional \$1.00 per month for their monthly newspaper subscription in exchange for an extra weekly section featuring stories about local businesses. The editor plans to survey a random sample of readers, and if the editor finds convincing statistical evidence that more than 75 percent of readers support the change, the editor will implement the plan of starting the extra weekly section.

In the context of this situation, which of the following describes a consequence of a Type I error?

- (A) The editor decides to implement the plan of starting the extra weekly section when the actual percentage of readers who support the change is greater than 75%.
 - (B) The editor decides to implement the plan of starting the extra weekly section when the actual percentage of readers who support the change is not greater than 75%.
 - (C) The editor decides to implement the plan of starting the extra weekly section when the actual percentage of readers who support the change is less than 50%.
 - (D) The editor decides not to implement the plan of starting the extra weekly section when the actual percentage of readers who support the change is greater than 75%.
 - (E) The editor decides not to implement the plan of starting the extra weekly section when the actual percentage of readers who support the change is not greater than 75%.
34. Which statement about the power of a significance test is true?
- (A) A test performed at a significance level of $\alpha = 0.01$ has a higher power than the same test performed at a $\alpha = 0.05$.
 - (B) If the probability of Type I error is decreased, the power of the test is increased.
 - (C) The power of a test is equal to the probability of making a Type II error.
 - (D) The power of a test is independent of the sample size.
 - (E) The power of a test at a particular parameter value increases as the distance from the hypothesized value to the parameter value increases.
35. A researcher in sports medicine is interested in comparing the amount of daily screen time between high school students who are athletes and high school students who are not athletes. A random sample of 100 high school students who are athletes and a random sample of 100 high school students who are not athletes will be selected. Each student will be asked “Do you typically have more than 3 hours of screen time each day?”

Which of the following is the most appropriate test for the researcher to use to investigate whether there is a difference in the amounts of daily screen time?

- (A) A two-sample z -test for a difference between population proportions
- (B) A two-sample t -test for a difference between population proportions
- (C) A two-sample z -test for a difference between population means
- (D) A two-sample t -test for a difference between population means
- (E) A chi-square goodness-of-fit test

Statistics 2021 Intl MCQ

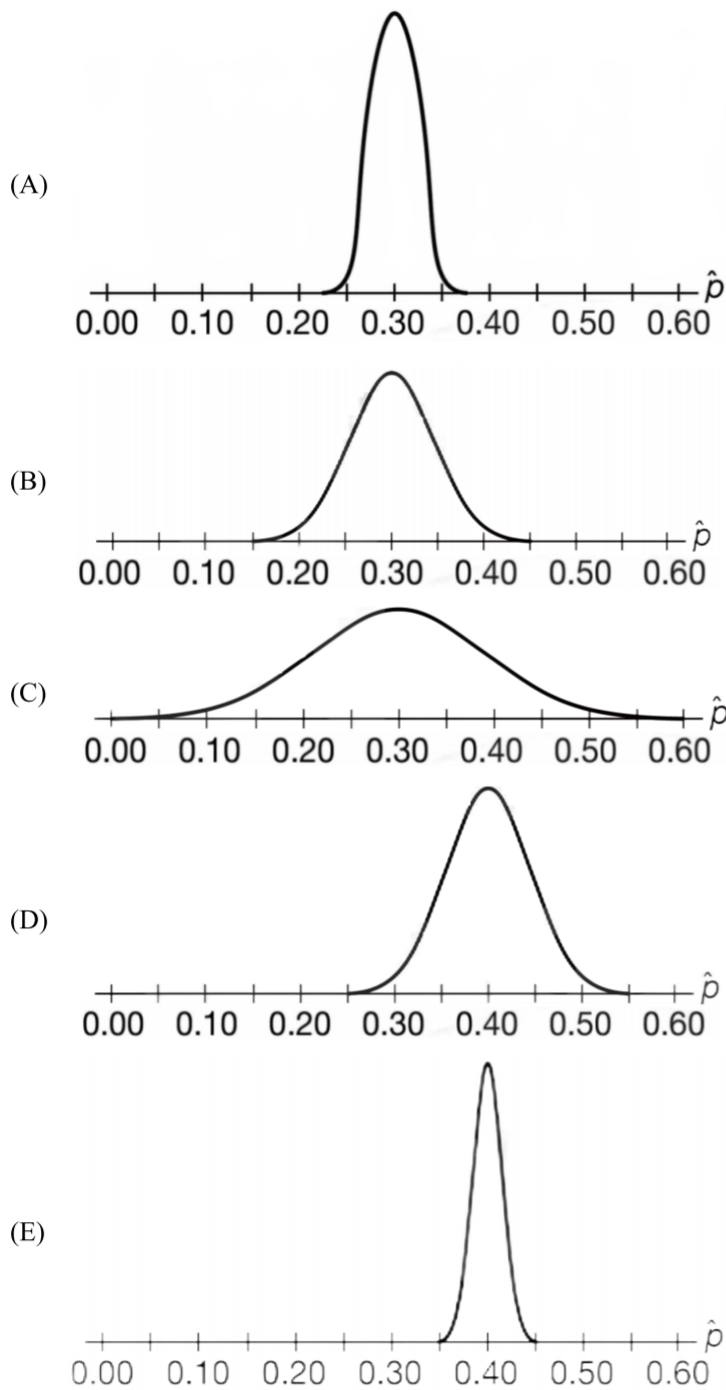
36. Researchers studying the relationship between sleep and memory obtained a sample of 28 adults to participate in an experiment. Half of the participants were randomly assigned to a full night's sleep, and the other half were awakened every two hours throughout the night. The following day, the participants were given a list of words to memorize and recall after one hour. Researchers recorded the number of words recalled by each participant. The p -value of an appropriate statistical test was 0.29.

Which of the following is a reasonable conclusion?

- (A) The data provide sufficient evidence that memory is related to sleep.
 - (B) The data provide sufficient evidence that memory is not related to sleep.
 - (C) A cause-and-effect relationship between memory and sleep cannot be established. However, the data provide sufficient evidence that there is an association between memory and sleep.
 - (D) The data do not provide convincing evidence that memory is related to sleep.
 - (E) The conditions for inference must not have been satisfied.
37. A recent report stated that 30 percent of smartphone users exceed their data limit in a given month. A consumer advocacy group questions the validity of the reported percent. The group obtained a random sample of 100 smartphone users and found that 40 of them exceeded their data limit in a certain month.

Of the following sampling distributions, which one should be used to test the hypothesis that the proportion is different than 30 percent?

Statistics 2021Intl MCQ



38. A college professor administered a survey to students from three different majors: engineering, nursing, and communications. The students were asked to identify how they prefer to be contacted: e-mail, telephone, or text message. The professor then used the survey data to conduct a chi-square test.

Which of the following best describes the null hypothesis for the test?

Statistics 2021 Intl MCQ

- (A) There is an association between major and preferred method of communication.
- (B) There is a positive association between major and preferred method of communication.
- (C) There is no association between major and preferred method of communication.
- (D) One-third of all students at the college are engineering majors, one-third are nursing majors, and one-third are communication majors.
- (E) One-third of all students at the college prefer to be contacted by e-mail, one-third prefer to be contacted by telephone, and one-third prefer to be contacted via text message.
39. A random sample of students was selected at each of two high schools to investigate whether the proportions of students who drive themselves to school are different for the two schools. Of the 100 students sampled at one high school, 25 reported driving themselves to school. Of the 200 students sampled at the other high school, 62 reported driving themselves to school.

Which of the following is equivalent to the p -value for a test to determine whether there is convincing statistical evidence of a difference between the proportions of students who drive themselves to school for the two schools?

- (A) $P\left(z > \frac{0.25 - 0.31}{\sqrt{\frac{(0.25)(0.75)}{100} + \frac{(0.31)(0.69)}{200}}}\right)$
- (B) $P\left(z < \frac{0.25 - 0.31}{\sqrt{\frac{(0.25)(0.75)}{100} + \frac{(0.31)(0.69)}{200}}}\right)$
- (C) $2 \cdot P\left(z < \frac{0.25 - 0.31}{\sqrt{\frac{(0.25)(0.75)}{100} + \frac{(0.31)(0.69)}{200}}}\right)$
- (D) $P\left(z < \frac{0.25 - 0.31}{\sqrt{(0.29)(0.71)\left(\frac{1}{100} + \frac{1}{200}\right)}}\right)$
- (E) $2 \cdot P\left(z < \frac{0.25 - 0.31}{\sqrt{(0.29)(0.71)\left(\frac{1}{100} + \frac{1}{200}\right)}}\right)$
40. A research agency surveyed 2,209 residents who live in one of three districts in a county: urban, suburban, or rural. The residents were asked to choose one of four categories that aligned with their opinion on how wireless infrastructure in the county should be paid for: wage tax, sales tax, property tax, or user fee.

When investigating whether opinion about how the wireless infrastructure should be paid for is independent of the district in which residents live, which of the following is the correct degrees of freedom for the appropriate chi-square test?

- (A) 6
- (B) 7
- (C) 9
- (D) 12
- (E) 16

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 .

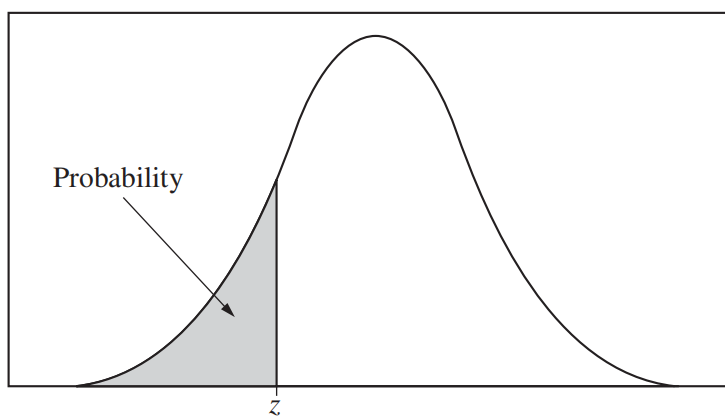


Table A Standard normal probabilities

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 .

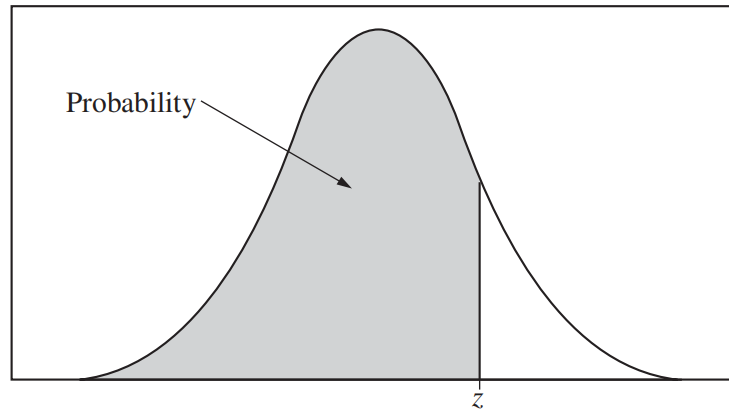


Table A (Continued)

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^* .

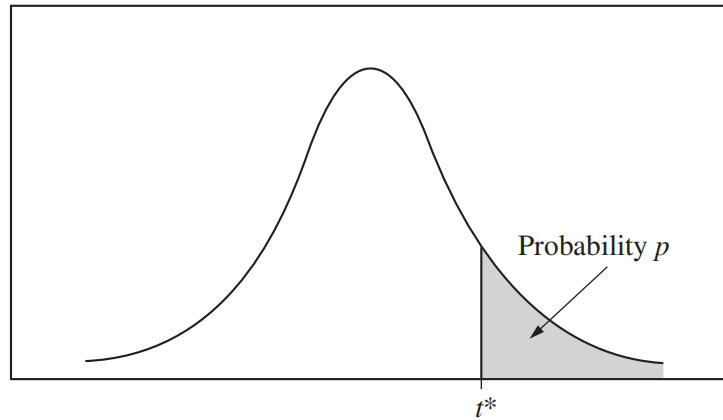


Table B t distribution critical values

df	Tail probability p											
	.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.

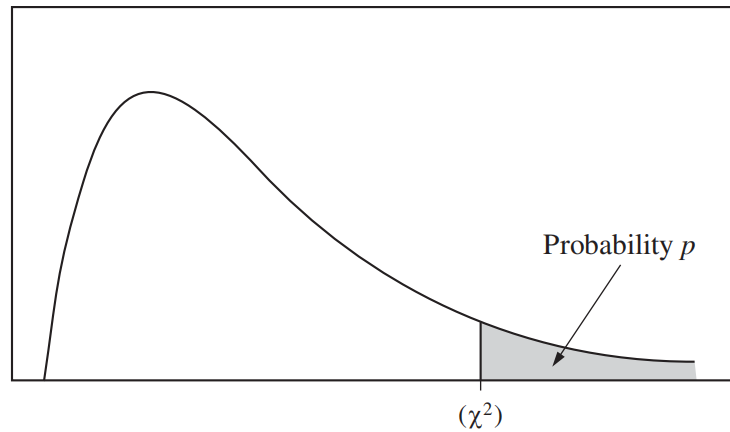


Table C χ^2 critical values

df	Tail probability p											
	.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

2021

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 4BPB4-S

90

Formulas begin on page 3.
Questions begin on page 6.
Tables begin on page 22.

Formulas

(I) Descriptive Statistics

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

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

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

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

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

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

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

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

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

(II) Probability

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

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

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

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

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: $\frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

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

Single-Sample

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

Two-Sample

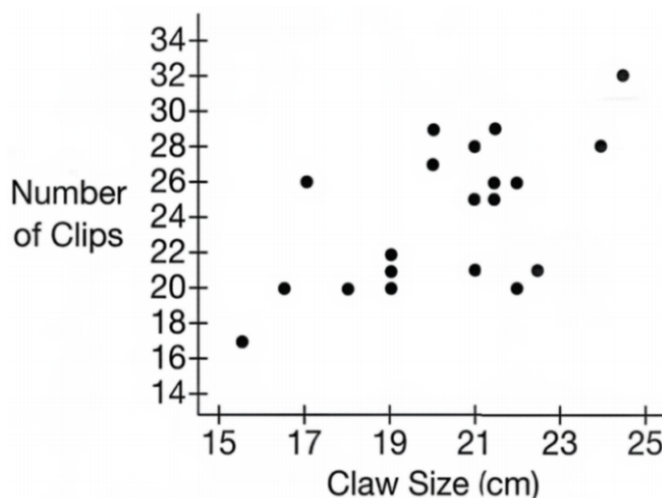
Statistic	Standard Deviation
Difference of sample means	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ Special case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

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

Statistics 2021 Intl MCQ

1. 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.

A robotics company is developing a robotic claw that can grab paper clips. The company is investigating the number of clips that can be grabbed by the claw at one time based on the size of the claw. The following scatterplot shows the number of clips grabbed by 20 claws, with claw size measured as the span of the claw in centimeters (cm).



Two models have been proposed to predict the number of clips from claw size. Model I is given as the following.

$$\text{predicted number of clips} = 3.179 + 1.032(\text{claw size})$$

- (a) Interpret the slope of Model I in context.
- (b) Explain why the intercept of Model I is not meaningful in context.
- (c) Model II has an intercept of 0 and is given as the following.

$$\text{predicted number of clips} = 1.190(\text{claw size})$$

- (i) Consider the claw of size 18cm that grabbed 20 clips. Calculate the residual for this observation using both models. Show your work.

Model I residual:

Model II residual:

- (ii) Which of the two models provides a better prediction for a claw of size 18cm that grabbed 20 clips? Justify your answer.

Statistics 2021Intl MCQ

2. 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.

Researchers at a large university with a population of 20,000 students are investigating student opinions about the university fitness center. A list of the names of students and how often they use the fitness center is available.

Two methods of stratification are being considered for selecting a sample of 120 students.

Method I - Stratify by how often students use the fitness center (no usage, infrequent usage, and frequent usage).

Method II - Stratify by first letter of students' last names (A to F, G to M, N to S, and T to Z).

(a) Which of the two methods for creating the strata, method I or method II, should the researchers use for their investigation? Explain your answer.

(b) Describe a procedure that could be used to select a stratified random sample of 120 students at the university for the method selected in part (a). The procedure should result in a sample that consists of an equal number of students from each stratum.

Statistics 2021 Intl MCQ

3. 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.

A farm produces olive oil that is put into bottles that are sealed with a cork. The diameter of the bottle opening is 1.03 inches. To provide an acceptable fit, the cork diameter should be from 0.99 inch to 1.03 inches. The corks are purchased in bulk from a supplier, and the distribution of cork diameters is approximately normal with mean 1.00 inch and standard deviation 0.05 inch.

(a) If one cork is selected at random, what is the probability that the cork selected will provide an acceptable fit for the bottle? Show your work.

A farm manager decides to purchase the corks from a new supplier. Seventy percent of corks produced by the new supplier have a diameter with an acceptable fit.

(b) The farm manager is investigating how many corks in a box of 20 randomly selected corks will provide an acceptable fit. Consider the probability that at least 19 of the corks will provide an acceptable fit for the bottle.

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

(ii) What is the probability that at least 19 of the corks will provide an acceptable fit for the bottle? Show your work.

(c) Suppose corks are inspected one at a time until an acceptable fit is found. Let the random variable W represent the number of corks inspected. What is the probability that W is less than or equal to 2 ? Show your work.

Statistics 2021 Intl MCQ

4. 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.

Disc golf is a sport in which players throw discs at a target. Atticus is investigating whether the weight of the disc used to play disc golf affects the accuracy of the throw. He obtained 15 standard discs weighing 175 grams each and 15 lightweight discs weighing 150 grams each. A player who is good at the game stood 40 yards from a target and threw each disc in random order. Atticus assessed the accuracy of each throw by measuring the distance, in feet, from where the disc landed to the target. The results and summary statistics are shown in the following table.

Accuracy at 175 Grams		
3.5	3.8	3.8
3.9	4.0	4.0
4.1	4.2	4.2
4.3	4.4	4.4
4.5	4.7	4.8

Accuracy at 150 Grams		
3.3	3.5	3.5
3.8	3.9	4.0
4.1	4.2	4.2
4.2	4.3	4.4
4.5	4.6	4.7

Accuracy at 175 Grams	Accuracy at 150 Grams
Summary Statistics	Summary Statistics
$n = 15$	$n = 15$
$\bar{x} = 4.173$ feet	$\bar{x} = 4.080$ feet
$s = 0.356$ feet	$s = 0.416$ feet

Do the data provide convincing statistical evidence, at the level of $\alpha = 0.05$, that players tend to have more accuracy with one disc weight than with the other? Complete the appropriate inference procedure to support your answer.

Statistics 2021 Intl MCQ

5. 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.

The instructor of a training course for a certain coding language hopes that a student's score on a pretest will predict the student's score on the certification final taken at the end of the course. The final scores indicate proficiency with the coding language. A random sample of 12 students who completed the pretest and the training course was selected, and the final scores of the students were recorded. The following tables show summary statistics and computer output for a regression analysis of the sample.

	Mean	Standard Deviation
Pretest	54.833	10.408
Final	78.867	6.946

Predictor	Coef	SE	T	P
Constant	49.2804	6.9199	7.12	0.0000
Pretest	0.5396	0.1242	4.35	0.0015
$S = 4.2682$		$R\text{-sq} = 65.38\%$		$R\text{-sq(adj)} = 61.92\%$

(a) Identify the equation of the least-squares regression line for predicting final score from pretest score. Identify the variables that appear in the equation in context.

(b) Consider the information presented in the tables and your answer from part (a).

(i) Determine the value of the predicted final score for students with a pretest score of 65.

(ii) Interpret $S = 4.2682$ in the context of the problem.

For any particular pretest score, two point estimates can be calculated from the given simple linear regression analysis.

Point Estimate I: An estimate of the mean final score among all students with a particular pretest score who complete the course.

Point Estimate II: A prediction for the final score of one student with a particular pretest score who completes the course.

The values of both point estimates are equal to the predicted value of the least-squares regression equation for the particular pretest score, such as the estimate from part (b-i) for a pretest score of 65. Although the values of the two point estimates are the same, the sampling variability is not the same.

(c) Which of the two point estimates has the larger sampling variability? Explain your reasoning.

(d) The value calculated in part (b-i) is an estimate of the mean final score for all students with a pretest score of 65 who completed the course. If a list of the pretest scores and final scores for all students who completed the course is obtained, explain how this information could be used to investigate the sampling distribution of the estimated mean final scores for all students with a pretest score of 65.

Statistics 2021 Intl MCQ

6. 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.

A study was conducted with 50 volunteers to investigate whether yawning is contagious. Each volunteer was randomly assigned to one of two groups: Group A or Group B. Each person in Group A was shown a video in which someone yawned, and each person in Group B was shown a video in which no one yawned. Of the 34 people assigned to Group A, 11 yawned, and of the 16 people assigned to Group B, 3 yawned.

- (a) Is this study an observational study or an experiment? Explain your choice in context.

Let p_A represent the proportion of all people who would yawn when shown the video in which someone yawned, and let p_B represent the proportion of all people who would yawn when shown the video in which no one yawned.

- (b) Explain why it is inappropriate to use a two-sample z-test for proportions to test the following hypotheses.

$$H_0 : p_A = p_B$$

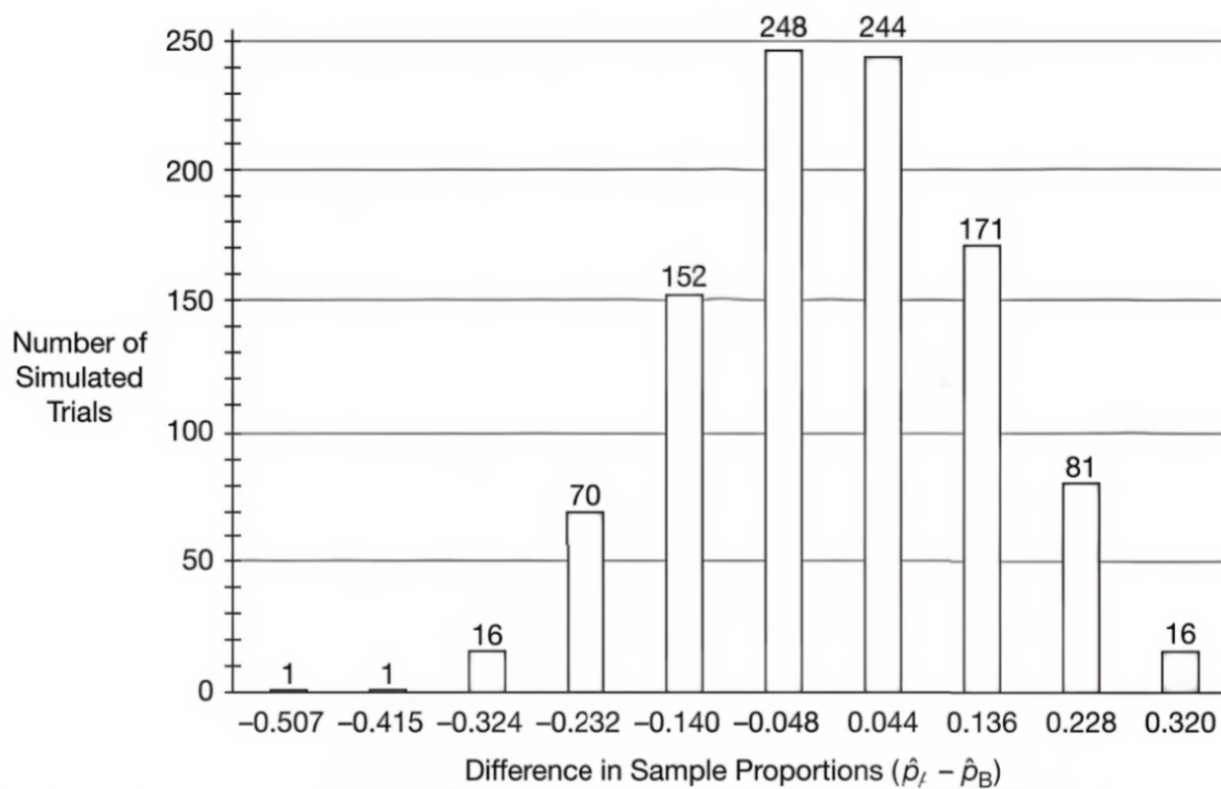
$$H_a : p_A > p_B$$

- (c) Based on the study data, determine the point estimate for $p_A - p_B$. Show your work.

- (d) A simulation to approximate the sampling distribution of the difference in sample proportions $\hat{p}_A - \hat{p}_B$ was conducted under the assumption that yawning is not contagious.

The graph shows the results of 1,000 trials of such a simulation.

Statistics 2021Intl MCQ



- (i) Based on your answer in part (c) and the simulated results, what is the approximate p -value for the test stated in part (b)? Show your work.
- (ii) Based on your answer to part (d-i) and using a significance level of 0.05, what conclusion is appropriate about whether yawning is contagious? Justify 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 .

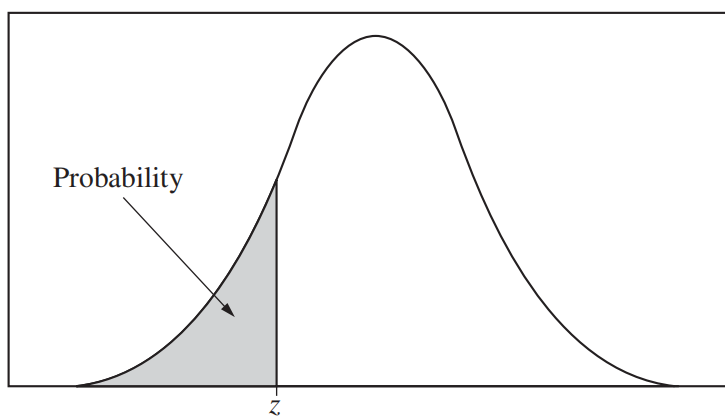


Table A Standard normal probabilities

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 .

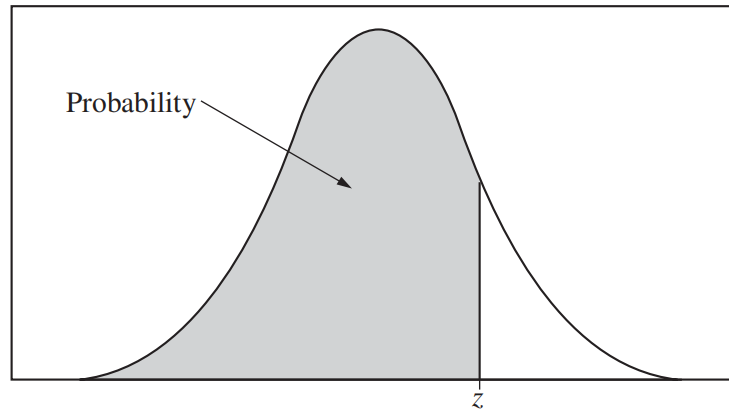


Table A (Continued)

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^* .

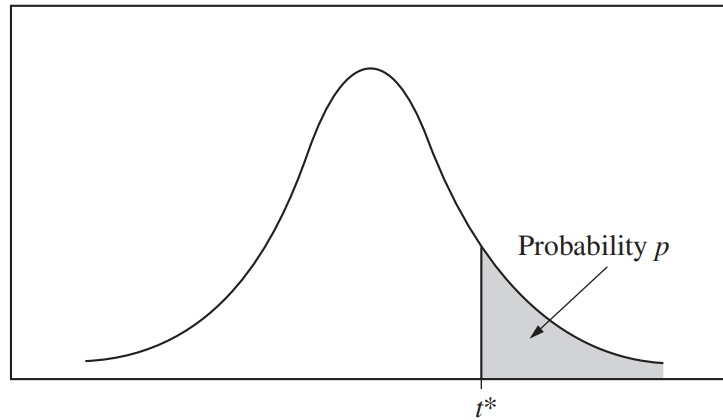


Table B t distribution critical values

df	Tail probability p											
	.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.

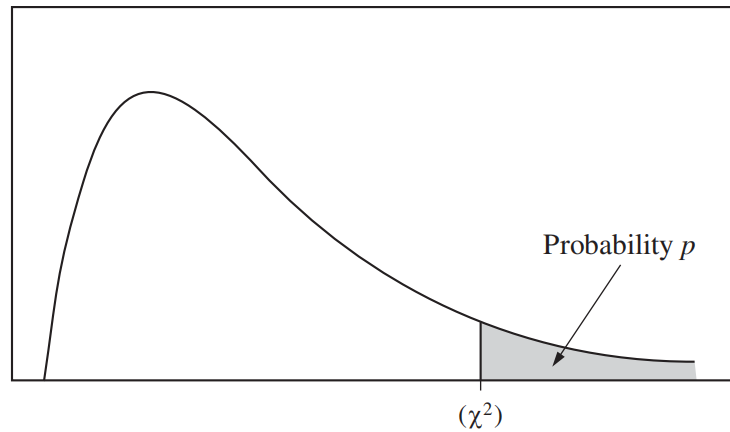


Table C χ^2 critical values

df	Tail probability p											
	.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

2021

APPENDIX - ANSWERS

Matching multiple-choice answers and free-response scoring follow.

Answer Key for AP Statistics

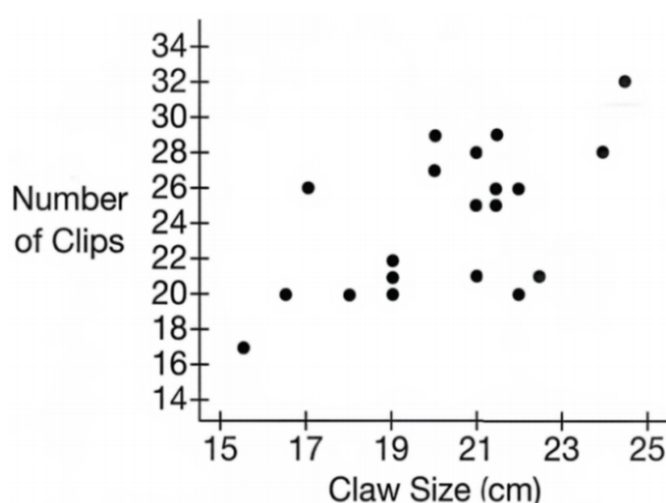
Practice Exam, Section I

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

Statistics 2021 Intl MCQ

1. 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.

A robotics company is developing a robotic claw that can grab paper clips. The company is investigating the number of clips that can be grabbed by the claw at one time based on the size of the claw. The following scatterplot shows the number of clips grabbed by 20 claws, with claw size measured as the span of the claw in centimeters (cm).



Two models have been proposed to predict the number of clips from claw size. Model I is given as the following.

$$\text{predicted number of clips} = 3.179 + 1.032(\text{claw size})$$

- (a) Interpret the slope of Model I in context.
- (b) Explain why the intercept of Model I is not meaningful in context.
- (c) Model II has an intercept of 0 and is given as the following.

$$\text{predicted number of clips} = 1.190(\text{claw size})$$

- (i) Consider the claw of size 18cm that grabbed 20 clips. Calculate the residual for this observation using both models. Show your work.

Model I residual:

Model II residual:

- (ii) Which of the two models provides a better prediction for a claw of size 18cm that grabbed 20 clips? Justify your answer.

Part A, Part B, Part C

Statistics 2021 Intl FRQ

Select a point value to view scoring criteria, solutions, and/or examples and to score the response

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4.
- The model solution represents an ideal response to each part of the question and the scoring criteria identify the specific components of the model solution that are used to determine the score.
- If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions.

Scoring

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

✓

0	1	2	3	4
---	---	---	---	---

All three parts essentially correct

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect

Solution

Part (a):

The least-squares estimate of the slope of Model I is 1.032 clips/cm. The predicted (or average) number of paper clips that can be grabbed by the claw increases by about 1.032 clips for each one-centimeter increase in the span of the claw.

Scoring

Part (a) is scored as follows:

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

Identifies the value of the slope as 1.032

Statistics 2021 Intl FRQ

Provides an interpretation that references an increase of the correct number of paper clips (e.g., 1.032) for each one-centimeter increase in the span of the claw, or an equivalent rate of increase (e.g., an increase of 10.32 clips for each 10-centimeter increase in the span of the claw)

Indicates that the slope is a change in the "average", "mean", "expected", or "predicted" value of the response variable (number of paper clips)

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.

Additional Notes:

· Component 2 is satisfied if the value for the rate of increase is consistent with the identified value of the slope in component 1, even if an incorrect value of the slope is identified.

· If the response does not identify a value for the slope, assume the correct value was identified when scoring component 2.

· “Claw size” and “span of the claw” should be scored as equivalent references to the explanatory variable.

Solution**Part (b):**

The least-squares estimate of the intercept of Model I is 3.179 clips. This corresponds to the predicted (or average) number of paper clips that can be grabbed by a claw with a span of zero centimeters. A predicted value of 3.179 clips is not sensible because a claw with a span of zero centimeters could not exist and would not be able to pick up any paper clips.

Scoring

Part (b) is scored as follows:

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

Identifies the value of the intercept as 3.179 OR indicates that the intercept is positive

Indicates that the intercept is the “predicted”, “average”, “mean”, or “expected” number of paper clips that would be grabbed by a claw with a span of zero centimeters

Explains that a claw with a span of zero centimeters could not exist

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.

Additional Notes:

Component 1 can be satisfied by stating that the claw could not grab any paper clips.

Solution**Part (c):**

(i) The Model I residual is

Statistics 2021 Intl MCQ

$$\begin{aligned} &20 - (3.179 + 1.032(18)) \\ &= 20 - 21.755 \\ &= -1.755 \end{aligned}$$

The Model II residual is

$$\begin{aligned} &20 - (1.190)(18) \\ &= 20 - 21.42 \\ &= -1.42 \end{aligned}$$

(ii) Model II provides a better prediction (a prediction closer to the 20 paper clips that were actually grabbed with a claw of size 18 cm in this study) because the residual for Model II is closer to zero than the residual for Model I.

Scoring

Part (c) is scored as follows:

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

Correctly calculates the Model I residual in part (c-i) as -1.755 AND provides supporting work or justification (e.g., showing that the residual is computed by subtracting the predicted number of clips from the actual number of clips grabbed by the claw with an 18 cm span)

Correctly calculates the Model II residual in part (c-i) as -1.42 AND provides supporting work or justification (e.g., showing that the residual is computed by subtracting the predicted number of clips from the actual number of clips grabbed by the claw with an 18 cm span)

Indicates that Model II provides a better prediction in part (c-ii)

Provides a justification for part (c-ii) that is based on which model has a residual closer to zero (or a predicted value closer to 20)

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

OR

if the response correctly calculates both residuals but does not provide supporting work or justification AND satisfies at least one of component 3 or component 4.

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

Additional Notes:

- If incorrect values of one or both residuals are reported in part (c-i), then component 3 is satisfied if the response indicates that the model with the residual closer to zero (or a predicted value closer to 20) provides a better prediction.
- Component 4 can be met by stating that the residual is smaller for Model II, even though -1.42 is larger than -1.755 .

Statistics 2021 Intl FRQ

2. 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.

Researchers at a large university with a population of 20,000 students are investigating student opinions about the university fitness center. A list of the names of students and how often they use the fitness center is available.

Two methods of stratification are being considered for selecting a sample of 120 students.

Method I - Stratify by how often students use the fitness center (no usage, infrequent usage, and frequent usage).

Method II - Stratify by first letter of students' last names (A to F, G to M, N to S, and T to Z).

(a) Which of the two methods for creating the strata, method I or method II, should the researchers use for their investigation? Explain your answer.

(b) Describe a procedure that could be used to select a stratified random sample of 120 students at the university for the method selected in part (a). The procedure should result in a sample that consists of an equal number of students from each stratum.

Part A, and Part B

Select a point value to view scoring criteria, solutions, and/or examples and to score the response

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4.
- The model solution represents an ideal response to each part of the question and the scoring criteria identify the specific components of the model solution that are used to determine the score.
- If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions.

Scoring

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

✓

0	1	2	3	4
---	---	---	---	---

Both parts essentially correct

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect

Statistics 2021 Intl MCQ

- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect

Solution**Part (a):**

The researchers should use method I for creating the strata in their investigation since it would be better to stratify by how often students use the fitness center than by first letter of last name. Stratified random sampling is best used with homogeneous strata, that is, when variation in responses is relatively small within strata. There is no reason to believe that forming groups of students based on the first letter of their last name would result in very homogeneous strata with respect to opinions about the fitness center. On the other hand, students who frequently use the fitness center may have similar opinions, and those opinions may differ substantially from opinions of students who never use it, or students who use it infrequently. Consequently, method I would produce strata with smaller within-strata variation in opinions about the fitness center.

Scoring

Part (a) is scored as follows:

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

Indicates that method I should be used

Addresses EITHER homogeneity within groups (e.g., "students in the same group have similar opinions") OR heterogeneity between groups (e.g., "students in different groups tend to have different opinions")

Relates the stratification variable to the response variable in context (e.g., "how often students use the gym may affect their opinion")

Partially correct (P) if the response satisfies component 1 AND component 2

OR

If the response satisfies component 1 AND 3,

OR

if the response does not select either method AND satisfies both component 2 and component 3.

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

Additional Notes:

· Examples of responses that earn a P

o "Method 1 because it creates homogenous groups" satisfies components 1 and 2.

o "Method 1 because the first letter of students' last names likely has nothing to do with opinions about the fitness center" satisfies components 1 and 3.

Statistics 2021 Intl MCQ

o "It is important to select the method that creates the most similar groups of students in terms of their opinions about the fitness center" (satisfies components 2 and 3)

· Note that a response that selects Method 2 does not meet the criteria for E or P and scores I.

Solution**Part (b):**

Break up the list of all 20,000 students at the university into three separate lists corresponding to the three strata for method I – one for students who never use the fitness center, one for students who infrequently use the fitness center, and one for students who frequently use the fitness center. Select separate random samples of size 40 from each of the three lists. Start with the list of students who never use the fitness center. If there are N_1 students on the list, number the students from 1 to N_1 based on their position on the list. Using a random number generator, generate 40 random numbers between 1 and N_1 , without repeats. These numbers indicate which students should be included in the sample from the first stratum. Repeat this procedure to draw a random sample of 40 students from the N_2 students on the second list and a random sample of 40 students from the N_3 students on the third list.

Scoring

Part (b) is scored as follows:

Essentially correct (E) if the response is consistent with the method selected in (a) or with the correct method AND satisfies the following four components:

Selects a separate sample from each stratum (3 strata for method I, or 4 strata for method II)

Identifies appropriate labels for students, such as numbering the students from 1 to N where N is the number of students in a stratum, for each sample (stratum)

Describes how to implement a process for randomly selecting students from a stratum, such that every student in the stratum has an equal chance of being selected

The random selection process results in the same number of students being selected from each stratum (40 per stratum for method I, or 30 per stratum for method II)

Partially correct (P) if the response is consistent with the method selected in (a) or with the correct method AND satisfies only two or three of the four components

OR

if the response is not consistent with the method in (a) AND stratifies on a variable other than frequency of gym usage AND satisfies components two and three. (E.g., stratifying on class standing – freshman, sophomore, junior senior.)

OR

if the response describes a sampling method other than stratified (e.g., systematic or SRS) that identifies appropriate labels for sampling units in context (e.g., students or groups of students) AND describes how to implement a process for randomly selecting units such that every unit has an equal chance of being selected.

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

Additional Notes:

Statistics 2021 Intl MCQ

· A response that selects method II in part (a) may score E on part (b) if the response satisfies all four components, such as the following example:

o Break up the list of all 20,000 students at the university into four separate lists corresponding to the four strata for method II – one for students whose first letter of their last name is A to F, one for students whose first letter of their last name is G to M, one for students whose first letter of their last name is N to S, and one for students whose first letter of their last name is T to Z. Select separate random samples of size 30 from each of the four lists. Start with the list of students whose first letter of their last name is A to F. If there are N_1 students on the list, number the students from 1 to N_1 based on their position on the list. Using a random number generator, generate 30 random numbers between 1 and N_1 , without repeats. These numbers indicate which students should be included in the sample from the first stratum. Repeat this procedure to draw a random sample of 30 students from the N_2 students on the second list, a random sample of 30 students from the N_3 students on the third list, and a random sample of 30 students from the N_4 students on the fourth list.

· To satisfy component 4, the response must be clear that within a stratum, each person can be selected at most once (e.g., repeats are ignored, sampling is done without replacement, must produce a sample of 40 people).

· A response that does not include context (e.g., students or fitness center usage groups) may score at most a P.

A response that does not select a method in part (a) or leaves part (a) blank may still score an E in part (b) if the response satisfies all four components.

3. 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.

A farm produces olive oil that is put into bottles that are sealed with a cork. The diameter of the bottle opening is 1.03 inches. To provide an acceptable fit, the cork diameter should be from 0.99 inch to 1.03 inches. The corks are purchased in bulk from a supplier, and the distribution of cork diameters is approximately normal with mean 1.00 inch and standard deviation 0.05 inch.

(a) If one cork is selected at random, what is the probability that the cork selected will provide an acceptable fit for the bottle?

A farm manager decides to purchase the corks from a new supplier. Seventy percent of corks produced by the new supplier have a diameter with an acceptable fit.

(b) The farm manager is investigating how many corks in a box of 20 randomly selected corks will provide an acceptable fit. Consider the probability that at least 19 of the corks will provide an acceptable fit for the bottle.

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

(ii) What is the probability that at least 19 of the corks will provide an acceptable fit for the bottle? Show your work.

(c) Suppose corks are inspected one at a time until an acceptable fit is found. Let the random variable W represent the number of corks inspected. What is the probability that W is less than or equal to 2? Show your work.

Part A, Part B, Part C

Select a point value to view scoring criteria, solutions, and/or examples and to score the response

Statistics 2021 Intl MCQ

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4.
- The model solution represents an ideal response to each part of the question and the scoring criteria identify the specific components of the model solution that are used to determine the score.
- If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions.

Scoring

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



0	1	2	3	4
---	---	---	---	---

All three parts essentially correct

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect

Solution

Part (a):

Let the random variable D represent the diameter of a randomly selected cork. The distribution of D is approximately normal with mean 1.00 inch and standard deviation 0.05 inch. The probability that a randomly selected cork will provide an acceptable fit is the probability that the diameter of a randomly selected cork is between 0.99 and 1.03 inches, which is:

$$\begin{aligned}
 &P(0.99 < D < 1.03) \\
 &= P\left(\frac{0.99-1.00}{0.05} < Z < \frac{1.03-1.00}{0.05}\right) \\
 &= P(-0.2 < Z < 0.6) \\
 &\approx 0.305
 \end{aligned}$$

Scoring

Part (a) is scored as follows:

Statistics 2021 Intl FRQ

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

Indicates the use of a normal distribution with mean 1.00 inch and standard deviation 0.05 inch

Specifies the correct event, including the correct boundary values for the cork diameter

Provides the correct probability of 0.305 (or probability consistent with components 1 and 2), with supporting work

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.

Additional Notes:

· A response may satisfy component 1 by any of the following:

o Graphical: Displaying a graph of a normal density function with the horizontal axis clearly labeled using the mean and standard deviation for the cork diameter distribution

o Calculator function syntax: Labeling correct values of the mean and standard deviation in a “normalcdf” statement, such as:

· normalcdf(lower = 0.99, upper = 1.03, mean = 1.00, standard deviation = 0.05)

· normalcdf(lower = 0, upper = 1.03, $\mu = 1.00$, $\sigma = 0.05$)–

· normalcdf(lower = 0, upper = 0.99, $\mu = 1.00$, $\sigma = 0.05$)

· normalcdf(lower = –0.2, upper = 0.6, $\mu = 0$, $\sigma = 1$)

o Words or standard notation: Using a statement such as “normal distribution with mean 1.00 and standard deviation 0.05”, or using standard notation such as $N(1.00, 0.05)$

o Z-score: Displaying the correct mean and standard deviation in a z-score calculation, such as:

$$\cdot Z = \frac{(\text{random variable}) - 1.00}{0.05}$$

$$\cdot \frac{0.99 - 1.00}{0.05} \leq Z \leq \frac{1.03 - 1.00}{0.05}$$

$$\cdot \frac{1.03 - 1.00}{0.05} = 0.60$$

· A response may satisfy component 2 by any of the following:

o Graphical: Displaying a graph of a normal density function with the horizontal axis clearly labeled and the region of interest clearly identified

o Calculator function syntax: Labeling the lower and upper bounds of the region of interest in a “normalcdf” statement, such as:

· normalcdf(lower = 0.99, upper = 1.03, mean = 1.00, standard deviation = 0.05)

· normalcdf(0, upper = 1.03, $\mu = 1.00$, $\sigma = 0.05$)–

· normalcdf(0, upper = 0.99, $\mu = 1.00$, $\sigma = 0.05$)

Statistics 2021 Intl FRQ

· $\text{normalcdf}(\text{lower} = -0.2, \text{upper} = 0.6, \mu = 0, \sigma = 1)$

o **Words:** Specifying the correct event in words with correct numerical values for the boundary values, such as $P(0.99 < \text{cork diameter} < 1.03)$

o **Random variable:** Specifying the correct boundary values using a random variable, such as:

· $P(0.99 < D < 1.03)$

· $P(-0.20 < Z < 0.60)$

Solution

Part (b):

(i) The random variable of interest, X , is the number of corks in a box of 20 randomly selected corks that provide an acceptable fit. The random variable X follows a binomial distribution with $n = 20$ and $p = 0.7$.

(ii) The probability that at least 19 of the corks will provide an acceptable fit for the bottle is:

$P(X \geq 19)$

$$= \binom{20}{19} (0.7)^{19} (0.3)^1 + \binom{20}{20} (0.7)^{20} (0.3)^0$$

≈ 0.00764

Scoring

Part (b) is scored as follows:

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

Defines an appropriate random variable (e.g., X) in part (b-i), in the context of the number of corks in a box of 20 randomly selected corks that provide an acceptable fit

States that the random variable in part (b-i) follows a binomial distribution with $n = 20$ and $p = 0.7$

Indicates the event of interest (i.e., identifies the correct boundary value AND direction for the event) in the calculation of the probability in part (b-ii)

Provides supporting work to identify the correct probability of 0.00764 (or 0.008, if rounded) OR a probability consistent with components 2 and 3

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.

Additional Notes:

· A response may satisfy component 1 if the random variable is defined correctly as part of the answer to (b-ii), as long as the random variable is not incorrectly defined elsewhere in the response.

· A response may satisfy the context requirement in component 1 by referring to the number of corks in a box that provide an acceptable fit; additional details about the box (box of 20, randomly selected) may be omitted without losing credit for component 1.

Statistics 2021 Intl MCQ

· A response may satisfy component 2 by any of the following:

o Calculator function syntax: Labeling correct values of n and p in a “binomcdf” or “binompdf” statement, such as the examples below. Note that it is not necessary to label the event boundary or specify the correct event in order to satisfy component 2.

· $1 - \text{binomcdf}(n = 20, p = 0.7, \text{upper bound} = 18)$

· $1 - \text{binomcdf}(n = 20, p = 0.7, 18)$

· $\text{binomcdf}(n = 20, p = 0.7, 18)$

· $\text{binompdf}(n = 20, p = 0.7, 19) + \text{binompdf}(n = 20, p = 0.7, 20)$

o Words or standard notation: Using a statement such as “binomial distribution with $n = 20$ and $p = 0.7$ ”, or using standard notation such as $B(20, 0.7)$

o Binomial probability formula: Displaying a binomial probability formula using $n = 20$ and $p = 0.7$, such as the examples below satisfies component 2. Note that it is not necessary to label the event boundary or specify the correct event in order to satisfy component 2.

· $\binom{20}{19} (0.7)^{19} (0.3)^1 + \binom{20}{20} (0.7)^{20} (0.3)^0$

· $1 - \binom{20}{18} (0.7)^{18} (0.3)^2$

· $\binom{20}{19} (0.7)^{19} (0.3)^1$

· A response that correctly calculates a normal approximation to the binomial probability may satisfy components 3 and 4 because $np = 14$ and $n(1 - p) = 6$ are both greater than 5. Note that the response must show the correct boundary value AND direction to satisfy component 3.

o Without using a continuity correction, the probability is $P\left(Z \geq \frac{19 - (20)(0.7)}{\sqrt{(20)(0.7)(0.3)}}\right) = P(Z \geq 2.44) \approx 0.007344$.

o Using a continuity correction, the probability is $P\left(Z \geq \frac{18.5 - (20)(0.7)}{\sqrt{(20)(0.7)(0.3)}}\right) = P(Z \geq 2.196) \approx 0.0140$.

· A response may satisfy component 3 by any of the following:

o Graphical: Displaying a clearly labeled bar graph with the bar for 19 corks AND the bar for 20 corks shaded

o Calculator function syntax: Labeling a “binomcdf” or “binompdf” statement to indicate the correct boundary value AND direction, such as the examples below. Note that it is not necessary to label $n = 20$ and $p = 0.7$ in order to satisfy component 3.

· $1 - \text{binomcdf}(n = 20, p = 0.7, \text{upper bound} = 18)$

· $\text{binompdf}(n = 20, p = 0.7, x = 19) + \text{binompdf}(n = 20, p = 0.7, x = 20)$

Statistics 2021 Intl FRQ

$$\cdot 1 - \text{binomcdf}(20, 0.7, \text{upper bound} = 18)$$

o Words: Using a statement such as “ $P(\text{at least 19 acceptable corks})$ ”

o Probability expression based on the defined random variable: Using the random variable identified in part (b-i) to write a probability expression, such as:

$$\cdot P(X \geq 19)$$

$$\cdot P(X > 18)$$

$$\cdot 1 - P(X \leq 18)$$

o Binomial probability formula: Displaying the binomial probability formula for the correct probability

$$\cdot \binom{20}{19} (0.7)^{19} (0.3)^1 + \binom{20}{20} (0.7)^{20} (0.3)^0$$

$$\cdot 1 - \sum_{k=0}^{18} \binom{20}{k} (0.7)^k (0.3)^{20-k}$$

$$\cdot 1 - \sum_{k=0}^{18} \binom{20}{k} (0.7)^k (0.3)^{20-k}$$

· A response may satisfy component 4 with a numerically incorrect result if the result is preceded immediately by a complete and arithmetically correct expression for computing the probability; this scenario will be treated as a transcription error, such as:

$$\cdot \binom{20}{19} (0.7)^{19} (0.3)^1 + \binom{20}{20} (0.7)^{20} (0.3)^0 \\ \approx 0.006834 + 0.000798 \approx 0.0148$$

Note: Just reporting $0.006834 + 0.000798 \approx 0.0148$ is not sufficient to satisfy component 4.

Solution**Part (c):**

The random variable W follows a geometric distribution with $p = 0.7$. The probability that W is less than or equal to 2 corks is:

$$\begin{aligned} P(W \leq 2) &= P(W = 1) + P(W = 2) \\ &= 0.7 + (0.3)(0.7) \\ &= 0.91 \end{aligned}$$

Scoring

Part (c) is scored as follows:

Statistics 2021 Intl FRQ

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

Indicates that W follows a geometric distribution with $p = 0.7$ with an explicit statement or a correct geometric probability formula

Indicates the event of interest (i.e., the correct boundary value AND direction), for the calculation of the probability

Provides supporting work to identify the correct probability of 0.91 OR a probability consistent with components 1 and 2

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.

Additional Notes:

· A response may satisfy component 1 by any of the following:

o Calculator function syntax: Labeling the correct value of p in a “geomcdf” or “geompdf” statement, such as the examples below. Note that it is not necessary to label the event boundary or specify the correct event in order to satisfy component 1.

- $\text{geomcdf}(p = 0.7, \text{upper bound} = 2)$
- $\text{geompdf}(p = 0.7, w = 1) + \text{geompdf}(p = 0.7, w = 2)$
- $\text{geomcdf}(p = 0.7, 2)$

o Words or standard notation: Using a statement such as “geometric distribution with $p = 0.7$ ”, or using standard notation such as $G(0.7)$

o Geometric probability formula: Displaying a geometric probability formula using $p = 0.7$, such as the examples below. Note that it is not necessary to label the event boundary or specify the correct event in order to satisfy component 1.

- $0.7 + (0.3)(0.7)$ satisfies components 1, 2 and 3.
- $(1 - 0.7)^{2-1}(0.7)$, satisfies component 1, but not component 2 or 3.

o Probability tree: Displaying a probability tree in which branches are labeled with the probabilities of 0.7 and 0.3, and the arithmetic operations produced by the tree are shown and are consistent with the geometric probability formula.

· If the response describes a geometric distribution with an incorrect probability value (e.g., uses $p = 0.305$ instead of $p = 0.7$), the response may satisfy component 1 if the same incorrect probability value was used for the binomial distribution given in the response to part (b).

· A response may satisfy component 2 by any of the following:

o Graphical: Displaying a clearly labeled bar graph with the bar for 1 cork AND the bar for 2 corks shaded

o Calculator function syntax: Labeling a “geomcdf” or “geompdf” statement to indicate the correct boundary value AND direction, such as the examples below. Note that it is not necessary to label $p = 0.7$ in order to satisfy component 2.

- $\text{geomcdf}(p = 0.7, \text{upper bound} = 2)$
- $\text{geompdf}(p = 0.7, w = 1) + \text{geompdf}(p = 0.7, w = 2)$

Statistics 2021 Intl FRQ

· `geomcdf(0.7, upper bound = 2)`

o Words: Using a statement such as “ $P(\text{no more than 2 corks inspected to find acceptable fit})$ ”

o Probability expression based on the given random variable: Using the given random variable W to write a probability expression, such as:

· $P(W \leq 2)$

· $P(W < 3)$

· $P(W = 1) + P(W = 2)$

o Geometric probability formula: Displaying the geometric probability formula for the correct probability, $0.7 + (0.3)(0.7)$

o Probability tree: Displaying a probability tree in which the paths representing $W = 1$ and $W = 2$ are clearly identified as the paths of interest

· A response may satisfy component 3 with a numerically incorrect result if the result is preceded immediately by a complete and arithmetically correct expression for computing the probability; this scenario will be treated as a transcription error, such as:

o $0.7 + (0.3)(0.7) = 0.28$

Statistics 2021 Intl MCQ

4. 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.

Disc golf is a sport in which players throw discs at a target. Atticus is investigating whether the weight of the disc used to play disc golf affects the accuracy of the throw. He obtained 15 standard discs weighing 175 grams each and 15 lightweight discs weighing 150 grams each. A player who is good at the game stood 40 yards from a target and threw each disc in random order. Atticus assessed the accuracy of each throw by measuring the distance, in feet, from where the disc landed to the target. The results and summary statistics are shown in the following table.

Accuracy at 175 Grams		
3.5	3.8	3.8
3.9	4.0	4.0
4.1	4.2	4.2
4.3	4.4	4.4
4.5	4.7	4.8

Accuracy at 150 Grams		
3.3	3.5	3.5
3.8	3.9	4.0
4.1	4.2	4.2
4.2	4.3	4.4
4.5	4.6	4.7

Accuracy at 175 Grams	Accuracy at 150 Grams
Summary Statistics	Summary Statistics
$n = 15$	$n = 15$
$\bar{x} = 4.173$ feet	$\bar{x} = 4.080$ feet
$s = 0.356$ feet	$s = 0.416$ feet

Do the data provide convincing statistical evidence, at the level of $\alpha = 0.05$, that players tend to have more accuracy with one disc weight than with the other? Complete the appropriate inference procedure to support your

Statistics 2021 Intl FRQ

answer.

3-part Inference Scoring

Select a point value to view scoring criteria, solutions, and/or examples and to score the response

· This question is scored in three sections. Each section is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The first section includes statements of the null and alternative hypotheses and identification of the appropriate hypothesis test. The second section includes verifying the conditions for the identified test, and calculating the value of the test statistic and the corresponding p-value. The third section includes the conclusion for the identified test. The response is then categorized based on the scores assigned to each section and awarded an integer score between 0 and 4 (see the table at the end of the question).

· The model solution represents an ideal response to each part of the question and the scoring criteria identify the specific components of the model solution that are used to determine the score.

· If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions.

Scoring

Scoring steps 1, 2, and 3 are each scored as essentially correct (E), partially correct (P), or incorrect (I).



0	1	2	3	4
---	---	---	---	---

All three scoring parts essentially correct

- ☐ Scoring Step 1 Essentially correct
- ☐ Scoring Step 1 Partially correct
- ☐ Scoring Step 1 Incorrect
- ☐ Scoring Step 2 Essentially correct
- ☐ Scoring Step 2 Partially correct
- ☐ Scoring Step 2 Incorrect
- ☐ Scoring Step 3 Essentially correct
- ☐ Scoring Step 3 Partially correct
- ☐ Scoring Step 3 Incorrect

Solution

Section 1:

Let μ_S represent the mean distance that standard discs weighing 175 grams land from the target for all possible throws, and let μ_L represent the mean distance that lightweight discs weighing 150 grams land from the target for all possible

Statistics 2021 Intl FRQ

throws.

The null hypothesis is $H_0: \mu_S = \mu_L$ and the alternative hypothesis is $H_a: \mu_S \neq \mu_L$.

An appropriate test is a two-sample t-test for a difference in population means.

Scoring

Section 1 is scored as follows:

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

States a correct null hypothesis of equal means AND a correct alternative hypothesis of unequal means

Provides sufficient context by including reference to:

· the two groups (types of discs),

AND

· *populations* (all possible throws for each type of disc),

AND

· the response variable (distance from the target, or accuracy)

Identifies a two-sample t-test for a difference in population means by name (e.g., "two-sample t-test") or by formula

Partially correct (P) if the response does not meet the criteria for E but satisfies either component 1 and/or component 3.

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

Additional Notes:

· The elements of component 2 do not have to be satisfied with the statement of the hypotheses. They may be satisfied by work presented anywhere in the response, most likely by the statement of the conclusion.

· If the statement of the hypotheses refers to population means and the conclusion refers to sample means (or vice versa), then the population aspect of component 2 is not satisfied.

· A response that states the hypotheses in words (e.g., "the null hypothesis is that the means are equal, and the alternative hypothesis is that the means are not equal") may satisfy component 1. Neither context nor the concept of *population* is required to satisfy component 1.

· To satisfy component 1, the hypotheses must be stated in terms of two means. Use of standard symbols for means, with group labels (such as μ_{Standard} and $\mu_{\text{Lightweight}}$, or μ_1 and μ_2 , or $\bar{X}_S$ and $\bar{X}_L$) is sufficient to satisfy component 1. Any other symbols must be clearly defined as a mean and include labels to indicate two different populations (but the labels do not need to reflect the context of standard discs and lightweight discs) in order for the response to satisfy component 1.

· The notation μ_{Standard} and $\mu_{\text{Lightweight}}$ satisfies the group and population aspects of component 2, because it is standard notation for population means and it reflects the context of standard discs and lightweight discs.

· Without additional description that provides links to standard discs and lightweight discs, notation such as μ_1 and μ_2 does not satisfy the group aspect of component 2.

Statistics 2021 Intl MCQ

- A response that states the hypotheses in words may satisfy some or all aspects of component 2. For example, “the null hypothesis is that the mean distance from the target for all possible throws of a standard disc is equal to the mean distance from the target for all possible throws of a lightweight disc and the alternative hypothesis is that the means are not equal”) satisfies all aspects of component 2.
- If the response clearly refers to *sample* means instead of *population* means using words or symbols (e.g., $\bar{X}_S$ and $\bar{X}_L$, or $\hat{\mu}_{\text{Standard}}$ and $\hat{\mu}_{\text{Lightweight}}$), then the population aspect of component 2 is not satisfied unless the symbols used are defined as *population* means.
- If the response identifies the correct test by name, but also states an incorrect formula or uses incorrect notation in the formula, then component 3 is not satisfied.
- A pooled two-sample *t*-test may satisfy component 3 if the response comments on the plausibility of the equal population variances assumption.
- If the response identifies a two-sample *z*-test for a difference in population means, then component 3 is not satisfied.

Confidence Interval Approach:

- o For a response using a confidence interval approach, component 1 is not considered in scoring. Such a response should be scored as follows:
 - Essentially correct (E) if the response satisfies components 2 and 3.
 - Partially correct (P) if the response does not meet the criteria for E and satisfies either component 2 or 3.
- o A response that uses a two-sample *t*-interval for a difference in population means must demonstrate all three aspects (groups, population, and response variable) to satisfy component 2. For example, component 2 is satisfied if the response indicates that it is a confidence interval for the difference between the mean distance from the target for all possible throws of a standard disc and the mean distance from the target for all possible throws of a lightweight disc, even if the hypotheses are not stated.
- o If a two-sample *t*-interval for a difference in population means is identified correctly by name or by formula, then component 3 is satisfied.
- o If the response uses individual one-sample *t*-intervals for the two means, which is not a correct approach, then component 3 is not satisfied.

Matched-Pairs Approach:

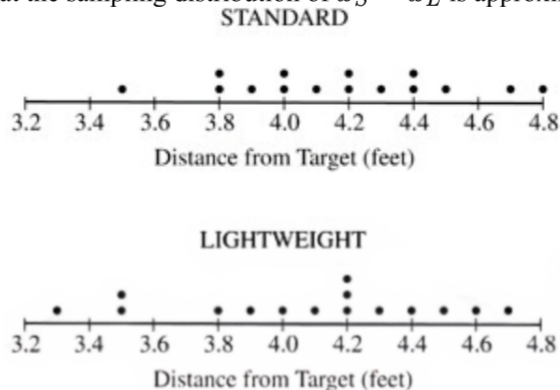
- o Because a response presenting a matched-pairs *t*-test approach cannot satisfy component 3, such a response cannot meet the criteria for essentially correct, E. If the response satisfies component 1, it may be scored P. Otherwise, it is scored I.
- o A response that attempts a matched-pairs approach using a confidence interval may be scored P if component 2 is satisfied and the response indicates that the confidence interval is for the mean difference. Otherwise, it is scored I.

Solution**Section 2:**

The randomization condition for performing the two-sample *t*-test for a difference in population means is satisfied because the 30 discs (15 standard and 15 lightweight) are thrown in a random order. The sample sizes are both 15, which is less than 30, however, the approximately symmetric shape of the dotplots of the sample data from both discs support an

Statistics 2021 Intl MCQ

assumption that the sampling distribution of $\bar{x}_S - \bar{x}_L$ is approximately normal.



$$t = \frac{\bar{x}_S - \bar{x}_L}{\sqrt{\frac{s_S^2}{n_S} + \frac{s_L^2}{n_L}}} = \frac{4.173 - 4.080}{\sqrt{\frac{(0.356)^2}{15} + \frac{(0.416)^2}{15}}} \approx 0.658.$$

The value of the t -statistic is

The corresponding p -value, using an approximate degrees of freedom of 27.3, is $P(|t| > 0.658) \approx 0.516$.

Scoring

Section 2 is scored as follows:

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

Checks the randomization condition by referring to the throwing of the discs in random order

Checks that the sampling distribution of $\bar{x}_S - \bar{x}_L$ is approximately normal or doesn't violate the assumption that it could be normal (e.g., by sketching and commenting on the approximately symmetric shape of the graphs of the sample data from both discs) OR states that this condition must be assumed because both sample sizes are less than 30

Correctly reports the value of the t -statistic

Correctly reports the p -value, consistent with the reported test statistic and stated alternative hypothesis

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.

Additional Notes:

· In order to satisfy the reference to the randomization condition in component 1, it is minimally acceptable to state "randomization – check." However, component 1 is not satisfied if the response implies that randomization was only used for one of the two types of discs or if the response implies that the discs were randomly sampled from a population of discs (because it is the order in which the discs were thrown that was randomized).

· The 10% condition (all possible throws of each type of disc > 150) does not need to be checked because the discs thrown were not randomly sampled from the population of all such discs. Because the inference conclusion is based on random ordering, it is a randomized experiment and not a random sample.

· Values for the test statistic and p -values depend on how the degrees of freedom were determined. The following combinations of t -values, p -values all satisfy components 3 and 4, respectively:

Statistics 2021 Intl MCQ

o The degrees of freedom for $t \approx 0.658$ from the model solution were computed with the Satterthwaite approximation used by many calculators:

$$df = \frac{\left(\frac{s_S^2}{n_S} + \frac{s_L^2}{n_L}\right)^2}{\frac{1}{n_S-1} \left(\frac{s_S^2}{n_S}\right)^2 + \frac{1}{n_L-1} \left(\frac{s_L^2}{n_L}\right)^2} = \frac{\left(\frac{(0.356)^2}{15} + \frac{(0.416)^2}{15}\right)^2}{\frac{1}{15-1} \left(\frac{(0.356)^2}{15}\right)^2 + \frac{1}{15-1} \left(\frac{(0.416)^2}{15}\right)^2} \approx 27.3$$

o Using $t \approx 0.658$ and $(15 - 1) = 14$ degrees of freedom, the resulting p -value is 0.521.

o Using a pooled t -test with $t \approx 0.658$ and 28 degrees of freedom (which would be appropriate because the standard errors for the two samples are similar), the pooled estimate of the standard error is 0.387 and the resulting p -value is 0.516.

· Discussion of degrees of freedom will be treated as extraneous in scoring for Section 2.

· If the response compares the value of the test statistic to a critical value instead of computing a p -value, then the critical value (2.051 for 27.3 df, 2.145 for 14 df, or 2.048 for 28 df), or a critical value consistent with the stated alternative hypothesis, satisfies component 4.

· If a one-tailed alternative hypothesis is stated, then the p -value must be consistent with the stated alternative hypothesis to satisfy component 4.

· A response that reports the correct value for the t -statistic but contains errors in supporting work may still satisfy component 3.

· If the response satisfies component 4, any supporting work for the p -value may be treated as extraneous.

Confidence Interval Approach:

o If the response correctly calculates a 95% two-sample t -interval for the difference between two population means, then both component 3 and component 4 are satisfied.

o Based on 27.3 degrees of freedom, the 95% confidence t -interval for the difference between two population means is

$$(\bar{x}_S - \bar{x}_L) \pm t_{(27.3), 0.975} \sqrt{\frac{s_S^2}{n_S} + \frac{s_L^2}{n_L}} = (4.173 - 4.080) \pm (2.051) \sqrt{\frac{(0.416)^2}{15} + \frac{(0.356)^2}{15}}$$

$$= (-0.196, 0.383)$$

o Based on $(15 - 1) = 14$ degrees of freedom, the 95% confidence t -interval for the difference between two population means is

$$(\bar{x}_S - \bar{x}_L) \pm t_{(14), 0.975} \sqrt{\frac{s_S^2}{n_S} + \frac{s_L^2}{n_L}} = (4.173 - 4.080) \pm (2.145) \sqrt{\frac{(0.416)^2}{15} + \frac{(0.356)^2}{15}}$$

$$= (-0.210, 0.396)$$

Based on a pooled estimate of the common standard deviation and 28 degrees of freedom, the 95% confidence t -interval for the difference between two population means is

$$(\bar{x}_S - \bar{x}_L) \pm t_{(28), 0.975} \sqrt{s_p^2 \left(\frac{1}{n_S} + \frac{1}{n_L} \right)} = (4.173 - 4.080) \pm (2.048) \sqrt{(0.387)^2 \left(\frac{1}{15} + \frac{1}{15} \right)}$$

$$= (-0.196, 0.383)$$

o If the response uses separate one-sample t -intervals for the individual means, which is not a correct approach, then component 3 is not satisfied but component 4 may be satisfied if both intervals are computed correctly.

Statistics 2021 Intl MCQ

Some commonly used percentiles of t -distributions are shown in the table:

	Upper Percentiles of t -distributions				
	0.90	0.95	0.975	0.99	0.995
$df = 14$	1.415	1.761	2.145	2.624	2.977
$df = 27.3$	1.313	1.702	2.051	2.471	2.768
$df = 28$	1.313	1.701	2.048	2.467	2.763
Confidence Level	80%	90%	95%	98%	99%

If the response incorrectly identified a two-sample z -test for a difference in population means, then Section 2 should be scored as follows:

- o Components 1 and 2 are scored as above.
- o Component 3 is satisfied if the response identifies the value of the test statistic as 0.658
- o Component 4 is satisfied if the response identifies a p -value of 0.5106.

Matched-Pairs t -test Approach:

Because it is unclear how to match, a response using a matched-pairs t -test or a matched-pairs confidence interval may be scored P, if it satisfies both of the following components:

- Component 1 is satisfied if the response checks the randomness condition by referring to the throwing of the discs in random order
- Component 2 is satisfied if the response includes a graph of DIFFERENCES in accuracy between the two types of discs that clearly labels the differences in accuracy or distance as the response variable AND provides a comment that does not rule out the application of an approximately normal model for the sampling distribution of DIFFERENCES.
- o Otherwise, the response is scored I.

Solution

Section 3:

Statistics 2021 Intl FRQ

Because the p -value is greater than any commonly used significance level, such as $\alpha = 0.05$ or $\alpha = 0.10$, the null hypothesis should not be rejected. The data do not provide convincing statistical evidence that mean distance from the target for all possible throws of a standard disc is not equal to the mean distance from the target for all possible throws of a lightweight disc, i.e., there is not convincing statistical evidence that the player tended to have more accuracy with one disc weight than with the other.

Scoring

Part (c) is scored as follows:

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

Provides a correct justification of the conclusion based on the size of the reported p -value (or a comparison of the value of the test statistic to an appropriate critical value)

States a correct conclusion about the two means consistent with the stated alternative hypothesis OR states a correct conclusion about the difference in means that answers the inference question.

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.

Additional Notes:

· Including proper context (identifying groups, the concept of population means, and referencing the response variable) is important in stating the conclusion. Context displayed in stating the conclusion is considered in scoring component 2 of Section 1.

· If the conclusion and justification are consistent with an incorrect p -value (or an incorrect value of the test statistic, or an incorrect confidence interval, or an incorrect critical value), the response may satisfy component 1 and component 2.

· If the statement of the conclusion is incorrect AND the justification is inconsistent with the stated conclusion, then neither component 1 nor component 2 is satisfied. For example, the response correctly indicates that the p -value is larger than any reasonable significance level and makes an incorrect conclusion that the null hypothesis should be rejected and there is sufficient evidence to conclude that the mean accuracies are different for the two types of discs. The statement about the p -value is correct, but the conclusion is incorrect and the conclusion is not justified by the statement about the p -value.

· If the response omits hypotheses, assume the correct alternative hypothesis, $H_a: \mu_S \neq \mu_L$, was provided when scoring component 2.

· If the response includes a statement that is equivalent to accepting the null hypothesis (e.g., “we conclude that the mean accuracies, or the mean distances from the target, are the same for the standard disc and the lightweight disc”), then component 2 is not satisfied.

· If the conclusion includes a definitive statement (such as “this proves that we do not have enough evidence to claim a difference” ...or “there is no evidence...”, then component 2 is not satisfied.

· If the response includes an incorrect interpretation of the p -value, then the score for Section 3 is lowered from E to P or from P to I.

· The conclusion may reference whether or not the types of discs *tend* to have different accuracy, or that “we cannot conclude that the two types of discs have different *tendencies*.” This can be read as “central tendency,” and therefore can be interpreted as the *population mean*.

Statistics 2021 Intl MCQ

· If the response incorrectly identified a two-sample z -test for a difference in population means, then section 3 should be scored using the reported p -value (or comparison of the z -value to a critical value obtained from the standard normal distribution).

· Confidence Interval Approach:

o If the alternative hypothesis is specified correctly as $H_a: \mu_S \neq \mu_L$, then component 1 is satisfied if the justification is based on whether 0 is included in the reported confidence interval.

o If no alternative hypothesis is specified in the response, then assume the correct alternative hypothesis is provided when scoring component 1.

o If the response includes an incorrect interpretation of the confidence interval, then the score for Section 3 is lowered from E to P or from P to I.

o If the response uses separate one-sample t -intervals for the individual means, which is not a correct approach, then component 1 is satisfied if the justification is based on whether the intervals overlap.

Rejection Region Approach:

o A response that uses a correct critical value in lieu of reporting and comparing the p -value to a significance level may satisfy component 1. Some commonly used percentiles of t -distributions are shown in the table:

	Upper Percentiles of t -distributions				
	0.90	0.95	0.975	0.99	0.995
$df = 14$	1.415	1.761	2.145	2.624	2.977
$df = 27.3$	1.313	1.703	2.051	2.471	2.768
$df = 28$	1.313	1.701	2.048	2.467	2.763

Matched-Pairs t -test Approach:

o Components should be scored according to the rubric using the reported p -value, or by comparing the t -value to an appropriate critical value.

Statistics 2021 Intl MCQ

5. 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.

The instructor of a training course for a certain coding language hopes that a student's score on a pretest will predict the student's score on the certification final taken at the end of the course. The final scores indicate proficiency with the coding language. A random sample of 12 students who completed the pretest and the training course was selected, and the final scores of the students were recorded. The following tables show summary statistics and computer output for a regression analysis of the sample.

	Mean	Standard Deviation
Pretest	54.833	10.408
Final	78.867	6.946

Predictor	Coef	SE	T	P
Constant	49.2804	6.9199	7.12	0.0000
Pretest	0.5396	0.1242	4.35	0.0015
$S = 4.2682$		$R\text{-sq} = 65.38\%$		$R\text{-sq(adj)} = 61.92\%$

(a) Identify the equation of the least-squares regression line for predicting final score from pretest score. Identify the variables that appear in the equation in context.

(b) Consider the information presented in the tables and your answer from part (a).

(i) Determine the value of the predicted final score for students with a pretest score of 65.

(ii) Interpret $S = 4.2682$ in the context of the problem.

For any particular pretest score, two point estimates can be calculated from the given simple linear regression analysis.

Point Estimate I: An estimate of the mean final score among all students with a particular pretest score who complete the course.

Point Estimate II: A prediction for the final score of one student with a particular pretest score who completes the course.

The values of both point estimates are equal to the predicted value of the least-squares regression equation for the particular pretest score, such as the estimate from part (b-i) for a pretest score of 65. Although the values of the two point estimates are the same, the sampling variability is not the same.

(c) Which of the two point estimates has the larger sampling variability? Explain your reasoning.

(d) The value calculated in part (b-i) is an estimate of the mean final score for all students with a pretest score of 65 who completed the course. If a list of the pretest scores and final scores for all students who completed the

Statistics 2021 Intl MCQ

course is obtained, explain how this information could be used to investigate the sampling distribution of the estimated mean final scores for all students with a pretest score of 65.

Part A, Part B, Part C, and Part D

Select a point value to view scoring criteria, solutions, and/or examples and to score the response

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4.
- The model solution represents an ideal response to each part of the question and the scoring criteria identify the specific components of the model solution that are used to determine the score.
- If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions.

Scoring

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

✓

0	1	2	3	4
---	---	---	---	---

Parts (a) through (d) sum to 4 points

OR

Parts (a) through (d) sum to $3\frac{1}{2}$ points AND a holistic approach is used to decide to score up

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect
- ☐ Part (d) essentially correct
- ☐ Part (d) partially correct
- ☐ Part (d) incorrect

Statistics 2021 Intl FRQ

Solution

Part (a):

The equation of the least-squares regression line for predicting final score from pretest score is

$\hat{y} = 49.2804 + 0.5396x$, where x is the pretest score and $\hat{y}$ is the predicted final score for students with a pretest score of x .

Scoring

Part (a) is scored as follows:

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

Identifies correct values for the slope and intercept in the equation of the least-squares regression line

Identifies the final score as the response variable AND pretest score as the explanatory variable

The concept of a “predicted value” or “estimated mean” is expressed either with words (e.g., “predicted final score”) OR using the $\hat{y}$ symbol in the equation

Partially correct (P) if the response satisfies component 2 AND component 3 but interchanges the values of the slope and intercept in the equation

OR

if the response satisfies component 1 AND only one of the other two components.

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

Additional Notes:

· In identifying the equation of the least-squares regression line, any symbols may be used for the variables.

· A response that reports an equation of the form $(\text{final score}) = 49.2804 + 0.5396(\text{pretest score})$ satisfies component 1 and component 2.

· A response that reports an equation of the form $(\text{predicted final score}) = 49.2804 + 0.5396(\text{pretest score})$ or $\widehat{\text{final score}} = 49.2804 + 0.5396(\text{pretest score})$ satisfies all three components and is scored E.

A response that reports an equation of the form $(\text{predicted final score}) = 0.5396 + 49.2804(\text{pretest score})$ satisfies the first set of criteria for P, but a response that reports an equation of the form $(\text{final score}) = 0.5396 + 49.2804(\text{pretest score})$ does not satisfy components 1 or 3 and is scored I.

Solution

Part (b):

(i) The value of the predicted final score for students with a pretest score of 65 is $49.2804 + 0.5396(65) \approx 84.354$.

(ii) The value $S = 4.2682$ is the standard deviation of the residuals from the regression for predicting final score from pretest score. When the regression line is used to predict final score, a typical prediction error would be about 4.2682.

Scoring

Part (b) is scored as follows:

Statistics 2021 Intl FRQ

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

1. Correctly calculates the value of the predicted final score for students with a pretest score of 65 in part (b-i) AND provides supporting work for the calculation of the value in part (b-i) using the equation identified in part (a)
2. Provides a correct interpretation in context of $S = 4.2682$ as a typical prediction error (difference between an observed and predicted final score) OR provides a correct definition in context of $S = 4.2682$ as the standard deviation (or estimate of the standard deviation) of the residuals.

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

OR

if the response to part (b-i) substitutes the value 65 into the correct equation, or an incorrect equation identified in part (a), but does not correctly calculate the value of the prediction AND provides a weak interpretation of $S = 4.2682$ in part (b-ii)

OR

if the response to part (b-i) correctly calculates the value of the prediction without supporting work AND provides a weak interpretation of $S = 4.2682$ in part (b-ii).

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

Additional Notes:

- Acceptable alternate wordings for interpretation of $S = 4.2682$ in part (b-ii):
- $S = 4.2682$ is the standard deviation (or the estimate of the standard deviation) of all final scores for students with a particular pretest score x
- A typical difference between an observed final score and the predicted final score is about $S = 4.2682$
- On average, we expect the difference between a predicted final score and the actual final score to be about $S = 4.2682$
- A difference between an observed final score and the predicted final score is about $S = 4.2682$
- A response that includes a correct definition of S in context and includes an interpretation of S that is considered weak still satisfies component 2
- An interpretation of $S = 4.2682$ is considered weak, if it exhibits at least one of the following characteristics:
- It does not reference final exam scores (no context)
- The interpretation refers to “prediction error” or “residual” without a qualifier such as “typical” or “on average”
- A response that includes a correct definition of S in context but includes an incorrect interpretation of S does not satisfy component 2.

Solution

Part (c):

Point Estimate II has the larger sampling variability. The sampling variability for Point Estimate I reflects how much the value on the least-squares regression line for a particular pretest score varies across all possible random samples of 12 students. The sampling variability for Point Estimate II is larger because it reflects student-to-student variability in final scores among all students with the same particular pretest score *in addition to* the sampling variability of Point Estimate I.

Scoring

Statistics 2021 Intl FRQ

Part (c) is scored as follows:

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

Identifies Point Estimate II as having the larger sampling variability

(A) Provides a valid explanation that includes the following two aspects:

- The sampling variability for both estimates includes the variation in the value on the least-squares regression line (the prediction) for a particular pretest score across all possible samples of 12 students

AND

- The sampling variability for Point Estimate II includes student-to-student variability in final scores among all students with the same particular pretest score

(B) Provides a valid explanation that includes the following two aspects:

- The sampling variability for an individual observation is greater than the sampling variability for an estimate of a mean

AND

- The response makes an explicit reference to the linear regression context.

Partially correct (P) if the response satisfies component 1 AND

- provides a weak explanation that satisfies only one of the two aspects of component 2(A)

OR

- includes only the first aspect of component 2(B).

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

Additional Notes:

- Examples of statements satisfying both aspects of component 2(B) are:
 - An estimated mean final score will tend to be closer to the fitted line than an individual predicted final score will be to the fitted line.
 - Sample means will not vary as much around a regression line as individual observations.

Solution

Part (d):

The sampling distribution of the conditional estimated mean final score can be simulated using the following steps:

Solution 1

1. Select a random sample of 12 students from the list of students who completed the course.
2. Compute a least-squares regression line using the information for the 12 students in the sample and use it to predict the final score for a student with a pretest score of 65.

Statistics 2021 Intl FRQ

3. Repeat the first two steps a large number of times, say M times.

Solution 2

1. From the list of all students who complete the course, create a second list of all students with a pretest score of 65. Select a random sample of 12 students from the second list.

2. Record the mean final score for the 12 students in the sample.

3. Repeat the first two steps a large number of times, say M times.

The resulting list of M values constitutes the simulated sampling distribution of estimated mean final score for all students with a pretest score of 65.

Scoring

Part (d) is scored as follows:

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

Describes selecting a random sample of 12 students from: the list of students who completed the course (Solution 1); OR from the list of all students with a pretest score of 65 (Solution 2).

Computes a least-squares regression line for the generated sample AND uses it to predict the final score for a student with a pretest score of 65 (Solution 1)

Calculates the mean final score of the 12 students in the generated sample (Solution 2)

Indicates that the steps described in components 1 and 2 should be performed a large number of times

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.

Additional Notes:

· In order to satisfy component 1, the response must include the following:

o Reference taking random samples of 12 students

AND

o Identify from what list the samples are being drawn.

· A response that describes taking all samples of size 12 (from either list) to construct the sampling distribution rather than random samples of size 12 has not addressed simulating the sampling distribution and therefore does not satisfy component 1, but does satisfy component 3.

· In order to satisfy component 2 using Solution 1, the response must indicate that a regression analysis is performed for the generated sample and used to compute the predicted final score for a student with a pretest score of 65. Note that the pretest score of 65 must be referenced.

· In order to satisfy component 2 using Solution 2, the response must indicate that the mean final score is calculated for the generated sample of 12 students.

Statistics 2021 Intl MCQ

6. 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.

A study was conducted with 50 volunteers to investigate whether yawning is contagious. Each volunteer was randomly assigned to one of two groups: Group A or Group B. Each person in Group A was shown a video in which someone yawned, and each person in Group B was shown a video in which no one yawned. Of the 34 people assigned to Group A, 11 yawned, and of the 16 people assigned to Group B, 3 yawned.

(a) Is this study an observational study or an experiment? Explain your choice in context.

Let p_A represent the proportion of all people who would yawn when shown the video in which someone yawned, and let p_B represent the proportion of all people who would yawn when shown the video in which no one yawned.

(b) Explain why it is inappropriate to use a two-sample z-test for proportions to test the following hypotheses.

$$H_0 : p_A = p_B$$

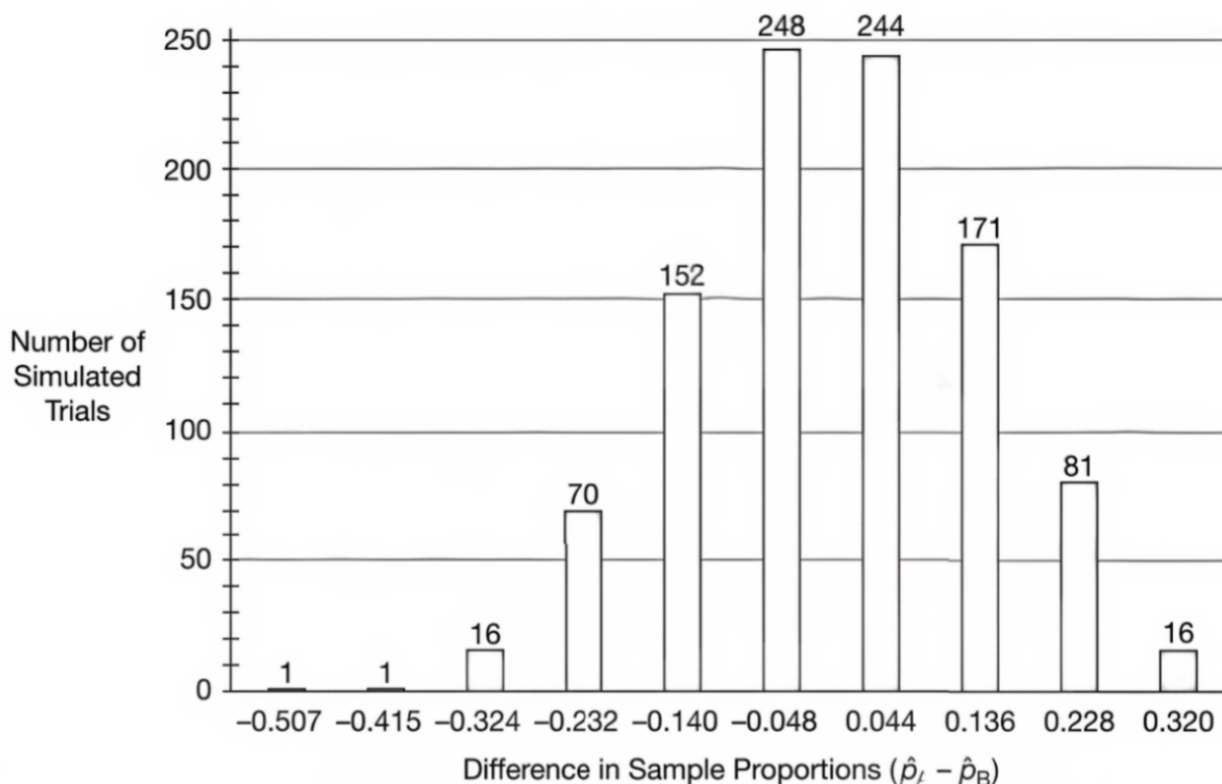
$$H_a : p_A > p_B$$

(c) Based on the study data, determine the point estimate for $p_A - p_B$. Show your work.

(d) A simulation to approximate the sampling distribution of the difference in sample proportions $\hat{p}_A - \hat{p}_B$ was conducted under the assumption that yawning is not contagious.

The graph shows the results of 1,000 trials of such a simulation.

Statistics 2021 Intl MCQ



(i) Based on your answer in part (c) and the simulated results, what is the approximate p -value for the test stated in part (b)? Show your work.

(ii) Based on your answer to part (d-i) and using a significance level of 0.05, what conclusion is appropriate about whether yawning is contagious? Justify your answer.

Part A, Part B, Part C, and Part D

Select a point value to view scoring criteria, solutions, and/or examples and to score the response

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4.
- The model solution represents an ideal response to each part of the question and the scoring criteria identify the specific components of the model solution that are used to determine the score.
- If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions.

Scoring

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

Statistics 2021 Intl MCQ



0	1	2	3	4
---	---	---	---	---

Parts (a) through (d) sum to 4 points

OR

Parts (a) through (d) sum to $3\frac{1}{2}$ points AND a holistic approach is used to decide to score up

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect
- ☐ Part (d) essentially correct
- ☐ Part (d) partially correct
- ☐ Part (d) incorrect

Solution

Part (a):

This study is an experiment. The researchers imposed a treatment (being shown a video in which someone yawned, or being shown a video in which no one yawned) on each volunteer in the study. Moreover, since the volunteers were randomly assigned to the treatments, this is a completely randomized experiment.

Scoring

Part (a) is scored as follows:

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

Identifies the study as an experiment (or completely randomized experiment)

Provides a valid explanation that indicates that the researchers imposed the treatments

Provides sufficient context, by including reference to the subjects (volunteers) AND the two treatments (being shown a video in which someone yawned and being shown a video in which no one yawned; or Group A and Group B)

Statistics 2021 Intl FRQ

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.

Additional Notes:

· A response may satisfy component 2 by indicating that the volunteers were assigned (or randomly assigned), or not allowed to select their treatment to the two treatment groups.

A response that provides an explanation that only refers to the use of randomization but does not also refer to the imposition of treatments on the volunteers (e.g., “this is an experiment because the researchers used randomization”) does not satisfy component 2. However, a response that specifically refers to the researchers’ use of “random assignment” (e.g., “this is an experiment because the researchers used random assignment”) may satisfy component 2 but may not satisfy component 3 without sufficient context.

Solution

Part (b):

In order to use the two-sample z-test for a difference in population proportions to test the given hypotheses, the sample sizes must be sufficiently large enough to ensure that the sampling distribution of $\hat{p}_A - \hat{p}_B$ is approximately normal. In other words, for the combined sample, the combined (or pooled) proportion $\hat{p}_c = \frac{n_A\hat{p}_A + n_B\hat{p}_B}{n_A + n_B}$ is used to verify that each of $n_A\hat{p}_c$, $n_A(1 - \hat{p}_c)$, $n_B\hat{p}_c$, and $n_B(1 - \hat{p}_c)$ are all greater than or equal to 10 (or 5). Since there were only 16 volunteers in Group B and the combined (or pooled) proportion of volunteers who yawned was $\hat{p}_c = \frac{11+3}{34+16} = \frac{14}{50} = 0.28$, the expected number of volunteers who would yawn in Group B is $n_B\hat{p}_c = (16)(0.28) = 4.48$. Because this expected count is less than 10 (or 5), the sample sizes are not sufficiently large enough to ensure that the sampling distribution of $\hat{p}_A - \hat{p}_B$ is approximately normal.

Scoring

Part (b) is scored as follows:

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

Explains that a condition for using the two-sample z-test for a difference in population proportions is that:

· The sampling distribution of $\hat{p}_A - \hat{p}_B$ is approximately normal
OR

· $n(\hat{p}_c)$ and $n(1 - \hat{p}_c)$ should be greater than 10 (or 5) for BOTH groups

Correctly calculates the expected number of volunteers who would yawn in Group B as $(16)(0.28) = 4.48$

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.

Additional Notes:

The second bullet of component 1 may be satisfied by indicating that the observed counts for BOTH groups, $n_A\hat{p}_A$, $n_A(1 - \hat{p}_A)$, $n_B\hat{p}_B$, and $n_B(1 - \hat{p}_B)$, must all be greater than 10 (or 5).

· Component 1 is not satisfied if the response indicates that it is checking for some condition other than the normality condition for the sampling distribution of $\hat{p}_A - \hat{p}_B$ (e.g. checking for independence). Other references to independence are considered extraneous.

Statistics 2021 Intl MCQ

- Component 2 may be satisfied by identifying Group B as the group with an observed count of 3.

Solution

Part (c):

The point estimate for $p_A - p_B$ is the estimated difference in proportions, or the observed difference in the sample proportions, $\hat{p}_A - \hat{p}_B = \frac{11}{34} - \frac{3}{16} \approx 0.136$.

Scoring

Part (c) is scored as follows:

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

Correctly calculates the point estimate for $p_A - p_B$ as 0.136

Provides supporting work for the calculation of the point estimate for $p_A - p_B$

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

OR

if the response shows the correct proportions but reverses the order of the subtraction resulting in an incorrect value of the point estimate of -0.136 .

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

Additional Notes:

- $\frac{11}{34} - \frac{3}{16} = \frac{74}{544}$ or $\frac{11}{34} - \frac{3}{16}$ or $0.3235 - 0.1875$ satisfies both component 1 and component 2.

· When calculating the value 0.136, ignore minor arithmetic errors or transcription errors if they can be identified by the work shown

Solution

Part (d):

(i) Using the point estimate for $p_A - p_B$ from part (c) of 0.136 and the simulated results, the approximate p -value for the test stated in part (b) is $0.171 + 0.081 + 0.016 = 0.268$.

(ii) Because the p -value from part (d-i) is greater than the significance level of 0.05, the null hypothesis from part (b) $H_0 : p_A = p_B$ should not be rejected. There is not convincing statistical evidence that the proportion of all people who would yawn when shown the video in which someone yawned is greater than the proportion of all people who would yawn when shown the video in which no one yawned. Therefore, the study does not provide convincing evidence that yawning is contagious.

Scoring

Part (d) is scored as follows:

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

Correctly calculates the p -value in part (d-i)

Provides appropriate supporting work for the calculation of the p -value in part (d-i)

Statistics 2021Intl FRQ

Indicates that the data do not provide sufficient evidence to conclude that yawning is contagious.

Provides a correct justification for the conclusion based on the size of the p -value calculated in part (d-i).

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.

Additional Notes:

- A response that excludes the bar from the graph at 0.136 in the calculation of the p -value in part (d-i) does not satisfy component 1 since the p -value is the probability of obtaining a test statistic as extreme or more extreme than the observed test statistic, when the null hypothesis and probability model are assumed to be true. However, such a response may satisfy component 2 if the response provides the supporting work $0.081 + 0.016 = 0.097$.

- A response that rounds the decimal approximations of the proportions from the graph to two places (e.g., $0.17 + 0.08 + 0.02 = 0.27$) in the calculation of the p -value in part (d-i) may still satisfy component 1 and component 2.

- A response that incorrectly calculates the p -value in part (d-i) may still satisfy components 2 through 4 if the response is consistent with the use of the incorrect p -value.