

Statistics Competency Exam Fall 2018

Name _____

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.

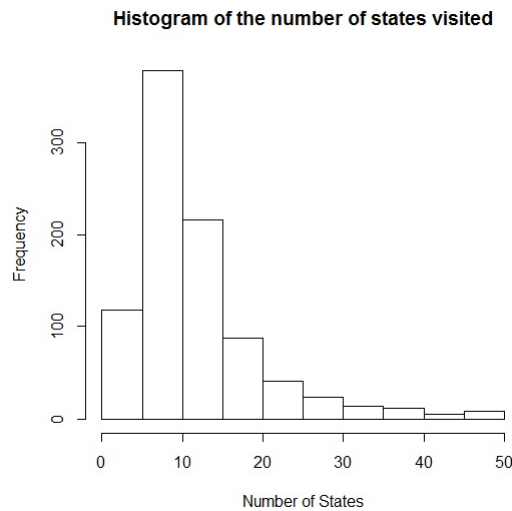
Stat Comp

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

- According to the College Board, SAT writing scores from the 2015 school year for high school students in the United States were normally distributed with a mean of 484 and a standard deviation of 115. What percent of high school students that took the SAT in 2015 scored above 369?
(A) 68% (B) 95% (C) 34% (D) 16% (E) 84%
- Suppose a recent plant journal indicated that the mean height of mature plants of a certain species of sunflower is 10.4 ft. A biologist observing a huge field of mature plants of this particular species of sunflower thinks that the mean height is shorter than reported. He measures the heights of 45 randomly selected sunflowers and finds the mean height to be 10.1 ft. The hypotheses of interest are
(A) $H_0 : \mu = 10.1$ vs $H_a : \mu < 10.1$ (C) $H_0 : \mu > 10.4$ vs $H_a : \mu = 10.4$ (E) $H_0 : \bar{x} = 10.1$ vs $H_a : \bar{x} > 10.1$
(B) $H_0 : \bar{x} = 10.1$ vs $H_a : \bar{x} < 10.1$ (D) $H_0 : \mu = 10.4$ vs $H_a : \mu < 10.4$
- Suppose the value of the sample standard deviation is zero for a set of data. Then we
(A) should recalculate the value of the sample standard deviation, because it can never be equal to 0
(B) can infer that all the data values are the same, but not necessarily zero
(C) cannot infer anything about the values of the data
(D) can say that all the data values are 0
(E) none of the above
- An educator wishes to study the effects of sleep deprivation on the ability to concentrate. He decides to study the students in a calculus class. Before the start of the class, each student is asked about the number of hours slept the previous night. Each student's eye movement is then tracked throughout the lecture. The amount of time is recorded whenever the student is not focused on either the instructor or taking notes. This is an example of:
(A) a sample survey (C) a randomized experiment (E) a completely randomized experiment
(B) an observational study (D) a matched pairs study
- Suppose that you calculate the sample Pearson correlation coefficient between X = weekly total amount of time exercising, and Y = resting heartbeat rate per minute. A value of $r = -1.18$ would be interpreted as
(A) a perfect negative correlation between exercise and heart rate
(B) a strong negative correlation between exercise and heart rate
(C) no correlation between exercise and heart rate
(D) an impossible value for r
(E) weak correlation between exercise and heart rate
- Seventy percent of adults own a car. Forty percent of adults own a house. Twenty percent of adults own both a house and a car. What is the probability that a randomly selected adult will own neither a house nor a car?
(A) 0.5 (B) 0.2 (C) 0.9 (D) 0.1 (E) 0.8

7. A vice president at a bank states that customers obtain service within an average time of 4.9 minutes of arriving at the bank's drive-up service window. However a long-time customer of the bank wants to test the hypothesis that the average time to obtain service exceeds 4.9 minutes, using a significance level of $\alpha = 0.01$. The customer collects data from 16 randomly selected customers, and calculates the average time to obtain service. The sample mean time is 5.3 minutes and the sample standard deviation is 1.2 minutes. If the researcher decides to reject the null hypothesis, the researcher would conclude that
- (A) at the 1% significance level, the average waiting time for these 16 bank customers to receive service does not exceed 4.9 minutes
 - (B) at the 1% significance level, the average waiting time for these 16 bank customers to receive service does exceed 4.9 minutes
 - (C) at the 1% significance level, the average waiting time for all of the bank customers to receive service does not exceed 4.9 minutes
 - (D) at the 1% significance level, the average waiting time for all of the bank customers to receive service does exceed 4.9 minutes
 - (E) at the 1% significance level, the total waiting time for these 16 bank customers to receive service does not exceed 4.9 minutes
8. The sampling distribution of the sample mean
- (A) describes the exact value of the population standard deviation for a sample of size n
 - (B) describes the distribution of the possible values that the sample mean may have for a sample of size n .
 - (C) describes the exact value of the sample mean for a sample of size n .
 - (D) describes the exact value of the population mean for a sample of size n .
 - (E) describes the value of the sample standard deviation for a sample of size n .
9. For a specific sample consisting of n observations, the width of a 90% confidence interval on μ
- (A) would be larger than the width of a 95% confidence interval on μ .
 - (B) would be smaller than the width of a 95% confidence interval on μ .
 - (C) would be the same as the width of a 95% confidence interval on μ .
 - (D) cannot be compared to the width of a 95% confidence interval on μ .
 - (E) none of the above
10. A sociologist studying freshmen at a major university carried out a survey, asking, among other questions, how often students went out per week, how many hours they studied per day, and how many hours they slept at night. The sociologist used an introductory sociology class to carry out the survey. The sociologist has learned from previous studies that females and males often behave differently regarding study and sleep patterns. She decides that she needs to ensure adequate numbers of females and males. She should take:
- (A) a multistage sample. (C) a multigroup sample. (E) two convenience samples, one of females and one of males.
 - (B) a stratified random sample. (D) a simple random sample.
11. A confounding variable is
- (A) one whose effect on a response variable cannot be distinguished from other variables in the experiment.
 - (B) one whose effect on a response variable can be distinguished from other variables in the experiment.
 - (C) one that is changed in an experiment.
 - (D) one that occurs in one of two possible states, labeled success and failure.
 - (E) an extraneous one which causes an explanatory and response variable to appear to be related in a way that they are not.

Use the following for the next 2 questions. A class survey asked “How many U.S. states have you visited?” The figure below represents the responses of 959 students.



12. Which of the following best describes this distribution? This distribution is
- (A) evenly-spaced from 0 to 50
 (B) left-skewed
 (C) right-skewed
 (D) symmetric around 10
 (E) multi-peaked with the bulk of the data from 5 to 15
13. Which of the following is TRUE for this distribution?
- (A) mean < median (C) mean = median (E) the mean is the best description of the center
 (B) mean > median (D) standard deviation = 0

Use the following for the next 2 questions. Suppose that a research and development scientist is studying the burn times in the current model and the new model of road flare that his company produces. He hopes to show that the average burn time of the new flare is higher than that of the current flare. First, he collects a random sample of 50 flares of the current model and records the burn time for each one. Next, he collects a random sample of 40 flares of the new model and records the burn time for each one. The first sample is roughly symmetric and bimodal, and the second is symmetric and single-peaked. Neither sample has outliers. Some of the summary statistics for his two samples are displayed below.

Population	Population description	Sample size	Sample mean	Sample standard deviation
1	current model	$n_1 = 50$	$\bar{x}_1 = 15.81$	$s_1 = 2.35$
2	new model	$n_2 = 40$	$\bar{x}_2 = 16.59$	$s_2 = 2.66$

14. Select the statement that accurately assesses whether or not a two-sample t -test is valid in this experiment.
- (A) A two-sample t -test is valid because the samples are independent, random, contain no outliers, and are large enough.
 (B) A two-sample t -test is not valid because the distribution of one sample is single-peaked but the other is not.
 (C) A two-sample t -test is not valid because it is not known whether the data comes from normal populations.
 (D) A two-sample t -test is not valid because the first sample has 50 observations, but the second sample has 40 observations.
 (E) A two-sample t -test is valid because the two samples are independent, random, and the burn times for the two types of flare are continuous, quantitative variables.
15. If we performed the two-sample t -test for this problem, what would be the appropriate degrees of freedom (using a conservative approximation)?
- (A) 49 (B) 40 (C) 39 (D) 50 (E) 45

16. Safety regulations state that roller coasters cannot operate safely in heavy rains. Throughout the day, ride operators measure the rainfall at several random locations along the track to verify that the rainfall is below an established threshold value, allowing the ride to operate safely. The collected data is used to conduct a one-sample z -test of the null hypothesis that the mean rainfall is minimal against the alternative that the mean rainfall is too heavy. If the null hypothesis is rejected in favor of the alternative, the ride is shut down due to unsafe conditions. Choose the correct description of a type I error in this context.
- (A) The rainfall is too heavy to operate safely, and the ride continues to operate.
 - (B) The rainfall is minimal enough to operate safely, and the ride continues to operate.
 - (C) The rainfall is too heavy to operate safely, and the ride is shut down.
 - (D) The rainfall is minimal enough to operate safely, and the ride is shut down.
 - (E) The ride continues to operate and the visitors are happy.
17. A scatterplot of the damage (in dollars) caused to houses by fire show a strong correlation with the number of firefighters at the scene. Does that mean that we shouldn't call firefighters, as they cause the damage?
- (A) No, the positive correlation is probably explained by the fact that more damage is done and more firefighters need to fight the fire at bigger blazes.
 - (B) Yes, the data show that more firefighters cause more weight in the buildings, which does more damage as it weakens the floors.
 - (C) Yes, the correlation can't be just by an accident.
 - (D) No, a negative correlation would allow that conclusion, but this correlation is positive.
 - (E) No, this is reverse cause-and-effect.
18. Based on meteorological data for the past century, a local TV weather forecaster estimates that the region's average winter snowfall is 23 inches, with a margin of error of ± 2 inches. Assuming he used a 95% confidence interval, how should viewers interpret this news?
- (A) During 95 of the past 100 winters, the region got between 21 inches and 25 inches of snow.
 - (B) There's a 95% chance the region will get between 21 inches and 25 inches of snow this winter.
 - (C) There will be between 21 inches and 25 inches of snow on the ground for 95% of the winter days.
 - (D) Residents can be 95% sure that the area's average snowfall is between 21 inches and 25 inches.
 - (E) Residents can be 95% confident that the average snowfall during the past century was between 21 inches and 25 inches per winter.
19. Only about 20% of people who try to quit smoking succeed. Sellers of a motivational tape claim that listening to the recorded messages can help people quit. What are the null and alternative hypotheses you would use to test this situation?
- (A) $H_0 : \mu = 0.2$ vs $H_a : \mu > 0.2$
 - (B) $H_0 : p < 0.2$ vs $H_a : p = 0.2$
 - (C) $H_0 : \hat{p} = 0.2$ vs $H_a : \hat{p} < 0.2$
 - (D) $H_0 : p = 0.2$ vs $H_a : p > 0.2$
 - (E) $H_0 : \hat{p} = 0.2$ vs $H_a : \hat{p} > 0.2$
20. The weight of potato chips in a medium-size bag is stated to be 10 ounces. The amount that the packaging machine puts in these bags is believed to have a Normal model with mean 10.2 ounces and standard deviation 0.12 ounces. Some of the chips are sold in "bargain packs" of 3 bags. What's the probability that the mean weight of the 3 bags is below 10 ounces?
- (A) 0.5 (B) 0.2 (C) 0.0019 (D) 0.0475 (E) 0.12

21. Will a fluoride mouthwash used after brushing reduce cavities? Twenty sets of twins were used to investigate this question. One member of each set of twins used the mouthwash after brushing, and the other did not. After six months, the difference in the number of cavities for those using the mouthwash was compared with the number of cavities for those who did not use the mouthwash. This experiment uses:
- (A) random placebos
 - (B) double-blinding
 - (C) a matched pairs design
 - (D) double replication
 - (E) an observational study
22. An educator wishes to study the effects of sleep deprivation on the ability to concentrate. He decides to study the students in a calculus class. After consultation with a statistician, the educator decides to randomly allocate students to either a group that will sleep for 8 hours the night before class or 6 hours. The educator does not know which group a student belongs to when she or he comes to class. This study is:
- (A) a single-blinded randomized study, because the educator does not know the treatment groups students belong to but the students know.
 - (B) double-blinded, because the educator does not know who belongs to the 6- or 8-hour group.
 - (C) a placebo controlled study, because 8 hours is normal sleeping time.
 - (D) double-blinded, because the students don't know that there are two different sleep times.
 - (E) a single-blinded randomized study, because the students don't know that there are two sleep times but the educator does.

Use the following for the next 2 questions. The Fall 2016 MATH1530 survey asked students "Should collegiate athletes be paid to play? (Yes, No)" In the sample of 902 students who responded to this question, 397 said "Yes" and 505 said "No."

23. The sample proportion \hat{p} that responded "Yes" is _____.
- (A) 505 (B) 0.786 (C) 0.560 (D) 0.440 (E) 397
24. A 90% confidence interval for p is
- (A) 40.8% to 47.2% (B) 41.3% to 46.7% (C) 372.5 to 421.5 (D) 52.8% to 59.2% (E) 53.3% to 58.7%
25. The table below represents the responses to a survey question "What is your favorite way of spending an evening?" Which type of graph is appropriate for these data?
- | Evening time | Reading | Resting/Relaxing | Staying at home | Visiting with friends | Watching Television | Other |
|--------------|---------|------------------|-----------------|-----------------------|---------------------|-------|
| Counts | 46 | 166 | 78 | 260 | 61 | 97 |
- (A) Histogram (B) Scatterplot (C) Boxplot (D) Bar chart (E) Stem plot
26. Which of these questions from a class survey produced variables that are categorical?
- i. What do you think is the ideal number of children for a family to have?
 - ii. Do you favor or oppose the death penalty for persons convicted of murder (Favor, Oppose)?
 - iii. How many pairs of shoes do you own?
 - iv. What is your opinion about a married person having sexual relations with someone other than their marriage partner (Always wrong, Almost always wrong, Wrong only sometimes, Not wrong at all)?
 - v. Usually, how many hours sleep do you get at night?
- (A) i, ii (B) iii, iv, v (C) ii, iv (D) iii (E) i, ii, iii, iv, v

27. Make _____ to display the relationship between two quantitative variables measured on the same subjects.
- (A) a scatterplot (C) a two sample t-test
 - (B) side-by-side boxplots (D) back-to-back stemplots
 - (E) the five-number summary

28. A recent study found that the following behaviors- smoking, drinking too much alcohol, not exercising, and not eating enough fruits and veggies- can reduce your lifespan (on average). Whether moderation of these behaviors will extend life is not clear. Suppose that the added life expectancy from moderation of these behaviours is just 3 months. A statistical test is more likely to find a significant increase in mean life if

- (A) The size of the sample doesn't have any effect on the significance of the test.
- (B) it is based on a very small random sample.
- (C) it is based on a very large random sample.
- (D) it is of practical significance.
- (E) the p-value is large.

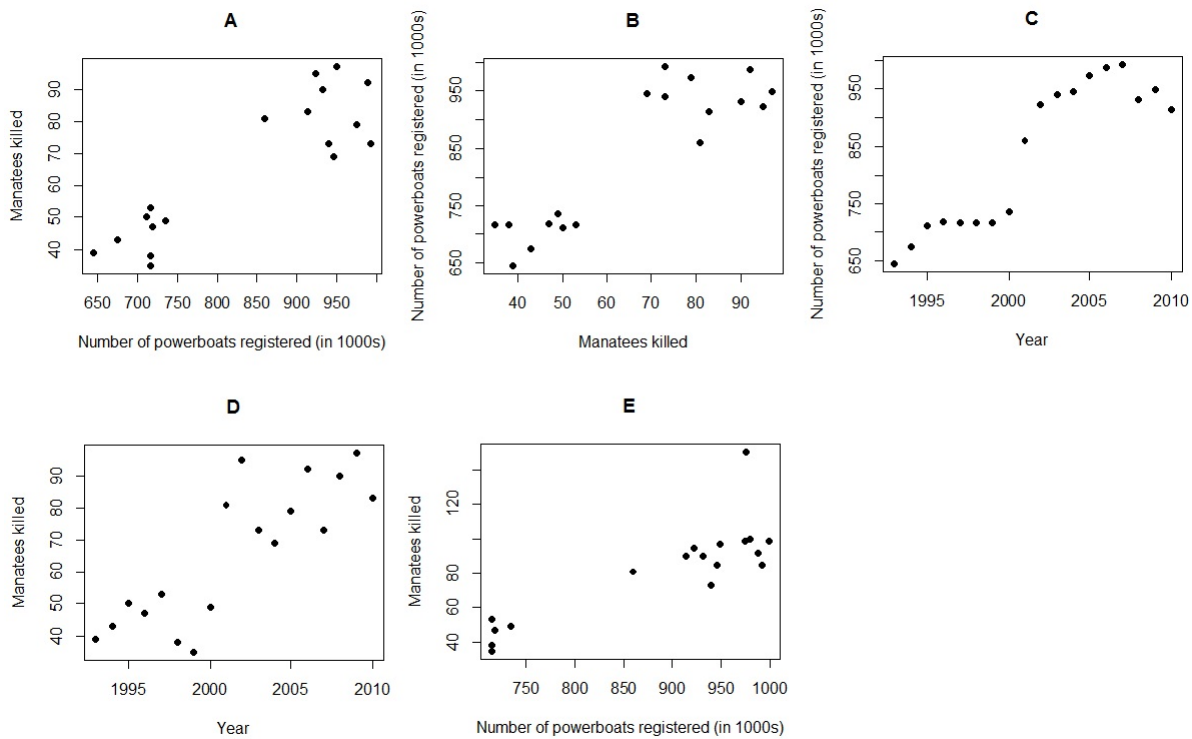
Use the following for the next 4 questions. Marine biologists warn that the growing number of powerboats registered in Florida threatens the existence of manatees. The data below comes from the Florida Fish and Wildlife Conservation Commission and the National Marine Manufacturers Association, the powerboats registered is in 1000s.

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Manatees Killed	39	43	50	47	53	38	35	49	81	95	73	69	79	92	73	90	97	83
Powerboat Register	646	675	711	719	716	716	716	735	860	923	940	946	974	988	992	932	949	914

29. In this context, which is the explanatory variable?

- (A) The year
- (B) The number of manatees killed
- (C) The number of powerboats registered in Florida
- (D) The marine biologists
- (E) The manatee food

30. Which option is a plausible scatterplot of the data?



31. Which of the following statements is true regarding the data?
- (A) The association between number of manatees killed and number of powerboats registered in Florida is nearly perfect.
 - (B) The association between number of manatees killed and number of powerboats registered in Florida is positive, linear, and strong.
 - (C) The association between number of manatees killed and number of powerboats registered in Florida has groups: small, medium and large.
 - (D) The association between number of manatees killed and number of powerboats registered in Florida is negative, slightly curved, and strong.
 - (E) The association between number of manatees killed and number of powerboats registered in Florida is positive, linear, and weak.
32. Which option is a plausible value for the correlation coefficient between powerboat registrations and manatees killed? (approximately)
- (A) 0 (B) 1.00 (C) -0.490 (D) 0.225 (E) 0.893

Use the following for the next 2 questions. According to the Pew Research Center, the average number of books read annually in the United States by persons aged 18-29 was 11 (in 2015). In the Spring 2016 Math 1530 Survey, students were asked "How many books (not including textbooks or other books for class) did you read in 2015?" Is there evidence that the average number of books read by ETSU students is less than 11? Use the output below to answer the next two questions.

One-Sample T: BOOKS

Test of $\mu = 11$ vs less than 11

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	T	P
Books	708	7.571	16.717	0.628	8.605	-5.46	0

33. Which of the following is the best conclusion?
- (A) No. There is weak evidence ($P = 0$) that, on average, the number of books read by ETSU students is less than 11.
 - (B) No. We fail to reject H_0 since the P is close to 0.
 - (C) There is strong evidence ($P = 0$) that, on average, the number of books read by ETSU students is less than 11.
 - (D) The probability that H_0 is true is 1 since the P -value is 0.
 - (E) The data are not statistically significant since the P -value is small.
34. The most important condition for sound conclusions from statistical inference is usually that
- (A) the t-test is used.
 - (B) the p-value is small.
 - (C) the population distribution is exactly Normal.
 - (D) the data can be thought of as a random sample from the population of interest.
 - (E) the number of observations is large.

Use the following for the next 3 questions. Computer games in which the players take the roles of characters are very popular. They go back to earlier tabletop games such as Dungeons & Dragons. These games use many different types of dice. A four-sided die has faces with one of the numbers 1, 2, 3, or 4 appearing at the bottom of each visible face. When you roll this die, the number appearing at the bottom of the face is the number you “roll.”

35. The sample space for rolling a four-sided die twice and counting the total you “rolled” is

- (A) $S = 1, 2, 3, 4, 5, 6$
 (B) $S = 1, 2, 3, 4$
 (C) $S = 2, 4, 6, 8$
 (D) $S = 2, 3, 4, 5, 6, 7, 8$
 (E) $S = 1, 2, 3, 4, 5, 6, 7, 8$

36. What is the assignment of probabilities to outcomes in this sample space?

(A)

Total	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(B)

Total	2	3	4	5	6	7	8
Probability	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{6}{16}$	$\frac{7}{16}$

(C)

Total	1	2	3	4
Probability	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

(D)

Total	1	2	3	4	5	6	7	8
Probability	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

(E)

Total	2	3	4	5	6	7	8
Probability	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

37. What is the probability that we “roll” a 4?

- (A) $\frac{3}{16}$ (B) $\frac{1}{6}$ (C) $\frac{4}{20}$ (D) $\frac{1}{4}$ (E) $\frac{1}{8}$

Use the following for the next 2 questions. According to genetic theory, the blossom color in the second generation of a certain cross of sweet peas should be red or white in a 3:1 ratio. That is, each plant has probability $\frac{3}{4}$ of having red blossoms, and the blossom colors of separate plants are independent.

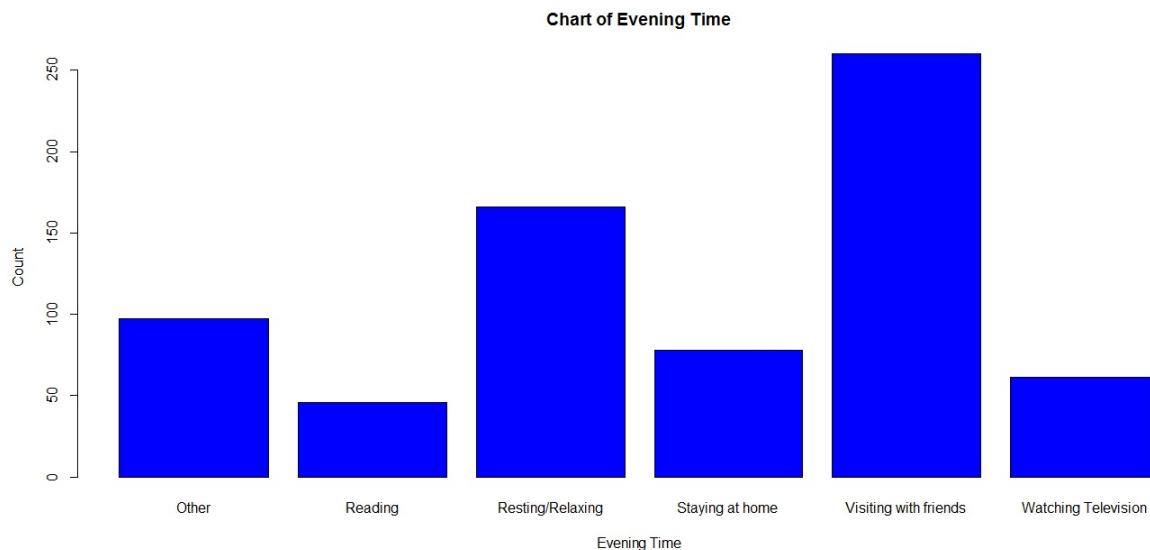
38. What is the probability that exactly three out of four of these plants have red blossoms?

- (A) 0.75 (C) 0.4219
 (B) 0.5 (D) 0.1055
 (E) 0.25

39. What is the mean number of red-blossomed plants when 60 plants of this type are grown from seeds?

- (A) 45 (B) 30 (C) 15 (D) 11.25 (E) 3.35

40. In the Spring 2016 Math 1530 Survey, students were asked “What is your favorite way of spending an evening? (Staying at home, Watching TV, Resting/Relaxing, Reading, Visiting with friends, Other)” The following graph displays the information on this variable:



Which of the following describes this distribution?

- (A) The distribution is multimodal and skewed left.
- (B) The distribution is evenly spaced from “Other” to “Watching Television.”
- (C) The distribution is skewed left and the range is from “Other” to “Watching Television.”
- (D) “Visiting with friends” is the favorite way of spending an evening while “Reading” is the least favorite way of spending an evening.
- (E) The distribution is irregular and centered around “Staying at home.”