Math 1530 Final Exam Spring 2019

Name ______________________
Section # ______________________
Instructor ______________________

There are five possible responses to each of the following multiple choice questions. There is only one "BEST" answer. Be sure to read all possible choices before selecting your answer. You may mark on this examination. You can use a calculator but a calculator manual cannot be used.

Form B

Please circle your answer to each question and fill the blank sheet. After you finish the exam, log into D2L and input your answers under the right quiz item. There are a total of 43 questions.

- If your test is Form A, take Quiz item *Final Form A*.
- If your test is Form B, take Quiz item *Final Form B*.
- If your test is Form C, take Quiz item *Final Form C*.
- If your test is Form D, take Quiz item *Final Form D*.

CONFIDENCE LEVEL	90%	95%	99%
z^*	1.645	1.96	2.576

1. Which of these questions from the Spring 2019 MATH1530 class survey produced variables that are quantitative?
   i. What is your age (in years)?
   ii. What is your race?
   iii. Do you prefer to read textbooks in print or on an electronic device (such as a smart phone, tablet, computer or e-reader)?
   iv. How much more money would you be willing to spend for a hard copy textbook instead of an electronic version of a textbook (in whole dollars)?
   v. Have you ever driven an automobile?

   (A) i. iv. (B) ii. iii. v. (C) i. (D) i. ii. iii. iv. v. (E) None

   Use the following for the next 3 questions. A student survey asked "What is the fastest you have driven an automobile on the highway or interstate (in mph)? " Here is the data:

   100 85 160 102 175 90 85 21 90

2. The mean of the sample is
   (A) 175 mph. (B) 132.5 mph. (C) 102 mph. (D) 100.9 mph. (E) 90 mph.

3. The standard deviation of the sample is 44.88 mph. If the standard deviation is computed without the 21 mph, the standard deviation of the eight speeds would be
   (A) greater than 44.88 mph. (C) equal to 44.88 mph.
   (B) less than 44.88 mph. (D) equal to 23.88 mph.
   (E) equal to 1276.13 mph.

4. If the speeds were recorded in kilometers per hour (kph) then the standard deviation of the nine observations would be (Note: 1 mph = 1.609 kph)
   (A) 116.186 kph. (B) 72.21 kph. (C) 46.489 kph. (D) 44.88 kph. (E) 27.89 kph.

5. If a distribution is skewed to the right,
   (A) the mean and median are equal. (C) the mean is greater than the median.
   (B) the mean is less than the median. (D) the standard deviation is small.
   (E) the standard deviation is large.

6. A _________________ is a number that can be computed from the sample data without making use of any unknown parameters. In practice, we often use a _________________ to estimate an unknown parameter.

   (A) statistic, parameter (C) mean, normal curve.

   (B) population, population (D) parameter, parameter

   (E) statistic, statistic

7. The _________________ is robust to outliers while the _________________ is not robust to outliers.

   (A) mean, standard deviation (C) standard deviation, median

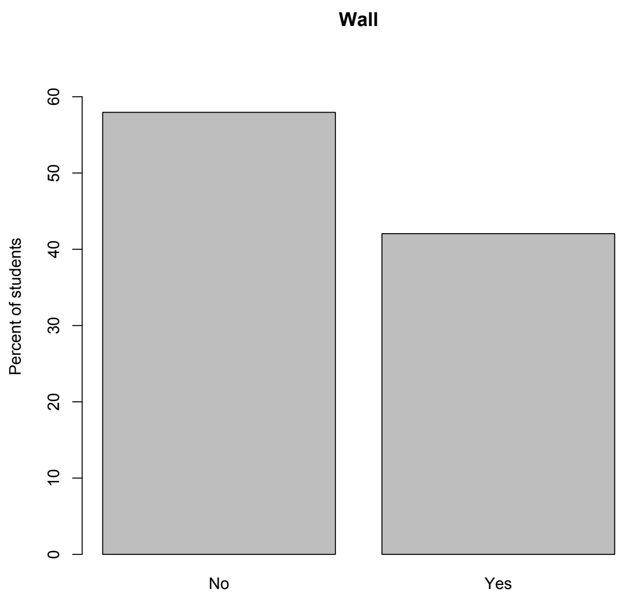
   (B) standard deviation, mean (D) mean, median

   (E) median, mean

8. The Spring 2019 MATH1530 survey asked "If you had to lose one of your five senses, which would you choose?" The responses from 629 students were: Smell 390; Taste 113; Touch 73; Hearing 29; Sight 24. To display this data you should construct a

   (A) time plot. (B) 5-number summary. (C) bar graph. (D) histogram. (E) stemplot.

9. The Spring 2019 MATH1530 survey asked "Do you support the idea of building a border wall between the U.S. and Mexico?" The following graph displays the information on this variable:



Which of the following best describes this distribution?

(A) The distribution is skewed right, the center is between 'No' and 'Yes.'
(B) The distribution is irregular and skewed right.
(C) The distribution is uniformly spaced between 'No' and 'Yes'.
(D) The distribution is bimodal and skewed right.
(E) About 40% of the students in the survey support building a border wall between U.S. and Mexico.

10. The Spring 2019 MATH1530 class survey asked “How many speeding tickets have you received?” The table below represents the responses of 621 students.

Number of Speeding Tickets	0	1	2	3 or more
Number of Students	388	139	50	44

The median number of speeding tickets for these 621 students is

(A) 311 (B) 94.5 (C) 50% (D) 1.5 (E) 0

Use the following for the next 2 questions The following data represent scores (out of 100 points) of ALL 50 students in a math course that took a test.

51	53	56	57	57	58	59	61	61	61	64	65	65	65	66
67	68	68	68	69	70	71	72	72	72	72	72	73	73	74
74	74	74	75	75	76	76	77	78	79	80	81	81	82	83
84	84	87	90	93										

11. Which of the following best describes the distribution of the math test scores?

(A) The distribution of test scores is fairly symmetric. The center is close to 72 and the test scores range from 51 to 93.

(B) The distribution is multimodal with no outliers. There are many centers and the variability in the test scores is quite small.

(C) The test scores are evenly spaced from 51 to 93.

(D) The distribution is single-peaked, right skewed, and has possible outliers.

(E) The distribution is single-peaked, left skewed, and has possible outliers.

12. The mean and standard deviation of all 50 test scores is 71.26 and 9.525, respectively. A 95% confidence interval for the mean score on this test is found to be 71.26 ± 2.71. Comment on this result.

(A) We can be 95% confident that the mean test score for all 50 students is 71.26.

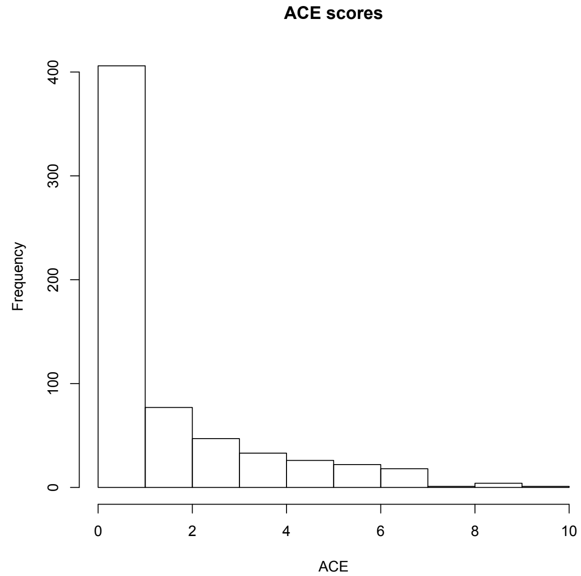
(B) 95% of all the scores of the 50 students lies between 71.26 ± 2.71.

(C) We are 95% confident that the mean test score for all 50 students lies between 68.55 to 73.97.

(D) The confidence interval makes no sense since we have the entire population of test scores.

(E) A larger sample should be taken to reduce the margin of error.

Use the following for the next 3 questions. The Spring 2019 MATH 1530 survey asked "What was your Adverse Childhood Experiences (ACE) quiz score?" The histogram below represents the responses of 635 students.

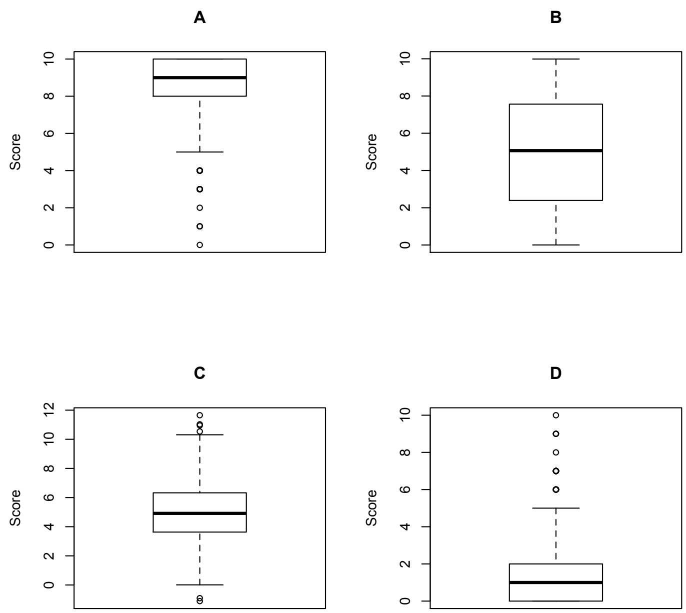


13. Which summary statistics would you use to summarize these data? Why?

   (A) The 5-number summary since it is strongly skewed.
   (B) The mean and standard deviation because the distribution is unimodal and symmetric.
   (C) The mean since it describes the center of the distribution for skewed distributions.
   (D) The mean or median since they would be approximately equal.
   (E) The mean and the interquartile range (IQR) because the distribution is symmetric with no outliers.

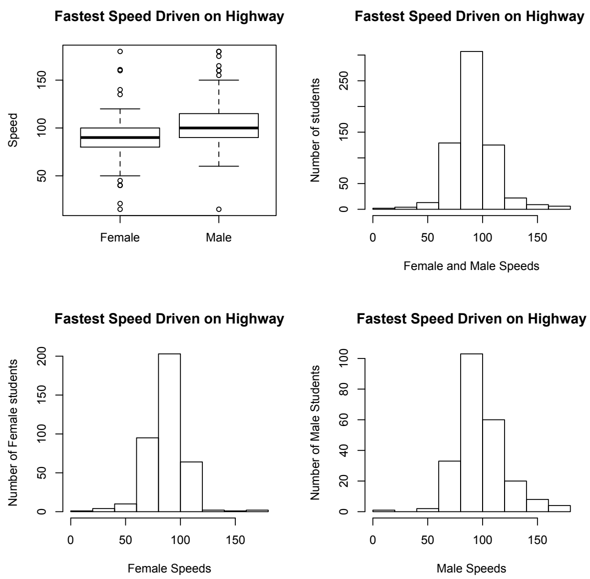
14. Approximately what percent of the students have an ACE score between 2 and 3?
   (A) 75% (B) 50% (C) 8% (D) 2.5% (E) 0%

15. Which of the following boxplots corresponds to the above histogram?



16. An insurance company is interested in comparing how fast males and females have driven an automobile on the highway or interstate (in mph)? Here are the summary statistics and graphs of the fastest speed driven by females and males from a survey.

GENDER	n	Mean	Standard Deviation	Minimum	Q1	Median	Q3	Maximum
FEMALE	382	90.96	17.16	15	80	90	100	180
MALE	231	102.7	21.82	15	90	100	115	180



Which of the following describes and compares the distributions of fastest speed driven by males and females?

(A) The distribution of speeds is symmetric around 95 mph with spread from 15 mph to 180 mph. There are outliers present and the five-number summary of all the speeds should be used.

(B) The five-number summaries could be used to describe both distributions since there are outliers. Both distributions are somewhat symmetric with the males tending to have driven the fastest.

(C) Both distributions are strongly skewed right with outliers. The male and female speeds range from 15 mph to 180 mph.

(D) Both distributions are uniformly distributed from 15 mph to 180 mph with outliers. The males drive faster than the females.

(E) Both distributions are somewhat symmetric about their respective medians with a number of outliers. We shouldn't compare the two distributions since the sample sizes are not equal.

Use the following for the next 2 questions. Assume that cholesterol levels of adult American women can be described by a Normal model with a mean of 180 mg/dL and a standard deviation of 25.

17. What percent of adult women do you expect to have cholesterol levels between 155 and 205 mg/dL?

(A) 100% (B) 99.7% (C) 95% (D) 68% (E) 50%

18. Below what value are the lowest 5% of the women's cholesterol levels?

(A) 131 (B) 139 (C) 212 (D) 221 (E) 229

19. How does drinking beer affect the level of alcohol in our blood? Student volunteers at Ohio State University drank different numbers of cans of beer. Thirty minutes later, a police officer measured their blood alcohol content. In this study, which is the explanatory variable and which is response variable?

(A) Thirty minutes is the explanatory variable, and the police officer is the response variable.

(B) Student volunteers is the explanatory variable, and the number of cans of beers they drank is the response variable.

(C) Number of beers consumed is explanatory variable, and percent of alcohol in the blood is the response variable.

(D) Thirty minutes is the explanatory variable, and the number of beers consumed is the response variable.

(E) Percent of alcohol in the blood is explanatory variable, and number of beers consumed is the response variable.

20. A study of consumer behavior finds a positive correlation between sales of ice cream (x) and sales of beer (y). What is a plausible explanation for the observed correlation?

    (A) Both x and y are changing with the lurking variable outdoor temperature.

    (B) Ice cream creates a thirst for beer.

    (C) The possible correlation can only be the result of an arithmetic mistake.

    (D) People generally have ice cream for dessert if they have drunk beer with a meal.

    (E) As ice cream sales go up beer sales go down and vice versa.

Use the following for the next 4 questions. Suppose you were to collect data for each pair of variables listed in I-V.

I: The number of hours a plane is in flight and the number of miles flown.
II: The number of hours in flight of a plane and the gallons of fuel remaining.
III: The shoe size and grade point average of college freshmen.
IV: The age and grip strength of adults.
V: The weight of apples in grams and ounces.

21. Which of the data sets would you expect to see a positive association?

    (A) II, IV (B) I,V (C) IV (D) III (E) I

22. Which of the data sets would you expect to see a negative association?

    (A) III, V (B) II, IV (C) V (D) II (E) I

23. Which of the data sets would you expect to see very little relationship or no association?

    (A) V (B) IV (C) III (D) II (E) I

24. Which of the data sets would you expect to see a perfect positive linear association?

    (A) V (B) IV (C) III (D) II (E) I

25. The deadly Ebola virus is a threat to both people and gorillas in Central Africa. An outbreak in 2002 and 2003 killed 91 of the 95 gorillas in 7 ranges in the Congo. To study the spread of the virus, measure "distance" by the number of home ranges separating a group of gorillas from the first group infected. Here are the data on distance and time in number of days until deaths began in each later group:

Distance	1	3	4	4	4	5
Time	4	21	33	41	43	46

Which of the following best describes the relationship between distance and time?

    (A) The correlation is probably close to 0 indicating there is not a linear relationship between distance and time.

    (B) The side-by-side boxplots show that time is higher than distance.

    (C) The median time is larger than the median distance.

    (D) There is a strong positive linear relationship between distance and time.

    (E) There is no relationship between distance and time.

26. You can summarize the data for two categorical variables x and y by

    (A) drawing a scatterplot of the x- and y-values.

    (B) calculating the correlation between x and y.

    (C) constructing a box plot for each variable.

    (D) calculating the least-squares regression line.

    (E) constructing a two-way table for the x- and y-values.

27. You would like to start a club for psychology majors on campus, and you are interested in finding out what proportion of psychology majors would join. The dues would be $35 and used to pay for speakers to come to campus. You ask 5 psychology majors from your senior psychology honors seminar whether they would be interested in joining this club and find that 4 of the 5 students questioned are interested. Are there any potential sources of bias and what conclusions can be drawn from this survey?

    (A) Yes. Since all the students surveyed are enrolled in a special senior honors class, these students may be more likely to be interested in joining the club (and more likely to pay $35 to do so). The direction of the bias is likely to overestimate the proportion of all psychology majors willing to pay to join this club. This is a convenience sample.

    (B) No, but we would need to know the margin of error before we can draw a conclusion from this survey.

    (C) No. We estimate that $80\% \pm 36\%$ of psychology majors on campus would be willing to join a new club.

    (D) No. The results will produce a yes or no response and the binomial distribution can be used to draw conclusions about the population with $n = 5$ and $p = .8$.

    (E) No. Since the sample size is small we are able to draw conclusions about the population. Hence, the club for psychology majors should be started since 4 out of 5 students said yes.

28. Researchers studied a group of 10,892 middle-aged adults over a period of nine years. They found that smokers who quit had a higher risk for diabetes within three years of quitting than either nonsmokers or continuing smokers. Does this show that stopping smoking causes the short-term risk for diabetes to increase?

    (A) Based on this research, we should tell a middle-aged adult who smokes that stopping smoking can cause diabetes and advise him or her to continue to smoke.

    (B) This is an observational study, so it not reasonable to conclude any cause-and-effect relationship.

    (C) Since the sample is large, it is reasonable to conclude that stopping smoking will increase risk for diabetes.

    (D) Since this is an experiment with a large sample, we can draw a strong conclusion that stopping smoking will cause an increase in short-term diabetes.

    (E) We would need to know if the results are statistically significant or not to make a cause-and-effect claim.

29. Archaeologists plan to examine a sample of two-meter-square plots near an ancient Greek city for artifacts visible in the ground. They choose separate samples of plots from floodplain, coast, foothills, and high hills. This is an example of

    (A) a simple random sample.
    (C) an observational sample.
    (B) a voluntary response sample.
    (D) a stratified random sample.
    (E) an experiment.

30. Which of the following questions does a test of significance answer?

    (A) Is the observed effect important?
    (C) Is the observed effect due to chance?
    (B) Is the sample or experiment properly designed?
    (D) Is the null hypothesis true?
    (E) Is it based on a very small random sample?

31. The most important condition for sound conclusions from statistical inference is usually

    (A) that the population distribution is exactly Normal.

    (B) that the data contain no outliers.

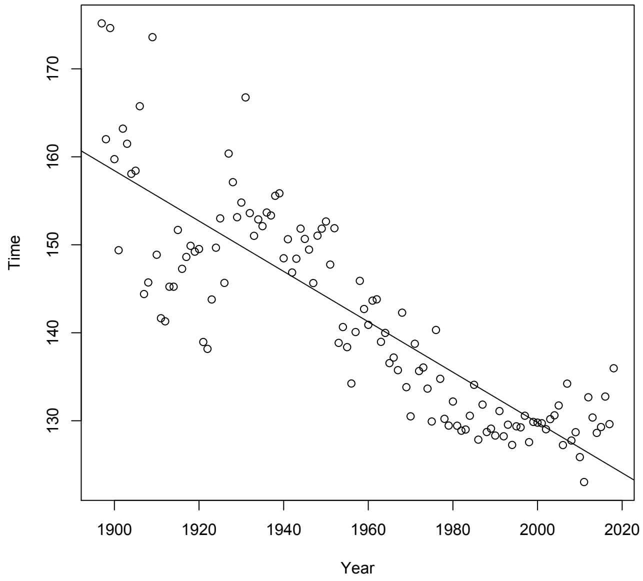
    (C) that the mean and standard deviation of the population are known.

    (D) that the p-value is smaller than 0.05.

    (E) that the data can be thought of as a random sample from the population of interest.

Use the following for the next 3 questions. Data has been collected on world record times (in minutes) for men running a marathon. The observations and least-squares regression appear in the scatterplot. The correlation between the two variables is $r = -0.85$ and the least-squares regression equation is

$$Time = 702.5 - 0.286 \times Year.$$



32. Which of the following statements is true regarding the data?

    (A) There is a strong negative linear relationship between year and record winning time (in minutes) for men running a marathon.

    (B) There is a strong positive linear relationship between year and record winning time (in minutes) for men running a marathon.

    (C) The relationship is weak between year and record winning time since the correlation coefficient is negative.

    (D) There is no association between year and record winning time.

    (E) The record winning times are scattered from 1900 to 2019.

33. Which is the most appropriate interpretation of the slope?

    (A) For every year increase, the record winning time increased by 0.286 minutes.

    (B) For every year increase, the record winning time decreased by 0.286 minutes.

    (C) For every minute decrease, years decreased 0.286.

    (D) For every year increase, the record winning time increased by 702.5 minutes.

    (E) For every year increase, the record winning time decreased by $r = -0.85$ minutes.

34. A researcher wants to predict a record winning time in 2050. Which of the following is correct?

    (A) The predicted record winning time in 2050 will be 100 minutes.

    (B) The predicted record winning time in 2050 will be 1289 minutes.

    (C) The predicted record winning time in 2050 will be 116.2 minutes.

    (D) The predicted winning record time in 2050 will be 702.2 minutes.

    (E) This is extrapolation and we shouldn't use the least-squares regression line for prediction for years that are far beyond 2019.

35. Which of the following statement(s) is TRUE?

    i. Because large random samples have small chance variation, very small population effects can be highly significant if the sample is large.

    ii. Statistical significance is the same thing as practical significance.

    iii. Because small random samples have a lot of chance variation, even large population effects can fail to be significant if the sample is small.

    iv. Statistical significance means the sample showed an effect larger than would often occur just by chance.

    (A) i. ii. iii. iv. (B) iv. (C) i. iii. iv. (D) ii. (E) ii. iv.

36. A medical experiment compared the herb echinacea with a placebo for preventing colds. One response variable was "volume of nasal secretions" (if you have a cold, you blow your nose a lot). Take the average volume of nasal secretions in people without colds to be $\mu = 1$. An increase to $\mu = 3$ indicates a cold. The significance level (α) of a test of $H_0 : \mu = 1$ versus $H_a : \mu > 1$ is defined as

    (A) the probability that the test fails to reject H_0 when $\mu = 3$ is true.

    (B) the probability that the test does not reject H_0 when $\mu = 1$ is true.

    (C) the probability that H_0 is false.

    (D) the probability that the test rejects H_0 when $\mu = 1$ is true.

    (E) the probability that the test rejects H_0 when $\mu = 3$ is true.

37. Motor vehicles sold in the United States (ignoring heavy trucks) are classified as either cars or light trucks and as either domestic or imported. "Light trucks" include SUVs and minivans. "Domestic" means made in Canada, Mexico, or the United States, so that a Toyota made in Canada counts as domestic. In 2010, 76% of the new vehicles sold to individuals were domestic, 50% were light trucks, and 43% were domestic light trucks. Choose a vehicle sale at random. What is the probability for choosing a vehicle that is a domestic or light truck?

    (A) 0.38 (B) 0.43 (C) 0.57 (D) 0.83 (E) 1.26

38. The Spring 2019 MATH1530 class survey asked "Do you prefer to read textbooks in print or on an electronic textbook?" In the sample survey 480 students responded "In print" and 155 students responded "On an electronic device." Assume that this sample represents all college students. A large-sample 95% confidence interval for estimating the proportion of all college students who prefer a printed textbook is

    (A) 72.2% to 78.9% (B) $76\% \pm 5\%$ (C) 74% to 77% (D) 31.1% to 37.8% (E) 76%

Use the following for the next 3 questions. Two questions asked on a survey were "What gender do you identify with?" and "Would you identify yourself as pro-choice or pro-life on abortion?" Here is a two-way table of the results:

	Abortion		
Gender	**Pro-Choice**	**Pro-Life**	**Total**
Female	205	186	391
Male	121	118	239
Other	4	1	5
Total	330	305	635

39. What percent of the students are Pro-Choice?

    (A) 61.57% (B) 51.97% (C) 50.63% (D) 36.67% (E) 19.06%

40. What percent of students are Male given they are Pro-Choice?

    (A) 61.57% (B) 51.97% (C) 50.63% (D) 36.67% (E) 19.06%

41. What percent of students are Pro-Choice given they are Male?

    (A) 61.57% (B) 51.97% (C) 50.63% (D) 36.67% (E) 19.06%

Use the following for the next 2 questions. In a recent student survey, males and females were asked "At what age do you expect to retire?" Assume the students who responded to the survey represent an SRS of all undergraduate students. Do males and females plan to retire at the same age? A two-sample t-test was performed using MINITAB:

Descriptive Statistics: Retire

Gender	N	Mean	StDev	SE Mean
Female	391	63.26	8.06	0.41
Male	239	61.14	11.99	0.78

Estimation for Difference

Difference	95% CI for Difference
2.12	$(0.396, 3.841)$

Test

Null hypothesis $H_0 : \mu_1 - \mu_2 = 0$

Alternative Hypothesis $H_1 : \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
2.418	370.21	0.016

42. Does the analysis provide sufficient evidence at the significance level $\alpha = 0.05$ to conclude that, on the average, females and males plan to retire at different ages? Why?

(A) No, since the p-value = 0.016 is less than $\alpha = 0.05$ we fail to reject H_0 and conclude that females and males plan to retire about the same age.

(B) No, there is weak evidence that the retirement ages are different since the p-value is small.

(C) Yes, there is strong evidence since the mean retirement ages for females and males are 63.26 and 61.14, respectively.

(D) Yes, there is both practical significance and statistical significance to suggest that all males will retire at an older age than females will.

(E) Yes, there is evidence (p-value = 0.016) that, on average, females and males plan to retire at different ages.

43. Which of these conclusions about the confidence interval are correct?

(A) We are 95% confident that females plan to retire anywhere from 0.396 to 3.841 years later than males on the average.

(B) We are 95% sure that the difference between the planned retirement ages of all females and males is between 0.396 and 3.841 years.

(C) 95% of all random samples are between 0.396 and 3.841 years.

(D) We are 95% confident the difference in the sample planned retirement means is between 0.396 and 3.841.

(E) There is a 95% chance that the difference in means is between 0.396 and 3.841 years.