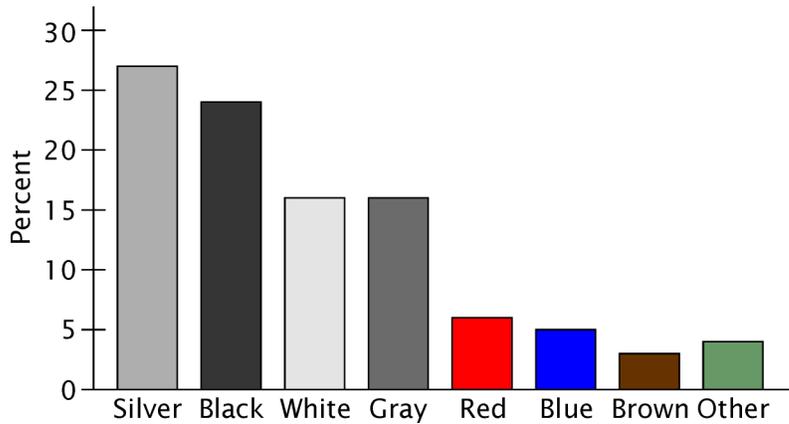


AP Statistics – Sample Problems

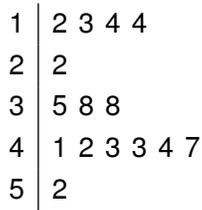
1. The bar graph below gives the distribution of the most popular colors for cars and light trucks sold globally.



Which of the following types of graphs is also appropriate for the data?

- (A) a boxplot
 - (B) a histogram
 - (C) a pie chart
 - (D) none of the above
 - (E) all of the above
2. Starting at $(0,0)$ on a coordinate grid and moving along the grid lines either to the right or up, how many different ways are there to reach $(6,4)$?
- (A) 45
 - (B) 84
 - (C) 126
 - (D) 210
 - (E) 252

3. The number of phone calls received by an internet service provider (ISP) were recorded for 15 days. The following stem-and-leaf plot was drawn for the data.



Which two of these statements could explain the gaps and clusters?

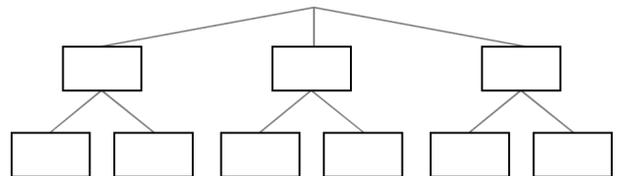
- I. The number of calls on weekends is different from weekdays.
- II. All payments must be made during the first 10 days of a month.
- III. The company had a two-week promotional offer for new subscribers.
- IV. Customers are not satisfied with the service.

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

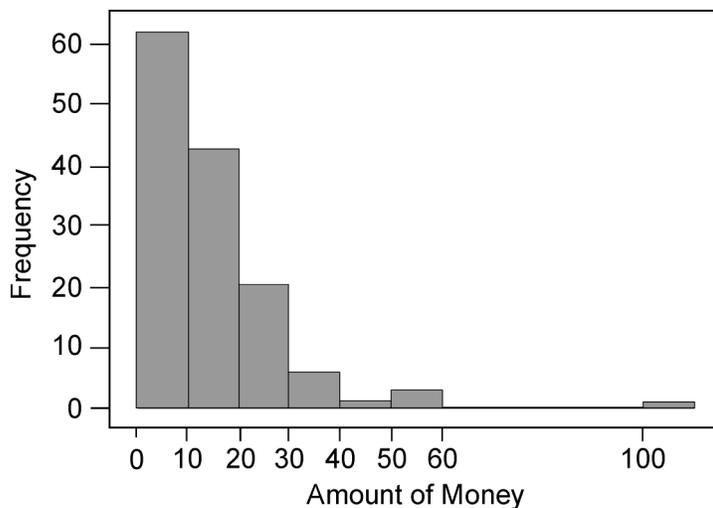
4. A jar contains three disks lettered 'o', 's' and 't'. Two of the disks are pulled out at random, and they are placed next to each other in the same order as they are pulled out.

Fill in the tree diagram and use it to find the probability that the letters on the disks form an actual word.

- (A) $\frac{1}{12}$
- (B) $\frac{1}{6}$
- (C) $\frac{1}{3}$
- (D) $\frac{1}{2}$
- (E) $\frac{2}{3}$



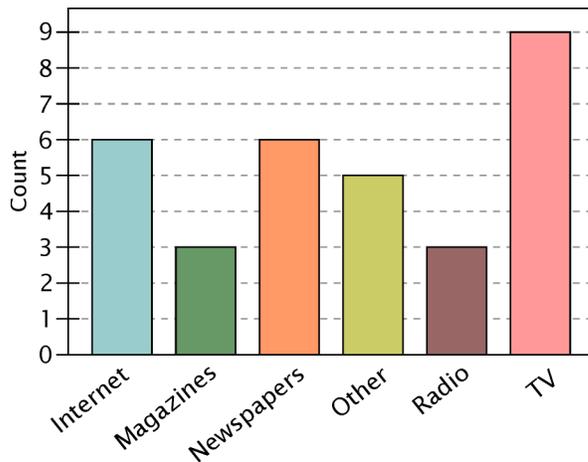
5. Following the analysis of a well-designed and completely randomized experiment, it was reported that the observed effect was statistically significant. Which of the following statements best explains the meaning of “statistically significant”?
- (A) The observed result made sense to the experimenter since it was what was hoped would happen.
 - (B) The observed effect happened because the experiment was properly designed and carried out without bias.
 - (C) The experimenter carefully employed the basic principles of experimental design in conducting the study.
 - (D) The observed effect was sufficiently large so that it would rarely occur simply by chance.
 - (E) The laws of probability say that this observed result would be expected to happen by chance.
6. In an introductory statistics class with 136 students, the teacher recorded how much money each student spent eating out last week. The histogram shows the data collected.



The mean for the distribution is:

- (A) to the right of the median
- (B) to the left of the median
- (C) the same as the median
- (D) greater than \$100.00
- (E) equal to \$45.00

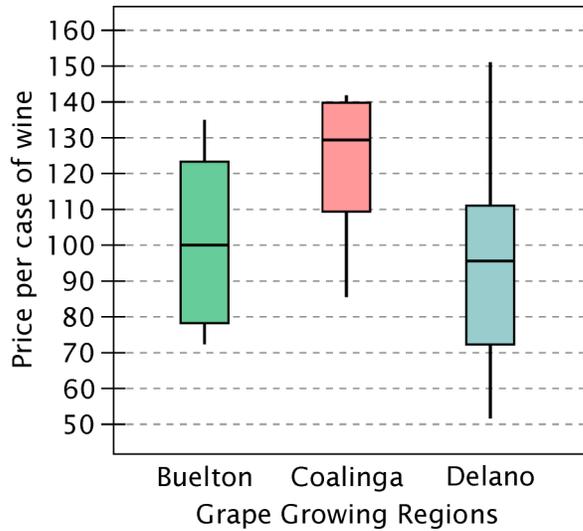
7. A sample of adults were surveyed and asked, "Is most of your information about politics from newspapers, magazines, the internet, television, radio, or some other source?" A bar chart of the survey results is shown below.



For what percentage of respondents is the internet their primary news source?

- (A) 0.1667
(B) 0.1875
(C) 0.667
(D) 6
(E) 32
8. Patty told Jie that she purchased 5 movie DVDs. If the mean price of the DVDs was 14 dollars, and the median price was 13, which of the following data sets represents her purchase?
- (A) 12, 19, 20, 12, 18
(B) 20, 20, 13, 12, 18
(C) 13, 12, 13, 13, 19
(D) 12, 13, 20, 17, 16
(E) 13, 17, 20, 12, 16

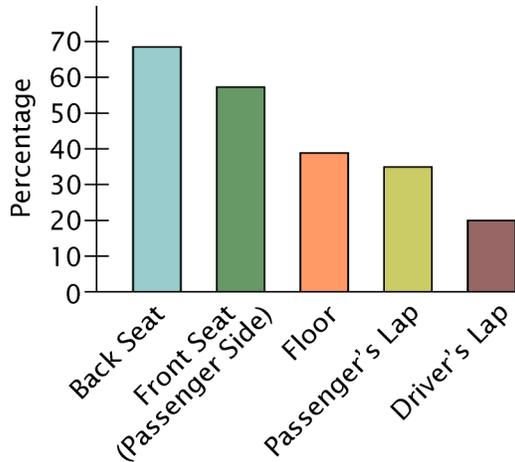
9. The boxplots below show the price per case of wine (in dollars) for three different grape growing regions.



Which price distribution is clearly skewed to the left?

- (A) Buelton
 - (B) Coalinga
 - (C) Delano
 - (D) none of them
 - (E) two or more of them
10. For a social club, the members had a mean age of 68 and a standard deviation of 10 years. What is the z-score for an age of 50?
- (A) -1.8
 - (B) -0.9
 - (C) 0.9
 - (D) 1.18
 - (E) 1.8

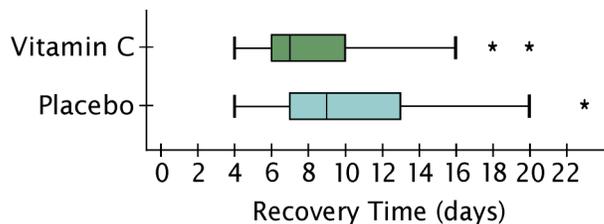
11. The bar graph below summarizes responses of dog owners to the question, "Where in the car do you let your dog ride?"



Which of the following statements is *false*?

- (A) About twice as many owners let their dog ride in the back seat as in a passenger's lap.
- (B) Some owners let their pets ride in more than one place in the car.
- (C) The least common place where dogs ride is in the driver's lap.
- (D) The data presented in the graph is categorical.
- (E) The data could also be presented in a pie chart.

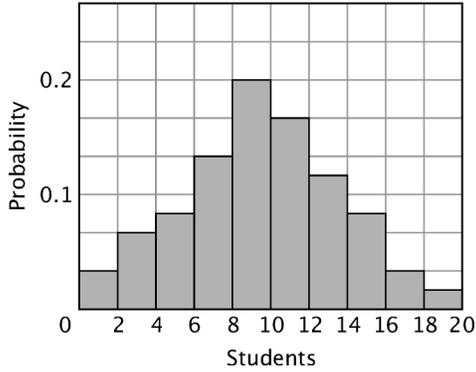
12. A clinical trial is conducted to determine the effectiveness of Vitamin C lozenges at reducing the duration of the common cold. 100 study participants are randomly assigned to begin taking Vitamin C lozenges or placebo lozenges within 24 hours of developing cold symptoms. The side-by-side boxplots below display the distributions of recovery time (in days) for the two groups.



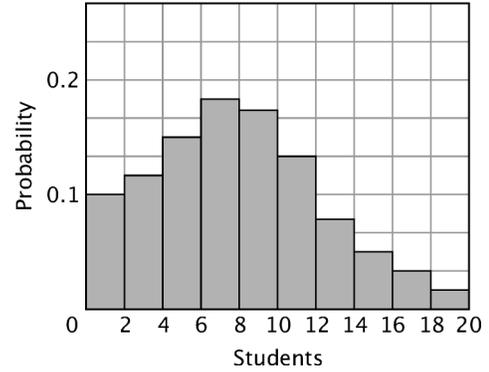
Write a short paragraph which compares the distribution of recovery time for the treatment group taking the Vitamin C lozenges and the distribution of recovery time for the placebo group.

13. Which set of data is likely to have the lowest standard deviation?

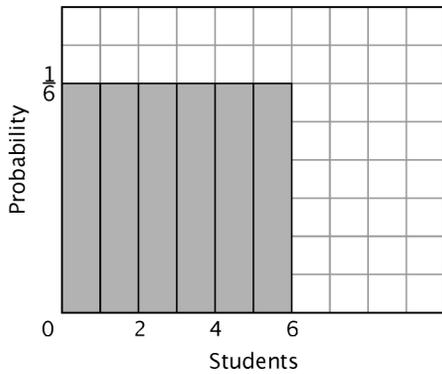
(A) Probability that Gr. 10 students in class of 20 are wearing running shoes



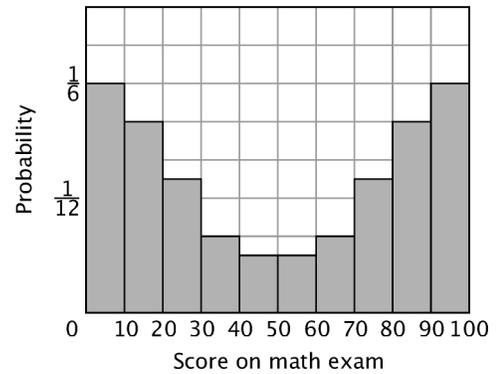
(B) Probability that Gr. 4 students in class of 20 are wearing glasses



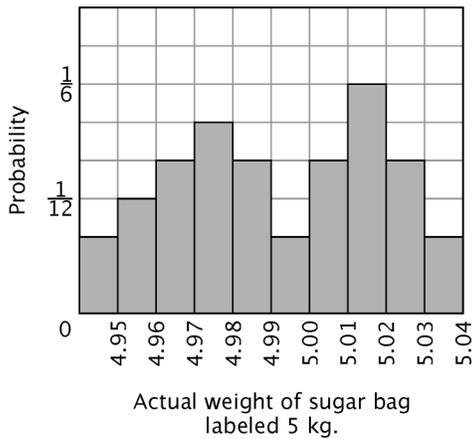
(C) Probability of rolling each number on a die



(D) Probability of obtaining a given percent on a math exam



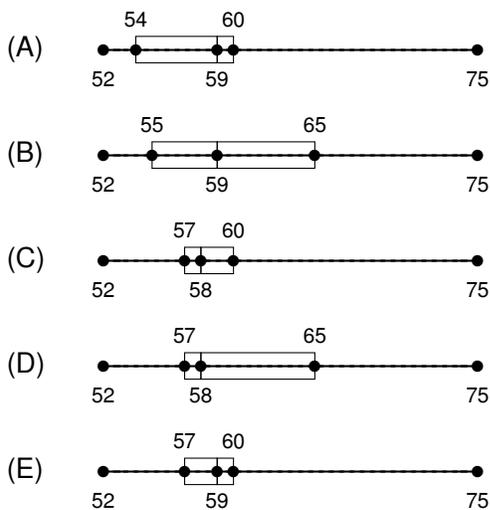
(E) Probability of actual weights of 5 kg bags of sugar



14. The following table shows the annual salaries for 21 executives at a company.

Salary in \$1000s	Number of Executives
\$75	4
\$60	6
\$58	5
\$57	4
\$52	2

Select the correct box-and-whisker plot for the table.



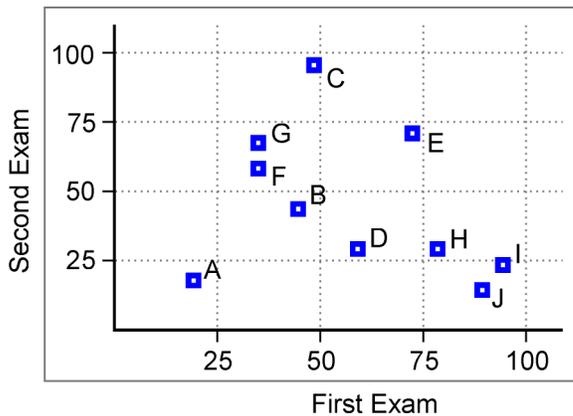
15. A police department wants to know if residents are satisfied with police services in their neighborhood. A questionnaire is used to determine if the residents are satisfied. Compared to other types of research, what is a specific *disadvantage* of using a questionnaire to collect data on this topic?

- (A) misinterpretation of questions
- (B) results are anonymous
- (C) not all questionnaires may be returned
- (D) tabulating results is expensive
- (E) personal contact is required

16. A research firm wanted to investigate the economic and humanitarian contribution of pharmaceutical companies and compare them to other types of businesses. The firm began by surveying pharmaceutical companies. After collecting information from 10 of them, someone at the firm suggested that the results be publicized. What could be a noticeable bias in publicizing the results at that point in time?

- (A) Undercoverage of the population
- (B) Non-response bias
- (C) Voluntary response bias
- (D) Response bias
- (E) There does not appear to be any bias.

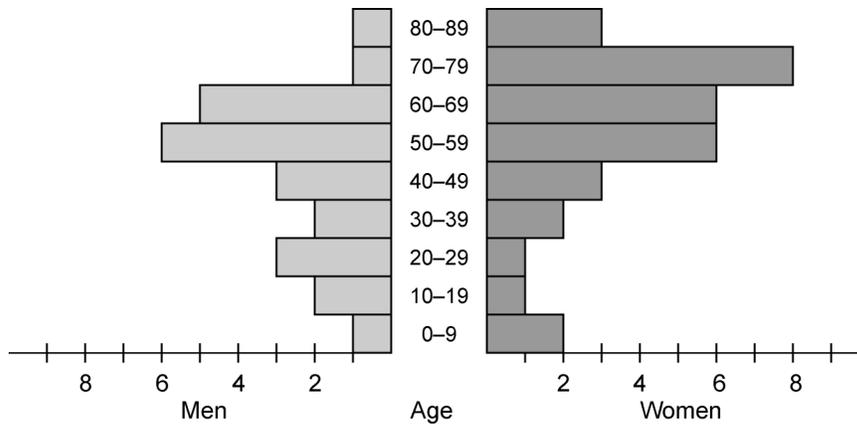
17. The scatterplot represents the scores of 10 students who took a first and second exam in a Biology course.



Which of the following is the best estimate of the mean of the second exam?

- (A) 25%
- (B) 40%
- (C) 50%
- (D) 60%
- (E) 75%

18. The population of a small town is represented by the back-to-back horizontal bar graphs below.



Based on the graph, how many more women than men are at least 50 years old?

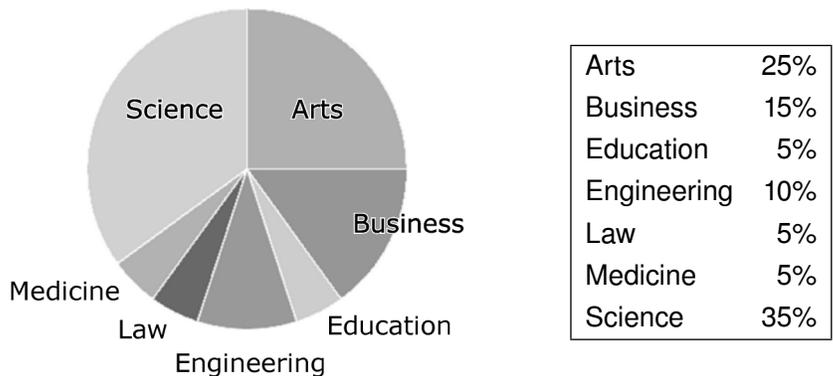
- (A) 6
 - (B) 10
 - (C) 13
 - (D) 17
 - (E) 46
19. A quarterback makes 22 passes in a game and you record the number of completions. The sample space for the number of completions you record is:
- (A) any number between 0 and 22.
 - (B) any integer between 1 and 22.
 - (C) the integers between 0 and 22.
 - (D) any integer greater than 0.
 - (E) any number between 0 and 22.

20. For their math project, two students conducted a survey on whether their school should expand the parking lot. During their presentation to the class, they explained that their data was obtained by going around the school and handing out surveys randomly. The teacher pointed out that the students really had a *convenience* sample rather than a *random* sample. What should the students have done to obtain a random sample?

21. A researcher wants to investigate whether different forms of exercise can be used to help hyperactive children. A group of 90 children is divided into two groups according to age: 5–9 and 10–12. Within each age division the children are randomly assigned to one of three groups. The first group will do just their normal exercise. The second group will be given an additional exercise routine, called moderate. The third group will be given an additional exercise routine, called strenuous. At the end of a four-month period parents will be asked to evaluate their children’s hyperactivity.

Determine whether the experiment is single blind, double blind, or neither. Justify your answer.

22. The pie chart below shows the percentage of graduating students with various majors at a university.



If there were 4800 students who majored in Arts and Business, then how many students majored in Science?

- (A) 3400
- (B) 3600
- (C) 3800
- (D) 4000
- (E) 4200

23. The following is a list of the salaries of employees at the Perfect Pop toaster company.

Employee Number	Salary	Employee Number	Salary	Employee Number	Salary
1	\$20,560	6	\$13,300	11	\$23,450
2	\$27,000	7	\$25,456	12	\$24,500
3	\$13,345	8	\$36,750	13	\$23,680
4	\$37,900	9	\$22,456	14	\$23,000
5	\$97,980	10	\$24,560	15	\$38,300

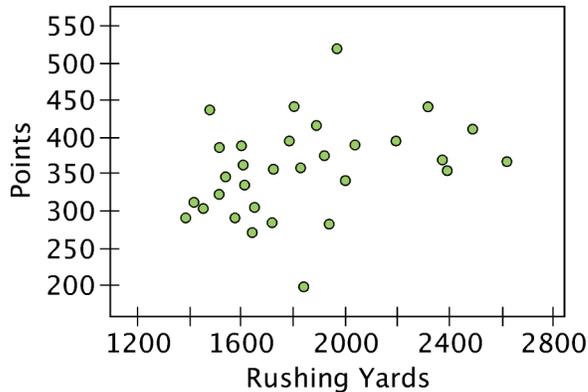
Draw a histogram using the following distribution of salaries.

- I. \$10000 to \$14999
- II. \$15000 to \$19999
- III. \$20000 to \$29999
- IV. \$30000 to \$34999
- V. \$35000 to \$39999
- VI. \$40000 to \$94999
- VII. \$95000 to \$99999

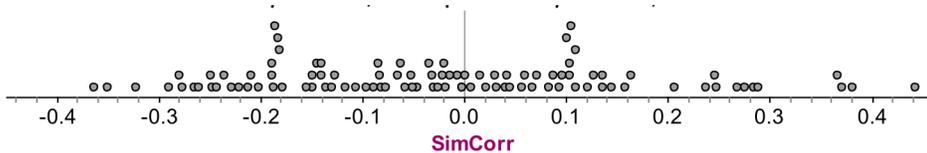
- a) What are the clusters of data observed from the graph?
- b) What are the gaps of data observed from the graph?
- c) Select the best reason for the gaps and clusters.
 - I. The Perfect Pop company tries to distribute the salaries evenly.
 - II. The Perfect pop company employs only the most experienced toaster makers.
 - III. Different positions receive different salaries.

24. A noodle machine in a spaghetti factory makes about 5 percent defective noodles even when properly adjusted. The noodles are then packed in boxes containing 1900 noodles each. A box is examined and found to contain 115 defective noodles. What is the approximate probability of finding at least this many defective noodles any other box selected at random?

25. The scatterplot shows the relationship between rushing yards and points scored for 32 NFL teams in the regular season. The correlation for these variables is $r = 0.35$.



- Describe the association between rushing yards and points scored.
- The team at the bottom of the scatterplot is the Panthers. What effect do the Panthers have on the correlation? Explain.
- A test is done on whether there is a positive association between rushing yards and points scored using the correlation as a test statistic. What hypotheses should we test?
- Describe how to simulate the distribution of the correlation. The true correlation is 0.
- Here are the results of 100 trials of the simulation described in the previous part. Use the results to estimate the p -value, interpret the p -value, and make a conclusion.



- If your conclusion was in error, which type of error did you make? Explain.

26. A juice drink manufacturer fills boxes advertised as containing 250 mL. A machine pours 251 mL into each box with a standard deviation of 0.3 mL. Any boxes containing less than 250.1 mL are rejected. How many boxes will be rejected in a production run of 25,000 boxes?

- 38
- 75
- 375
- 625
- 750

27. Compute the correlation coefficient for the following set of data.

Cost of burger (cents)	97	147	198	225
Amount of meat (oz.)	2	2.4	4	4.2

- (A) 0.997
- (B) 0.987
- (C) 0.977
- (D) 0.967
- (E) 0.957

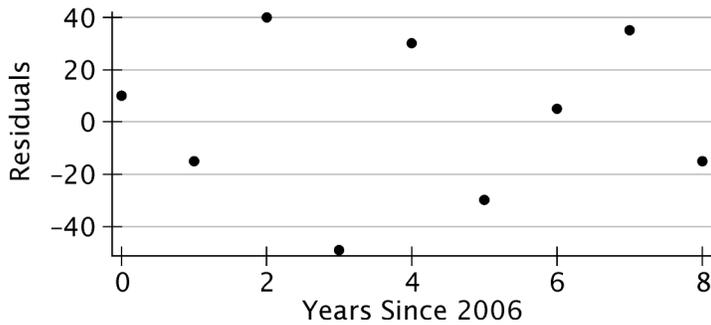
28. A recent study of literacy rates collected data from 107 different countries. Statistical results are shown in the tables below. X = percentage of people living in cities and Y = percentage of people who can read. Correlation is significant at the 0.01 level.

		People living in cities (X)	People who read (Y)
X	Pearson Correlation	1	0.650
	Sig. (2-tailed)		0.000
	N	107	107
Y	Pearson Correlation	0.650	1
	Sig. (2-tailed)	0.000	
	N	107	107

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	SE	Beta		
(Constant)	42.928	4.384		9.793	0.000
people living in cities (%)	0.622	0.071	0.650	8.754	0.000

- a) What is the value of the linear correlation coefficient between X and Y ?
- b) Is the linear correlation between X and Y significant? Explain. Use $\alpha = 0.05$
- c) Write the equation for the regression line.
- d) Find the coefficient of determination. Interpret.
- e) Find the percentage of people who can read in a country with 80% of population living in cities.

29. Bill and Janelle were searching online to find information on air travel in England. They found data on the number of commercial planes flying in England during the years 2006–2014. The dates were recorded as years since 2006. They fit a least squares regression line to the data. The graph of the residuals and partial computer output for the regression analysis are given below.



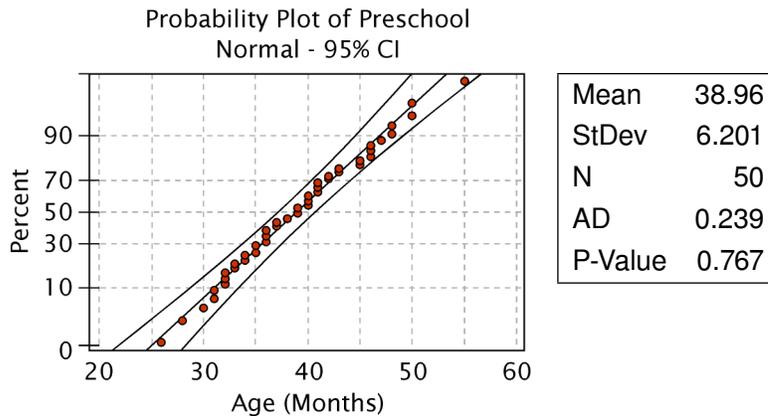
Variable	Coefficient	Std. Dev.	T	P
Constant	2939.93	20.55	143.09	0.000
Years	233.517	4.316	54.11	0.000

What is the actual number of commercial planes flying in 2008 (rounded to the nearest whole number)?

- (A) 2,472 flights
 (B) 2,513 flights
 (C) 3,022 flights
 (D) 3,367 flights
 (E) 3,447 flights
30. A company advertises that the overhead projector bulbs they produce have a life of 250 hr. A study of these bulbs indicates that their lives are normally distributed with a mean of 266.8 hr and a standard deviation of 8.4 hr. What percent of the bulbs do *not* last as long as the manufacturer claims?
- (A) 2.5%
 (B) 5%
 (C) 13.5%
 (D) 27%
 (E) 47.5%

31. A preschool recorded the ages (in months) when children were first enrolled. Data was collected for 50 children.

A normal quartile plot is shown below for the preschool's data.



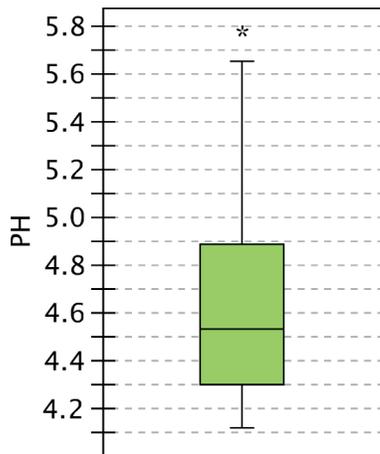
What can you conclude from the plot?

- (A) The data are skewed to the left.
 - (B) There is a strong linear relationship between age and gender.
 - (C) A normal distribution describes the data reasonably well.
 - (D) The data are skewed to the right.
 - (E) There are outliers.
32. Two drugs A and B, used in the treatment of glaucoma, were tested for effectiveness on 10 diseased dogs. Drug A was administered to one eye of each dog and drug B to the other eye. Pressure measurements were taken 1 hour later on both eyes of each dog. The smaller the measurement, the less serious the eye disease.

In the study described, which of the following is the most important?

- (A) We need to randomize the assignment of drugs to eyes.
- (B) We need to pair the dogs based on some relevant criteria related to the response.
- (C) We need to stratify the dogs before assigning the drugs.
- (D) We need to select the dogs randomly from a larger population.
- (E) We need to randomize the assignment of dogs to drugs.

33. Researchers measured the pH of water collected from precipitation in an Ohio town. pH is a measure of acidity; a value of 7 is neutral, and values below 7 are acidic. The results are shown in the boxplot below.



Describe the shape of the distribution of pH values.

- (A) Skewed to the left
 - (B) Skewed to the right
 - (C) Approximately symmetric
 - (D) Like a normal curve
 - (E) Cannot conclude anything about shape from a boxplot
34. A farmer grows watermelons. When they are ripe, his watermelons have weights that vary according to a normal distribution with mean 21 pounds and standard deviation 2.4 pounds. A supermarket will only buy watermelons that weigh between 18 and 23 pounds. What proportion of the farmer's watermelons will the supermarket buy?
- (A) 0.1
 - (B) 0.2
 - (C) 0.3
 - (D) 0.7
 - (E) 0.8

35. A researcher at a company wants to determine how employees regard the use of smartphones at work. The researcher emails 100 randomly selected employees of the company, and all 100 reply. The employees are asked, "Is it appropriate to use smartphones during work for non-work related purposes?" The survey results are shown in the table.

Question: Is it appropriate to use smartphones?	Number of employees
<i>No, unless it is an emergency</i>	27
<i>Yes, in small quantities</i>	21
<i>Yes, it's not a big deal</i>	52

The result that 52% believe that smartphone use at work is *not a big deal* is a:

- (A) parameter, because it summarizes information about the 100 employees in the sample.
 - (B) parameter, because it summarizes information about the population of all employees.
 - (C) statistic, because it summarizes information about the 100 employees in the sample.
 - (D) statistic, because it summarizes information about the population of all employees.
 - (E) conclusion, because it is the response given by the most number of employees.
36. It is known that 20% of all students in a particular school are varsity athletes. Their eating habits appear to be different. On any particular weekday, 50% of varsity athletes eat breakfast, compared to only 25% of all other students.
- A = the event that a student is a varsity athlete
 B = the event that a student eats breakfast on a particular weekday
- a) Are the events A and B independent? Justify your answer.
 - b) Are the events A and B disjoint? Justify your answer.
 - c) Find the overall proportion of students that eat breakfast on any particular weekday.
 - d) Given a student is eating breakfast on a particular weekday, what is the probability that that student is a varsity athlete?
 - e) In actuality, the non-varsity-athlete students are comprised of two further subgroups: 30% of them are club athletes and 70% are non-athletes. So there are actually 3 distinct groups in the student population: varsity athletes, club athletes, and non-athletes. Of the club athletes, 40% eat breakfast.
 - (i) What percent of non-athletes eat breakfast?
 - (ii) Given a student is eating breakfast, what is the probability of the student being a non-athlete?

37. A recent Gallup Poll found that only 17% of those surveyed approved of the job that Congress was doing. At the bottom of the webpage discussing the survey were these notes:

Results for this Gallup poll are based on telephone interviews conducted Nov. 4–7, 2010, with a random sample of 1,021 adults, aged 18 and older, living in the continental U.S., selected using random-digit-dial sampling.

For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error is ± 4 percentage points.

Which of the following is the best conclusion from the Gallup Poll notes?

- (A) Not everyone responds to surveys. This causes error in surveys. It is entirely plausible, because of this, that the true percentage might be as low as 13% or as high as 21%.
- (B) Somewhere between 133 (13% of 1021) and 214 (21% of 1021) of the people in their survey responded that they approved of the job Congress was doing.
- (C) If they were to repeat the survey again, using exactly the same procedures but with a different random-digit-dial sample, then they probably would get a result between 13% and 21%.
- (D) The practical difficulties of reaching a cross-section of the country's population means that no survey is completely accurate. Hence there will always some uncertainty, but it is 95% likely that somewhere between 13% and 21% of those surveyed approve of the job Congress is doing.
- (E) The surveyors are 95% confident that the statistic is accurate.

38. Which of the following are true?

- I. You can appropriately apply the binomial distribution if $p = 0.57$ and $q = 0.43$.
- II. You can appropriately apply the binomial distribution when $p = q$.
- III. The probability of 4 p -events when $n = 10$ and $p = 0.6$ is the same as 7 q -events if $q = 0.4$ and $n = 10$.
- IV. For fair coins, the probability of getting 5 heads out of a toss of 5 coins equals the probability of getting 5 tails out of a toss of 5 coins.

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

39. The houses in a city are sold at a rate of 1.02 per day and, on average, 13.3% of those houses are “old”, meaning they were built more than three decades ago.

A real estate agent noticed that, in comparison to the number of sales that occur each month, the number of old and new houses that are put on the market and the number of buyers and sellers are very large. Therefore, the agent adopts the following mathematical modeling assumptions:

- sales of old and new houses are independent of each another, and
- the number of sales, and the time until the next sale, are independent across time periods.

- a) Find the probability that exactly 1 of the next 7 houses sold will be “old”.
- b) Find the probability that exactly 8 houses will be sold in the next 7 days.
- c) What is the probability that it will be at least 7 days before the next house is sold?
- d) No houses are sold in April. What is the probability that no house will be sold in the first 7 days of May?

40. Individuals in a sample were classified by Age and Body Mass Index (BMI), with the following results.

Age in years	BMI		
	below healthy (<i>B</i>)	healthy (<i>H</i>)	above healthy (<i>A</i>)
18–44 (<i>Y</i>)	21	295	471
45–64 (<i>M</i>)	12	283	807
65 and over (<i>O</i>)	8	179	413

The events *Y* and *H* are:

- (A) independent, because Age and BMI are unrelated.
- (B) independent, because $P(Y \text{ and } H) \neq P(Y)P(H)$.
- (C) independent, because $P(Y) + P(H) \neq 1$.
- (D) dependent, because $P(Y \text{ and } H) \neq P(Y)P(H)$.
- (E) dependent, because $P(Y) + P(H) \neq 1$.

41. There are two scales for measuring weights in a chemistry lab. Both scales give answers that vary a little with repeated weighings of the same item. If the true weight of a compound is 2.00 grams, the first scale produces readings (X) that have a mean of 2.000 grams and a standard deviation of 0.002 grams. The second scale's readings (Y) have a mean of 2.001 grams and standard deviation of 0.001 grams.
- What are the mean and standard deviation of the difference $y - x$ between the readings? (The readings X and Y are independent.)
 - If an item is measured once with each scale and the weight is averaged, the result is $Z = \frac{(X + Y)}{2}$. What are the mean and standard deviation of Z ?
42. A test is used to help predict the performance of college applicants. The test is designed so that the mean score is 550 and the standard deviation is 100. A certain college will only accept applicants whose score on the test is in the top 3%. What score does an applicant need to achieve to be in the top 3%?
- 550 or above
 - 900 or above
 - 600 or above
 - 360 or above
 - 740 or above
43. To get a job as a firefighter, the height requirements are set at 165 cm to 198 cm for males, and 152 cm to 183 cm for females. Assume the populations are normally distributed. Men's heights have a mean of 178 cm with a standard deviation of 8 cm, and women's heights have a mean of 167 cm with a standard deviation of 7 cm. When applying for a firefighter job, which gender has the greater percentage meeting the height requirement, and what is the percent?
- women, 3.1%
 - women, 97.3%
 - women, 98.9%
 - men, 94.2%
 - men, 99.4%

44. Which *three* of the following statements may be properly justified by the Central Limit Theorem?
- I. The mean weight of apples in a simple random sample of 72 boxes of apples from a large population will be approximately normally distributed.
 - II. 2000 students take an exam. The IQR of the test scores for a simple random sample of 100 students selected from all 2000 students has an approximately normal distribution.
 - III. The proportion of part-time students at a university, as determined by a simple random sample of 800 students, is approximately normally distributed.
 - IV. A simple random sample of families is selected from all families in a city, and family income for the year 2016 is recorded for each. In the city, family incomes possess a moderately skewed distribution. If we select a simple random sample of 400 families and record each family's income, the histogram of these 400 incomes will be approximately normal.
 - V. The number of people who own a cellular phone in a simple random sample of 550 people, selected from all adults residing in a city, will be approximately normally distributed.
- (A) I, II, and III
 (B) II, IV, and V
 (C) I, IV and V
 (D) II, III and V
 (E) I, III and V

45. A study was conducted of the number of dens occupied by a particular species of fox over its lifetime. It was found that x , the number of dens occupied by an individual fox, has a distribution as follows:

x	1	2	10
$p(x)$	0.4	0.3	0.3

- a) For a given fox, what is the expected number of dens that it occupies, and what is the variance of the number of dens that it occupies?
- b) Foxes are called "unitary" if they occupy exactly one den during their lifetime. 75 foxes were sampled at random, and it was observed, for each fox, whether it was unitary or not. For these 75 foxes, find the approximate probability that 28 or fewer foxes were unitary.
- c) State the assumptions that you made for part (b), and verify them if possible.

46. The time it takes students to complete a biology final exam is a random variable having a normal distribution with mean 160 minutes and standard deviation 15 minutes.

Person A and Person B are two friends taking this exam. What is the probability that Person A will complete the exam at least ten minutes before Person B? (Assume all students start the exam at the same time and that their completion times are independent.)

- (A) less than 0.01
- (B) between 0.01 and 0.35
- (C) between 0.35 and 0.45
- (D) between 0.45 and 0.55
- (E) greater than 0.55

47. In metropolitan areas, it is unusual for more than 25% of the adult population to possess a current passport. Cities A and B have the highest percentages of adults who do possess a current passport. A survey by a travel agency was conducted to determine if there is a significant difference (at a 5% level) between the two cities. A random sample of 300 adults from each city revealed that 105 from City A and 90 from City B hold a current passport. The survey data is shown below.

Z Test for Two Proportions

	A	B
Sample Proportion	0.350000	0.300000
Number of Observations	300	300
H_0 : XXXX	H_a : XXXX	
Z^*	1.307441	
$2 * P[Z \geq Z^*]$ two tail	0.191063	
Z Critical, $\alpha = 0.05$	XXXX	
95% CI for $p_1 - p_2$	-0.024972 to 0.124972	

What is the point estimate of the proportion of adults in City B that do *not* hold a passport?

- (A) 0.3
- (B) 0.325
- (C) 0.65
- (D) 0.7
- (E) 1.307

48. A reporter suspects that gender and voting preferences are related, and that there are significant differences between men and women. He collects data from a random sample of 500 voters and classifies the respondents by gender (male or female) and by voting preference (A, B, or C). He runs a Chi-square test of independence at a 0.05 level of significance to test his hypothesis. Below is the sample data.

	A	B	C	Row total
Male	100	75	25	200
Female	125	150	25	300
Column total	225	225	50	500

What is the *alternative* hypothesis?

- (A) H_a : Voting preferences are dependent on gender
 - (B) H_a : Voting preferences are independent of gender
 - (C) H_a : gender and voting preference are dependent
 - (D) H_a : gender and voting preference are independent
 - (E) none of the above
49. A previous study by the Centers for Disease Control (CDC) reported that 15% of American adults smoke cigarettes. A researcher suspects that the percentage who smoke has decreased in the last 10 years. The researcher takes a random sample of 1,000 adults and uses the sample proportion of smokers to test the hypotheses:

$$H_0: p = 0.15$$

$$H_a: p < 0.15$$

In these hypotheses, what does p represent?

- (A) sample proportion of adults who quit smoking
- (B) sample proportion of adults who currently smoke
- (C) population proportion of adults who quit smoking
- (D) population proportion of adults who currently smoke
- (E) proportion of the 1,000 randomly selected adults who smoke

50. A random sample of 24 college students were surveyed to determine how long it takes to walk from the library to the education building. The survey team found a sample mean of 12.3 minutes with a standard deviation of 3.2 minutes. Assuming that walking times are normally distributed for the population, which of the following is a 95% confidence interval for the population mean of walking times?

(A) $12.3 - 1.645 \frac{3.2}{\sqrt{24}} \leq \mu \leq 12.3 + 1.645 \frac{3.2}{\sqrt{24}}$

(B) $12.3 - 2.069 \frac{3.2}{\sqrt{24}} \leq \mu \leq 12.3 + 2.069 \frac{3.2}{\sqrt{24}}$

(C) $12.3 - 2.500 \frac{3.2}{\sqrt{24}} \leq \mu \leq 12.3 + 2.500 \frac{3.2}{\sqrt{24}}$

(D) $12.3 - 2.064 \frac{3.2}{\sqrt{24}} \leq \mu \leq 12.3 + 2.064 \frac{3.2}{\sqrt{24}}$

(E) $12.3 - 1.96 \frac{3.2}{\sqrt{24}} \leq \mu \leq 12.3 + 1.96 \frac{3.2}{\sqrt{24}}$

51. Biologists have theorized that male Monarch butterflies have, on average, a larger thorax than do females. A sample of male and female butterflies yielded the data below. For these samples you can use $df = 22$.

	Male	Female
N	12	15
\bar{x}	70	65
s	9	8
SE	2.60	2.07

- a) Find a 90% confidence interval for the difference in the population means. At 90% confidence, can you conclude that male thorax size is different than female?
- b) At $\alpha = 0.10$ significance, use a t -test (and find the p -value) to determine whether we should reject the hypothesis that $H_0: \mu_{male} = \mu_{female}$ and accept the non-directional alternative $H_a: \mu_{male} \neq \mu_{female}$.
- c) How large does t_s need to be to conclude that $\mu_{male} \neq \mu_{female}$?
- d) At $\alpha = 0.10$ significance, use a t -test (and find the p -value) to determine whether we should reject the hypothesis that $H_0: \mu_{male} = \mu_{female}$ and accept the directional alternative $H_a: \mu_{male} > \mu_{female}$

52. The following set of data was collected to determine the effect of sleep deprivation on a student's ability to solve problems. The amount of sleep deprivation varied, with 8, 12, 16, 20 and 24 hours without sleep. A total of ten subjects participated in the study. A set of simple addition problems was administered to each subject after his or her sleep deprivation period, and the number of errors recorded. These results were obtained:

Number of Errors	8	6	6	10	8	14	14	12	16	12
Hours without Sleep	8	8	12	12	16	16	20	20	24	24

Perform a linear regression and analysis of variance for these data and fill in the blank spaces.

a) The formula for the regression line: $\hat{y} = \text{---} + \text{---}x$

b)

Source	DF	SS	MS
Regression			
Error			
Total		112.4	•

c) The coefficient of determination is $r^2 = \text{---}$

d) Determine a 95% confidence interval for the slope β of the linear model:

$\text{---} < \beta < \text{---}$

e) Compute the relevant t-statistic and p -value for the two-sided test of the hypothesis that $\beta = 0$:

$t = \text{---}$ $df = \text{---}$ $p\text{-value} = \text{---}$

f) Do you think these data support the thesis that sleep deprivation leads to decreased mental acuity?

53. A team of newspaper reporters studied the connection between the age in years of a licensed driver (x) and the percentage of fatal accidents for drivers of that age which were caused by speeding (y). They collected the following data:

x	17	27	37	47	57	67	77
y	36	25	20	12	10	7	5

$\sum x = 329$ $\sum y = 115$ $\sum x^2 = 18263$

$\sum y^2 = 2639$ $\sum xy = 4015$ $s_e = 3.455$

a) Find the equation of the regression (best fit) line relating x to y .

b) Find the coefficient of correlation and interpret its value in context.

c) Construct a 95% prediction interval for the percentage of fatal accidents caused by speeding for a 30-year-old driver.

54. Recently, the price of gasoline was near the four-dollar mark in many parts of the United States. A survey indicated that 32% of Americans curtailed their travel plans because of the price of gasoline. In fact, some experts claim that the number was even higher. A survey was done to test this claim. The survey revealed that 175 of 500 respondents were planning on traveling less because of high gasoline prices.

What conclusion can be drawn at the 10% level of significance?

- (A) Reject H_a . There is evidence that supports the claim.
- (B) Reject H_0 . The claim is not supported.
- (C) Reject H_0 . There is evidence that supports the claim.
- (D) Fail to reject H_0 . There is evidence that supports the claim.
- (E) Fail to reject H_a . The claim is not supported.

55. An economist working for the government wants to test if there is any difference in income between County A and County B. To do this, she collects information from random samples of residents from both counties. The information is tabulated below. Assume that both populations are normally distributed with *equal* population variances.

County A (Population 1)	County B (Population 2)
$n_1 = 11$	$n_2 = 15$
$\bar{x}_1 = \$36,700$	$\bar{x}_2 = \$34,700$
$s_1 = \$7800$	$s_2 = \$7375$

At 5% significance level, find the rejection region and state your decision.

- (A) Reject H_0 if $t < -1.711$. Decision: Reject H_0 .
- (B) Reject H_0 if $t < -2.064$. Decision: Reject H_0 .
- (C) Reject H_0 if $t < -2.064$ or $t > 2.064$. Decision: Do not reject H_0 .
- (D) Reject H_0 if $t < -1.711$ or $t > 1.711$. Decision: Do not reject H_0 .
- (E) Reject H_0 if $t > 2.064$. Decision: Do not reject H_0 .

56. The table below shows the number of individuals visiting a doctor's office who have either an ear, nose or throat irritation. (None have more than one.)

Irritation	Age 18–44	Age 45 and older
Ear	80	60
Nose	200	90
Throat	40	30

Is the type of irritation independent of age group? Test at $\alpha = 0.05$.

57. For the past few years the average score on a common final exam in Psychology has been 80. In a sample of 92 students, the mean score on this year's exam was 82 with standard deviation 8.7. Test at a 5% level of significance whether the mean score on this year's exam is higher than that of previous years.

- State the null and the alternative hypotheses.
- Write down the formula of the test statistics and find its value.
- Determine the rejection region and make a decision.
- State your conclusion in context.

58. Consider the following pairs of observations on the variables x and y :

x	0	1	1	2	3	4	5	6	6	7
y	6	6	7	4	5	2	3	0	1	1

From the data one obtains:

$$SS_{xx} = 54.5, \quad SS_{yy} = 54.5, \quad SS_{xy} = -50.5, \quad s_e = 0.98148$$

- Draw a scatter diagram for the data.
- Calculate b the slope of the regression line.
- Perform a hypothesis test to determine if B , the slope of the population regression line, is different from 0.

59. The United Nations collected data on “school life expectancy”, which is the expected number of years of formal education, from a sample of 54 countries. The data is shown in the table below.

11	8	15	12	11	12
11	14	15	5	6	11
16	14	13	14	13	9
11	19	12	10	13	5
13	7	3	14	13	16
20	15	5	15	15	5
15	7	9	11	17	18
10	14	9	7	4	15
13	12	16	8	12	12

- a) Construct a frequency distribution with the following characteristics:
- A “reasonable” number of classes. Describe briefly how you arrived at this number. Mathematically defensible methods are preferred to “ad hoc” methods.
 - Give the *frequency, relative frequency, cumulative frequency, and relative cumulative frequency* for each class. (i.e., a standard complete frequency distribution)
 - Make a complete stem-and-leaf diagram for these data.
- b) Draw one of the following graphs to represent the data. Why did you chose to use this type of graph? What does it show about the data and the data-generating process?
- histogram
 - frequency polygon
 - less-than-cumulative frequency ogive
 - more-than-cumulative frequency ogive
 - pie graph
 - box-and-whisker plot

60. Which of the following statements about the t -distribution is *not* correct?

- The t -distribution has a mean of 0.
- The t -distribution is symmetric.
- The t -distribution can be used when samples come from a population with approximately normal distribution.
- The t -distribution has less variability than the standard normal distribution.
- The t -distribution approaches the standard normal distribution as the degrees of freedom increases.