Sample Term Test 1A

1. Which of the following variables is **not** categorical and ordinal?

   (A) Purity of gold (10 karat, 14 karat, 18 karat, 24 karat, etc.)

   (B) First three numbers of a phone number (338, 667, 791, etc.)

   (C) Education Level (High School, Bachelor's Degree, Master's, Phd, etc.)

   (D) Colour of medal won by an athlete (Gold, Silver, Bronze)

   (E) Size of bed (Twin, Double, Queen, King, etc.)

2. A golfer plays a particular course 50 times one summer. His scores are represented in the frequency distribution below:

Score	Frequency
72–74	1
75–77	3
78–80	4
81–83	8
84–87	10
88–90	15
91–93	9

   The distribution of scores is:

   (A) skewed to the left and so the mean is greater than the median.

   (B) skewed to the right and so the mean is greater than the median.

   (C) skewed to the left and so the median is greater than the mean.

   (D) skewed to the right and so the median is greater than the mean.

   (E) approximately symmetric and so the mean and median are approximately equal.

The next **three** questions (**3, 4 and 5**) refer to the following:

The house values (in thousands of dollars) for samples of homes from three Calgary neighbourhoods are recorded. Some summary statistics are shown in the table below:

	# of homes	min.	Q1	med.	Q3	max.	mean	std. dev.
Evergreen:	180	160	300	355	480	615	367	122
Royal Oak:	153	250	495	630	715	820	581	160
McKenzie Lake:	174	170	320	370	400	590	355	89

3. How many homes in the sample from the neighbourhood of Evergreen are worth at least $300,000?

   (A) 75 (B) 135 (C) 45 (D) 120 (E) 150

4. What is the mean value of all homes in the samples from all three neighbourhoods combined?

   (A) $438,704
   (B) $440,928
   (C) $434,333
   (D) $443,136
   (E) $427,462

5. If we were to construct a modified (outlier) boxplot for the home values in McKenzie Lake, the value of a house would be labeled as a suspected outlier if it was:

   (A) less than $320,000 or greater than $400,000.
   (B) less than $250,000 or greater than $490,000.
   (C) less than $200,000 or greater than $520,000.
   (D) less than $180,000 or greater than $540,000.
   (E) less than $240,000 or greater than $480,000.

6. The stemplot below displays the number of points scored by the Toronto Raptors for each of their 41 homes games in the 2008/2009 NBA season.

7	6
8	5 6 7 8 9
9	0 1 1 3 3 3 4 5 6 7 8 8
10	0 1 1 2 3 3 3 7 7 7 8
11	0 0 1 1 2 2 3 3 5 8
12	7
13	4

What is the interquartile range for points scored by the Raptors in these games?

(A) 14.5 (B) 16 (C) 16.5 (D) 17 (E) 17.5

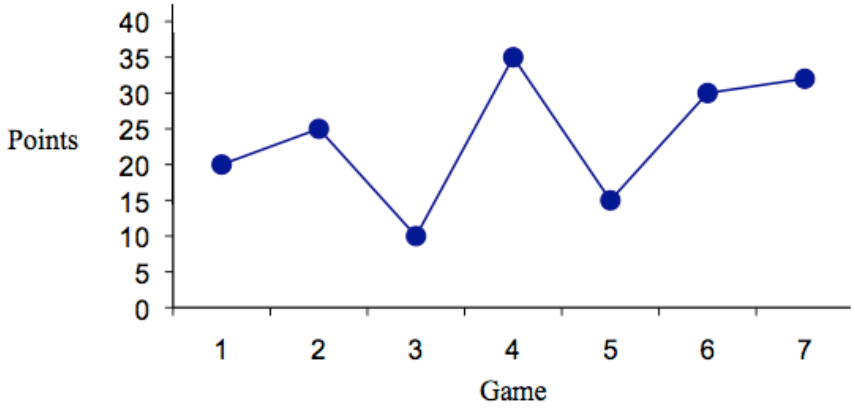
7. Big Blossom Greenhouse was commissioned to develop an extra large rose for the Rose Bowl Parade. A random sample of five blossoms yielded the following diameters (in inches) for mature peak blossoms:

6 10 3 7 4

What is the value of the sample standard deviation of diameters for these blossoms?

(A) 2.74 (B) 3.06 (C) 7.50 (D) 2.45 (E) 3.37

8. The timeplot below displays the number of points scored by a basketball player in her first seven games of the season:



The median number of points scored by the player in her first seven games is:

(A) 35 (B) 20 (C) 15 (D) 25 (E) 30

9. There are four people in a room. The mean and standard deviation of the ages of the four people are 32 years and 8.62 years, respectively. Two new people, aged 20 and 44, enter the room. What is the new standard deviation of the ages of all people now in the room?

(A) 8.62 (B) 9.47 (C) 10.11 (D) 11.36 (E) 12.04

10. Jakob, a student in a large class with 200 students, receives a score of 80% on his exam. Jakob learns that his score is at the 70^{th} percentile in the class. Which of the following statements is true?

(A) About 40 students scored lower than Jakob.

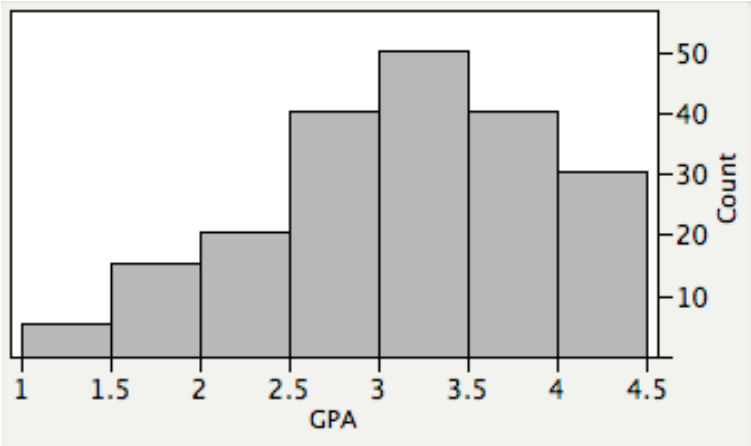
(B) About 70 students scored lower than Jakob.

(C) About 140 students scored higher than Jakob.

(D) About 160 students scored lower than Jakob.

(E) About 60 students scored higher than Jakob.

11. The histogram shown below displays the GPAs for a class of 200 students:

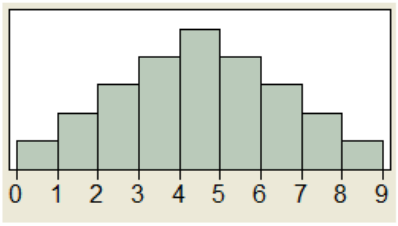


What proportion of students in the class have a GPA less than 3.0?

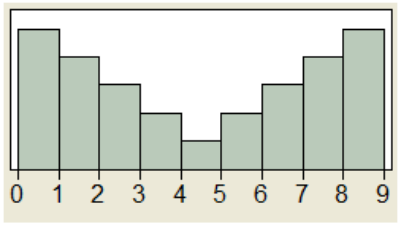
(A) 0.20 (B) 0.30 (C) 0.35 (D) 0.40 (E) 0.45

12. Which of the following histograms displays the dataset with the smallest standard deviation?

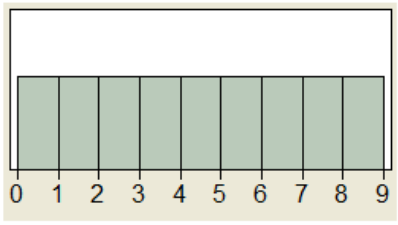
(A)



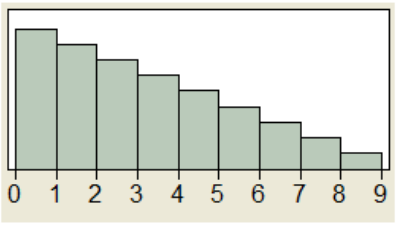
(B)



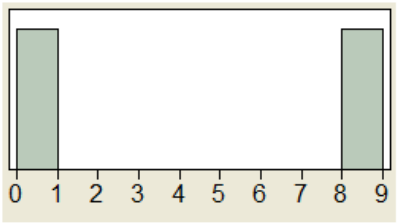
(C)



(D)



(E)



13. The mean ($\bar{x}$), median (M), standard deviation (s), range (R), interquartile range (IQR) and correlation (r) are all measures of center, spread or association. Which of these are resistant to the presence of outliers?

(A) R and IQR

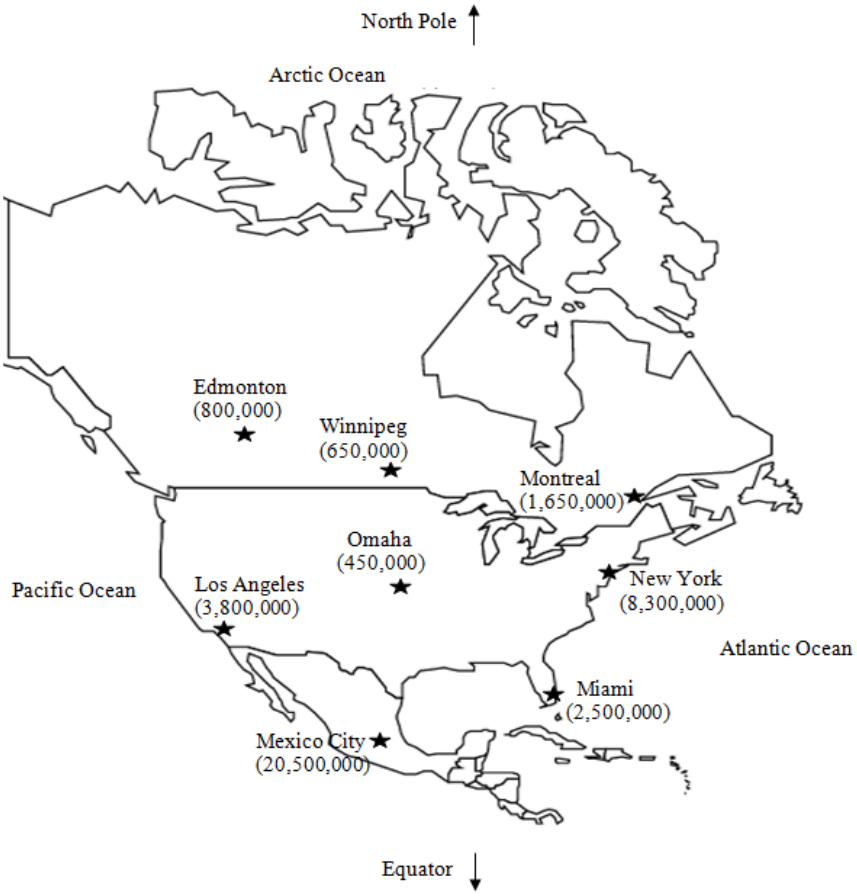
(B) M and R

(C) $\bar{x}$, s and r

(D) M, IQR and r

(E) M and IQR

14. A random sample of eight North American cities is selected. The cities (with their populations) are shown on the map below:



Consider the following pairs of variables measured for these cities:

(I) X = distance from the North Pole
Y = average January temperature

(II) X = distance from the equator
Y = average July temperature

(III) X = distance from the nearest ocean
Y = population

The correlations for these three pairs of variables for these cities are, respectively:

(A) negative, positive, negative

(B) negative, positive, positive

(C) positive, positive, negative

(D) positive, negative, positive

(E) positive, negative, negative

15. Which of the following studies is the only one that could produce a correlation close to $r = 0.5$?

(A) Select a sample of commercial airline flights leaving from the airport one day:
X = flight distance in kilometers; Y = flight distance in miles

(B) Select a sample of North American cities:
X = distance from the equator; Y = average January temperature

(C) Select a sample of STAT 1000 students:
X = number of incorrect answers on this test; Y = score on the test

(D) Select a sample of adults in Winnipeg:
X = IQ; Y = weight in kilograms

(E) Select a sample of male students from the University of Manitoba:
X = height in inches; Y = shoe size

16. A random sample of adults is selected and their height X (in cm) and their weight Y (in kg) are measured. The correlation between height and weight is calculated to be $r = 0.54$. What would be the value of the correlation if height had instead been measured in inches (1 cm = 0.4 inches) and weight had instead been measured in pounds (1 kg = 2.2 pounds)?

(A) 0.54 (B) 0.48 (C) 0.72 (D) 0.61 (E) 0.58

17. A sample of U of M students is selected. The Distance X (in km) between a student's place of residence and the Time Y (in minutes) it takes them to get to the university are recorded. The least squares regression line is calculated to be $\hat{y} = 2.3 + 1.7x$. The correlation between Distance and Time is calculated to be $r = 0.83$. If we had instead measured Distance in miles, which of the following values would change? (1 mile = 1.61 km)

(I) slope
(II) intercept
(III) correlation

(A) I only

(B) II only

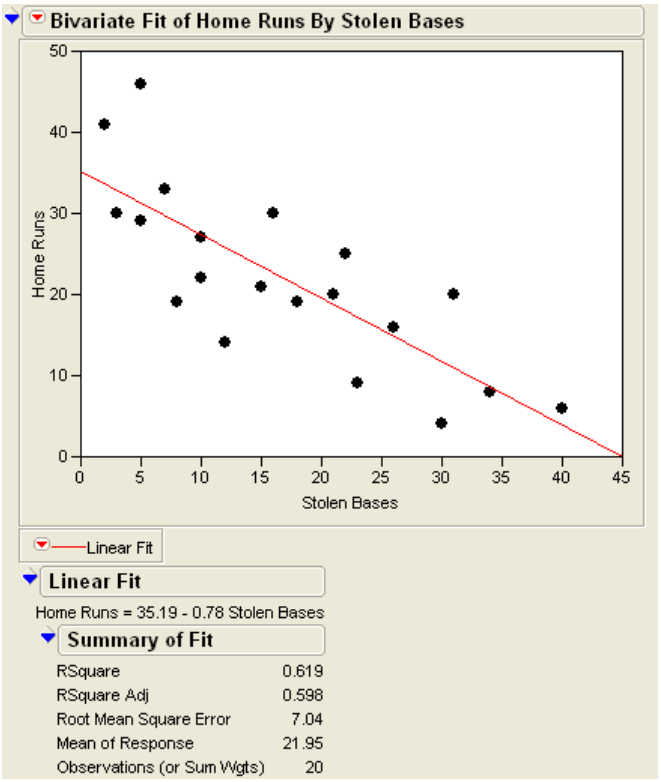
(C) I and II only

(D) I and III only

(E) II and III only

The next **four** questions (**18** to **21**) refer to the following:

We would like to examine the relationship between the power and speed of professional baseball players. Two important statistics in baseball are home runs and stolen bases. Strong hitters will hit many home runs, and fast runners will steal many bases. A random sample of 20 Major League Baseball players is selected, and the number of bases the player stole (X) and the number of home runs the player hit (Y) last season are recorded. The data are shown in the scatterplot below, as well as some regression output from *JMP*:



18. What is the correct interpretation of the slope of the least squares regression line?

(A) When the number of stolen bases increases by one, we predict an increase of 0.78 home runs.

(B) When the number of home runs increases by one, we predict a decrease of 0.78 stolen bases.

(C) When the number of stolen bases increases by 0.78, we predict a decrease of one home run.

(D) When the number of home runs increases by 0.78, we predict a decrease of one stolen base.

(E) When the number of stolen bases increases by one, we predict a decrease of 0.78 home runs.

19. What is the correlation between number of stolen bases and number of home runs?

(A) 0.787 (B) −0.619 (C) −0.598 (D) 0.619 (E) −0.787

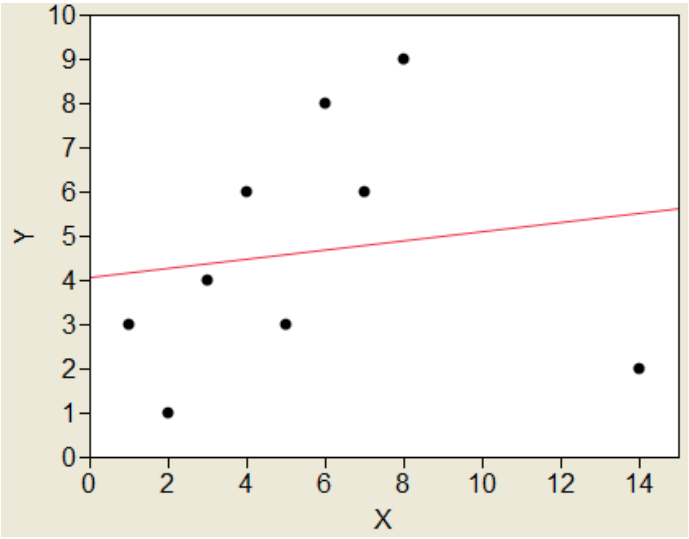
20. What percentage of the variation in number of home runs can be accounted for by its least squares regression on number of stolen bases?

(A) 78.7% (B) 59.8% (C) 35.2% (D) 61.9% (E) 78.0%

21. One player in the sample stole 30 bases and hit 4 home runs. What is the value of the residual for this player?

(A) 7.79 (B) –7.79 (C) –2.07 (D) 11.79 (E) –11.79

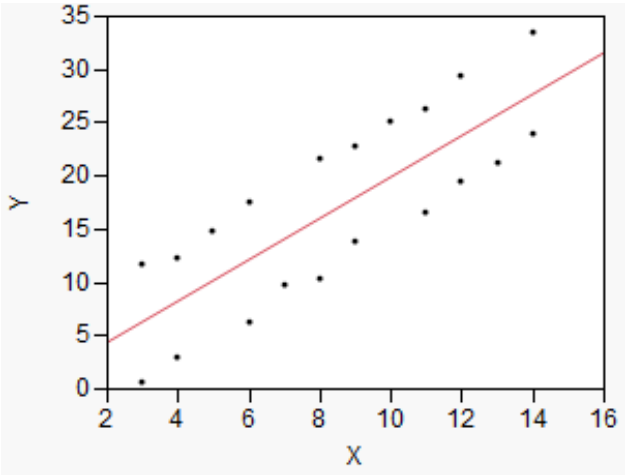
22. We measure the values of an explanatory variable X and a response variable Y on a sample of individuals. The scatterplot and the least squares regression line are shown below:



If we remove the influential observation and recalculate the regression line, what will happen to the values of the slope and the intercept?

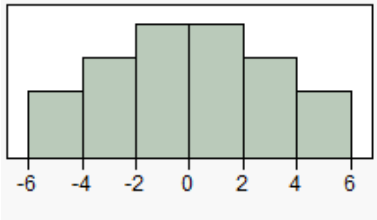
(A) The slope will increase and the intercept will decrease.
(B) Both the slope and the intercept will decrease.
(C) The slope will decrease and the intercept will increase.
(D) Both the slope and the intercept will increase.
(E) The slope will increase and the intercept will stay the same.

23. The values of an explanatory variable X and a response variable Y are measured on a sample of individuals. The scatterplot of the data is shown below:

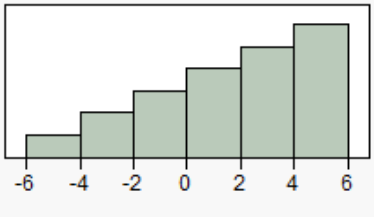


Which of the following is a histogram of the residuals?

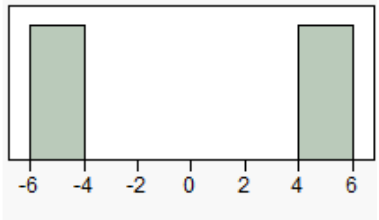
(A)



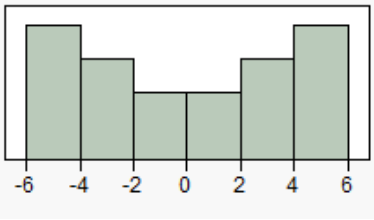
(B)



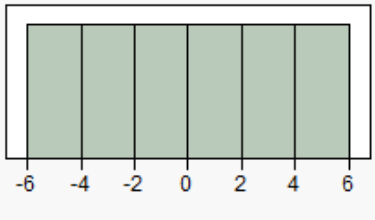
(C)



(D)



(E)



24. Consider the following cartoon:

Suppose that a table of random digits had instead been used to select a random six-digit number. What can be said about the possibility of obtaining the string of digits 999999?

(A) The six-digit number 999999 would never be found on a random digits table, as it is not random.

(B) The six-digit number 999999 could be obtained, as it is still more random than some six-digit numbers, like 123456.

(C) If we did see six consecutive 9's, then the next digit would probably be a 9 as well.

(D) If we did see six consecutive 9's, then the next digit would have almost no chance of being a 9.

(E) Although you might be surprised to see the six-digit number 999999, it is just as likely as any other six-digit number.

The next **two** questions (**25** and **26**) refer to the following:

Former Manitoba Premier Gary Doer was recently appointed as Canadian Ambassador to the United States. His office commissions a telephone survey in order to study the relations between the people of the two countries. A simple random sample of 20 people is contacted from each of the 50 American states. Respondents are asked whether they have a favourable opinion of Canadians.

25. The population of interest in this study is:

(A) all Canadians.

(B) all Americans.

(C) all Americans who participate in the survey.

(D) all Americans with a favourable opinion of Canadians.

(E) the 50 American states.

26. The resulting sample of 1000 Americans is a:

(A) stratified sample.
(B) multistage sample.
(C) simple random sample.
(D) randomized block sample.
(E) systematic sample.

27. A teacher is travelling to Europe for her summer holidays, but she wont have enough time to visit all the countries she would like to see. She estimates that she will have enough time to visit six countries. She makes a list of all the countries she would like to visit. They are numbered and shown below:

01 – Austria	11 – Hungary	21 – Romania
02 – Belgium	12 – Iceland	22 – Slovakia
03 – Bulgaria	13 – Ireland	23 – Spain
04 – Croatia	14 – Italy	24 – Sweden
05 – Czech Republic	15 – Latvia	25 – Switzerland
06 – Denmark	16 – Moldova	26 – Turkey
07 – Finland	17 – Netherlands	27 – Ukraine
08 – France	18 – Norway	28 – United Kingdom
09 – Germany	19 – Poland	
10 – Greece	20 – Portugal	

You will use the string of random digits below (starting at the left) to select a simple random sample of six countries for the teacher to visit:

29581 04637 51192 56043 86104 22736 02195 31522 37640 89045

Which is the last (sixth) country selected into the sample?

(A) Latvia (B) Slovakia (C) Poland (D) France (E) Belgium

28. A conservative organization contacts a simple random sample of 500 Manitoba voters, asking them the following question:

"Given the failures of the current NDP government on health care and education, will you support the New Democrats in the next election?"

34% of those responding answered yes. Which of the following statements is **true**?

(A) This survey is reasonably accurate, as it used a fairly large, random sample of individuals.

(B) This survey probably overestimates the true proportion of Manitobans who support the NDP.

(C) This survey probably underestimates the true proportion of Manitobans who support the NDP.

(D) This survey is very inaccurate, but neither overestimates nor underestimates the true proportion of Manitobans who support the NDP. Because simple random sampling was used, the results are unbiased.

(E) The results of this survey cannot be trusted, as not all Manitoba voters were contacted.

29. A pharmaceutical company would like to compare the effect of different dosages of a certain medication on the alleviation of headache pain in patients. One third of the 240 patients who volunteer for the study will be given a pill which contains 20 mg of the drug, one third will be given a pill with 10 mg of the drug, and one third will be given a sugar pill – one with no drug present. All of the pills will look the same and the doctor administering the medication will not tell the patients which pill they are getting. After one week of taking the medication, the patients reported pain relief will be compared for the three groups. The company anticipates that adults of different ages may respond differently to the medication, and so patients are separated prior to the study into two groups, those over 40 and those 40 and under. The experiment is then conducted separately for the two groups.

Which of the following is **not** used in this experiment?

(A) a placebo

(B) blocking

(C) double-blinding

(D) repetition

(E) All of the above are used in this experiment.

The next **two** questions (**30** and **31**) refer to the following:

A tire manufacturer would like to determine how weather conditions and speed affect the stopping distance of a vehicle using a certain type of tire. An experiment will be conducted on a test track under simulated weather conditions (either dry, rain or snow) and at different speeds (40 km/h, 60 km/h, 80 km/h or 100 km/h). Each combination of factor levels will be tested in three trial runs, and the same vehicle and the same set of tires will be used for each run.

30. What is/are the factor(s) in this experiment?

(A) stopping distance

(B) weather conditions and speed

(C) dry, rain, snow, 40 km/h, 60 km/h, 80 km/h, 100 km/h

(D) type of vehicle

(E) type of tire

31. How many treatments are used in this experiment?

(A) 12 (B) 7 (C) 36 (D) 10 (E) 15

32. Every time the fire department receives a 911 call, the number of fire fighters responding and the amount of damage (in $) are recorded. There is a high positive correlation between the two variables. Does this mean it is better to have fewer fire fighters respond to a fire call? Of course not! Which of the following is a lurking variable in this case?

(A) size of the fire

(B) distance from the fire hall to the fire

(C) neighbourhood

(D) value of the property

(E) all of the above

33. Gordon and his wife Michelle are avid gardeners who take pride in maintaining a healthy green lawn. Over the course of the last summer, Gordon was in charge of maintaining the grass in the back yard of their home and Michelle cared for the lawn in the front yard. Gordon used one brand of fertilizer (Brand A) and watered the grass once a week. Michelle used a different fertilizer (Brand B) and watered the grass twice a week. At the end of the summer, it was apparent that the grass in the front yard looked much greener and healthier than the grass in the back yard.

Which of the following can be concluded?

(I) Fertilizer B is better than Fertilizer A **and** watering the grass twice a week is more effective than watering the grass once a week.
(II) Either Fertilizer B is better than Fertilizer A **or** watering the grass twice a week is more effective than watering the grass once a week (but not necessarily both).
(III) Michelle cares for the lawn better than Gordon does.
(IV) Nothing meaningful can be concluded because this is an observational study.

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

34. A study was conducted and it was reported that a new treatment had been found to be significantly better than an existing treatment. Later, a critic commented that, in fact, nothing could be usefully determined from the study because the effect of the treatment on the response variable was confounded. What did the critic mean by this?

(A) The results were so badly described that the critic was unable to determine what was being concluded.
(B) The critic felt confusion as to whether the study was an observational study or a real experiment.
(C) Because of the way the study was conducted, the effect of the treatment described was mixed up with one or more other variables.
(D) The purpose of the study was not clearly described, so the results were meaningless.
(E) The results of this study contradict the results of a previous study.

35. Two drugs, Drug A and Drug B, are intended to lower the blood pressure in hypertensive patients. Suppose that you are going to use a matched pairs design to compare the effectiveness of the two drugs, and that you have 60 subjects available for your experiment. In this case, you would:

(A) randomly divide the 60 subjects into two groups, giving Drug A to the subjects in one group and Drug B to the subjects in the other group.

(B) randomly divide the 60 subjects into 30 pairs, and then flip a coin decide which patient in each pair would receive Drug A or Drug B.

(C) randomly divide the 60 subjects into 30 pairs, and then randomly select 15 of the pairs to receive Drug A, with the remaining 15 pairs receiving Drug B.

(D) subjectively divide the 60 subjects into two groups according to their characteristics (such as degree of hypertension), and then randomly give Drug A to half of the subjects in each group, and Drug B to the other half.

(E) use characteristics of the subjects (such as degree of hypertension) to subjectively divide the 60 subjects into 30 pairs (making the subjects within a pair as similar as possible), and then flip a coin to decide which patient in each pair would receive Drug A or Drug B.

Sample Term Test 1A Answer Key

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

Sample Term Test 1B

1. Consider the following variables:

   (i) Area code (Edmonton – 780, Winnipeg – 204, Montreal – 514, etc.)
   (ii) Weight class of a professional boxer (lightweight, middleweight, heavyweight, etc.)
   (iii) Office number of a Statistics professor in Machray Hall

   These three variables are, respectively:

   (A) categorical and ordinal, categorical and ordinal, categorical and nominal
   (B) categorical and nominal, categorical and ordinal, categorical and nominal
   (C) categorical and nominal, categorical and nominal, categorical and ordinal
   (D) categorical and ordinal, categorical and nominal, quantitative
   (E) categorical and nominal, categorical and ordinal, categorical and ordinal

2. The mean GPA in a class of 14 students is 3.12. One student, whose GPA is 2.41, drops the class. Two new students, whose GPAs are 3.97 and 4.26, join the class. What is the new mean GPA of the class?

   (A) 3.54 (B) 3.30 (C) 3.44 (D) 3.23 (E) 3.61

3. The following stemplot displays the number of forest fires that occurred each year over the period from 1970–2000 in Canada:

5	2
6	0 1 3 7
7	3 6
8	0 1 2 5 9 9 9
9	1 1 2 2 3 4 5 8
10	0 1 2 2 3 7
11	2 3
12	2

where 5|2 represents 5,200 fires. What is the value of the third quartile for this data set?

(A) 10,050 (B) 10,200 (C) 10,150 (D) 10,000 (E) 10,100

4. A physician wants to study the fasting plasma glucose levels of all his male patients who weigh over 200 pounds. To plot a histogram (with vertical bars) of the glucose levels, the horizontal and vertical axes would represent, respectively:

(A) glucose level and count.

(B) weight and count.

(C) weight and glucose level.

(D) glucose level and weight.

(E) count and glucose level.

5. Which of the following statements about the standard deviation is **false**?

(A) The standard deviation has the same units of measurements as the original observations.

(B) The standard deviation of a set of observations is the average of the squares of the deviations from their mean.

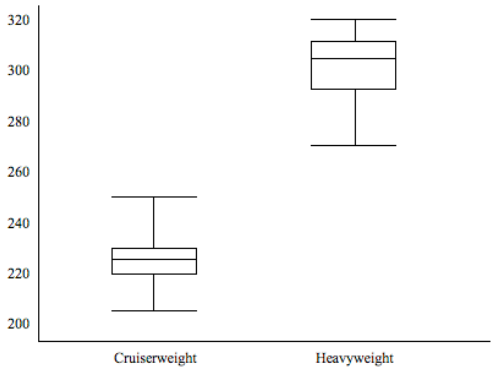
(C) If the standard deviation is zero then all observations have the same value.

(D) The standard deviation should be used as the measure of variability when the mean is chosen as the measure of center of the distribution.

(E) The standard deviation should not be used as a measure of spread when the distribution is strongly skewed.

The next **two** questions (**6** and **7**) refer to the following:

The sport of boxing divides its athletes into different weight classes in order to make the competition fairer. The side-by-side basic (quantile) boxplots shown below display the weights (in pounds) of a random sample of 16 Cruiserweight boxers and 17 Heavyweight boxers.



The five number summaries for the two weight classes are shown below:

	Minimum	Q1	Median	Q3	Maximum
Cruiserweight	204	220	226	230	250
Heavyweight	270	294	304	312	320

6. Which of the following statements is/are true?

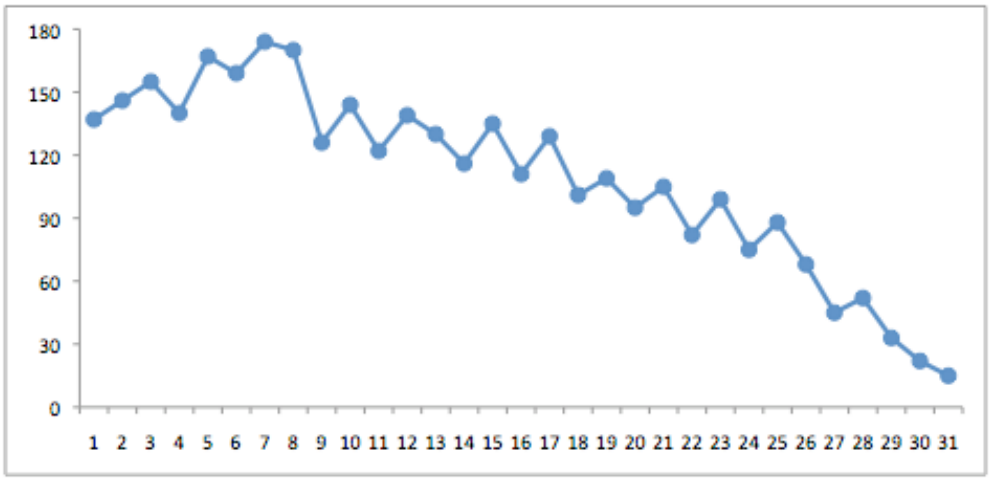
I. The distribution of weights for the Heavyweights is skewed to the left.
II. There are 12 Cruiserweights in the sample who weigh at least 220 pounds.
III. The mean weight for the Heavyweights is likely greater than the median weight.

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

7. What is the median weight of all of the boxers in the sample (Cruiserweight and Heavyweight) combined?

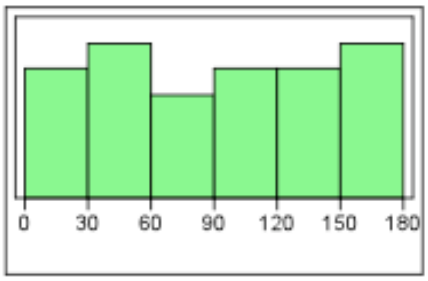
(A) 260 (B) 262 (C) 265 (D) 268 (E) 270

8. A university student records the time (in minutes) she spent studying each day for a month. The data are displayed in the time plot shown below:

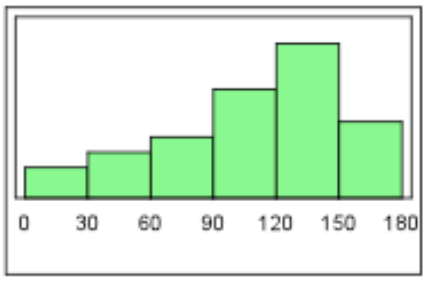


Which of the following is the appropriate histogram for this data set?

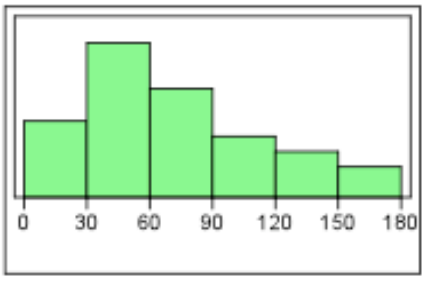
(A)



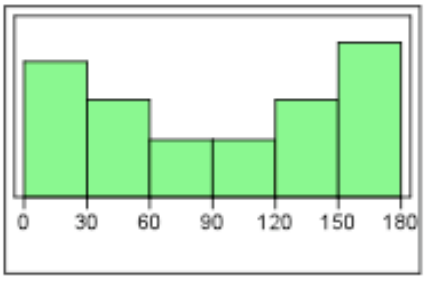
(B)



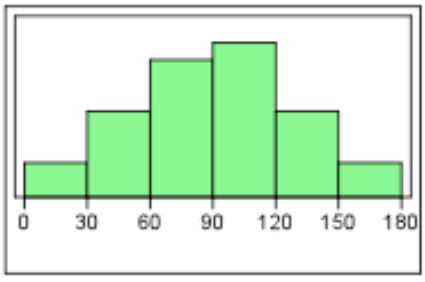
(C)



(D)



(E)



9. The weights (in kilogrtams) of ten boxes being shipped by a furniture company are shown below:

144 132 H 151 127 138 H 123 H 136

where H indicates that the weight of the box is too high to be recorded by the scale. What is the median weight (in kilograms) of the boxes?

(A) 132.5 (B) $\frac{951 + 3H}{10}$ (C) 141 (D) 138 (E) cannot be computed

10. The University of Manitoba uses a grade point system with a maximum possible Grade Point Average of 4.5. For each course, a student is awarded the following number of grade points per credit hour for each of the possible letter grades:

Letter Grade	Grade Points
A^+	4.5
A	4.0
B^+	3.5
B	3.0
C^+	2.5
C	2.0
D	1.0
F	0.0

One year, a student took the following courses, and received the letter grades shown:

Course	Credit Hours	Letter Grade
CIVL 2830	2	B^+
HIST 1260	3	A^+
ECE 2130	4	A
ECE 2160	5	C
COMP 2061	6	B

What was the student's GPA for the year?

(A) 3.265 (B) 3.375 (C) 3.400 (D) 3.225 (E) 3.175

11. The shower flow rates (in L/min) for a sample of 40 houses are ordered and shown below:

2.7	2.8	3.1	3.5	4.6	5.0	5.7	5.7	5.8	5.9
6.1	6.1	6.1	6.3	6.3	6.4	6.6	6.8	6.9	7.0
7.0	7.1	7.1	7.2	7.2	7.4	7.5	7.5	7.5	7.6
8.0	8.2	8.6	9.0	9.5	9.9	10.2	10.7	10.8	11.2

The five-number summary for this data set is as follows: 2.7 6.0 7.0 7.8 11.2

If we constructed an outlier boxplot for this data set, the lines coming out from the box (the whiskers) would extend to which values?

(A) 3.5 and 10.2

(B) 3.3 and 10.5

(C) 2.7 and 11.2

(D) 4.2 and 9.6

(E) 3.1 and 10.7

12. Determine whether the correlation for each of the following pairs of variables is positive or negative:

I. X = Speed of wind in a snowstorm
   Y = Visibility

II. X = Global supply of oil
   Y = Price of gasoline

III. X = Number of people in line at a bank when you arrive
   Y = Time spent waiting in line

(A) I. negative, II. positive, III. positive

(B) I. positive, II. positive, III. positive

(C) I. negative, II. negative, III. positive

(D) I. positive, II. negative, III. negative

(E) I. negative, II. negative, III. negative

The next **two** questions (**13** and **14**) refer to the following:

We would like to determine whether a man's shoe size X can be used to predict his height Y. The shoe sizes and heights of a random sample of eight men are shown below:

Shoe Size	11	10	9.5	12	11	11.5	10.5	10
Height (inches)	69	70	67	74	72	70	71	68

The correlation between Shoe Size and Height is calculated to be $r = 0.78$, and the equation of the least-squares regression line is calculated to be $\hat{y} = 50 + 2x$.

13. What is the correct interpretation of the slope of the least-squares regression line?

(A) When a man's shoe size increases by one, we predict his height to increase by two inches.

(B) When a man's height increases by two inches, we predict his shoe size to increase by one.

(C) When a man's shoe size increases by two, we predict his height to increase by one inch.

(D) When a man's height increases by one inch, we predict his shoe size to increase by two.

(E) When a man's shoe size increases by one, his height will always increase by exactly two inches.

14. Which of the following statements is **false**?

(A) The predicted height of a man with a size 11 shoe is 72 inches.

(B) It would not be appropriate to use this regression line to predict the height of a man with a size 8 shoe.

(C) About 78% of the variation in height is accounted for by its regression on shoe size.

(D) The high correlation between shoe size and height does not indicate a causal relationship.

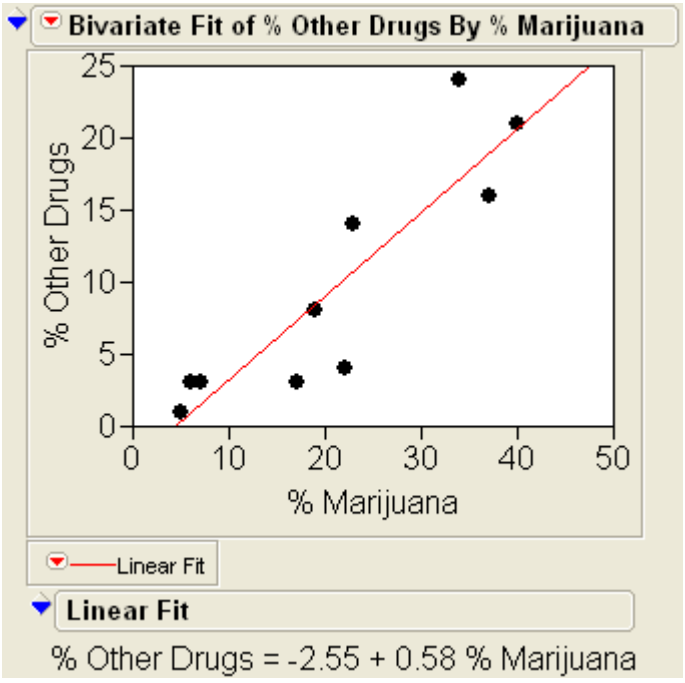
(E) It would not be appropriate to use this regression line to predict the height of a woman from her shoe size.

The next **two** questions (**15** and **16**) refer to the following:

Is there a relationship between marijuana use and use of other illegal drugs? A survey was conducted in the United States and nine countries in Western Europe. The percentage X of teenagers who have used marijuana and the percentage Y who have used other illegal drugs are as follows:

Country	% Marijuana	% Other Drugs
Czech Republic	22	4
Denmark	17	3
England	40	21
Finland	5	1
Ireland	37	16
Italy	19	8
Northern Ireland	23	14
Norway	6	3
Portugal	7	3
United States	34	24

The correlation between marijuana use and other drug use is calculated to be $r = 0.89$. The scatterplot is shown below, including the least squares regression line.



15. Which of the following statements is **true**?

(A) A good prediction for the percentage of teens who have used other drugs in Scotland (where 53% have used marijuana) is 28.19%.

(B) About 79% of the variation in other drug use can be accounted for by its regression on marijuna use.

(C) It is clear that marijuana is a"gateway drug"; that is, marijuana use leads to use of other illegal drugs.

(D) It is clear that most teens who have used other illegal drugs have also used marijuana.

(E) all of the above

16. What is the value of the residual for Denmark?

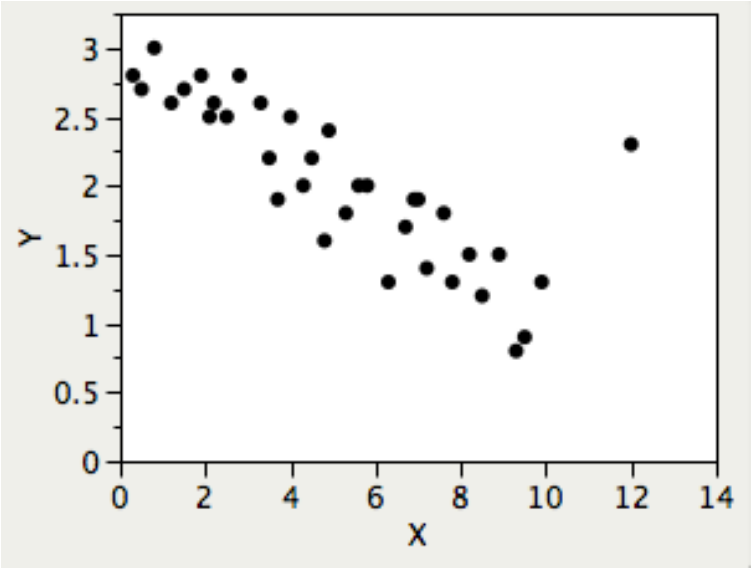
(A) -7.31 (B) 7.31 (C) 3.81 (D) -4.31 (E) 4.31

17. If a grocery store always prices its generic "store brand" products at 90% of the name brand products, what would be the correlation between the prices of the store brand products and the name brand products?

(A) 0.10 (B) -0.10 (C) 0.90 (D) -0.90 (E) 1

The next **two** questions (**18** and **19**) refer to the following:

The scatterplot below displays the relationship between an explanatory variable X and a response variable Y for a sample of individuals:



18. The point in the upper right corner of the scatterplot is:

    (A) an influential observation.
    (B) the predicted value of Y when $X = 12$.
    (C) an outlier in the y-direction.
    (D) a lurking variable.
    (E) the residual value when $X = 12$.

19. If we remove the point in the upper right corner of the scatterplot and fit the least-squares regression line to the data, the equation of the regression line would be:

    (A) $\hat{y} = 0.5 - 0.2x$
    (B) $\hat{y} = 3 - 0.5x$
    (C) $\hat{y} = 3 - 0.2x$
    (D) $\hat{y} = 0.5 + 3x$
    (E) $\hat{y} = 3 + 0.2x$

20. Which of the following statements about the least squares regression line is (are) **true**?

(I) The slope of the least squares regression line always has the same sign as the correlation.
(II) The least squares regression line is the line that minimizes the sum of residuals.
(III) The least squares regression line is the line that maximizes the value of the correlation.

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

21. Two quantitative variables X and Y are measured on a sample of five individuals. Consider the following (incomplete) table of values for this data set.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
			-5	25
3	3	-4	-3	12
6	7	-1	1	-1
10	8	3	2	6
	11		5	

The means and standard deviations are calculated to be $\bar{x} = 7$, $\bar{y} = 6$, $s_x = 5$, $s_y = 4$.

What is the value of the correlation between X and Y for this data set?

(A) 0.9950 (B) 0.9775 (C) 0.9625 (D) 0.9850 (E) 0.9575

22. A random sample of University of Manitoba students is asked whether their parents have a university degree. What is the population of interest for this survey?

(A) all University of Manitoba students.
(B) all University of Manitoba students whose parents have a university degree.
(C) all parents of University of Manitoba students.
(D) all parents of University of Manitoba students who have a university degree.
(E) all University of Manitoba students who were not included in the sample.

23. A university student is planning to take a course on Shakespeare next year and decides to prepare by reading some of his plays over the summer. She makes a list of the plays she would like to read:

01 – Merchant of Venice	09 – Julius Caesar
02 – Othello	10 – Macbeth
03 – Taming of the Shrew	11 – Titus Andronicus
04 – Midsummer Night's Dream	12 – Romeo & Juliet
05 – Hamlet	13 – King Lear
06 – The Tempest	14 – Much Ado About Nothing
07 – Twelfth Night	15 – Troilus & Cressida
08 – Henry VIII	16 – As You Like It

The student estimates that she will have time to read four plays during the summer. Use the following string of random digits (starting from the left) to select a simple random sample of four of the plays for the student to read during the summer.

27350 76724 19501 25531 58080 73915 01487 91333

Which of the following plays will the student **not** read during the summer?

(A) Merchant of Venice

(B) Twelfth Night

(C) Henry VIII

(D) Romeo & Juliet

(E) Troilus & Cressida

24. A STAT 1000 instructor would like to choose a sample of five students who attend her lecture on a given day to participate in a class activity. Suppose she places all the students' names in a hat and draws one name. The student she selects then chooses a second name from the hat. The second student chosen selects a third name from the hat, and so on until five names have been chosen. Selections are made without replacement. The resulting sample is a:

(A) multistage sample.

(B) simple random sample.

(C) stratified random sample.

(D) biased sample.

(E) systematic random sample.

25. Which of the following statements is **false**?

(A) A voluntary response sample is not reliable, as individuals who feel strongly about the issue are overrepresented.

(B) Undercoverage occurs when some individuals in the population have no chance of being selected into the sample.

(C) In multistage sampling, individuals in the population don't always have the same chance to be selected into the sample.

(D) Nonresponse is a frequent source of bias in telephone surveys.

(E) A convenience sample is reliable as long as the sample size is large.

26. Canada has ten provinces and three territories. Suppose we want to select a simple random sample of 130 Canadians. This sampling design ensures that:

(A) the sample will contain 10 individuals from each province and territory.

(B) there will be at least some individuals from each province and territory.

(C) more Quebecers will be chosen than Manitobans, because Quebec has a much higher population.

(D) someone from Yukon has a greater chance of being selected than someone from Ontario, because Ontario has a much higher population.

(E) all possible samples of 130 Canadians are equally likely.

27. A professor needs to conduct a survey of university students for his research paper. It is not realistic to select a simple random sample of students, and the professor would like responses from students in a wide variety of academic programs. The professor selects a simple random sample of faculties at the university. In each selected faculty, he selects a simple random sample of departments, and within each selected department, he selects a simple random sample of classes. He then goes to each of those classes and asks each student in the class to fill out the survey.

    This is an example of:

    (A) multistage sampling.

    (B) stratified sampling.

    (C) systematic sampling.

    (D) a completely randomized design.

    (E) a poorly designed study.

28. Consider the following cartoon:

    Which of the following are present in this survey?

    (I) simple random sampling
    (II) convenience sampling
    (III) leading question

    (A) II only

    (B) III only

    (C) I and II only

    (D) I and III only

    (E) II and III only

29. For the purpose of Canadian federal elections, the country is divided up into 308 electoral ridings, 14 of which are in Manitoba. In each riding, voters select one Member of Parliament to represent then in Ottawa. A survey is to be conducted to gauge the support of Manitobans for the various federal political parties. Which of the following sampling strategies employs the use of stratified random sampling?

(A) Select a simple random sample of five ridings. In each riding, select a simple random sample of 200 voters and administer the survey to them.

(B) Randomly select a sample of 1000 Manitobans from a list of all eligible voters in the province and contact them to answer the survey.

(C) Select a simple random sample of 100 voters from each of the 14 ridings and administer the survey to them.

(D) Select a simple random sample of three ridings, and survey all voters in those ridings.

(E) Select a simple random sample of four ridings. In each of the selected ridings, select a simple random sample of ten city blocks. Go door-to-door administering the survey to the residents of each house on the selected blocks.

The next **two** questions (**30** and **31**) refer to the following:

A popcorn company would like to determine the optimum microwave settings for popping a bag of its popcorn. They will test two different cooking times (4 minutes or 5 minutes) and three different temperature settings (low, medium or high). Three bags of popcorn will be popped at each combination of factor levels in the same microwave, with the order randomly determined. The percentage of unpopped kernels will then be compared for each treatment.

30. What is/are the factor levels in this experiment?

(A) time and temperature

(B) 4 minutes, 5 minutes, low temperature, medium temperature, high temperature

(C) percentage of unpopped kernels

(D) 4 min on low, 4 min on medium, 4 min on high, 5 min on low, 5 min on medium, 5 min on high

(E) bags of popcorn

31. How many bags of popcorn are required for this experiment?

(A) 9 (B) 15 (C) 6 (D) 18 (E) 12

32. A golfer bought new golf clubs this year to try to improve his game. His scores were lower this summer than last summer at all three golf courses he plays. This is an example of:

(A) an experiment with a control.
(B) a completely randomized design.
(C) a randomized block design with three blocks.
(D) a matched pairs design.
(E) an observational study.

33. Randomization is used in a matched pairs experiment to:

(A) select the individuals to participate in the experiment.
(B) select which treatments will be compared.
(C) place the individuals in pairs.
(D) assign the treatments to the individuals within each pair.
(E) all of the above.

34. An experiment is being conducted to compare the effects of five different diet plans on the weight loss of overweight adults. Fifty men and fifty women volunteer to participate in the experiment. It is believed that males and females may respond differently to the various treatments, and so a randomized block design is used. The experiment should use:

(A) ten blocks – five with 10 males and five with 10 females.

(B) ten blocks – each with 5 males and 5 females.

(C) five blocks – each with 10 males and 10 females.

(D) two blocks – one with all 50 males and one with all 50 females.

(E) two blocks – each with 25 males and 25 females.

35. A researcher is studying the relationship between sugar consumption and weight gain. Twelve volunteers were randomly assigned to one of two groups. The first group of five participants was put on a diet in which they were told they could eat only foods low in sugar. The second group of the remaining seven participants received 10% of their calories from sugar. After 8 weeks, weight gain was recorded for each participant.

This is an example of:

(A) an observational study.

(B) a double-blind experiment.

(C) an experiment, but not a double-blind experiment.

(D) a matched-pairs experiment.

(E) a randomized block design.

Sample Term Test 1B Answer Key

1. E
2. B
3. E
4. A
5. B

6. C
7. E
8. B
9. C
10. D

11. A
12. C
13. A
14. C
15. B

16. D
17. E
18. A
19. C
20. A

21. C
22. A
23. A
24. B
25. E

26. E
27. A
28. E
29. C
30. B

31. D
32. E
33. D
34. D
35. C