Sample Term Test 1 – A

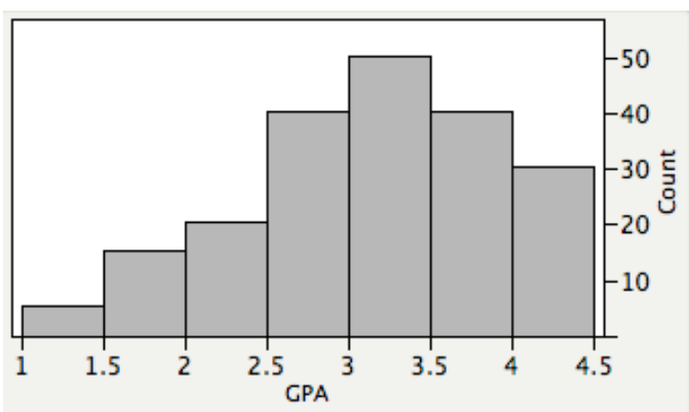
1. Consider the following variables:

   (i) Area code (Vancouver – 604, Edmonton – 780, Winnipeg – 204, etc.)
   (ii) Weight class of a professional boxer (lightweight, middleweight, heavyweight, etc.)
   (iii) Office number of a Statistics professor in Machray Hall

   These two variables are, respectively:

   (A) (i) categorical and nominal (ii) categorical and nominal (iii) categorical and ordinal
   (B) (i) categorical and ordinal (ii) categorical and nominal (iii) categorical and nominal
   (C) (i) categorical and ordinal (ii) quantitative (iii) categorical and ordinal
   (D) (i) categorical and nominal (ii) categorical and ordinal (iii) quantitative
   (E) (i) categorical and nominal (ii) categorical and ordinal (iii) categorical and ordinal

2. 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

The next **three** questions (**3** to **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. We would like to make a histogram (with vertical bars) of the number of vehicle thefts last year in all North American cities with populations over one million people. The horizontal and vertical axes represent, respectively:

   (A) number of thefts and frequency.

   (B) population and number of thefts.

   (C) number of thefts and population.

   (D) population and frequency.

   (E) city and number of thefts.

7. 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**

8. 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

9. The amount of damages (in $) for a sample of 12 Autopac claims are shown below:

$$1400 \quad 500^- \quad 2300 \quad 1700 \quad 15000^+ \quad 7200 \quad 1100 \quad 500^- \quad 4500 \quad 500^- \quad 8800 \quad 3300$$

where 500^- indicates that a claim value was below a driver's $500 deductible and 15000^+ indicates that one driver's damage was greater than the $15000 value of his car and so it was written off.

What is the median amount of damages for the claims in this sample?

(A) $2000

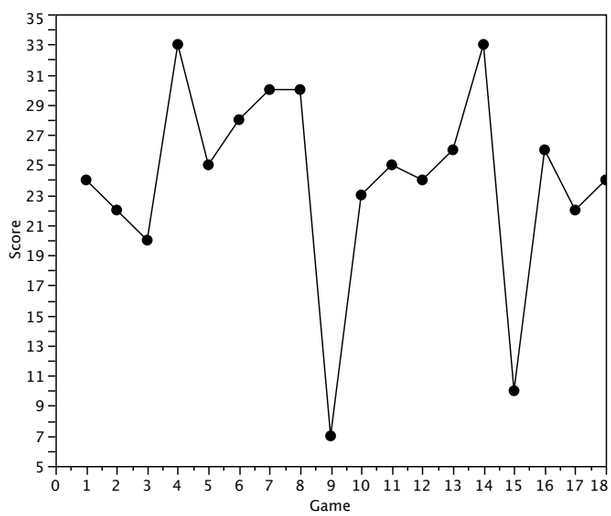
(B) $2800

(C) $3900

(D) $4300

(E) impossible to determine without the exact values of all claims.

10. Consider the following timeplot displaying the number of points scored by the Winnipeg Blue Bombers in each of their 18 games last season.



What is the third quartile of the number of points scored by the team?

(A) 20 (B) 23 (C) 25 **(D) 28** (E) 33

11. 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.

12. The frequency distribution below displays the GPAs of a sample of 135 university students who took Advanced Placement or International Baccalaureate programs in high school:

GPA	Frequency
$0.00 - 0.50$	1
$0.50 - 1.00$	1
$1.00 - 1.50$	4
$1.50 - 2.00$	6
$2.00 - 2.50$	13
$2.50 - 3.00$	18
$3.00 - 3.50$	25
$3.50 - 4.00$	31
$4.00 - 4.50$	36

Which interval contains the first quartile of GPA's for this sample?

(A) $1.00 - 1.50$ (B) $1.50 - 2.00$ (C) $2.00 - 2.50$ **(D)** $2.50 - 3.00$ (E) $3.00 - 3.50$

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 credit card company uses the Air Miles rewards system. The company offers new customers 150 Air Miles for signing up and 0.15 Air Miles for every dollar the customer spends using the credit card. The number of Air Miles Y earned by customers in their first year is therefore

$$Y = 150 + 0.15X$$

where X is the amount (in \$) spent by the customer using their credit card in the first year. In a sample of new customers, the mean amount spent using their credit card in the first year is \$2,500 and the sample standard deviation is \$1,200. The mean and standard deviation of the number of Air Miles Y earned in the first year are, respectively:

(A) 525 and 330

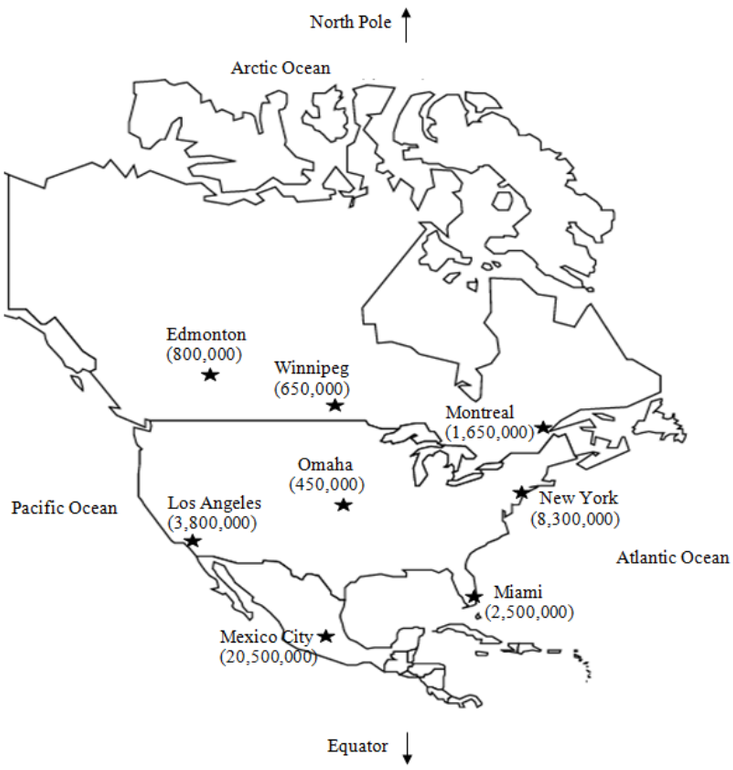
(B) 375 and 330

(C) 525 and 180

(D) 375 and 180

(E) 525 and 375

15. 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

16. We want to calculate the correlation r between two variables X and Y. Which of the following conditions are **necessary** for r to be a meaningful measure of association?

(I) X and Y are both quantitative variables.

(II) The relationship between X and Y is linear.

(III) X is an explanatory variable and Y is a response variable.

(A) I only

(B) II only

(C) I and II only

(D) I and III only

(E) I, II and III

17. Which of the following studies is most likely to 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 cars driving down the highway:
X = age of the driver; Y = speed of the car

(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**

18. Suppose we have gathered data from a sample of individuals for some explanatory variable X and some response variable Y. We plot the data on a scatterplot and we see that a linear relationship is a reasonable assumption.

We fit the least squares regression line to the data. This is the line that:

(A) minimizes the sum of the residuals.

(B) maximizes the value of the correlation.

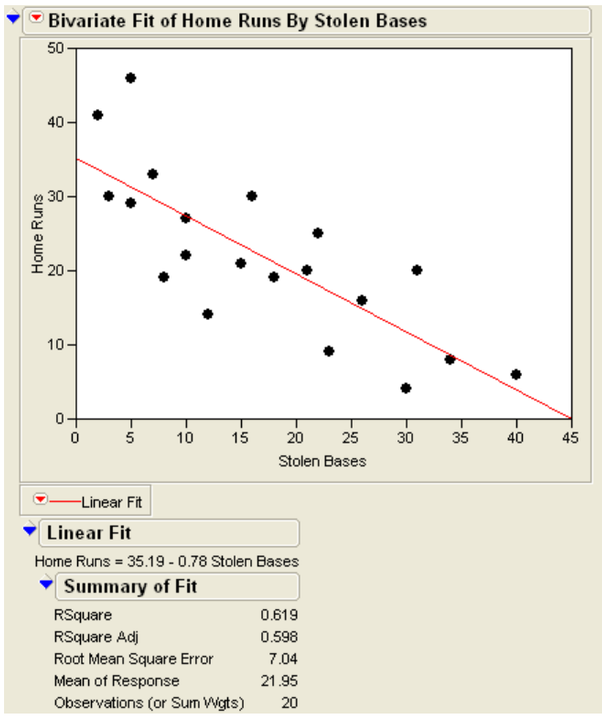
(C) minimizes the sum of deviations from the points to the line in the horizontal direction.

(D) minimizes the sum of squared residuals.

(E) minimizes the sum of squared deviations from the points to the line in the horizontal direction.

The next **four** questions (**19** to **22**) 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*:



19. 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.

20. 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

21. 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%

22. 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

23. 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. 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

24. A researcher would like to determine whether student performance in an introductory math course differs depending on the professor of the course, and the time of day the course is given. Two math professors, Dr. Smith and Dr. Johnson, are both teaching two sections of the same introductory math course this semester. Dr. Smith teaches sections A01 (at 9:30 a.m.) and A02 (at 1:30 p.m.) and Dr. Johnson teaches sections A03 (at 9:30 a.m.) and A04 (at 1:30 p.m.). At the end of the semester, the researcher will compare the average grades of the students in the four sections.

This is an example of:

(A) a completely randomized design with four treatments.

(B) a randomized block design with two blocks and two treatments.

(C) a randomized block design with four blocks and four treatments.

(D) a matched pairs design with two treatments.

(E) an observational study.

25. 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.

26. 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.

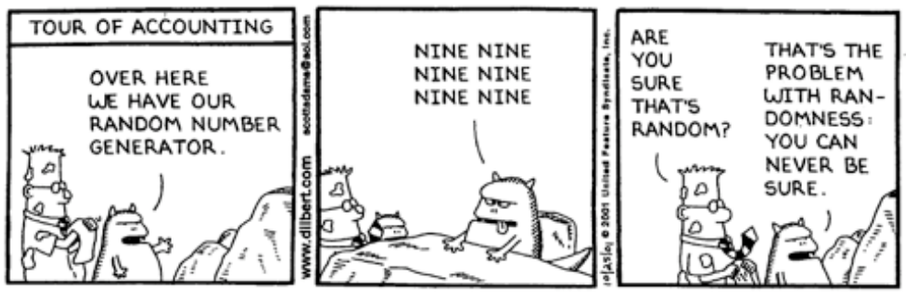
(B) ten blocks – each with 5 males and 5 females.

(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.

27. 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 (**28** and **29**) 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.

28. 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.

29. 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.

30. In order to assess the opinion of students at the University of Manitoba on increasing tuition fees, a reporter for *The Manitoban* newspaper stands in University Center over lunch hour and interviews the first 25 students he meets who are willing to express their opinions. The method used by the reporter is an example of:

(A) simple random sampling.
(B) an observational study.
(C) convenience sampling.
(D) a completely randomized design.
(E) volunteer response sampling.

31. A student has enough space in her schedule to take four electives one year. She makes a list of all the courses she would like to take. The courses are numbered and are shown below:

01 – Biology	11 – Geography
02 – Calculus	12 – Chemistry
03 – Anthropology	13 – French
04 – Spanish	14 – Statistics
05 – History	15 – Computer Science
06 – Physics	16 – World Religions
07 – Economics	17 – Sociology
08 – Linear Algebra	18 – Philosophy
09 – Philosophy	19 – German
10 – English	20 – Environmental Science

Use the following string of random digits (starting from the left) to select a simple random sample of four courses for the student to take this year.

32651 97208 55234 82719 47260 53779 41162 08466

Which of the following courses will the student **not** take this year?

(A) History

(B) Linear Algebra

(C) World Religions

(D) German

(E) Environmental Science

32. A sociology professor wants to estimate the true proportion of young adults who have used illegal drugs. She asks her class of 150 students to raise their hand if they have used illegal drugs. Which of the following statements is (are) **true**?

(I) The results of this survey are unreliable, as the sample is not representative of the population of interest.

(II) This survey probably overestimates the true proportion of young adults who have used illegal drugs.

(III) This survey probably underestimates the true proportion of young adults who have used illegal drugs.

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

(A) I only

(B) I and II only

(C) III only

(D) I and III only

(E) IV only

33. A simple random sample of size n is the **only** type of sample that **guarantees** that:

(A) every individual in the population has a known chance of being selected into the sample.

(B) every individual in the population has an equal chance of being selected into the sample.

(C) every group of n individuals has an equal chance of being selected into the sample.

(D) results will not be biased.

(E) all of the above

Sample Term Test 1 – B

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) License plate number (AWH 514, BYK 602, CJB 728, 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. 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. 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.

5. The five-number summary for some data set is calculated to be

   12 20 33 39 48

   The sample mean and sample standard deviation are calculated to be $\bar{x} = 30$ and $s = 8$. It is discovered that a data value of 2 was accidentally recorded as 12. After we make the appropriate correction and change the 12 to a 2, the median will _____, the mean will _____, and the standard deviation will _____.

   (A) decrease, decrease, decrease.

   (B) remain the same, decrease, decrease.

   (C) decrease, remain the same, decrease.

   (D) remain the same, decrease, increase.

   (E) remain the same, remain the same, increase.

6. The University of Manitoba uses a grade point system in which it awards the following number of grade points per credit hour for each possible letter grade:

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

A University of Manitoba student took four courses one semester and received the following grades:

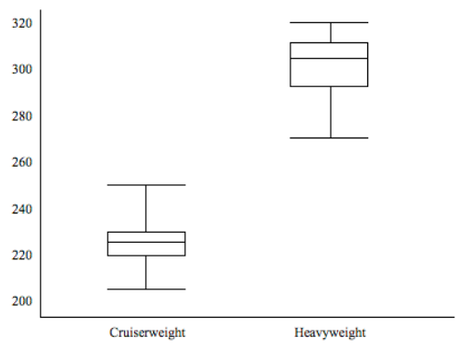
	Credit Hours	Letter Grade
Course 1	3	C^+
Course 2	6	A^+
Course 3	4	B^+
Course 4	3	D

What is the student's Grade Point Average for the term?

(A) 3.22 (B) 3.14 (C) 2.88 (D) 3.06 (E) 2.96

The next **two** questions (**7** and **8**) 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

7. 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

8. 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**

9. The scores for 14 amateur golfers are ordered and shown below:

80 81 82 86 86 87 87 87 88 89 89 91 94 96

If we constructed an outlier boxplot for these data, the lines coming out of the box (the whiskers) would extend to the values:

(A) 82 and 91
(B) 81 and 94
(C) 81.5 and 93.5
(D) 83.5 and 90.5
(E) 80 and 96

10. The temperature X (in degrees Celsius) at noon is recorded each day for one week in early November in downtown Winnipeg. The mean noon-hour temperature is calculated to be 2.8°C and the standard deviation of temperatures is calculated to be 5.6°C. What are the mean and standard deviation, respectively, of the weeks noon-hour temperatures in degrees Fahrenheit? (Recall that Y = 32 + 1.8X, where Y is the temperature in °F.)

(A) 26.96 and 10.08
(B) 37.04 and 42.08
(C) –5.04 and 42.08
(D) 26.96 and 42.08
(E) 37.04 and 10.08

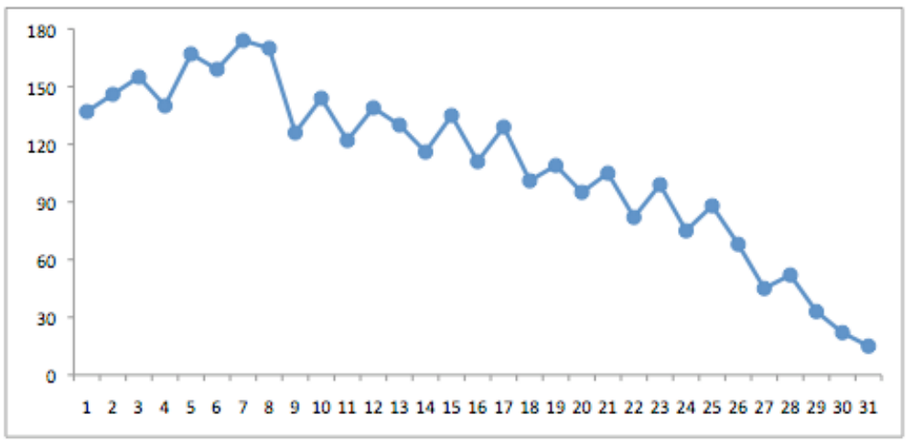
11. The weights (in kilograms) of ten boxes being shipped by a furniture company are shown below:

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?

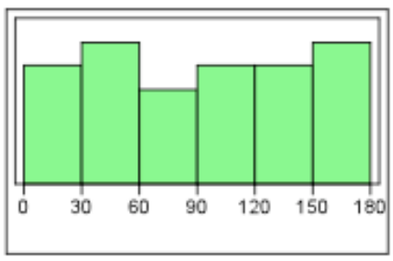
(A) 132.5 (B) $\frac{951 + 3H}{10}$ **(C) 141** (D) 138 (E) cannot be computed

12. 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:

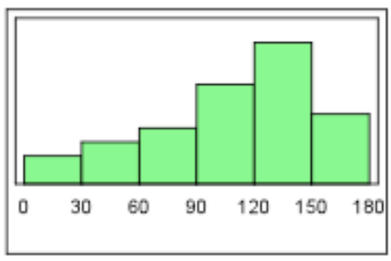


Which of the following is the appropriate histogram for this data set?

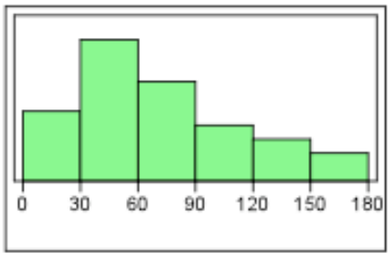
(A)



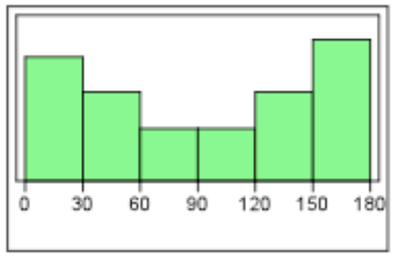
(B)



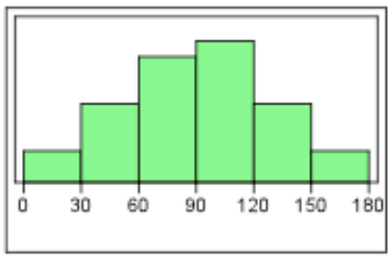
(C)



(D)



(E)



13. 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

14. A small graduate class of three students writes a math test. The student who finished writing the fastest got the highest score in the class. The student who finished second got the second highest score, and the student who took the longest to write the test got the lowest score. Let X be the time it takes for a student to write the test and let Y be the students test score. What can be said about the correlation r between X and Y for this class?

(A) There is a perfect negative linear relationship between X and Y, and so $r = -1$.

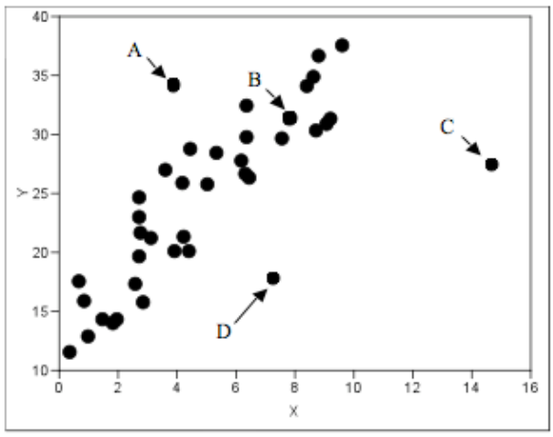
(B) The correlation between X and Y is negative, but not necessarily equal to -1.

(C) There is no linear relationship between X and Y, and so $r = 0$.

(D) There is a perfect positive linear relationship between X and Y, and so $r = 1$.

(E) The correlation between X and Y is positive, but not necessarily equal to 1.

15. Consider the scatterplot shown below, displaying the relationship between some explanatory variable X and some response variable Y:



(A) Point A is the residual when X = 4.
(B) Point B is the predicted value of Y when X = 8.
(C) Point C is an influential observation.
(D) Point D is an outlier in the y–direction, as its y–value lies out of the range of the data.
(E) All of the above are true.

16. The same math test is given to a sample of elementary school students in Grades 1 through 6. The correlation between a students height and his or her score on the test is calculated to be r = 0.8. This does not, however, indicate a causal relationship between the two variables. Which of the following is likely a lurking variable in this case?

(A) gender
(B) teacher
(C) age
(D) study time
(E) all of the above

The next **three** questions (**17** to **19**) refer to the following:

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 $\hat{y} = 2.3 + 1.7x$. It is reported that three-quarters of the variation in Time can be accounted for by its regression on Distance.

17. What is the value of the correlation between Distance and Time?

    (A) 0.250 **(B) 0.866** (C) 0.625 (D) 0.750 (E) 0.563

18. One student lives 5 kilometres from the university and takes 8 minutes to get there. What is the value of the residual for this student?

    (A) -2.8 (B) 2.8 (C) -10.9 (D) 10.9 (E) 10.8

19. 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

20. 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

21. A golfer bought new golf clubs this year to try to improve his game. After the golf season, he examined all of his scores for last summer and this summer, and he noticed his scores were lower this year at all three golf courses he played. 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.

The next **two** questions (**22** and **23**) 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.

22. What is/are the factor(s) 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

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

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

24. An experiment is to be conducted to determine the effect of temperature on the adhesive strength of a certain type of glue. The glue will be applied to a material at either 10°C, 20°C or 30°C. Two different types of material (metal and plastic) will be used. Six pieces of plastic and six pieces of metal are available for the experiment. The effect of temperature is expected to differ for the two materials, and so a randomized block design will be used. The experimenter should use:

    (A) three blocks (one for each temperature), where each block consists of two pieces of plastic and two pieces of metal.
    (B) two blocks, each consisting of three pieces of plastic and three pieces of metal.
    (C) six blocks, each consisting of one piece of plastic and one piece of metal.
    (D) two blocks, one consisting of the six pieces of plastic and the other consisting of the six pieces of metal.
    (E) three blocks (one for each temperature), where the pieces of plastic and metal are randomly assigned to the blocks.

25. 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.

26. An experiment is being conducted to study the effectiveness of different brands of sunscreen and SPF (sun protection factor) levels. Volunteers will be randomly assigned to apply either Coppertone or Ombrelle sunscreen, with an SPF level of either 30 or 60. Subjects will spend eight hours outside in the sun, and the degree of sunburn (if any) will be compared for all treatments. What is/are the treatment(s) in this experiment?

(A) degree of sunburn
(B) sunscreen brand and SPF level
(C) Coppertone, Ombrelle, SPF 30, SPF 60
(D) Coppertone/SPF 30, Ombrelle/SPF 30, Coppertone/SPF 60, Ombrelle/SPF 60
(E) time spent outside

27. 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.

28. Which of the following studies uses/use a matched pairs design?

(I) Which cola tastes better – Pepsi or Coke? In a blind taste test, volunteers were asked to taste each of the colas (in random order) and to score the taste of each of them out of 10. The colas were compared by tabulating the differences in scores for each volunteer.

(II) Do students perform better on multiple-choice or long-answer exams? A math professor teaches two sections of an introductory course. He randomly chooses one of the sections to write a multiple-choice exam and the other to write a long-answer exam. Scores for the two sections will be compared.

(III) Is it easier to remember words or images? Ten sets of twins are available to participate in a study. One twin is randomly assigned to study a list of 50 words and the other studies a page with 50 images. After 30 minutes, they are asked to list as many words or images as they can remember. The number of correct answers will be compared for each set of twins.

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

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.

30. 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.

31. 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.

32. 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 III only
(D) II and III only
(E) I, II and III

33. When a polling firm randomly selects people to call to participate in a telephone survey, there are some people who have no chance of being selected – those who don't own a telephone. This type of sampling bias is known as:

(A) non-response bias.
(B) undercoverage.
(C) response bias.
(D) underrepresentation.
(E) confounding.

Sample Term Test 1 – Solutions

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