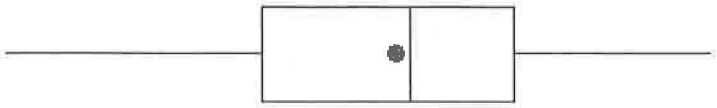
MA 1530 Final

Name _____

Questions 1 through 4, where the large dot represents the mean:



Boxplot 1

Boxplot 2

1. Boxplot 1 has a smaller range than Boxplot 2

a	b	c
False	True	cannot tell

2. Boxplot 1 is probably skewed right

a	b	c
False	True	cannot tell

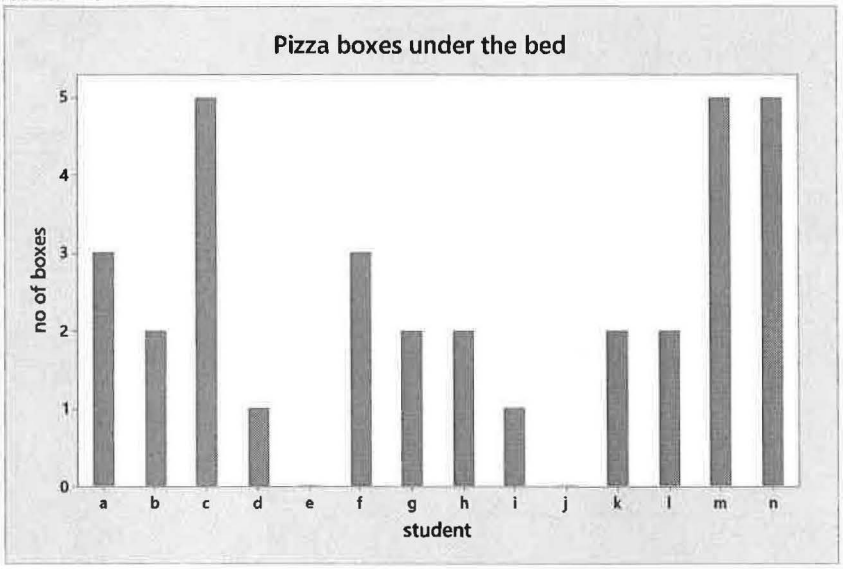
3. Boxplot 1 has a greater median than Boxplot 2

a	b	c
False	True	cannot tell

4. Boxplot 1 has a larger interquartile range than Boxplot 2

a	b	c
False	True	cannot tell

- Question 5 through 8: The number of pizza boxes college students have under the bed is represented below



5. What is the probability that a student chosen at random from this group had at most two boxes?

a	b	c	d	e
9	10/14	9/14	4/14	7/14

6. What proportion of this group of students had no boxes?

a	b	c	d	e
1/7	1/6	1/14	0	1/5

7. What is the mode?

a	b	c	d	e
No mode	0, 1, and 3	2 and 5	2 only	5 only

8. What is the average number of boxes per student?

a	b	c	d	e
1.26	2.47	2.33	2	2.357

9. The number of M&Ms in a bag is approximately a normal random variable with a mean of **510** and standard deviation of **5.50**. Garth bought a bag with **490** M&Ms. Does Garth's bag have an unusually low number of M&Ms?

a	No, because 490 is only 20 M&Ms below average
b	Yes because 490 is more than 3 standard deviations below average
c	Yes, because 490 is more than 20 standard deviations below average
d	No, because 490 is more than 3 standard deviations below average
e	No, because 490 is 96% of average

10. Choose the true statement

a	If you toss a fair coin 10 times and observe 10 heads, you are very likely to observe a tail on toss #11
b	If you toss a fair coin 10 times and observe 10 heads, you are very likely to observe a head on toss #11
c	If you toss a coin 5 times and get 5 heads, the coin cannot possibly be fair
d	If you toss a fair coin 10,000 times, the proportion of heads you observe will probably be closer to 0.5 than the proportion of heads you observe after just 50 tosses
e	If you toss a fair coin 50 times, you will get 25 heads and 25 tails

11. If a fair die is tossed 4 times, what is the probability of observing exactly 3 fives? Hint: n = ?, p = ?

a	b	c	d	e
0.4823	0.3858	0.0154	0.1157	0.0008

12. From a sample of **200** automobiles, 5% have bad tires. This means…

a	5 cars in 200 have bad tires
b	5 cars in 20 have bad tires
c	1 car in 20 has bad tires
d	1 car in 5 has bad tires
e	1 car in 10 has bad tires

- Question 13 and 14: Sally drove two different routes to work five times each. Her trip times are recorded below in minutes. Compute the means and sample standard deviations for the times for the 2 different routes and answer the following questions

Route 1	15.5	16	16.25	14.75	17
Route 2	14	15	15.5	18	16

13. Which route takes less time on average?

a	b	c	d	e
Same	None of these	Route 1	Route 2	Route 3

14. Which route seems to have less variation in time?

a	b	c	d	e
Same	None of these	Route 1	Route 2	Route 3

- For questions 15 through 20: The number of cups of coffee sold at the APSU library on 9 consecutive days are

10	79	85	85	86	94	98	110	122

15. What is the mean?

a	b	c	d	e
29.64	86.00	85.00	31.43	85.44

16. What is the median?

a	b	c	d	e
31.43	86.00	29.64	85.44	85.00

17. What is the sample standard deviation?

a	b	c	d	e
85.44	29.64	31.43	85.00	86.00

18. What is the range?

a	b	c	d	e
22.00	112.00	137.00	120.00	49.00

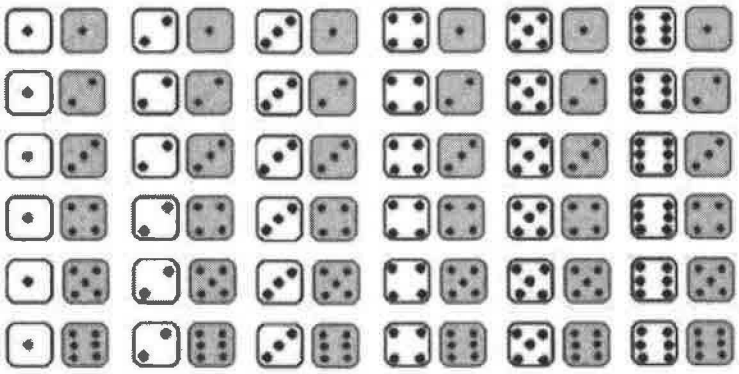
19. What is the 3rd quartile?

a	b	c	d	e
137.00	120.00	112.00	104.00	22.00

20. Which of the 9 data points is an outlier? Use the "1.5 x IQR" rule.

a	b	c	d	e
49.00	122.00	10.00	137.00	There is no outlier

- For questions 21 through 24: Consider the sample space containing all the possible outcomes of rolling two 2 dice



21. Let A represent the event that the sum of the two faces showing is less than or equal to 5. Find P(A).

A	b	c	d	e
12/36	26/36	5/36	10/36	6/36

22. Let B represent the event that the left die shows a 3. What is P(B)?

a	b	c	d	e
30/36	14/36	15/36	6/36	21/36

23. What is the probability that B occurs given that A has occurred?

a	b	c	d	e
8/10	2/10	1/6	7/36	6/10

24. The event A and the event B are _____ because _____.

a	Dependent; because the $P(B\|A) \neq P(A)$
b	Dependent; because the $P(B\|A) \neq P(B)$
c	Independent; because the $P(B\|A) = P(B)$
d	Independent; because the $P(B\|A) = P(A)$
e	cannot be determined

- Question 25 through 29: Given: P(A) = **0.4**, P(B) = **0.7**, and P(A AND B) = **0.2**, calculate

25. $P(A^C)$

a	b	c	d	e
0.60	0.40	0.30	0.50	0.70

26. P(A OR B)

a	b	c	d	e
1.10	0.90	0.20	0.28	0.82

27. $P((A \text{ AND } B)^C)$

a	b	c	d	e
0.82	0.80	0.20	0.72	0.28

28. P(A | B)

a	b	c	d	e
0.2857	0.7000	0.4000	0.5000	0.5714

29. Regarding the events above, which of the following is true?

a	A and B are disjoint and independent
b	A and B are disjoint and not independent
c	A and B are not disjoint but are independent
d	A and B are not disjoint and are not independent

- For questions 30 through 34: Axl guesses randomly at 6 multiple-choice problems on an exam. Each problem has four potential answers.

30. This scenario can be modeled as a…

a	Binomial experiment with 4 trials and success probability of $1/6$ per trial
b	Binomial experiment with 6 trials and success probability of $1/6$ per trial
c	Binomial experiment with 6 trials and success probability of ¼ per trial
d	Normal experiment with 6 trials and success probability of ¼ per trial
e	Binomial experiment with 4 trials and success probability of ¼ per trial

31. What is the probability Axl will answer exactly 4 questions correctly?

a	b	c	d	e
0.0044	0.3560	0.2966	0.1318	0.0329

32. What is the probability Axl will answer fewer than 4 questions correctly?

a	b	c	d	e
0.9624	0.0329	0.1318	0.9954	0.8306

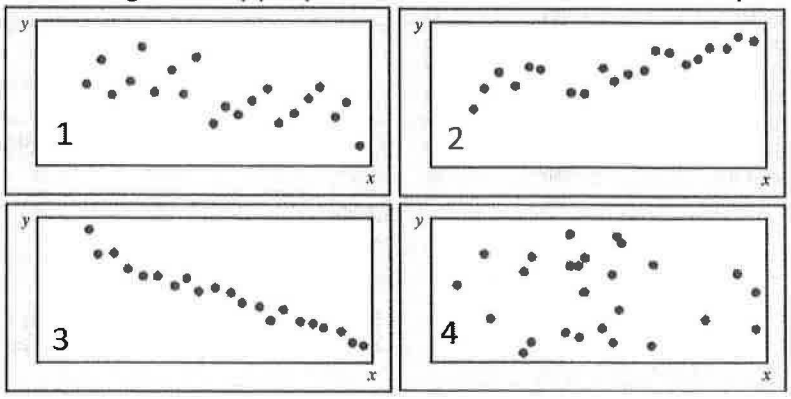
33. What is the expected number of questions Axl will answer correctly?

a	b	c	d	e
1.0607	1.1180	2.0000	1.5000	1.2500

34. What is the standard deviation of the number of questions Axl will answer correctly?

a	b	c	d	e
1.5000	1.1180	1.1250	1.2500	1.0607

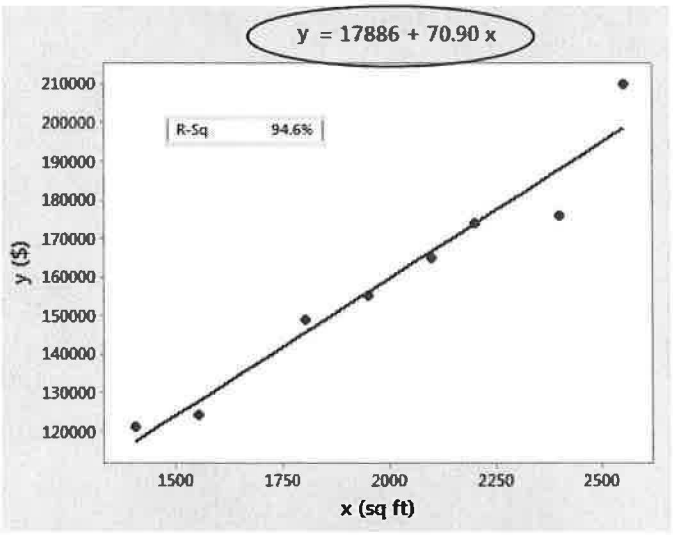
35. Assign the appropriate r values to the 4 scatter plots below



Answer choice	r value for scatterplot 1	r value for scatterplot 2	r value for scatterplot 3	r value for scatterplot 4
a	0.5	- 0.6	0.9	0
b	- 0.5	0.6	-0.9	0
c	0.5	- 0.5	-0.9	0
d	- 0.5	- 0.9	0.6	0

Questions 36 through 39: The price of the house is thought to be related to its square footage. The following table summarizes data for 8 houses, and the fitted line plot graphs the data points, gives the equation of the regression line, and the coefficient of determination, r2.

x (sq ft)	1400	1550	1800	1950	2100	2200	2400	2550
y ($)	121,000	124,000	149,000	155,000	165,000	174,000	176,000	210,000



36. Would this regression line be safe for predicting the price of the **3000** square foot house?

a	Yes, because r^2 is moderately high
b	Yes, because regression lines are used to make predictions
c	Not necessarily, because this would be extrapolation, and the price might not behave linearly as square footage increases above 2550
d	No, because the price of the house really has nothing to do with the square footage
e	No, because the regression line does not intersect all of the data points between 1500 ft.² and 2550 ft.²

37. For square footage value of **2200**, the regression line, $\hat{y} = 17886 + 70.90x$, predicts the price of

a	b	c	d	e
$39,349,271	$173,866	$17,957	- $138,094	$174,000

38. Adding a new data point at (**2200** sq ft, **$250,000**) would probably

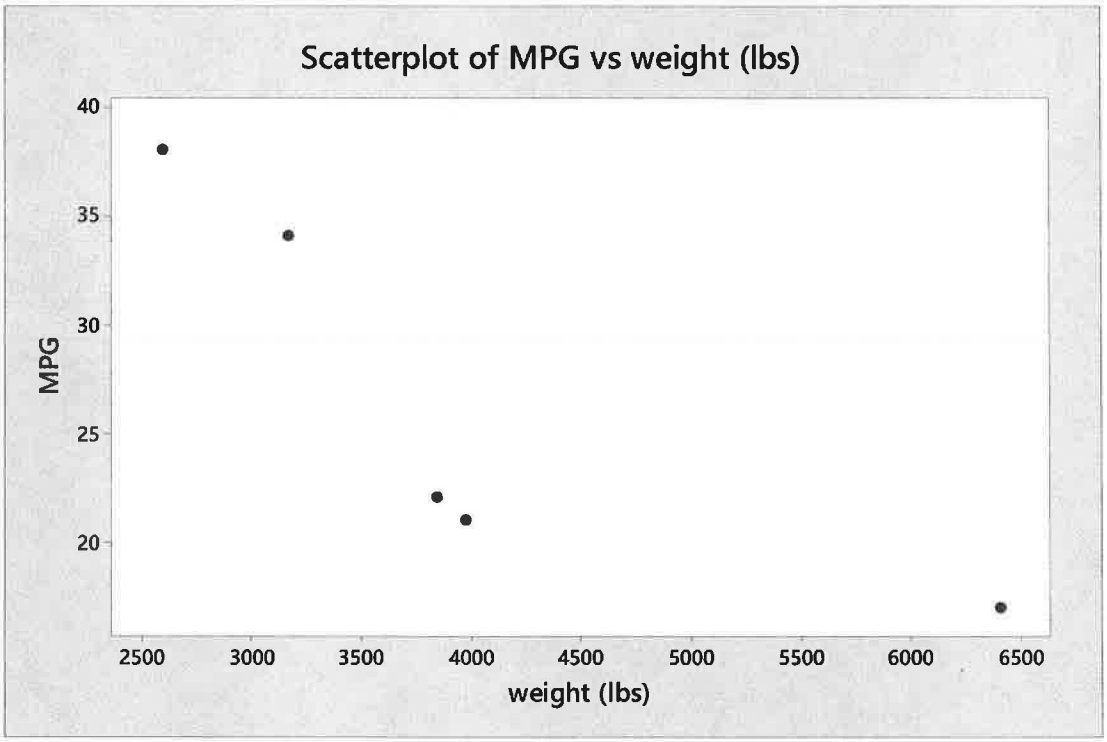
a	Influence some of the other data values
b	Be a good idea in case the houses are undervalued
c	Have little or no effect on the regression line
d	Cause r^2 to decrease
e	Be sufficient for pricing the home of that size

39. Regarding the scenario from above, which of the following is ***not*** true?

a	As square footage increases the price generally increases
b	94.6% of the variation in price can be explained by the regression line
c	As square footage increases, the price generally decreases
d	Least squares regression is appropriate for this scenario
e	There seems to be a strong linear relationship between the price and square footage of houses

- Questions 40 through 43: The weights and miles per gallon (MPG) of five 2018 automobile models appear in the table below, along with a scatter plot.

Model	Accord	Corolla	Canyon Club Cab	Grand Cherokee	Ram 1500 HD
Weight	3164	2590	3838	3970	6400
MPG	34	38	22	21	17



40. Regarding the table and graph above, which is true?

a	The response (or dependent) variable responds to the value of MPG
b	The explanatory (or independent) variable is weight and the response (or dependent) variable is MPG
c	There are no explanatory variables in this problem
d	The explanatory (or independent) variable is MPG and the response (or dependent) variable is weight

41. Which is ***not*** true from the scenario above?

a	MPG decreases as weight increases
b	There seems to be a moderately strong negative linear relationship between the weight and the MPG of cars
c	As weight increases, the MPG increases
d	The Ram 1500 HD data point is potentially an influential observation

42. What is the equation of the regression line from the scenario above?

a	$\hat{y} = 47.32x - 0.0052$
b	$\hat{y} = -134.86x + 7552.6$
c	$\hat{y} = -0.0052x + 47.32$
d	$\hat{y} = 7552.6x - 134.86$

43. What is the correct interpretation of the slope from the scenario above?

a	Weight increases by 134 mpg for each additional pound
b	MPG drops by 134 mpg for each additional pound
c	MPG drops by 0.0052 mpg for each additional pound
d	Weight drops by 0.0052 mpg for each additional pound

- Questions 44 through 48: Cherry tomatoes are put in packages averaging **454** grams with a standard deviation of **28** grams. If the Department of Agriculture inspector repeatedly takes random samples of **64** packages each…

44. What is the mean of the sampling distribution (aka the mean of the $\bar{x}$)?

a	b	c	d	e
64.000	56.750	3.500	454.000	0.438

45. How are the sample means distributed?

a	b	c	d	e
454	454 and 28	Bimodally	Binomially	Approx. normally

46. What is the standard deviation of the sampling distribution (aka the standard deviation of the $\bar{x}$)?

a	b	c	d	e
4.000	3.500	56.750	0.438	454.000

47. What is the probability that any sample mean, $\bar{x}$, will be between **447.00** grams and **461.00** grams?

a	b	c	d	e
19.74%	68.27%	95.45%	99.73%	essentially 100%

48. From the scenario above, what is **70th** percentile of the sampling distribution (i.e. 70% of the $\bar{x}$ will be below this number)?

a	b	c	d	e
468.6800	0.6800	455.8354	456.0976	439.3200

49. Calculate the minimum sample size required to make a 95% confidence interval for a population mean mass for newborn elephants if the margin of error is to be **3** kg and the standard deviation is known to be **4.8** kg.

$z_{0.10}$	$z_{0.05}$	$z_{0.10}$	$z_{0.025}$	$z_{0.005}$
1.282	1.645	1.960	2.326	2.576

a	b	c	d	e
30	10	2	4	6

50. Calculate the minimum sample size required to make a 95% confidence interval for the proportion of the population who prefers Coke to Pepsi if the margin of error is to be **0.03**.

a	b	c	d	e
545	1068	1066	2135	33

- Questions 51 through 53: For <u>any</u> confidence interval…

51. Increasing the confidence level

a	decreases the sample mean
b	has no effect on the confidence interval
c	increases the width of the confidence interval
d	decreases the width of the confidence interval
e	increases the sample mean

52. Increasing the sample size causes the margin of error to

a	do a little dance
b	do nothing
c	get smaller
d	get larger
e	none of the above

53. Changing which of the following will NOT affect the margin of error?

a	Sample size
b	Sample mean
c	Confidence level
d	Sample standard deviation

54. If a hypothesis test results in a p-value of **0.045** and α was specified at 5%,

a	There is a 4.5% chance that H_0 will be rejected
b	You should reject H_0
c	You should fail to reject H_0
d	You should accept H_0
e	None of the above is correct

55. Which one of the following statements is NOT true?

a	As the degrees of freedom of the t-distribution increases, the t distribution curve looks more like that of the standard normal curve
b	The total area under the t-distribution curve is one
c	The t-distribution is important for making hypothesis tests and confidence intervals for population means
d	The t-distribution curve is symmetric about the vertical line at 1
e	As the sample size *n* increases, the degrees of freedom increases

- Questions 56 through 58: In a simple random sample of **450** APSU students, **203** of them drive school.

56. A 95% confidence interval for the actual proportion of students who drive school is

a	There is not enough information to construct a 95% confidence interval
b	$(-1.5089 < p < 2.4111)$
c	$(0.4277 < p < 0.4746)$
d	0.4511
e	$(0.4051 < p < 0.4971)$

57. The correct interpretation of the previous confidence interval is

(i) 95% of the data are within the confidence interval
(ii) we are 95% confident that the population proportion lies in the confidence interval
(iii) we are 95% confident that the sample proportion lies in the confidence interval

a	b	c	d	e
ii only	i, ii, iii	i only	i and ii	ii and iii

58. Regarding the sample of APSU students from above, suppose the alternative hypothesis is that the population percentage is < 50%. In this case, the value of the test statistic is z = -2.074. The correct decision at a 5% significance level (aka $\alpha = 0.05$) is...

a	Fail to reject H_0
b	There is not enough information given to run the test
c	Reject H_0 in favor of the alternative hypothesis H_a
d	Fail to reject 0.4511
e	Accept H_0

- Questions 59 through 62: The average number of miles Clarksville commuters drove to work each day was **14.4** miles last year. City officials wonder if that average has changed. In a recent simple random sample of **81** commuters, the mean and sample standard deviation of the number of miles they drive to work each day were **13.8** miles and **3.9** miles respectively. Is the average different from what it was a year ago?

59. The correct null and alternate hypothesis statements for this problem are

a	$H_0: \mu = 13.8$	$H_a: \mu \neq 13.8$
b	$H_0: \mu \neq 14.4$	$H_a: \mu = 14.4$
c	$H_0: \mu = 14.4$	$H_a: \mu = 13.8$
d	$H_0: \mu = 14.4$	$H_a: \mu \neq 14.4$
e	$H_0: \mu = 13.8$	$H_a: \mu = 14.4$

60. What is the value of the test statistic?

a	b	c	d	e
-1.3846	13.3897	-19.4308	-1.96	1.3846

61. Using the correct t value, if the significance level is $\alpha = 0.05$, city officials should conclude

a	There is not enough statistical evidence to suggest that the mean is different
b	There is enough statistical evidence to suggest that the mean is different
c	There is enough statistical evidence to suggest that the mean has decreased
d	The mean is not as variable as it was in the past
e	The mean is probably less than 13.8 miles

62. A 95% confidence interval for the actual mean miles Clarksville residents commute

a	(13.967 < μ < 14.833 miles)
b	(13.367 < μ < 14.233 miles)
c	(11.810 < μ < 15.790 miles)
d	(12.938 < μ < 14.662 miles)
e	(12.951 < μ < 14.649 miles)

63) A set of data points and the equations of two lines are given. For each line, determine $\sum e^2$ which is the sum of squared residuals. Then, determine which line fits the set of data points better, according to the least-squares criterion.

x	1	2	4	4
y	2	3	5	4

Line A: $y = 1 + 0.9x$

Line B: $y = 0.8 + 1.1x$

A) Line A: $\sum e2 = 0.57$

Line B: $\sum e2 = 1.49$

Line A fits the set of data points better.

B) Line A: $\sum e2 = 1.31$

Line B: $\sum e2 = 1.57$

Line B fits the set of data points better.

C) Line A: $\sum e2 = 0.57$

Line B: $\sum e2 = 1.49$

Line B fits the set of data points better.

D) Line A: $\sum e2 = 1.31$

Line B: $\sum e2 = 1.57$

Line A fits the set of data points better.

64) A sample mean, sample standard deviation, and sample size are given. Use the one-mean t-test to perform the required hypothesis test about the mean, μ, of the population from which the sample was drawn. Use the critical-value approach.

$\bar{x} = 3.23$, $s = 0.59$, $n = 9$, $H_0: \mu = 2.85$, $H_a: \mu > 2.85$, $\alpha = 0.01$

A) Test statistic: t = 1.93. Critical value: t = 2.896. Reject H0. There is sufficient evidence to support the claim that the mean is greater than 2.85.

B) Test statistic: t = 1.93. Critical value: t = 2.33. Do not reject H0. There is not sufficient evidence to support the claim that the mean is greater than 2.85.

C) Test statistic: t = 1.93. Critical value: t = 2.896. Do not reject H0. There is not sufficient evidence to support the claim that the mean is greater than 2.85.

D) Test statistic: t = 1.93. Critical value: t = 2.821. Do not reject H0. There is not sufficient evidence to support the claim that the mean is greater than 2.85.

65) For the given hypothesis test, explain the meaning of a Type I error, a Type II error, or a correct decision as specified.

The recommended dietary allowance (RDA) of vitamin C for women is 75 milligrams per day. A hypothesis test is to be performed to decide whether adult women are, on average, getting less than the RDA of 75 milligrams per day. The hypotheses are

$H0 : \mu = 75$ mg

$Ha : \mu < 75$ mg

where μ is the mean vitamin C intake (per day) of all adult females. Explain the meaning of a Type II error.

A) A Type II error would occur if, in fact, $\mu < 75$ mg, and the results of the sampling lead to rejection of the null hypothesis that $\mu = 75$ mg.

B) A Type II error would occur if, in fact, $\mu = 75$ mg, and the results of the sampling do not lead to rejection of that fact.

C) A Type II error would occur if, in fact, $\mu = 75$ mg, but the results of the sampling lead to the conclusion that $\mu < 75$ mg

D) A Type II error would occur if, in fact, $\mu < 75$ mg, but the results of the sampling fail to lead to that conclusion.