

Sample Midterm A

1. Consider the following:

- Film ratings exist to inform parents about the content of movies, so they can decide which films are appropriate for their children to see, and at what age. Examples of film ratings include General (appropriate for all audiences), Parental Guidance (some material not suitable for children), Restricted (under 17 years old requires parental accompaniment) and NC-17 (no one under 17 years old allowed).
- Students of the martial art karate wear different coloured belts, depending on their skill levels. There are nine belt colours in total. Students start with a white belt (beginner) and work their way up to a black belt (expert).
- When making an international phone call, you must dial the country code, followed by the person's phone number. For example, the country code for France is 33, the country code for Australia is 61 and the country code for Mexico is 52.

The variables **Film Rating**, **Belt Colour** and **Country Code** are, respectively:

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

The next **three** questions (**2, 3 and 4**) 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

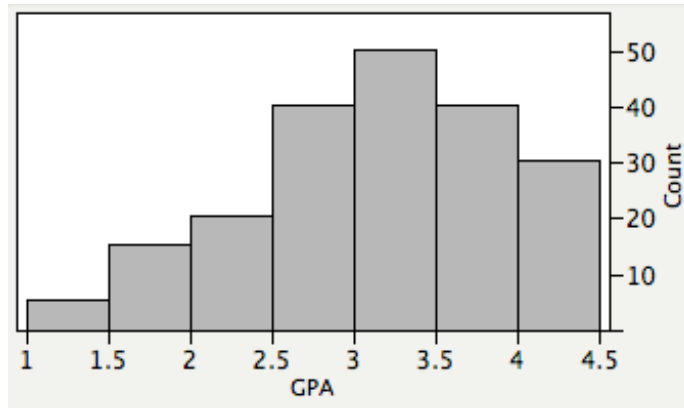
- How many homes in the sample from the neighbourhood of Evergreen are worth at least \$300,000?

(A) 75 (B) 135 (C) 45 (D) 120 (E) 150
- 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
- If we were to construct an 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.

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



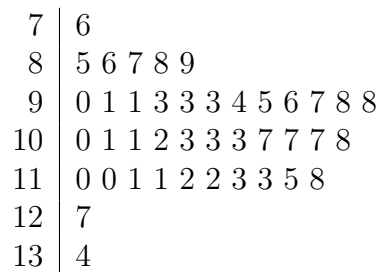
What proportion of students in the class have a GPA less than 3.0?

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

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 home games in the 2008/2009 NBA season.



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

1400 500⁻ 2300 1700 15000⁺ 7200 1100 500⁻ 4500 500⁻ 8800 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?

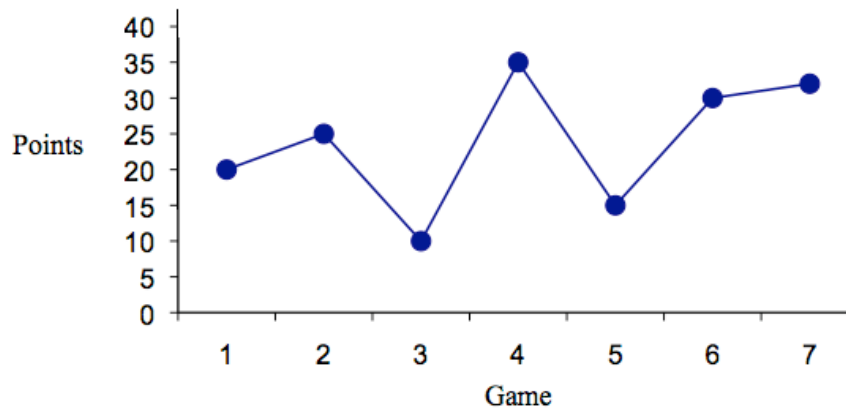
- (A) \$2000
 - (B) \$2800
 - (C) \$3900
 - (D) \$4300
 - (E) impossible to determine without the exact values of all claims.
9. The following frequency distribution summarizes the GPAs of a random sample of 149 students:

GPA	# of students
0.00 – 0.50	1
0.50 – 1.00	1
1.00 – 1.50	4
1.50 – 2.00	11
2.00 – 2.50	19
2.50 – 3.00	31
3.00 – 3.50	39
3.50 – 4.00	28
4.00 – 4.50	15

Which of the following statements regarding the distribution of GPAs is **true**?

- (A) The distribution is skewed to the left and so the median is greater than the mean.
- (B) The distribution is skewed to the right and so the mean is greater than the median.
- (C) The distribution is skewed to the left and so the mean is greater than the median.
- (D) The distribution is skewed to the right and so the median is greater than the mean.
- (E) The distribution is approximately symmetric and so the mean and median are approximately equal.

10. 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:

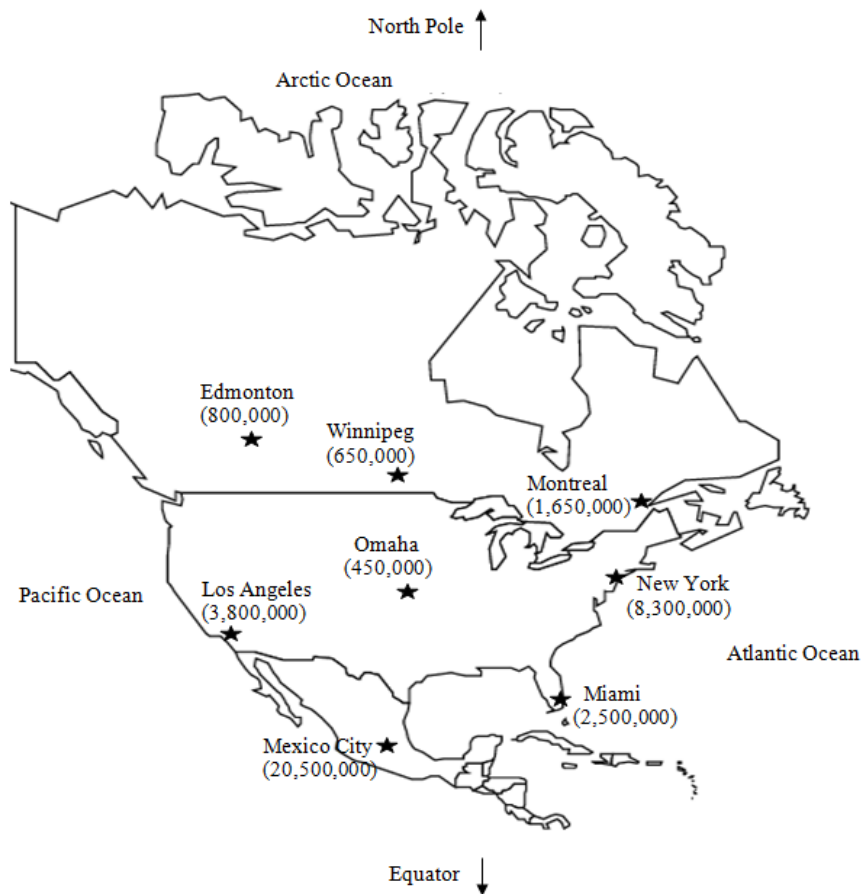
- (A) 35 (B) 20 (C) 15 (D) 25 (E) 30
11. The systolic blood pressures are recorded for a sample of four patients at a clinic and are shown below:

119 98 133 106

What is the standard deviation of the systolic blood pressures for this sample of patients?

- (A) 13.29 (B) 14.61 (C) 15.34 (D) 176.50 (E) 235.33

12. 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

13. 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 days throughout the year:
 $X =$ global supply of oil; $Y =$ price of gasoline
 - (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
14. 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

The next **three** questions (**15** to **17**) refer to the following:

A coffee shop manager would like to determine how the outside temperature (in $^{\circ}\text{C}$) affects the sales of hot chocolate (in \$). She records both variables for a sample of ten days. The least squares regression line is calculated to be $\hat{y} = 40.29 - 1.37x$. It is also reported that 46% of the variation in Y can be accounted for by its regression on X .

15. What is the correct interpretation of the slope of the least squares regression line?
- (A) When the temperature increases by 1°C , the predicted sales of hot chocolate increase by \$1.37.
 - (B) When the sales of hot chocolate increase by \$1.37, the predicted temperature decreases by 1°C .
 - (C) When the temperature increases by 1.37°C , the predicted sales of hot chocolate decrease by \$1.
 - (D) When the sales of hot chocolate increase by \$1, the predicted temperature decreases by 1.37°C .
 - (E) When the temperature increases by 1°C , the predicted sales of hot chocolate decrease by \$1.37.
16. What is the value of the correlation between hot chocolate sales and temperature?
- (A) 0.46 (B) -0.68 (C) -0.21 (D) -0.46 (E) 0.68
17. On one of the days in the sample, the temperature was -22°C and the coffee shop sold \$82.59 worth of hot chocolate. What is the value of the residual for this day?
- (A) 12.16 (B) 70.43 (C) 72.44 (D) -12.16 (E) -70.43

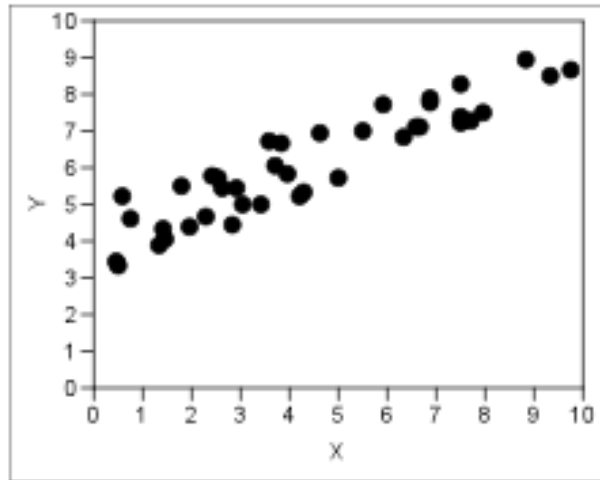
18. Can the speed of a vehicle be predicted by the age of its driver? The speeds (in km/h) of a sample of nine vehicles, on the same stretch of highway, are recorded along with the ages of their drivers:

Age	48	61	26	53	75	39	19	44	35
Speed	99	104	128	108	89	120	116	107	110

The equation of the least squares regression line is calculate to be $\hat{y} = 133.72 - 0.56x$ and the correlation is calculated to be $r = -0.836$. Which of the following statements is/are true?

- (I) A reliable prediction for the speed of an 88 year old driver is 84.44 km/h.
 - (II) If we had measured speed in miles per hour, the correlation would not change.
 - (III) It is clear that getting older causes people to drive slower.
- (A) I only
 - (B) II only
 - (C) I and II only
 - (D) I and III only
 - (E) I, II and III

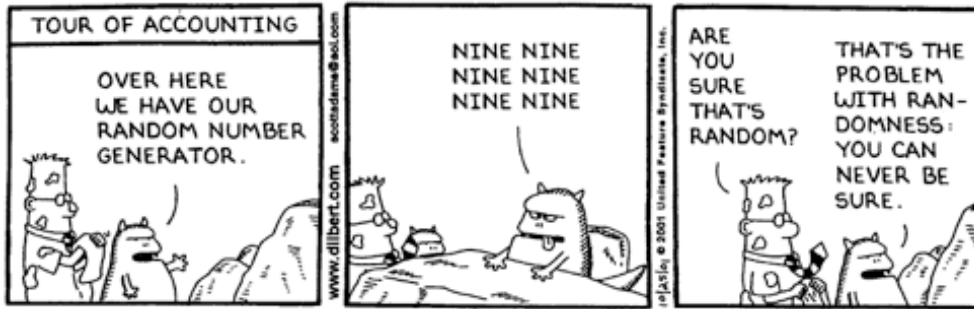
19. Data are collected for some explanatory variable X and some response variable Y and are displayed in the scatterplot below:



What would be the equation of the least squares regression line for these data?

- (A) $\hat{y} = 0.5 + 4x$
(B) $\hat{y} = 4 + x$
(C) $\hat{y} = 4 + 0.5x$
(D) $\hat{y} = 0.5 + x$
(E) $\hat{y} = 4 + 2x$
20. In order to assess the opinion of students at the University of Manitoba on increasing tuition fees, a reporter for *The Manitoban* stands in University Centre 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.

21. 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 (**22** and **23**) refer to the following:

Former Manitoba Premier Gary Doer later served 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.

22. 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.

23. 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.

24. A teacher is travelling to Europe for her summer holidays, but she won't 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
25. 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.

26. 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) replication
 - (E) All of the above are used in this experiment.
27. 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) use characteristics of the subjects (such as degree of hypertension) to subjectively divide the subjects into two groups, 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.

The next **two** questions (**28** and **29**) 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.

28. 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) dry/40, dry/60, dry/80, dry/100, rain/40, rain/60, rain/80, rain/100, snow/40, snow/60, snow/80, snow/100
- (E) type of tire

29. How many trial runs will be used in this experiment?

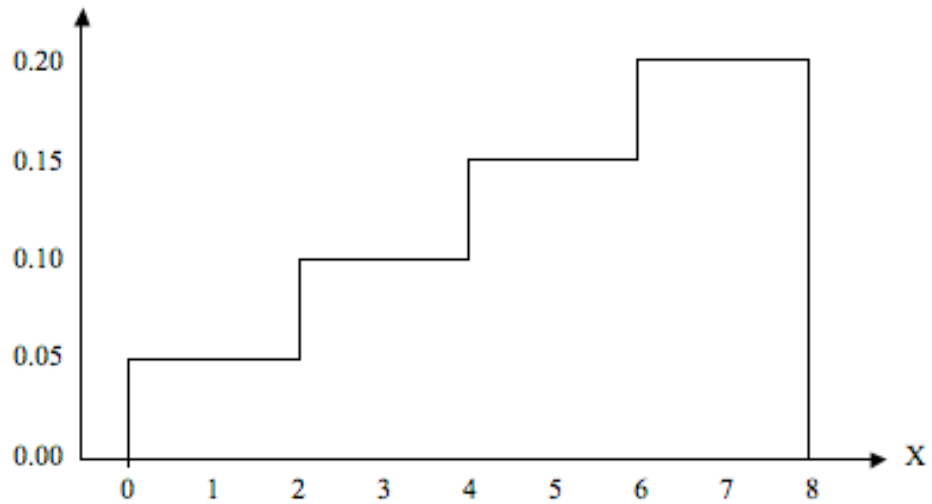
- (A) 12 (B) 7 (C) 21 (D) 36 (E) 18

30. Grapes grown in a vineyard are used for making wine. The owners of the vineyard would like to conduct an experiment to examine the effect of storage time (2 or 3 years) and temperature (5°C or 10°C) on the taste of their wine. They believe different types of wine (red and white) will react differently to the various treatments, and so the a randomized block design is used. 60 bottles of red wine and 60 bottles of white wine are available for the experiment.

What are the treatments in this experiment?

- (A) storage time and temperature
- (B) 2 years, 3 years, 5°C and 10°C
- (C) 2 years, 3 years, 5°C, 10°C, red and white
- (D) 2 years/5°C, 2 years/10°C, 3 years/5°C, 3 years/10°C
- (E) 2 years/5°C/red, 2 years/5°C/white, 2 years/10°C/red, 2 years/10°C/white, 3 years/5°C/red, 3 years/5°C/white, 3 years/10°C/red, 3 years/10°C/white

31. A variable X has a distribution which is described by the density curve shown below:



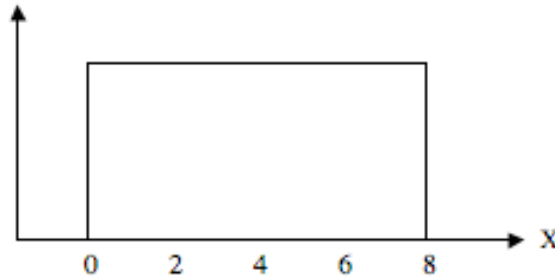
What proportion of values of X fall between 1 and 6?

- (A) 0.550 (B) 0.575 (C) 0.600 (D) 0.625 (E) 0.650

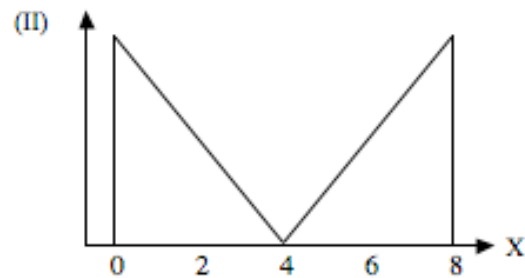
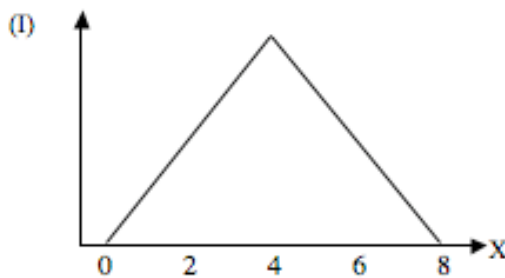
32. Which of the following statements about a normal distribution is **true**?

- (A) The value of μ must always be positive.
(B) The value of σ must always be positive.
(C) The shape of a normal distribution depends on the value of μ .
(D) The possible values of a standard normal variable range from -3.49 to 3.49 .
(E) The area under a normal curve depends on the value of σ .

33. A variable X follows a uniform distribution, as shown below:



The distribution of X has an interquartile range equal to 4 (since the middle 50% of the data are contained between the values 2 and 6). Consider the variables with the distributions shown below (assume that the heights of the curves are such that they are both valid density curves):



The interquartile range of density curve (I) is _____ and the interquartile range of density curve (II) is _____.

- (A) (I) less than 4, (II) greater than 4
- (B) (I) greater than 4, (II) less than 4
- (C) (I) equal to 4, (II) equal to 4
- (D) (I) less than 4, (II) less than 4
- (E) (I) greater than 4, (II) greater than 4

34. A variable Z has a standard normal distribution. What is the value z such that $P(-0.37 \leq Z \leq z) = 0.5749$?

- (A) 2.02
- (B) 1.48
- (C) 0.97
- (D) 0.63
- (E) 1.72

The next **three** questions (**35** to **37**) refer to the following:

The sport of women's gymnastics consists of four events. Suppose it is known that scores for each event follow a normal distribution with the following means and standard deviations:

Event	Mean	Std. Dev.
Balance Beam	8.3	0.3
Uneven Bars	8.6	0.5
Vault	8.2	0.4
Floor Exercise	9.0	0.2

35. What proportion of gymnasts receive a score between 8.2 and 8.7 on the uneven bars?
- (A) 0.2088 (B) 0.3674 (C) 0.6000 (D) 0.6837 (E) 0.3085
36. The top 6% of gymnasts in each event earn a trip to the national championships. What is the minimum vault score required to make it to the nationals?
- (A) 8.44 (B) 7.58 (C) 8.90 (D) 8.41 (E) 8.82
37. Julie receives a score of 9.0 on the balance beam, 9.2 on the uneven bars, 9.1 on the vault and a 9.3 on floor exercise. In which event did Julie do the best relative to other gymnasts?
- (A) Balance Beam
(B) Uneven Bars
(C) Vault
(D) Floor Exercise
(E) Julie did equally well on all events.

38. Speeds of vehicles on a highway follow a normal distribution with mean 106.2 km/h and standard deviation 8.7 km/h. What proportion of vehicles on this highway are travelling above the 100 km/h speed limit?
- (A) 0.7126 (B) 0.2612 (C) 0.7910 (D) 0.2874 (E) 0.7611
39. Percentage grades in a large Calculus class follow a normal distribution with mean 60 and standard deviation 10. Percentage grades in a large English class follow a normal distribution with mean 68 and standard deviation 8. Katherine received a grade of 85 in Calculus. What grade does she need in English to be at the same percentile?
- (A) 87 (B) 88 (C) 89 (D) 90 (E) 91
40. Bottles of a certain brand of apple juice are filled automatically by a machine. Fill volumes are normally distributed with a mean of μ and standard deviation of 1.46 ml. The label on the bottles claim that the bottles contain 500 ml of juice. What value should μ be set at so that only 2% of bottles will be underfilled?
- (A) 497 ml (B) 498 ml (C) 502 ml (D) 503 ml (E) 504 ml

Sample Midterm A Answer Key

- | | |
|-------|-------|
| 1. A | 21. E |
| 2. B | 22. B |
| 3. E | 23. A |
| 4. C | 24. A |
| 5. D | 25. E |
| 6. A | 26. C |
| 7. E | 27. E |
| 8. A | 28. B |
| 9. A | 29. D |
| 10. D | 30. D |
| 11. C | 31. A |
| 12. E | 32. B |
| 13. E | 33. A |
| 14. C | 34. B |
| 15. E | 35. B |
| 16. B | 36. E |
| 17. A | 37. A |
| 18. B | 38. E |
| 19. C | 39. B |
| 20. C | 40. D |

Sample Midterm B

1. Consider the following variables:

- (i) Area code of a Canadian city (e.g., Calgary 403, Winnipeg 204, Toronto 416, etc.)
- (ii) Weight class of an Ultimate Fighting Championship fighter (lightweight, middleweight, heavyweight, etc.)
- (iii) Time of day that a nurse's shift starts (morning, afternoon, evening, night)

These three variables are, respectively:

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

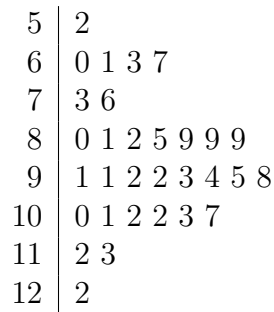
The next **two** questions (**3** and **4**) refer to the following:

The following frequency distribution displays the speeds (in km/h) for a random sample of vehicles driving on University Crescent.

Speed	Frequency
40 – 45	7
45 – 50	15
50 – 55	24
55 – 60	18
60 – 65	6
65 – 70	5
70 – 75	3
75 – 80	2
80 – 85	2

3. The distribution of speeds is
- (A) skewed to the right and so the median is greater than the mean.
 - (B) skewed to the left and so the median is greater than the mean.
 - (C) skewed to the right and so the median is less than the mean.
 - (D) skewed to the left and so the median is less than the mean.
 - (E) approximately symmetric and so the median is equal to the mean.
4. Which interval contains the third quartile of speeds?
- (A) 55 – 60
 - (B) 60 – 65
 - (C) 65 – 70
 - (D) 70 – 75
 - (E) impossible to determine with the information given.

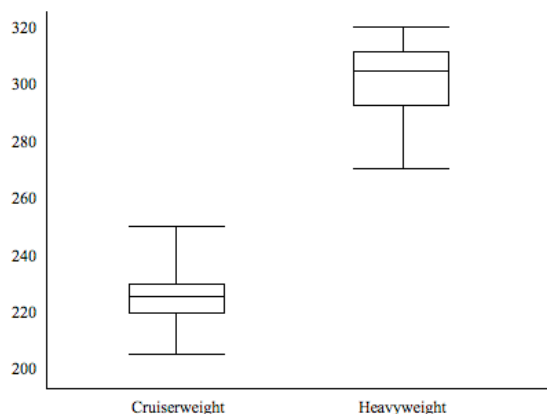
5. The following stemplot displays the number of forest fires that occurred each year over the period from 1970–2000 in Canada:



- where 5|2 represents 5,200 fires. What is the value of the first quartile for this data set?
- (A) 7,600 (B) 7,800 (C) 8,000 (D) 8,050 (E) 8,100
6. 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.
7. 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

The next **two** questions (8 and 9) refer to the following:

The sport of boxing divides its athletes into different weight classes. The side-by-side 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

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

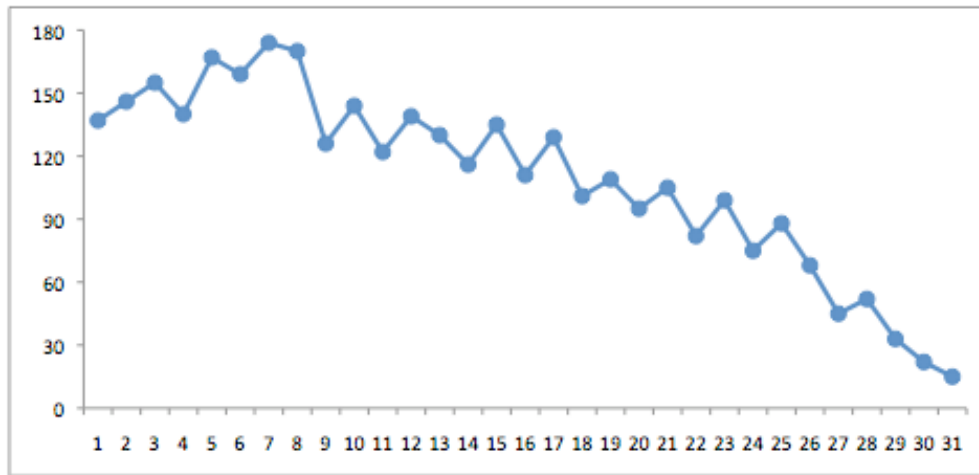
- I. The distribution of weights for the heavyweights is skewed to the left.
- II. There are 12 cruiserweights in the sample who weigh at least 220 pounds.
- III. The mean weight for the heavyweights is likely greater than the median weight.

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

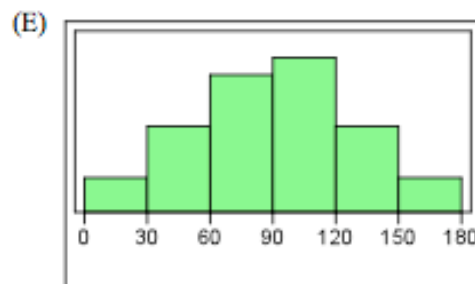
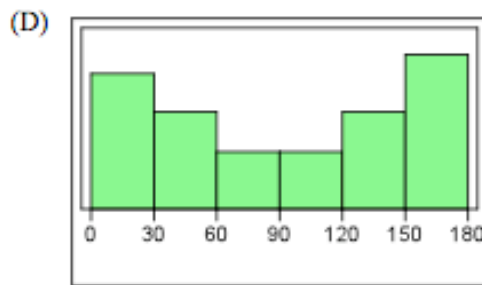
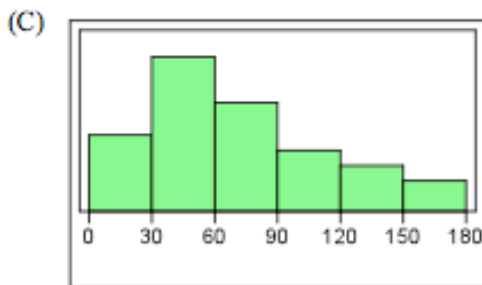
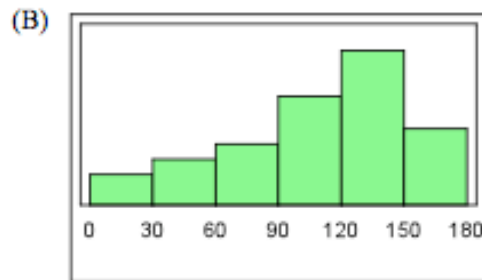
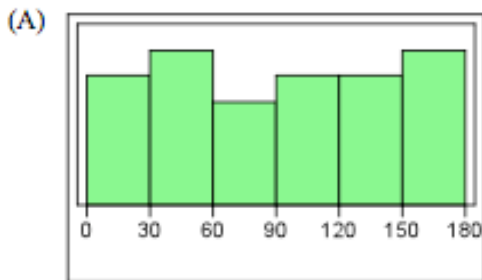
9. What is the median weight of all of the boxers in the sample (cruiserweight and heavy-weight) combined?

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

10. 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?



11. The speeds (in km/h) of a random sample of 25 vehicles driving on a highway are ordered and shown below:

67	75	82	86	93	100	104	104	105	106
106	107	108	108	109	110	112	113	113	115
126	130	135	141	152					

The five-number summary is: 67 102 108 114 152

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

- (A) 84 and 132
 - (B) 102 and 114
 - (C) 82 and 135
 - (D) 86 and 130
 - (E) 93 and 126
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 = Weight of a vehicle on the highway
Y = Speed of the vehicle
 - III. X = Number of people in line at a bank when you arrive
Y = Time spent waiting in line until you are served by a teller
- (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

13. If a grocery store always prices its generic “store brand” products at 80% of the name brand products (i.e., the store brand products are 20% off), what would be the correlation between the prices of the store brand products and the name brand products?

(A) -1 (B) -0.20 (C) 0.80 (D) -0.80 (E) 1

14. 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?

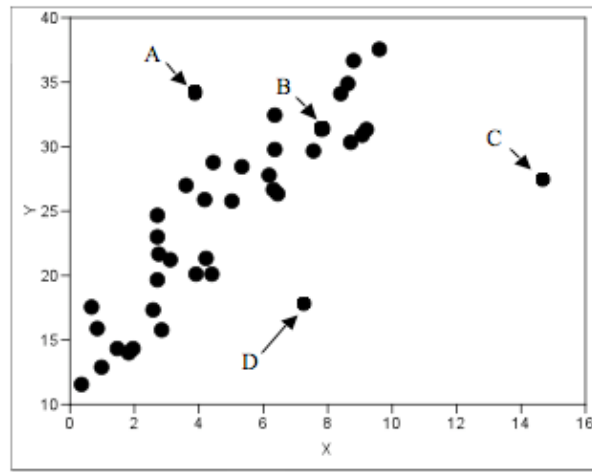
(A) 0.9850 (B) 0.9775 (C) 0.9625 (D) 0.9950 (E) 0.9575

15. 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

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



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

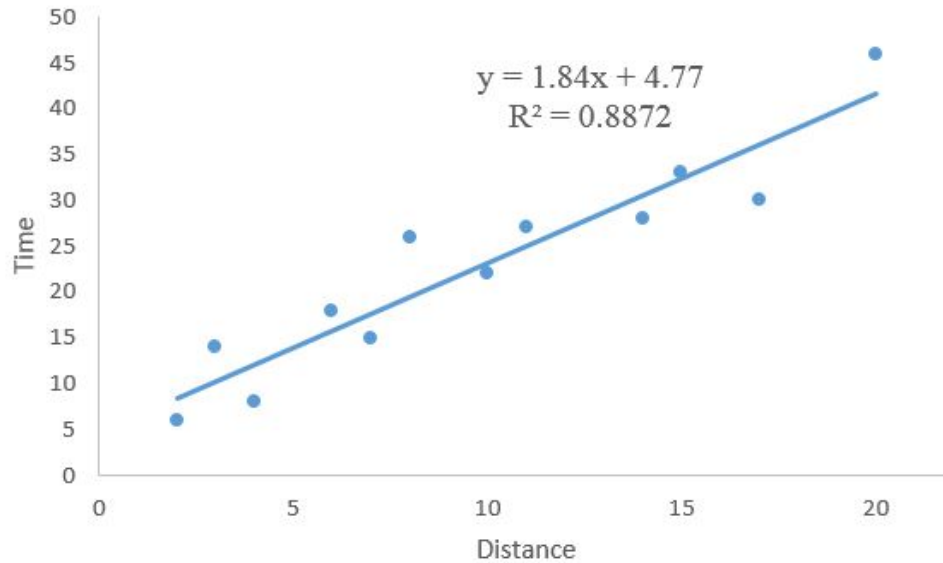
- (A) Point A is the residual when $X = 4$.
 - (B) Point B is the predicted value of Y when $X = 8$.
 - (C) Point C is an influential observation.
 - (D) Point D is an outlier in the y -direction.
 - (E) All of the above are true.
17. A professor would like to conduct a regression analysis to determine whether a student's final exam score in his class can be predicted from the student's midterm score. She records the midterm and final exam scores of a sample of students in her class. The midterm scores have a mean of 71.0 and a standard deviation of 9.4. The final exam scores have a mean of 66.2 and a standard deviation of 10.3. The correlation between midterm and final exam score is calculated to be 0.84.

What is the predicted increase in final exam score when a student's midterm score increases by one?

- (A) 0.71
- (B) 0.76
- (C) 0.84
- (D) 0.88
- (E) 0.92

The next **two** questions (**18** and **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 university and the time Y (in minutes) it takes them to get to the university are recorded. The *Microsoft Excel* output is shown below:



18. One student in the sample lives 5 kilometres from the university and takes 17 minutes to get there. What is the value of the residual for this student?

- (A) -3.03 (B) -8.69 (C) 13.97 (D) 3.03 (E) 8.69

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. 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 with the student **not** read during the summer?

- (A) Merchant of Venice
 - (B) Twelfth Night
 - (C) Henry VIII
 - (D) Romeo & Juliet
 - (E) Troilus & Cressida
21. 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

22. In 2006, Major League Baseball (MLB) set up the Joint Drug Prevention and Treatment Program in an effort to eliminate the widespread use of performance enhancing drugs, such as anabolic steroids, from the game. MLB does random tests on its players to make sure that they are not using such drugs. Ten of the 30 teams are randomly selected. From each of these ten teams, four players are randomly selected. Testers go out to each of the ten selected teams and collect urine and blood samples from the players that have been selected. If a player tests positive for performance enhancing drugs, he is suspended for 50 games. The resulting sample of 40 players is a:

- (A) simple random sample
- (B) stratified random sample
- (C) systematic sample
- (D) multistage sample
- (E) convenience sample

23. Consider the following cartoon:



Which of the following are present in this survey?

- (I) voluntary response bias
 - (II) simple random sampling
 - (III) convenience sampling
 - (IV) leading question
- (A) I and III only
 - (B) II and IV only
 - (C) I and IV only
 - (D) III and IV only
 - (E) I, III and IV only

24. For the purpose of Canadian federal elections, the country is divided up into 338 electoral ridings, 14 of which are in Manitoba. In each riding, voters select one Member of Parliament to represent them 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.
25. For a sample of 911 calls to the fire department, 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

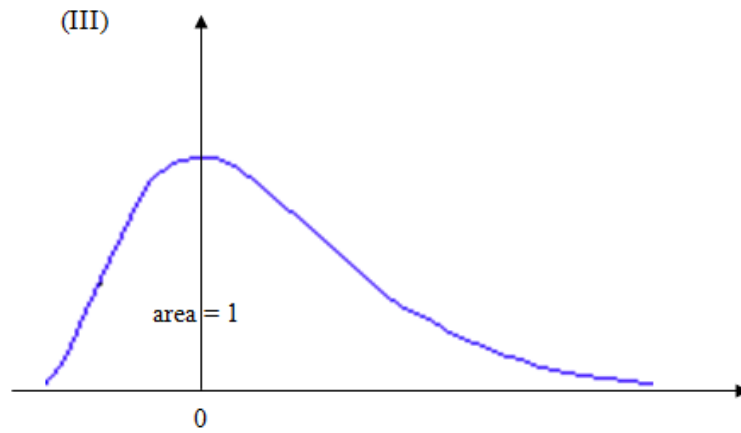
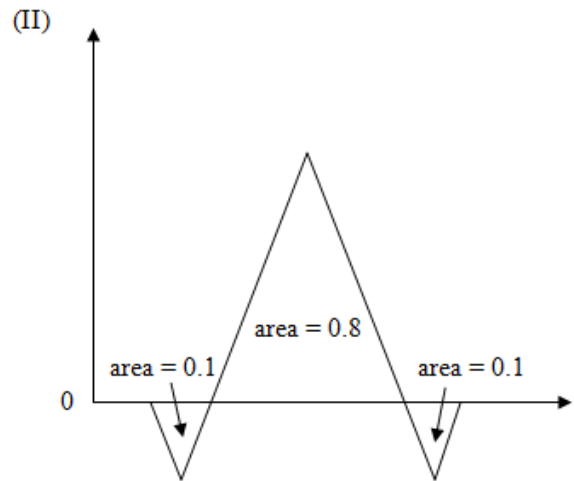
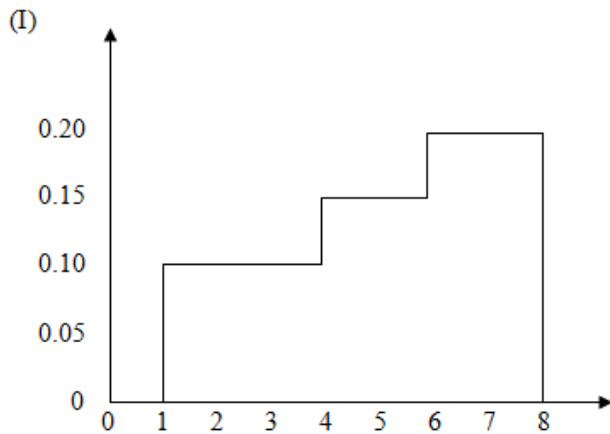
26. A popcorn company would like to determine the optimum microwave settings for popping a bag of its popcorn. They will test two different cooking times (4 minutes or 5 minutes) and three different temperature settings (low, medium or high). Three bags of popcorn will be popped at each combination of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

What is/are the factor level(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) brand of popcorn
27. A psychologist is conducting an experiment to compare the effects of five treatments, and 80 university students have volunteered to participate. Twenty volunteers are from the Faculty of Arts, twenty are from the Faculty of Science, twenty are from the Faculty of Management and twenty are from the Faculty of Engineering. The psychologist believes that students' responses may differ, depending on the student's faculty, so she decides to use a randomized block design. She should use:
- (A) four blocks – each consisting of five Arts students, five Science students, five Engineering students and five Management students.
 - (B) four blocks – one with all 20 Arts students, one with all 20 Science students, one with all 20 Engineering students and one with all 20 Management students.
 - (C) five blocks – each consisting of 16 Arts students, 16 Science students, 16 Engineering students and 16 Management students.
 - (D) twenty blocks – each consisting of one Arts student, one Science student, one Engineering student and one Management student.
 - (E) sixteen blocks – four with five Arts students, four with five Science students, four with five Engineering students and four with five Management students.

28. A study is being conducted to compare the effect of two different weight loss pills (A and B) on the weight loss of overweight adults. Researchers rank the 200 volunteer subjects from heaviest to lightest. For the two heaviest subjects, one is randomly assigned Pill A and one is randomly assigned Pill B. The same is done for the next two heaviest subjects, and the next two, and so on. The subjects are not told which pill they are receiving. Suppose at the end of the one-year study, subjects taking Pill A have lost consistently more weight than their counterparts taking Pill B. Can we say that the pill was the **cause** of the greater weight loss?
- (A) Yes, because this is a properly conducted matched-pairs experiment.
 - (B) Yes, because this is a properly designed double-blind experiment.
 - (C) No, because no control group was used and the difference may be due to the placebo effect.
 - (D) No, because association can never imply causation.
 - (E) No, because there are likely many lurking variables affecting weight loss for the two groups.
29. Mrs. Peterson teaches a high school English class and Mr. Jenkins teaches math to the same class of students. The average score for the class on the provincial English exam is 78%. The average score for the class on the provincial math exam is 62%. What can be concluded from this information?
- (A) Mrs. Peterson is a better teacher than Mr. Jenkins.
 - (B) The math exam was likely more difficult than the English exam.
 - (C) Both (A) and (B).
 - (D) Little if anything can be concluded because difficulty of the exams and quality of the teachers are confounding variables.
 - (E) Little if anything can be concluded because the intelligence of the students is a lurking variable.
30. Which of the following statements about experimental design is/are **true**?
- (I) Randomization is used to assign individuals to blocks.
 - (II) Replication is used to conduct the experiment again when the results of the first experiment are not conclusive.
 - (III) When comparing responses for two or more treatment groups, a control group should always be included.
- (A) none (B) I only (C) II only (D) III only (E) II and III only

31. Which of the following represent a legitimate density curve for some variable X ?



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

32. A random variable X follows a uniform distribution on the interval from 10 to 15. What proportion of values of X are greater than 13.2?

- (A) 0.18
- (B) 0.88
- (C) 0.36
- (D) 0.12
- (E) 0.28

33. What is the value z such that $P(-z < Z < z) = 0.7776$, where Z has a standard normal distribution?

- (A) 0.97 (B) 1.59 (C) 1.34 (D) 0.76 (E) 1.22

34. A variable X follows a normal distribution with mean 10 and standard deviation 5. Another variable Y follows a normal distribution with mean 25 and standard deviation 10. The maximum height of the density curve for X is _____ (i) _____ the maximum height for the density curve for Y , and the area under the density curve for X is _____ (ii) _____ the area under the density curve for Y .

- (A) (i) greater than, (ii) less than
(B) (i) less than, (ii) greater than
(C) (i) equal to, (ii) equal to
(D) (i) greater than, (ii) equal to
(E) (i) less than, (ii) less than

35. Weights of apples grown in an orchard are known to follow a normal distribution with mean 160 grams. It is known that approximately 99.7% of apples have weights between 124 and 196 grams. What is the standard deviation of weights of all apples grown in the orchard?

- (A) 9 grams (B) 12 grams (C) 18 grams (D) 24 grams (E) 36 grams

The next **two** questions (**36** and **37**) refer to the following:

Percentage grades in a large geography class follow a normal distribution with mean 67.5 and standard deviation 12.5.

36. What proportion of students in the class receive percentage grades above 73.4?

- (A) 0.2061 (B) 0.3192 (C) 0.4720 (D) 0.5280 (E) 0.6808

37. The professor decides to assign a grade of A+ to the students with the top 8% of the grades, and a grade of A to the next best 12%. What is the minimum percentage a student needs to earn a grade of A?

- (A) 77 (B) 78 (C) 79 (D) 80 (E) 81

38. The contents of bottles of water follow a normal distribution with mean μ and standard deviation 4 ml. What proportion of bottles have fill volumes within 1 ml of the mean?
- (A) 0.1974
 - (B) 0.5987
 - (C) 0.6826
 - (D) 0.4013
 - (E) unable to calculate without the value of μ
39. Diameters of spruce trees in a forest are known to follow a normal distribution with a mean of 31.3 cm. Suppose it is known that 33% of the spruce trees have diameters above 35.7 cm. What is the standard deviation of heights of all spruce trees in the forest?
- (A) 6 cm
 - (B) 7 cm
 - (C) 8 cm
 - (D) 9 cm
 - (E) 10 cm
40. Exam scores in a large course are known to follow a normal distribution with mean 64 and standard deviation 12. Sarah received an exam score that placed her at the 98th percentile of the class. Her friend Hayden received an exam score only half as high as Sarah's. At what percentile was Hayden's score?
- (A) 5th percentile
 - (B) 18th percentile
 - (C) 36th percentile
 - (D) 49th percentile
 - (E) 67th percentile

Sample Midterm B Answer Key

- | | |
|-------|-------|
| 1. E | 21. C |
| 2. B | 22. D |
| 3. C | 23. D |
| 4. A | 24. C |
| 5. C | 25. A |
| 6. B | 26. B |
| 7. E | 27. B |
| 8. C | 28. A |
| 9. E | 29. D |
| 10. B | 30. A |
| 11. D | 31. D |
| 12. C | 32. C |
| 13. E | 33. E |
| 14. C | 34. D |
| 15. A | 35. B |
| 16. C | 36. B |
| 17. E | 37. B |
| 18. D | 38. A |
| 19. A | 39. E |
| 20. A | 40. A |