

Sample Final Exam 1

Part A

1. James is taking a computer science course. The mark breakdown is as follows:

- Midterm Exam 1 – 20%
- Midterm Exam 2 – 30%
- Final Exam – 50%

All exams are out of 100 marks. James received scores of 70 on Midterm Exam 1 and 60 on Midterm Exam 2. To receive a final grade of B⁺ in the course, a student requires a final grade of 75% or higher. What is the minimum score James must get on the final exam in order to receive a final grade of B⁺?

- (A) 78 (B) 83 (C) 86 (D) 90 (E) 95

2. A researcher is studying how the height of children during early adolescence is affected by milk consumption. She plots the heights (in inches) and the milk consumption (in cups per day) for a sample of children and wants to fit a least squares regression line to the data. She calculates the following for this sample of children:

- The correlation between milk consumption and height is 0.4.
- The mean milk consumption is 4.0 cups per day.
- The mean height is 60.0 inches.
- The standard deviation of milk consumption is 1.5 cups per day.
- The standard deviation of heights is 3.0 inches.

What is the predicted height of a child who consumes 5 cups of milk per day?

- (A) 60.4 inches
(B) 60.8 inches
(C) 61.2 inches
(D) 61.6 inches
(E) 62.0 inches

3. A biologist is studying the mercury levels in fish from three different lakes in the area. She selects random samples of ten fish from each lake. The sample of 30 fish is a:
- (A) stratified sample.
 - (B) simple random sample.
 - (C) convenience sample.
 - (D) multistage sample.
 - (E) systematic sample.
4. The head of the Department of Mathematics 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 offered. Two math professors, Dr. Smith and Dr. Johnson, are each teaching two sections of the same introductory math course this semester. Dr. Smith teaches sections A01 (at 9:30 a.m.) and A03 (at 1:30 p.m.) and Dr. Johnson teaches sections A02 (at 9:30 a.m.) and A04 (at 1:30 p.m.). At the end of the semester, the department head 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.

The next **two** questions (**5** and **6**) refer to the following:

The following five games are scheduled to be played at the World Curling Championships one morning. The values in parentheses are the probabilities of each team winning their respective game.

Game 1:	Finland (0.43)	vs.	Germany (0.57)
Game 2:	USA (0.28)	vs.	Switzerland (0.72)
Game 3:	Japan (0.11)	vs.	Canada (0.89)
Game 4:	Denmark (0.33)	vs.	Sweden (0.67)
Game 5:	France (0.18)	vs.	Scotland (0.82)

5. The outcome of interest is the set of winners for each of the five games. How many outcomes are contained in the appropriate sample space?
- (A) 5 (B) 10 (C) 25 (D) 32 (E) 64
6. In a sports game, the “favourite” is the team with the higher probability of winning and the “underdog” is the team that is less likely to win. What is the probability that at least one underdog wins their game?
- (A) 0.93 (B) 0.74 (C) 0.80 (D) 0.67 (E) 0.59

The next **five** questions (**7** to **11**) refer to the following:

Suppose we have the following facts about customers buying alcohol at the Liquor Mart:

- 55% buy wine (W).
- 40% buy beer (B).
- 22% buy wine **and** beer.
- 6% buy wine **and** vodka (V).
- 5% buy beer **and** vodka.
- 50% buy beer **or** vodka.
- 2% buy wine and beer and vodka.

7. If we randomly select a customer, what is the probability the customer buys vodka?
- (A) 0.10 (B) 0.15 (C) 0.19 (D) 0.25 (E) 0.28

8. Which of the following statements is **true**?
- (A) Events W and B are independent.
 - (B) Events W and B are mutually exclusive.
 - (C) Events W and V are independent.
 - (D) Events W and V are mutually exclusive.
 - (E) Events B and V are independent.
9. If we randomly select a customer, what is the probability the customer buys **only** wine?
- (A) 0.14 (B) 0.17 (C) 0.25 (D) 0.29 (E) 0.34
10. In a random sample of 17 customers, what is the probability that exactly 10 of them buy wine?
- (A) 0.1153 (B) 0.1367 (C) 0.1495 (D) 0.1629 (E) 0.1841
11. In a random sample of 300 customers, what is the approximate probability that less than 38% of them buy beer?
- (A) 0.0808 (B) 0.1492 (C) 0.1736 (D) 0.2389 (E) 0.2929
12. In a particular election, 40% of voters voted for the NDP, 35% voted for the Liberals and 25% voted for the Conservatives. If we take a random sample of two voters, what is the probability that they voted for different parties?
- (A) 0.765 (B) 0.345 (C) 0.525 (D) 0.655 (E) 0.485
13. There are four patients on the neonatal ward of a local hospital who are monitored by two nurses. Suppose the probability (at any one time) of a patient requiring attention by a nurse is 0.3. Assuming the patients behave independently, what is the probability at any one time that there will not be enough nurses to attend to all patients who need them? (i.e., what is the probability that at least three patients require attention at the same time?)
- (A) 0.0756 (B) 0.1104 (C) 0.0837 (D) 0.0463 (E) 0.2646

14. The yearly rainfall in Vancouver, B.C. follows a normal distribution with standard deviation 172 mm. In 20% of years, the city gets over 1200 mm of rain. What is the mean annual rainfall in Vancouver (in mm)?

(A) 1055.5 (B) 1165.6 (C) 982.6 (D) 1344.5 (E) 1234.4

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

The time X taken by a cashier in a grocery store express lane to complete a transaction follows a normal distribution with mean 90 seconds and standard deviation 20 seconds.

15. What is the first quartile of the distribution of X (in seconds)?

(A) 73.8 (B) 85.0 (C) 69.4 (D) 81.2 (E) 76.6

16. What is the probability that the average service time for the next three customers is between 80 and 100 seconds?

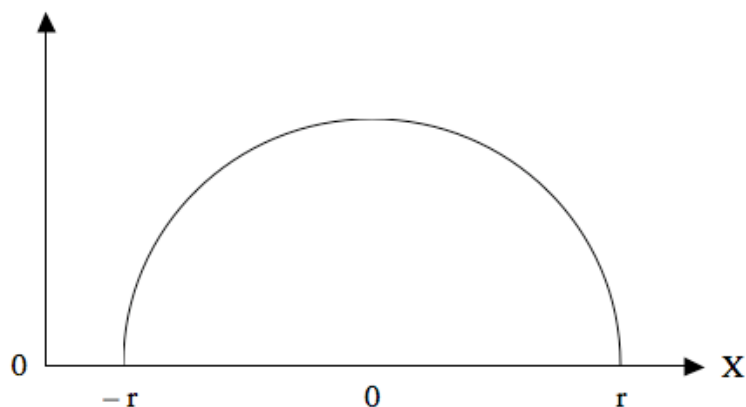
(A) 0.6156
(B) 0.4893
(C) 0.7212
(D) 0.5559
(E) impossible to calculate with the information given

The next **two** questions (**17** and **18**) refer to the following:

The amount X spent (in \$) by customers in the grocery store express lane follows some right-skewed distribution with mean \$24 and standard deviation \$15.

17. What is the probability that the average amount spent by the next three customers is more than \$20?
- (A) 0.4619
(B) 0.6772
(C) 0.8186
(D) 0.7673
(E) impossible to calculate with the information given
18. What is the approximate probability that the next 40 customers spend less than \$1,000 in total?
- (A) 0.5199 (B) 0.6064 (C) 0.6628 (D) 0.5784 (E) 0.6331
19. An economist calculates that, in order to estimate the true mean amount spent per year by Canadians on Christmas presents to within \$60 with 95% confidence, she requires a sample of 90 Canadians. What sample size would be required to estimate the true mean amount spent per year by Canadians on Christmas presents to within \$20 with 95% confidence?
- (A) 10 (B) 30 (C) 156 (D) 270 (E) 810
20. The breaking strength of yarn used in the production of woven carpet material is normally distributed with standard deviation 2.4 psi. A random sample of 4 specimens of yarn from a production run is measured for breaking strength, and a confidence interval for μ is calculated to be (128.609, 134.191). What is the confidence level for this interval?
- (A) 0.90 (B) 0.95 (C) 0.96 (D) 0.98 (E) 0.99

21. A variable X is described by a semi-circular density curve with mean 0 and standard deviation 0.4, as shown below:



We take a simple random sample of 100 individuals from the semi-circular distribution and calculate the sample mean $\bar{x}$. The sampling distribution of $\bar{X}$ is:

- (A) approximately normal with mean 0 and standard deviation 0.004.
 - (B) semi-circular with mean 0 and standard deviation 0.04.
 - (C) approximately normal with mean 0 and standard deviation 0.4.
 - (D) semi-circular with mean 0 and standard deviation 0.004.
 - (E) approximately normal with mean 0 and standard deviation 0.04.
22. We would like to conduct a hypothesis test to examine whether there is evidence that the true mean amount spent on textbooks by U of M students in one semester is greater than \$400. A random sample of 50 students is selected and the mean amount spent on textbooks for one semester is calculated to be \$430. Assume the population standard deviation is known to be \$165. What is the P-value for the appropriate hypothesis test?

- (A) 0.0985 (B) 0.0409 (C) 0.0764 (D) 0.1112 (E) 0.0630

23. A man accused of committing a crime is taking a polygraph (lie detector) test. The polygraph is essentially testing the hypotheses

H_0 : The man is telling the truth. vs. H_a : The man is lying.

Suppose we use a 5% level of significance. Based on the man's responses to the questions asked, the polygraph determines a P-value of 0.08. We conclude that:

- (A) there is insufficient evidence that the man is telling the truth.
- (B) there is sufficient evidence that the man is telling the truth.
- (C) there is insufficient evidence that the man is lying.
- (D) the probability that the man is lying is 0.08.
- (E) the probability that the man is telling the truth is 0.08.

24. A statistician conducted a test of $H_0: \mu = 10$ vs. $H_a: \mu > 10$ for the mean μ of some normally distributed variable X . Based on the gathered data, the statistician calculated a sample mean of $\bar{x} = 12$ and concluded that H_0 could be rejected at the 5% level of significance. Using the same data, which of the following statements **must be true**?

- I. A test of $H_0: \mu = 10$ vs. $H_a: \mu > 10$ at the 1% level of significance would also lead to rejecting H_0 .
- II. A test of $H_0: \mu = 9$ vs. $H_a: \mu > 9$ at the 5% level of significance would also lead to rejecting H_0 .
- III. A test of $H_0: \mu = 10$ vs. $H_a: \mu \neq 10$ at the 5% level of significance would also lead to rejecting H_0 .

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

25. A sample of 30 vehicles is outfitted with snow tires. The vehicles travel 80 km/h in winter driving conditions and apply the brakes. The sample mean and standard deviation of stopping distances for these 30 vehicles are calculated to be 162 metres and 35 metres, respectively. We would like to test whether the true mean stopping distance differs from 150 metres. At the 1% level of significance, we should:

- (A) reject H_0 , since the P-value is between 0.02 and 0.025.
- (B) fail to reject H_0 , since the P-value is between 0.025 and 0.05.
- (C) reject H_0 , since the P-value is between 0.025 and 0.05.
- (D) fail to reject H_0 , since the P-value is between 0.05 and 0.10.
- (E) reject H_0 , since the P-value is between 0.05 and 0.10.

The next **two** questions (**26** and **27**) refer to the following:

Researchers are studying the impact of alcohol consumption on a person's ability to perform a simple task. Five subjects complete a small puzzle while sober, and then again after drinking three glasses of wine. The times (in seconds) are shown below, as well as some summary statistics:

Subject	1	2	3	4	5	mean	std. dev.
Sober (S)	109	132	248	155	202	169.2	55.9
After Wine (AW)	141	128	290	174	260	198.6	72.5
Diff. ($d = S - AW$)	-32	4	-42	-19	-58	-29.4	23.5

Assume that all necessary normality conditions are satisfied. We would like to conduct a hypothesis test to determine whether there is evidence that the true mean time to complete a puzzle is longer after consuming three glasses of wine.

26. What are the hypotheses for the appropriate test of significance?

- (A) $H_0: \mu_d = 0$ vs. $H_a: \mu_d < 0$
- (B) $H_0: \mu_S = \mu_{AW}$ vs. $H_a: \mu_S > \mu_{AW}$
- (C) $H_0: \bar{X}_d = 0$ vs. $H_a: \bar{X}_d < 0$
- (D) $H_0: \mu_d = 0$ vs. $H_a: \mu_d > 0$
- (E) $H_0: \bar{X}_S = \bar{X}_{AW}$ vs. $H_a: \bar{X}_S < \bar{X}_{AW}$

27. The P-value of the appropriate test of significance is:

- (A) between 0.01 and 0.02.
- (B) between 0.02 and 0.025.
- (C) between 0.025 and 0.05.
- (D) between 0.05 and 0.10.
- (E) between 0.10 and 0.15.

28. You want to estimate the proportion of Canadians who would support a merger of the federal NDP and Liberal parties to within 0.035 with 92% confidence. What sample size is required?

- (A) 625
- (B) 784
- (C) 466
- (D) 816
- (E) 553

29. We would like to estimate the true proportion p of Manitobans who are bilingual (fluent in two languages). We take a random sample of 200 Manitobans and find that 42 of them are bilingual. What is the margin of error for a 95% confidence interval for p ?
- (A) 0.0565 (B) 0.0353 (C) 0.0693 (D) 0.0288 (E) 0.0404
30. The Acme Car Company claims that less than 8% of its new cars have a manufacturing defect. A quality control inspector randomly selects 300 new cars and finds that 15 have a defect. He conducts a hypothesis test to examine the significance of the car company's claim. What is the P-value of the appropriate hypothesis test?
- (A) 0.0359 (B) 0.0485 (C) 0.0162 (D) 0.0571 (E) 0.0274

Sample Final Exam 1

Part B

1. The shower flow rates (in L/min) for a sample of 30 houses are ordered and shown below:

2.2	2.8	3.8	4.2	4.6	5.0	5.7	6.0	6.2	6.2
6.5	6.5	6.7	6.8	7.0	7.2	7.3	7.4	7.4	7.6
7.7	7.9	8.0	8.1	8.5	8.6	8.7	8.8	8.8	11.4

- (a) Find the five-number summary for this data set.
- (b) Construct an outlier boxplot for this data set. Show any necessary calculations. What is the shape of the distribution, excluding outliers?
2. We would like to examine how the age X of a certain model of car affects its selling price Y . The age (in years) and price (in \$) for a sample of 10 cars of the same make and model are recorded from the classified ads in the newspaper one weekend. The least squares regression line is calculated to be $\hat{y} = 18,600 - 1,700x$. It is also determined that 81% of the variation in a car's price can be accounted for by its regression on age.
- (a) Provide an interpretation of the slope of the least squares regression line in this example.
- (b) What is the value of the correlation between age and price for these cars?
- (c) One car in the sample is 5 years old and has a price of \$15,500. Calculate the value of the residual for this car.

3. Three hats each contain ten coins. Hat 1 contains two gold coins, five silver coins and three copper coins. Hat 2 contains four gold coins and six silver coins. Hat 3 contains three gold coins and seven copper coins. We randomly select one coin from each hat.
- (a) The outcome of interest is the colour of each of the three selected coins. List the complete sample space of outcomes and calculate the probability of each.
 - (b) Let X be the number of gold coins selected. Find the probability distribution of X .
4. State the Central Limit Theorem.
5. We take a random sample of 27 trees in a large forest and measure their heights. The sample mean is calculated to be 28.4 feet and the sample standard deviation is calculated to be 5.3 feet. Heights of trees in the forest are known to follow a normal distribution.
- (a) Construct a 98% confidence interval for the true mean height of all trees in the forest.
 - (b) Provide an interpretation of the confidence interval in (a).
 - (c) Conduct a hypothesis test at the 2% level of significance to determine if there is evidence that the true mean height of all trees in the forest is less than 30 feet. Show all of your steps, including the hypotheses, test statistic, P-value and a properly-worded conclusion.
 - (d) Provide an interpretation of the P-value of the test in (c).
 - (e) Could the confidence interval in (a) have been used to conduct the test in (c)? Why or why not? If it could have been used, explain what the conclusion would be, and why.

Sample Final Exam 2

Part A

1. The receipt totals (rounded to the nearest dollar) for a sample of 25 customers in the express lane at a supermarket are ordered and shown below:

5	8	9	13	16	18	19	19	21	23
24	26	30	34	35	37	42	45	48	50
56	56	62	64	73					

What is the interquartile range for this data set?

- (A) 29 (B) 29.5 (C) 30.5 (D) 31 (E) 32
2. The table below displays the number of pass completions X and the number of passing yards Y for six Winnipeg Blue Bomber quarterbacks during the 2015 football season:

Quarterback	1	2	3	4	5	6
Completions	107	34	16	51	149	1
Passing Yards	1434	346	169	453	1757	6

The least squares regression line is calculated to be $\hat{y} = -53.51 + 12.53x$.

What is the value of the residual for Quarterback 5 (Matt Nichols)?

- (A) -109.97 (B) -56.46 (C) -14.46 (D) 14.46 (E) 56.46

3. A farmer grows pumpkins, whose weights follow a normal distribution with mean 18 pounds and standard deviation 3 pounds. A supermarket will only buy those pumpkins that weigh between 15 and 25 pounds. What proportion of the farmer's pumpkins will the supermarket buy?

(A) 0.7850 (B) 0.8314 (C) 0.8876 (D) 0.9772 (E) 0.9996

4. The yearly rainfall in Regina, Saskatchewan follows a normal distribution with mean 384 mm and standard deviation σ . In 10% of years, the city gets less than 320 mm of rain. What is the standard deviation of the amount of annual rainfall in Regina?

(A) 40 mm (B) 50 mm (C) 60 mm (D) 70 mm (E) 80 mm

5. From past records, the professor of a large university course has established the following distribution of grades received by students in the course (with some values missing):

Grade	A ⁺	A	B ⁺	B	C ⁺	C	D	F
Probability	0.08	0.17	???	0.13	???	0.22	0.09	0.07

A student requires a grade of C or better to pass the course. What is the probability that a randomly selected student passes the course?

- (A) 0.76
 (B) 0.62
 (C) 0.84
 (D) impossible to calculate without at least one of the missing probabilities
 (E) impossible to calculate without both of the missing probabilities

6. Over the last year, suppose it is known that 32% of Winnipeggers have been to a Jets hockey game, 25% of Winnipeggers have been to a Blue Bombers football game, and 63% of Winnipeggers have been to **neither** a Jets game nor a Bombers game. What is the probability that a randomly selected Winnipegger has been to **both** a Jets game and a Bombers game over the last year?

(A) 0.08 (B) 0.12 (C) 0.15 (D) 0.18 (E) 0.20

7. A recently married couple plans to have two children. The outcome of interest is the gender of each of the two children. Consider the event that exactly one of the couple's children will be a boy. Which of the following is the complement of this event?
- (A) two boys
 - (B) two girls
 - (C) one girl
 - (D) at least one girl
 - (E) zero or two girls

The next **three** questions (**8 to 10**) refer to the following:

We have a small deck of ten cards. Five of the cards are red, three are blue and two are green. We randomly select four cards from the deck **with replacement**. That is, after we select a card and record the colour, we replace the card in the deck and thoroughly shuffle it before we select another card.

8. Let X be the number of blue cards that are selected. The distribution of X is:
- (A) binomial with parameters $n = 4$ and $p = 0.1$.
 - (B) binomial with parameters $n = 10$ and $p = 0.3$.
 - (C) binomial with parameters $n = 10$ and $p = 0.4$.
 - (D) binomial with parameters $n = 4$ and $p = 0.3$.
 - (E) not binomial.
9. Let A be the event that the first selected card is the only red card in our four selections. Which of the following events is mutually exclusive of the event A ?
- (A) second card selected is blue
 - (B) no green cards are selected
 - (C) third selected card is the only green card
 - (D) same number of red and blue cards are selected
 - (E) same number of blue and green cards are selected
10. Now suppose that, instead of selecting four cards, we repeatedly select cards with replacement (replacing the card and shuffling the deck after each draw) until we draw a green card for the first time. What is the probability that we draw our first green card on the fifth draw?
- (A) 0.0819 (B) 0.4096 (C) 0.1342 (D) 0.0554 (E) 0.1746

11. A backpacking party carries five emergency flares, each of which will light with a probability of 0.93. What is the probability that exactly four of the flares will light?
- (A) 0.1271 (B) 0.0524 (C) 0.2947 (D) 0.2618 (E) 0.1835
12. Suppose it is known that 11% of students at a large university live in dorms. In a random sample of 500 students at the university, what is the approximate probability that at least 14% of them live in dorms?
- (A) 0.0735 (B) 0.0418 (C) 0.0162 (D) 0.0028 (E) 0.0968
13. A variable X follows some left-skewed distribution with mean 100 and standard deviation 50. We take a random sample of 2,500 observations from this distribution and create a histogram of the data values. We expect the histogram to be:
- (A) approximately normal with mean close to 100 and standard deviation close to 1.
(B) skewed to the left with mean close to 100 and standard deviation close to 1.
(C) approximately normal with mean close to 100 and standard deviation close to 50.
(D) skewed to the left with mean close to 100 and standard deviation close to 50.
(E) approximately normal with mean close to 100 and standard deviation close to 0.02.
14. A manufacturer of automobile batteries claims that the distribution of battery lifetimes has a mean of 54 months and a variance of 36 months squared. A consumer group decides to check the claim by purchasing a sample of 50 of these batteries and subjecting them to tests to determine their lifetime. Assuming the manufacturer's claim is true, what is the probability that the sample has a mean lifetime less than 52 months?
- (A) 0.1292 (B) 0.3707 (C) 0.0091 (D) 0.4909 (E) 0.3483
15. A recycling plant compresses aluminum cans into bales. The weights of the bales are known to follow a normal distribution with mean 100 pounds standard deviation 8 pounds. What is the probability that a random sample of 64 bales has a mean weight between 99 and 101 pounds?
- (A) 0.3413 (B) 0.4772 (C) 0.5561 (D) 0.6826 (E) 0.7485

16. We would like to estimate the value of the mean μ of some population. Ten statisticians each take a separate random sample of 100 individuals from the population, and each of them calculates a 90% confidence interval for μ . What is the probability that exactly eight of their confidence intervals will contain the value of μ ?
- (A) 0.1445
(B) 0.1937
(C) 0.2324
(D) 0.2891
(E) depends on the value of μ
17. We would like to construct a confidence interval to estimate the true mean systolic blood pressure of all healthy adults to within 4 mm Hg (millimetres of mercury). We have a sample of 25 adults available. Systolic blood pressures of healthy adults are known to follow a normal distribution with standard deviation 8.6 mm Hg. What is the maximum confidence level that can be attained for our interval?
- (A) 80% (B) 90% (C) 95% (D) 96% (E) 98%
18. Lumber intended for building houses and other structures must be monitored for strength. A random sample of 25 specimens of Southern Pine is selected, and the mean strength is calculated to be 3700 pounds per square inch. Strengths are known to follow a normal distribution with standard deviation 500 pounds per square inch. An 85% confidence interval for the true mean strength of Southern Pine is:
- (A) (3615, 3785)
(B) (3671, 3729)
(C) (3556, 3844)
(D) (3544, 3856)
(E) (3596, 3804)

19. The University of Manitoba uses thousands of fluorescent light bulbs each year. The brand of bulb it currently uses has a mean lifetime of 900 hours. A manufacturer claims that its new brand of bulbs, which cost the same as those the university currently uses, has a mean lifetime of more than 900 hours. It is known that the standard deviation of lifetimes of the new brand of bulb is 80 hours. The university has decided to purchase the new brand of bulb if a hypothesis test gives significant evidence supporting the manufacturer's claim at the 3% level of significance. A random sample of 64 bulbs were tested and their mean lifetime was 920 hours. Based on these findings:
- (A) the university will not purchase the new brand of bulb, as the P-value is 0.0228.
 - (B) the university will not purchase the new brand of bulb, as the P-value is 0.0456.
 - (C) the university will not purchase the new brand of bulb, as the P-value is 0.9772.
 - (D) the university will purchase the new brand of bulb, as the P-value is 0.0228.
 - (E) the university will purchase the new brand of bulb, as the P-value is 0.0456.
20. We would like to determine whether the true mean pH level of a lake differs from 7.0. Lake pH levels are known to follow a normal distribution. We select a sample of ten water specimens from random locations in the lake. The sample mean pH level is calculated to be 6.8. A 98% confidence interval for μ is calculated to be (6.5, 7.1). Based on this confidence interval, our conclusion is to:
- (A) fail to reject H_0 at the 2% level of significance, since the value 7.0 is contained in the 98% confidence interval.
 - (B) fail to reject H_0 at the 1% level of significance, since the value 6.8 is contained in the 98% confidence interval.
 - (C) fail to reject H_0 at the 4% level of significance, since the value 7.0 is contained in the 98% confidence interval.
 - (D) reject H_0 at the 2% level of significance, since the value 7.0 is contained in the 98% confidence interval.
 - (E) reject H_0 at the 4% level of significance, since the value 6.8 is contained in the 98% confidence interval.

21. We select a random sample of ten oranges from trees grown in a large orchard. The standard deviation of the weights of these ten oranges is calculated to be 19 grams. Based on this sample, a confidence interval for the true mean weight of all oranges grown in the orchard is calculated to be (126.41, 153.59). What is the confidence level of this interval?
- (A) 90% (B) 95% (C) 96% (D) 98% (E) 99%

The next **two** questions (**22** and **23**) refer to the following:

Carbon Monoxide (CO) is a colourless and odourless gas. Even at low levels of exposure, carbon monoxide can cause serious health problems. A home is considered safe if the mean CO concentration is 4.8 parts per million (ppm) or lower. We take a random sample of eight readings in various locations of a home. These readings have a mean of 5.0 ppm and a standard deviation of 0.5 ppm. Assume that CO readings follow a normal distribution. We would like to conduct a hypothesis test at the 1% level of significance to determine whether the home is unsafe.

22. The P-value for the appropriate hypothesis test is:

- (A) between 0.02 and 0.025.
(B) between 0.025 and 0.05.
(C) between 0.05 and 0.10.
(D) between 0.10 and 0.15.
(E) between 0.15 and 0.20.

23. We conclude that:

- (A) there is proof that the home is safe.
(B) there is sufficient evidence that the home is safe.
(C) there is insufficient evidence that the home is unsafe.
(D) there is sufficient evidence that the home is unsafe.
(E) there is insufficient evidence that the home is safe.

24. The heights (in inches) of all 24 players on the Winnipeg Jets hockey team are ordered and shown below:

69 70 70 71 72 72 73 73 73 74 74 75
75 75 75 75 75 75 75 76 77 77 77 80

We would like to determine if the true mean height of Winnipeg Jets hockey players differs from 75 inches. Which of the following statements is **true**?

- (A) We should conduct a z test, as it is reasonable to assume heights follow a normal distribution.
- (B) We should conduct a t test, as we have no idea of the form of the distribution of heights.
- (C) We should conduct a z test, as the sample size is fairly high, so $\bar{X}$ will have an approximate normal distribution.
- (D) We should conduct a t test, as the population standard deviation is unknown.
- (E) A hypothesis test is unnecessary in this situation.

The next **three** questions (**25** to **27**) refer to the following:

The high school and university GPAs for random samples of students from two universities are shown below:

<u>University A</u>									
Student	1	2	3	4	5	6	7	$\bar{x}$	s
High School (H)	3.76	3.64	4.15	4.07	4.20	3.56	3.71	3.87	0.26
University (U)	3.74	4.30	4.09	3.90	3.46	3.37	3.04	3.70	0.44
Diff. ($d = H - U$)	0.02	-0.66	0.06	0.17	0.74	0.19	0.67	0.17	0.47

<u>University B</u>										
Student	1	2	3	4	5	6	7	8	$\bar{x}$	s
High School (H)	3.60	2.79	4.21	3.75	2.58	4.00	4.25	3.14	3.54	0.64
University (U)	3.02	2.42	4.23	3.87	2.52	3.15	4.03	3.00	3.28	0.69
Diff. ($d = H - U$)	0.58	0.37	-0.02	-0.12	0.06	0.85	0.22	0.14	0.26	0.33

25. We would like to construct a 95% confidence interval for the true mean difference μ_d in high school GPA and university GPA for all students at University A. Which of the following statements are **true**?
- (I) For any University A student, high school GPA and university GPA are independent.
 - (II) For any two University A students, high school GPAs are independent.
 - (III) In order to construct the confidence interval, we must assume that differences in high school GPA and university GPA follow a normal distribution.
 - (IV) In order to construct the confidence interval, we must assume that high school GPAs and university GPAs both follow normal distributions.
- (A) I and III only
(B) I and IV only
(C) II and III only
(D) II and IV only
(E) I, II and III only
26. Refer to the previous question. Assuming the necessary normality conditions are satisfied, the 95% confidence interval for the true mean difference in high school GPA and university GPA for all students at University A is:
- (A) $(-0.12, 0.46)$
(B) $(-0.47, 0.81)$
(C) $(-0.03, 0.37)$
(D) $(-0.26, 0.60)$
(E) $(-0.38, 0.72)$
27. We would like to conduct a hypothesis test to determine whether high school GPAs differ on average from university GPAs for all students at University B. Assuming the proper normality conditions are satisfied, the P-value for the appropriate test of significance is:
- (A) between 0.01 and 0.02.
(B) between 0.02 and 0.025.
(C) between 0.025 and 0.05.
(D) between 0.05 and 0.10.
(E) between 0.10 and 0.15.

28. The North Dakota Department of Tourism is interested in estimating the true proportion p of Manitobans visit the state in any given year. What sample size is required in order to estimate p to within 0.06 with 95% confidence?

(A) 234 (B) 245 (C) 256 (D) 267 (E) 278

29. In a random sample of 1,000 university students, 837 of them say they have a Facebook account. A 90% confidence interval for the true proportion of all university students who have a Facebook account is:

(A) (0.818, 0.856)

(B) (0.783, 0.891)

(C) (0.829, 0.845)

(D) (0.794, 0.880)

(E) (0.806, 0.868)

30. In a random sample of 175 university students, 70 have student loans. We would like to test the claim that more than 30% of all university students have student loans. The test statistic for the appropriate test of significance is:

(A) $\frac{0.30 - 0.40}{\sqrt{\frac{(0.30)(0.70)}{175}}}$

(D) $\frac{0.40 - 0.30}{\sqrt{\frac{(0.40)(0.60)}{175}}}$

(B) $\frac{0.40 - 0.30}{\sqrt{\frac{(0.30)(0.70)}{175}}}$

(E) $\frac{0.30 - 0.40}{\sqrt{\frac{(0.30)(0.40)}{175}}}$

(C) $\frac{0.40 - 0.30}{\sqrt{\frac{(0.35)(0.65)}{175}}}$

Sample Final Exam 2

Part B

1. A dentist would like to conduct an experiment to determine the effect of the brand of toothpaste (Crest or Colgate) and the type of toothbrush (regular or electric) his patients use on the health of their teeth. The dentist has 240 patients (80 children and 160 adults) who volunteer to participate in the experiment. The treatments will be randomly assigned to the subjects. The dentist believes the effect of the treatments will differ for children and adults, so she conducts a randomized block design.

(a) Identify the following in this experiment:

- i. factors
- ii. factor levels
- iii. treatments
- iv. response variable
- v. blocking variable

(b) How is control achieved in this experiment?

(c) Suppose at the end of the experiment that it is determined that subjects who used Crest toothpaste and an electric toothbrush had significantly healthier teeth than subjects in the other treatment groups. Can we conclude that the treatment was likely the **cause**?

2. The Winnipeg Jets' next three scheduled games are shown below, together with the probabilities that each team will win their respective game, as determined by odds-makers. Note that tied games are not possible. We will assume the outcome of one game is independent of any other.

Game 1:	Winnipeg Jets (0.6)	vs.	Pittsburgh Penguins (0.4)
Game 2:	Winnipeg Jets (0.7)	vs.	Montreal Canadiens (0.3)
Game 3:	Winnipeg Jets (0.8)	vs.	Buffalo Sabres (0.2)

- (a) The outcome of interest is the set of winners of each of the three games. List the complete sample space of outcomes and calculate the probability of each.
- (b) Let X be the number of games the Winnipeg Jets win. Find the probability distribution of X .

3. The weights of adobe bricks used for construction follow a normal distribution with mean 5.0 pounds and standard deviation 0.2 pounds.
- (a) Only 12% of bricks are heavier than what weight?
 - (b) We take a random sample of four bricks. Can you calculate the probability that the average weight of these four bricks is greater than 5.1 pounds? If so, calculate the probability. If not, explain why not.
 - (c) What is the probability that the total weight of a random sample of 50 bricks is less than 248 pounds?
 - (d) Is the probability you calculated in (c) exact or approximate? Explain.
4. A random sample of 20 students at a large university has a mean GPA of 3.12. GPAs at the university are known to follow a normal distribution with standard deviation 0.47.
- (a) Calculate a 95% confidence interval for the true mean GPA of all students at the university.
 - (b) Provide an interpretation of the confidence interval in (a).
 - (c) Conduct an appropriate hypothesis test at the 5% level of significance to determine whether there is evidence that the true mean GPA of students at the university differs from 3.00. Show all of your steps, including the statement of hypotheses, the calculation of the appropriate test statistic and P-value, and a carefully-worded conclusion.
 - (d) Provide an interpretation of the P-value of the test in (c).
 - (e) Could the confidence interval in (a) have been used to conduct the test in (c)? Why or why not? If it could have been used, explain what the conclusion would be, and why.

Sample Final Exam 3

Part A

1. The five-number summary for the heights (in cm) of the players on a men's hockey team is as follows:

162 174 176 181 194

The three goalies on the hockey team are Darren, who is 191 cm tall, Brandon, who is 165 cm tall, and Kyle, who is 193 cm tall. If an outlier boxplot for the heights of the hockey team was constructed, which of these players' heights would be labelled as outliers?

- (A) only Brandon
 - (B) only Kyle
 - (C) Darren and Kyle, but not Brandon
 - (D) Brandon and Kyle, but not Darren
 - (E) Darren, Brandon and Kyle
2. Researchers want to determine how the height of a mountain can help explain the temperature at the top of the mountain. The two variables were measured for a sample of mountains and the least squares regression line was calculated. It was also reported that 58% of the variation in temperature at the top of a mountain can be explained by its regression on the mountain's height. What is the value of the correlation between the two variables?

(A) 0.76 (B) 0.58 (C) -0.34 (D) -0.76 (E) -0.58

3. How fast do icicles grow? Researchers measured the growth rate of icicles in a cold chamber subjected to different temperatures (-20°C or -30°C) and different wind speeds (30, or 40 km/h). What is/are the factor level(s) in this experiment?
- (A) length of the icicles
 (B) growth rate of the icicles
 (C) temperature and wind speed
 (D) -20°C , -30°C , 30 km/h, 40 km/h
 (E) $-20^{\circ}\text{C}/30$ km/h, $-20^{\circ}\text{C}/40$ km/h, $-30^{\circ}\text{C}/40$ km/h, $-30^{\circ}\text{C}/40$ km/h
4. There are three coloured coins in a hat – one gold coin, one silver coin and one copper coin. You will randomly select coins from the hat, one at a time without replacement, until the gold coin is selected, and then you will stop. The outcome of interest is the sequence of colours that are selected during this process. How many outcomes are contained in the appropriate sample space?
- (A) 3 (B) 4 (C) 5 (D) 6 (E) 9
5. Consider two events A and B . We know $P(A) = 0.37$ and $P(A \cup B) = 0.7354$. If A and B are independent, then what is $P(B)$?
- (A) 0.47 (B) 0.39 (C) 0.62 (D) 0.44 (E) 0.58
6. A six-sided die has faces numbered 1, 2, 3, 4, 5 and 6. Suppose the die is “loaded”, so that any particular even-numbered face is twice as likely to land face up as any particular odd-numbered face. If we roll this die once, what is the probability that the number showing is greater than 3?
- (A) $\frac{1}{2}$ (B) $\frac{2}{3}$ (C) $\frac{5}{9}$ (D) $\frac{7}{12}$ (E) $\frac{9}{16}$

7. Two friends are playing a game of Rock, Paper, Scissors. To play this game, each player simultaneously and independently selects one of the three items. From prior experience, it is known that each of the two friends will select the items with the following probabilities:

Player	Rock	Paper	Scissors
Joe	0.4	0.2	0.4
Tom	0.5	0.2	0.3

In this game, rock beats scissors, scissors beats paper and paper beats rock. The two friends will play one game. What is the probability Joe wins against Tom?

- (A) 0.30 (B) 0.35 (C) 0.40 (D) 0.45 (E) 0.50

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

- (A) If the probability of A is 0.6 and the probability of B is 0.5, then A and B cannot be mutually exclusive.
- (B) If the probability of A is 0.4 and the probability of B is 0.6, and if A and B are independent, then $P(A \cap B)$ must be equal to 0.24.
- (C) If $P(A) = 0.3$, $P(B) = 0.6$, and $P(A \cup B) = 0.72$, then A and B must be independent.
- (D) If A and B are mutually exclusive, and if A and C are mutually exclusive, then B and C must be mutually exclusive.
- (E) Two events that are mutually exclusive cannot be independent.

9. Which of the following variables has a binomial distribution?

- (A) You roll five fair dice, each with face values of 1 through 6.
 X = total number of dots facing up on the five dice
- (B) You monitor the weather statistics every Saturday in Winnipeg for a year.
 X = number of Saturdays during the year that it snows
- (C) A paper boy delivers the newspaper to every house on your block.
 X = number of houses that get their newspaper on time tomorrow morning
- (D) You repeatedly flip two quarters simultaneously until both quarters land on Heads.
 X = number of flips required for both quarters to land on Heads
- (E) An unprepared student randomly guesses the answer to each of the 30 multiple-choice questions on the final exam.
 X = number of multiple-choice answers the student gets correct

10. During Tim Hortons' annual "Roll Up the Rim to Win" promotion, customers who purchase a cup of coffee check under the rim of the cup to see if they have won a prize. It is known that 16.7% of all cups are winners. If you buy one cup of coffee every day for a week (7 days), what is the probability you win at least two times?

(A) 0.1892 (B) 0.2349 (C) 0.2857 (D) 0.3311 (E) 0.3836

The next **three** questions (**11** to **13**) refer to the following:

The amount of soap per bottle for a particular brand of dish soap follows a normal distribution with mean 828 ml and standard deviation 4 ml.

11. What is the probability that a random sample of 10 bottles of dish soap contain a mean amount greater than 830 ml?

(A) 0.3085 (B) 0.1539 (C) 0.9429 (D) 0.2296 (E) 0.0571

12. There is an approximate 99.7% chance that a random sample of 16 bottles of dish soap will contain an average fill volume between:

(A) 825 and 831 ml.
(B) 824 and 832 ml.
(C) 826 and 830 ml.
(D) 816 and 840 ml.
(E) 820 and 836 ml.

13. What amount should be placed on the label of the bottles so that only 4% of bottles contain less than that amount?

(A) 820 ml (B) 821 ml (C) 822 ml (D) 828 ml (E) 829 ml

14. When an archer shoots her arrow, she hits the bullseye on the target 78% of the time. The result of each of her shots is independent of any other. If she shoots 300 arrows, what is the approximate probability she hits the bullseye at least 240 times?
- (A) 0.0162 (B) 0.0436 (C) 0.0719 (D) 0.1401 (E) 0.2005
15. The Central Limit Theorem states that:
- (A) when n gets large, the sample mean $\bar{X}$ gets closer and closer to the population mean μ .
- (B) if a variable X has a normal distribution, then for any sample size n , the sampling distribution of $\bar{X}$ is also normal.
- (C) if a variable X follows a normal distribution, then when n gets large, the sampling distribution of $\bar{X}$ is exactly normal.
- (D) when n gets large, the standard deviation of the sample mean $\bar{X}$ gets closer and closer to $\sigma/\sqrt{n}$.
- (E) regardless of the population distribution of a variable X , when n gets large, the sampling distribution of $\bar{X}$ is approximately normal.

The next **two** questions (**16** and **17**) refer to the following:

The fill volume per bottle for a certain brand of beer follows a normal distribution with mean 341 ml and standard deviation 3 ml.

16. If you buy a case of 24 bottles of beer, what is the probability that the bottles will contain an average of exactly 339.5 ml?
- (A) 0.0034 (B) 0.0207 (C) 0.0000 (D) 0.0071 (E) 0.0122
17. If you buy a twelve-pack of beer, what is the probability that the total volume of beer will exceed 4.1 litres?
- (A) 0.0838 (B) 0.2206 (C) 0.1397 (D) 0.4443 (E) 0.3632

18. Weights of loaves of bread made at a bakery follow a normal distribution with a mean of 450 grams and standard deviation 12 grams. A random sample of five loaves of bread is selected. There is only a 2.5% chance that the average weight of the sample will be below:
- (A) 439.48 grams.
 - (B) 441.20 grams.
 - (C) 442.54 grams.
 - (D) 438.62 grams.
 - (E) 440.76 grams.
19. Sizes of apartments in a large city follow a normal distribution with mean 846 square feet and standard deviation 160 square feet. We will take a simple random sample of four apartments in the city and calculate their average size $\bar{x}$. The sampling distribution of $\bar{X}$ is:
- (A) approximately normal with mean 846 and standard deviation 40.
 - (B) exactly normal with mean 846 and standard deviation 40.
 - (C) approximately normal with mean 846 and standard deviation 80.
 - (D) exactly normal with mean 846 and standard deviation 80.
 - (E) impossible to determine with the information given.
20. Consider two Canadian universities. University A has a student population of 10,000 and University B has a student population of 30,000. A statistician, interested in the average age of first-year students, collects the general records of samples of 500 first-year students from University A and 600 first-year students from University B. She assumes the standard deviations of ages are equal for the two universities. Based on her samples, she constructs a 95% confidence interval for the true mean age of first-year students at each university. Which of the following statements is/are **true**?
- (I) The interval for University A will be narrower, as a higher percentage of students was sampled from University A.
 - (II) The interval for University B will be narrower, as a higher number of students was sampled from University B.
 - (III) If the statistician calculated 90% intervals instead, the intervals would be wider.
 - (IV) If the statistician doubled her sample sizes, the lengths of her intervals would be reduced by half.
- (A) I only (B) II only (C) II and III (D) I and IV (E) II and IV

The next **two** questions (**21** and **22**) refer to the following:

City engineers in Hamilton, Ontario would like to estimate the true mean commuting distance of all workers in the city between home and their principal place of business. They calculate that, in order to estimate this mean to within ± 1 kilometre with 99% confidence, they require a sample of 120 workers.

21. What sample size would be required to estimate the true mean commuting distance for all workers in Hamilton to within ± 2 kilometres with 99% confidence?

(A) 30 (B) 60 (C) 85 (D) 240 (E) 480

22. The city of Montreal has a population five times greater than that of Hamilton (and so it would be logical to assume that it has five times as many workers). Suppose we wanted to estimate the true mean commuting distance between home and work for all workers in Montreal to within ± 1 kilometre with 99% confidence. Assuming equal standard deviations for the two cities, we would require a sample of how many Montreal workers?

(A) 24 (B) 120 (C) 269 (D) 600 (E) 3000

23. We take a random sample of n individuals and measure the value of some variable X . We conduct a hypothesis test of $H_0: \mu = 30$ vs. $H_a: \mu \neq 30$ at the 5% level of significance. We calculate a sample mean of $\bar{x} = 33$ and we reject the null hypothesis. Which of the following **must** be true?

- (I) A test of $H_0: \mu = 30$ vs. $H_a: \mu \neq 30$ at the 10% level of significance would also result in the rejection of H_0 .
- (II) A test of $H_0: \mu = 30$ vs. $H_a: \mu > 30$ at the 5% level of significance would also result in the rejection of H_0 .
- (III) A test of $H_0: \mu = 31$ vs. $H_a: \mu \neq 31$ at the 5% level of significance would also result in the rejection of H_0 .

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

24. Prior to distributing a large shipment of bottled water, a beverage company would like to determine whether there is evidence that the true mean fill volume of all bottles differs from 600 ml, which is the amount printed on the labels. Fill volumes are known to follow a normal distribution with standard deviation 2.0 ml. A random sample of 25 bottles is selected. The sample has a mean fill volume of 598.8 ml and a standard deviation of 3.0 ml. What is the value of the test statistic for the appropriate test of significance?
- (A) $t = -0.50$ (B) $z = -2.00$ (C) $t = -2.00$ (D) $z = -3.00$ (E) $t = -3.00$
25. We would like to conduct a hypothesis test at the 5% level of significance to determine whether there is evidence that the true mean amount of jam per jar for a certain brand of jam differs from 350 ml, the amount stated on the label. We take a simple random sample of 16 jars and measure the amount of jam in each of them. The sample mean is calculated to be 348.7 ml. Suppose it is known that the amount of jam per jar follows a normal distribution with standard deviation 2.8 ml. The correct conclusion is to:
- (A) reject H_0 , since the P-value is 0.0157.
 (B) reject H_0 , since the P-value is 0.0314.
 (C) fail to reject H_0 , since the P-value is 0.0314.
 (D) reject H_0 , since the P-value is 0.0628.
 (E) fail to reject H_0 , since the P-value is 0.0628.
26. We would like to conduct a hypothesis test of $H_0: \mu = 50$ vs. $H_a: \mu \neq 50$ for the mean μ of some normally distributed variable X . A random sample of 25 observations is taken from the population. A 90% confidence interval for μ is calculated to be (44.46, 49.58). A 96% confidence interval for μ is calculated to be (43.82, 50.22). The P-value of the appropriate test of significance must be:
- (A) less than 0.04.
 (B) greater than 0.10.
 (C) between 0.02 and 0.05.
 (D) between 0.04 and 0.10.
 (E) equal to 0.06.

The next **three** questions (**27** to **29**) refer to the following:

Coke and Pepsi are the two most popular colas on the market. Do consumers prefer either one of the two brands of cola over the other? We conduct a matched pairs experiment as follows: 20 volunteers participate in a blind taste test. Each volunteer tastes both Coke and Pepsi (in random order) and scores the taste of each cola on a scale from 0 to 100. Some information that may be helpful is shown in the table below:

Scores for Coke	Scores for Pepsi	Difference ($d = \text{Coke} - \text{Pepsi}$)
mean = 78	mean = 83	mean = -5
std. dev. = 27	std. dev. = 24	std. dev. = 13

27. Which of the following statements is/are **true**?

- (I) The scores for Coke and Pepsi for each individual are independent.
- (II) The scores for Coke and Pepsi for each individual are dependent.
- (III) In order to conduct the matched pairs t test, we must assume that scores for Coke and scores for Pepsi both follow normal distributions.
- (IV) In order to conduct the matched pairs t test, we must assume that differences in scores (Coke $-$ Pepsi) follow a normal distribution.

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

28. What are the hypotheses for the appropriate test of significance?

- (A) $H_0: \mu_d = 0$ vs. $H_a: \mu_d < 0$
- (B) $H_0: \bar{X}_d = 0$ vs. $H_a: \bar{X}_d \neq 0$
- (C) $H_0: \mu_C = \mu_P$ vs. $H_a: \mu_C < \mu_P$
- (D) $H_0: \bar{X}_d = 0$ vs. $H_a: \bar{X}_d < 0$
- (E) $H_0: \mu_d = 0$ vs. $H_a: \mu_d \neq 0$

29. Assuming the appropriate assumptions are satisfied, what is the value of the test statistic for the appropriate test of significance?

- (A) -0.38 (B) -7.69 (C) -2.85 (D) -1.72 (E) -0.88

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

- (A) The P-value in a hypothesis test is calculated under the assumption that the null hypothesis is true.
- (B) The lower the P-value of a hypothesis test, the greater the strength of evidence against the null hypothesis.
- (C) The level of significance in a hypothesis test is the highest P-value for which the null hypothesis will be rejected.
- (D) A hypothesis test is designed to assess the evidence in favour of the null hypothesis.
- (E) For normally distributed populations, the most important assumption in a hypothesis test for μ is that the sample is a simple random sample.

Sample Final Exam 3

Part B

1. Researchers are studying the effect of alcohol consumption on a person's short-term memory. Six volunteer subjects are assigned to consume a certain number of drinks X . A short time later, the subjects are given a memory test, and their score Y on the test is recorded. The data are shown below, together with the means and standard deviations:

Subject	1	2	3	4	5	6	mean	std. dev.
No. of Drinks	4	2	8	3	1	6	4	2.61
Test Score	81	90	22	58	80	53	64	25.07

The correlation between number of drinks and test score is calculated to be $r = -0.869$.

- What fraction of the variation in test score can be accounted for by its regression on number of drinks consumed?
 - Find the equation of the least squares regression line.
 - Find the predicted test scores for a subject who consumes 5 drinks and for a subject who consumes 11 drinks.
 - Is one of your predictions in (c) more reliable than the other? Explain.
2. Consider the following information about three events A , B and C :
- $P(A) = 0.43$
 - $P(C) = 0.28$
 - $P(A \cap B) = 0.19$
 - $P(B \cap C) = 0.12$
 - $P(A \cup B) = 0.61$
 - $P(A \cup C) = 0.71$
- What is $P(B)$?
 - Are any two of the three events mutually exclusive?
 - Are any two of the three events independent?
 - What is the probability that exactly one of the three events occurs?

3. We would like to estimate the true mean size μ (in square feet) of all two-bedroom apartments in Winnipeg. A random sample of 30 two-bedroom apartments in the city is selected, and the mean size of these apartments is calculated to be 1000 square feet. Suppose it is known that sizes of two-bedroom apartments in the city follow a normal distribution with standard deviation 200 square feet.
 - (a) Calculate a 93% confidence interval for the true mean size of all two-bedroom apartments in Winnipeg.
 - (b) Interpret the meaning of the interval you calculated in (a).
4. An apple grower selects a random sample of 30 apples from this year's crop. The mean and standard deviation of weights of these apples are calculated to be 156 grams and 14 grams, respectively. Weights of apples are known to follow a normal distribution.

The apple grower suspects that the true mean weight of apples from this year's crop differs from that of last year's crop, when the true mean weight was 150 grams. Conduct an appropriate hypothesis test at the 5% level of significance to investigate the grower's suspicion. Show all of your steps.

5. We would like to estimate the true proportion p of Canadians who support Donald Trump. In a random sample of 400 Canadians, 80 of them say they support the American president.
 - (a) Calculate a 90% confidence interval for the true proportion of Canadians who support Donald Trump.
 - (b) Conduct a hypothesis test at the 10% level of significance to determine whether there is evidence that the true proportion of Canadians who support Donald Trump is less than 0.25.
 - (c) Provide an interpretation of the P-value of the test in (b).

Sample Final Exam 4

Part A

1. The average height of all 23 students in a class is 132.5 cm. The average height of the 14 boys in the class is 135.2 cm. What is the average height of the girls in the class?

(A) 129.1 cm (B) 128.6 cm (C) 129.8 cm (D) 127.9 cm (E) 128.3 cm

2. The batting averages of 34 Major League Baseball players are ordered and shown below:

0.178	0.202	0.210	0.219	0.222	0.237	0.245	0.250	0.256
0.258	0.258	0.260	0.261	0.263	0.267	0.268	0.271	0.275
0.277	0.279	0.279	0.281	0.283	0.284	0.286	0.288	0.294
0.299	0.305	0.311	0.313	0.327	0.338	0.340		

The five-number summary is: 0.178 0.256 0.273 0.288 0.340

We construct an outlier boxplot for this data set. To what values do the lines coming out from the box (i.e., the whiskers) extend?

- (A) 0.202 and 0.338
(B) 0.210 and 0.327
(C) 0.208 and 0.336
(D) 0.237 and 0.313
(E) 0.178 and 0.340

3. A police officer would like to determine how the number of alcoholic beverages consumed by a person can predict his or her blood alcohol level. The officer measures the values of both variables on a sample of people leaving a bar one night. The correlation between the two variables is calculated to be 0.88 and the equation of the least squares regression line is calculated to be $\hat{y} = 0.003 + 0.012x$. Which of the following statements is **true**?
- (A) 88% of the variation in blood alcohol level can be accounted for by its regression on number of alcoholic beverages consumed.
 - (B) 77% of the variation in number of alcoholic beverages consumed can be accounted for by its regression on blood alcohol level.
 - (C) 94% of the variation in blood alcohol level can be accounted for by its regression on number of alcoholic beverages consumed.
 - (D) 88% of the variation in number of alcoholic beverages consumed can be accounted for by its regression on blood alcohol level.
 - (E) 77% of the variation in blood alcohol level can be accounted for by its regression on number of alcoholic beverages consumed.
4. 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) sunscreen brand, SPF level and degree of sunburn
 - (D) Coppertone, Ombrelle, SPF 30, SPF 60
 - (E) Coppertone/SPF 30, Ombrelle/SPF 30, Coppertone/SPF 60, Ombrelle/SPF 60

5. The professor of a large class decides to grade the final exam on a curve. He will use the following criteria:

- 15% of students will get a grade of A⁺.
- 17% of students will get a grade of A.
- 11% of students will get a grade of B⁺.
- 18% of students will get a grade of B.
- 9% of students will get a grade of C⁺.
- 12% of students will get a grade of C.
- 10% of students will get a grade of D.
- 8% of students will get a grade of F.

Scores on the exam are known to follow a normal distribution with mean 67 and standard deviation 11. What is the minimum score required to obtain a grade of C?

- (A) 56.88 (B) 57.26 (C) 58.90 (D) 59.32 (E) 60.44

6. Labels on the boxes for a certain brand of soap claim the bars of soap weigh 150 grams. Weights of bars of soap are in fact known to be normally distributed with a standard deviation of 4 grams. What should be the mean weight of all bars of soap if we want only 7% of bars to weigh less than the amount stated on the label?

- (A) 144.08 grams
(B) 145.83 grams
(C) 153.28 grams
(D) 154.75 grams
(E) 155.92 grams

The next **five** questions (**7** to **11**) refer to the following:

Winnipeg has three professional sports teams – the Winnipeg Jets hockey team, the Winnipeg Blue Bombers football team and the Winnipeg Goldeyes baseball team.

Suppose we have the following information about Winnipeg residents:

- 55% are Jets fans (J).
- 20% are Goldeyes fans (G).
- 50% are Bombers fans (B) or Goldeyes fans.
- 27% are Jets fans and Bombers fans.
- 11% are Jets fans and Goldeyes fans.
- 10% are Bombers fans and Goldeyes fans.
- 6% are fans of all three teams.

7. What is the probability that a randomly selected Winnipeg resident is a Bombers fan?

- (A) 0.30 (B) 0.35 (C) 0.40 (D) 0.45 (E) 0.50

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

- (A) Events J and B are independent.
(B) Events J and G are mutually exclusive.
(C) Events J and G are independent.
(D) Events B and G are mutually exclusive.
(E) Events B and G are independent.

9. What is the probability that a randomly selected Winnipeg resident is a fan of none of the three teams?

- (A) 0.21 (B) 0.23 (C) 0.25 (D) 0.27 (E) 0.29

10. If we take a random sample of 12 Winnipeg residents, what is the probability that exactly three of them are Goldeyes fans?

- (A) 0.2084 (B) 0.2137 (C) 0.2256 (D) 0.2362 (E) 0.2419

11. If we take a random sample of 500 Winnipeg residents, what is the approximate probability that less than half of them are Jets fans?

(A) 0.0122 (B) 0.0287 (C) 0.0351 (D) 0.0446 (E) 0.0594

The next **three** questions (**12** to **14**) refer to the following:

We have four boxes, each of which contains 10 coloured balls:

- Box 1 contains 4 red balls, 5 green balls and 1 yellow ball.
- Box 2 contains 3 red balls, 5 green balls and 2 yellow balls.
- Box 3 contains 2 red balls, 5 green balls and 3 yellow balls.
- Box 4 contains 1 red ball, 5 green balls and 4 yellow balls.

12. If we randomly select one ball from each box, what is the probability of selecting **exactly** one yellow ball?

(A) 0.44 (B) 0.40 (C) 0.36 (D) 0.48 (E) 0.32

13. If we randomly select one ball from each box, what is the probability of selecting **at least** one yellow ball?

(A) 0.5 (B) 0.6 (C) 0.7 (D) 0.8 (E) 0.9

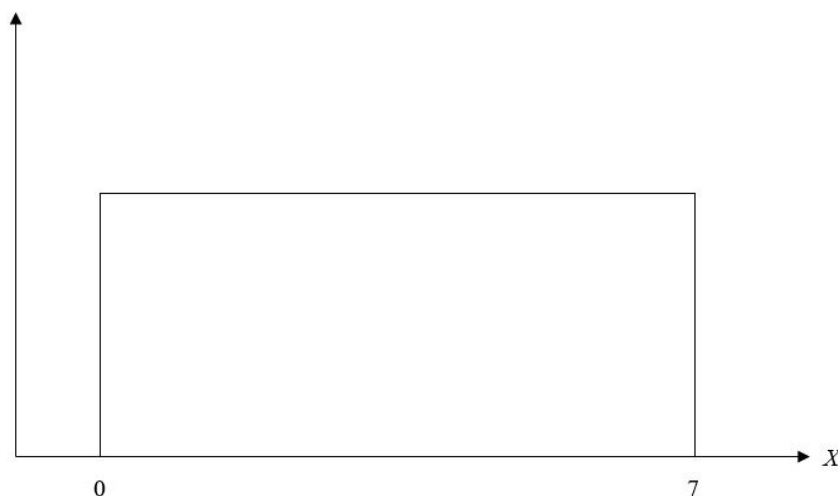
14. Which of the following variables have a binomial distribution?

- (I) Randomly select three balls from Box 1 with replacement.
 X = number of red balls selected
- (II) Randomly select one ball from each of the four boxes.
 X = number of yellow balls selected
- (III) Randomly select three balls from Box 1 without replacement.
 X = number of red balls selected
- (IV) Randomly select one ball from each of the four boxes.
 X = number of green balls selected

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

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

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



The mean and standard deviation of this distribution are 3.5 and 2, respectively.

15. We take a random sample of 400 individuals from this distribution and create a histogram of the data values. We expect this histogram to be:
- (A) approximately normal with mean close to 3.5 and standard deviation close to 0.1.
 - (B) approximately uniform with mean close to 3.5 and standard deviation close to 2.
 - (C) approximately normal with mean close to 3.5 and standard deviation close to 0.005.
 - (D) approximately uniform with mean close to 3.5 and standard deviation close to 0.1.
 - (E) approximately normal with mean close to 3.5 and standard deviation close to 2.
16. We take a random sample of 400 individuals from this distribution and calculate the value of the sample mean $\bar{x}$. The sampling distribution of $\bar{X}$ is:
- (A) approximately normal with mean 3.5 and standard deviation 0.1.
 - (B) uniform with mean 3.5 and standard deviation 2.
 - (C) approximately normal with mean 3.5 and standard deviation 0.005.
 - (D) uniform with mean 3.5 and standard deviation 0.1.
 - (E) approximately normal with mean 3.5 and standard deviation 2.

The next **two** questions (**17** and **18**) refer to the following:

Credit card balances for the I. O. U. Credit Corporation follow a normal distribution with mean \$800 and the standard deviation is \$200.

17. What is the probability that the mean balance for a random sample of 100 customers is greater than \$780?

(A) 0.7157 (B) 0.5793 (C) 0.9484 (D) 0.8413 (E) 0.6368

18. What is the probability that the total balance for a random sample of 40 customers is less than \$30,000?

(A) 0.0162 (B) 0.0287 (C) 0.0336 (D) 0.0446 (E) 0.0571

19. A variable X follows a normal distribution with mean 43 and variance 144. We take a random sample of n individuals from this distribution and calculate the sample mean $\bar{x}$. What must the sample size be so that the standard deviation of the sample mean $\bar{X}$ is equal to 0.8?

(A) 15 (B) 64 (C) 120 (D) 180 (E) 225

20. The number of undergraduate students at the University of Winnipeg is approximately 9,000, while the University of Manitoba has approximately 27,000 undergraduate students. Suppose that, at each university, a simple random sample of 3% of the undergraduate students is selected and the following question is asked: “Do you approve of the provincial government’s decision to lift the tuition freeze?”. Suppose that, within each university, approximately 20% of undergraduate students favour this decision. What can be said about the sampling variability associated with the two sample proportions?

- (A) The sample proportion for the U of W has less sampling variability than that for the U of M.
(B) The sample proportion for the U of W has more sampling variability than that for the U of M.
(C) The sample proportion for the U of W has approximately the same sampling variability as that for the U of M.
(D) It is impossible to make any statements about the sampling variability of the two sample proportions without taking many samples.
(E) It is impossible to make any statements about the sampling variability of the two sample proportions because the population sizes are different.

21. The sizes of farms in a U.S. state follow a normal distribution with standard deviation 30 acres. We would like to take a sample of farms large enough to estimate the true mean size of all farms in the state to within 5 acres with 94% confidence. What sample size is required?
- (A) 126 (B) 128 (C) 130 (D) 132 (E) 134
22. A researcher would like to estimate the true mean daily water intake μ for all Canadians. She measures the daily water intake (in ml) for a sample of Canadians and calculates a 95% confidence interval for μ to be (1771.6, 1928.4). Daily water intake for Canadians is known to follow a normal distribution with standard deviation 280 ml. What sample size did the researcher take?
- (A) 16 (B) 25 (C) 36 (D) 49 (E) 64
23. We would like to test whether the true mean IQ of all adult Canadians is less than 110. Suppose that IQs of adult Canadians follow a normal distribution with standard deviation 17. A random sample of 30 adult Canadians has a mean IQ of 108. What is the P-value for the appropriate test of $H_0: \mu = 110$ vs. $H_a: \mu < 110$?
- (A) 0.6444
 (B) 0.2090
 (C) 0.3556
 (D) 0.2611
 (E) impossible to determine because the level of significance was not given
24. Which of the following statements comparing the standard normal distribution and the t distributions is **false**?
- (A) The density curve for Z is taller at the centre than the density curve for T .
 (B) The t distributions have more area in the tails than the standard normal distribution.
 (C) In tests of significance for μ , Z should be used as the test statistic when the distribution of X is normal, and T should be used in other cases.
 (D) As the sample size increases, the t distributions approach the standard normal distribution.
 (E) In tests of significance for μ , T should be used as the test statistic when the population standard deviation is unknown.

25. A sample of five Major League Baseball games has a mean duration of 168 minutes and a standard deviation of 17 minutes. Game durations are known to follow a normal distribution. A 90% confidence interval for the true mean duration of all Major League baseball games is:
- (A) (155.49, 180.51)
 - (B) (154.18, 181.82)
 - (C) (153.10, 182.90)
 - (D) (152.68, 183.32)
 - (E) (151.79, 184.21)
26. We would like to conduct a hypothesis test at the 10% level of significance to determine whether the true mean score of all players in a bowling league differs from 150. The mean and standard deviation of the scores of 12 randomly selected players are calculated to be 170 and 16, respectively. Scores of all players in the league are known to follow a normal distribution. The P-value of the appropriate test of significance is:
- (A) between 0.0005 and 0.001.
 - (B) between 0.001 and 0.002.
 - (C) between 0.0025 and 0.005.
 - (D) between 0.005 and 0.01.
 - (E) between 0.01 and 0.02.
27. For which of the following situations is a matched pairs t test appropriate? (Assume all appropriate normality conditions are satisfied.) A researcher wants to know:
- (I) whether hypnosis reduces pain. 20 volunteers place their hand in ice water and the time until the pain is unbearable is recorded. They are hypnotized and the exercise is repeated.
 - (II) whether Shell sells more gasoline than Petro Canada. The weekly gas sales are recorded for samples of 5 Shell gas stations and 5 Petro Canada gas stations.
 - (III) whether husbands spend more household money than wives on average. The amount spent on the husband's credit card and wife's credit card are compared for a sample of 10 married couples.
- (A) I only (B) III only (C) I and II (D) I and III (E) II and III

28. On the evening news, a reporter states, “In a recent poll of likely voters, 50% of respondents indicated that they plan to support the Conservative Party in the upcoming provincial election. The results are accurate to within 3.1%, 19 times out of 20”. What sample size was used for the survey?
- (A) 600 (B) 800 (C) 1,000 (D) 1,200 (E) 1,400
29. We would like to estimate some population proportion p . We take a random sample of n individuals and calculate an 80% confidence interval for p to be **(0.33, 0.37)**. Using the same sample, which of the following could be a 99% confidence interval for p ?
- (A) (0.36, 0.38)
(B) (0.32, 0.40)
(C) (0.34, 0.36)
(D) (0.36, 0.44)
(E) (0.31, 0.39)
30. We would like to conduct a hypothesis test at the 5% level of significance to determine whether the true proportion of Canadians who support Prime Minister Justin Trudeau differs from 0.50. In a random sample of 300 Canadians, 165 of them say they support the prime minister. The correct conclusion is to:
- (A) reject H_0 , since the P-value is 0.0418.
(B) fail to reject H_0 , since the P-value is 0.0836.
(C) fail to reject H_0 , since the P-value is 0.9582.
(D) reject H_0 , since the P-value is 0.0836.
(E) fail to reject H_0 , since the P-value is 0.0418.

Sample Final Exam 4

Part B

1. A student driving to university must pass through two sets of traffic lights. The first set of lights is green 50% of the time, yellow 10% of the time, and red 40% of the time. The second set of lights is green 60% of the time, yellow 10% of the time, and red 30% of the time. It is also known that the lights function independently.
 - (a) The outcome of interest is the colour of each of the two lights when the student arrives at the intersections. List the complete sample space of outcomes, and calculate the probability of each.
 - (b) What is the probability that both sets of lights are the same colour when the student arrives at the respective intersections?
 - (c) What is the probability that the first light is green or the second light is red?
 - (d) Let X be the number of red lights encountered by the student on her way to university. Find the probability distribution of X .
2.
 - (a) List the four conditions of the binomial setting.
 - (b) Heights of professional basketball players follow a normal distribution with mean 196 cm and standard deviation 8 cm. What proportion of professional basketball players are taller than 200 cm?
 - (c) If we take a random sample of ten professional basketball players, what is the probability that exactly four of them are taller than 200 cm?
3.
 - (a) Give the definition of a lurking variable.
 - (b) Give the definition of a simple random sample.
 - (c) Explain the difference between a statistic and a parameter. Give three examples of each.
 - (d) Explain the difference between two events being mutually exclusive and two events being independent.
 - (e) Explain what it means for a result to be statistically significant.

4. One measure of the quality of education provided by a university is the number of students per class. A random sample of 30 third-year classes is selected at a large university. The number of students in each of the classes are ordered and shown below:

10	11	14	14	15	17	18	20	22	22
22	23	26	26	27	28	31	33	34	34
36	42	44	49	50	55	62	68	77	82

From these data, the sample mean is calculated to be 33.73. The population standard deviation of class sizes is known to be 18.50.

- Construct a histogram for this data set. What is the shape of the distribution of class sizes for this sample?
 - We would like to estimate and conduct a hypothesis test for the true mean class size of all third-year classes at the university. Despite the fact that class size does not appear to follow a normal distribution, explain why it is nevertheless appropriate to use inference methods which rely on the assumption of normality.
 - Construct a 99% confidence interval for the true mean class size of all third-year classes at this university.
 - Provide an interpretation of the confidence interval in (c).
 - Conduct an appropriate hypothesis test at the 1% level of significance to determine whether there is evidence that the true mean third-year class size at this university differs from the national average of 40. Show all of your steps.
5. Sixteen people (eight males and eight females) volunteered to be part of an experiment. All 16 people were between the ages of 25 and 35. The question of interest in this experiment was whether females receive faster service at nightclubs than males. Each of the eight male participants was randomly assigned a bar, and each of the eight females was randomly assigned to one of these same eight bars. One Friday night, all 16 people went out to the bar. The male and female assigned to the same bar would arrive within five minutes of each other. Each person then waited at the bar and ordered a similar drink. The time (in seconds) until the drink was served was recorded. Some information that may be helpful is shown below:

Females	Males	Difference ($d = F - M$)
mean = 275	mean = 302	mean = -27
std. dev. = 37	std. dev. = 48	std. dev. = 29

- Conduct an appropriate hypothesis test, at the 1% level of significance. Show all of your steps, including the statement of hypotheses, the calculation of the appropriate test statistic and P-value, and a carefully-worded conclusion.
- Provide an interpretation of the P-value of the test in (a).

Sample Final Exam Answers

Part A

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

Sample Final 1

Part B Solutions

1. (a) The minimum is 2.2 and the maximum is 11.4.

The median is in position $(n + 1)/2 = (30 + 1)/2 = 15.5$, so the median is the average of the 15th and 16th ordered data values, i.e., $(7.0 + 7.2)/2 = 7.1$.

The first quartile is the median of all data values lower in position than the median, so $Q1$ is in position $(15 + 1)/2 = 8$. Therefore, $Q1 = 6.0$.

The third quartile is in position 8 above the median (or equivalently, 8 positions down from the maximum), and so $Q3 = 8.0$.

The five-number summary is: 2.2 6.0 7.1 8.0 11.4

- (b) We calculate the fences:

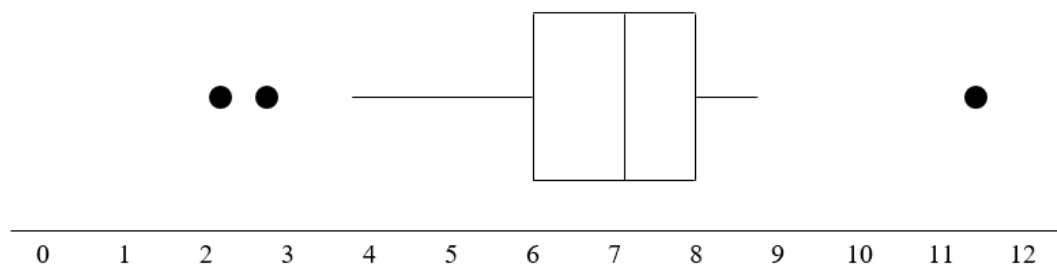
$$LF = Q1 - 1.5IQR = 6.0 - 1.5(8.0 - 6.0) = 6.0 - 3.0 = 3.0$$

$$UF = Q3 + 1.5IQR = 8.0 + 1.5(8.0 - 6.0) = 8.0 + 3.0 = 11.0$$

There are two values less than the lower fence (2.2 and 2.8) and one value greater than the upper fence (11.4). These outliers will be plotted as points on the outlier boxplot.

The whiskers extend to the lowest and highest data values that are not outliers (the “new minimum” and “new maximum”), i.e., 3.8 and 8.8.

The outlier boxplot is shown below:



Excluding outliers, the distribution is skewed to the left.

2. (a) When the age of a car increases by one year, we predict its selling price to decrease by \$1,700.
- (b) We know that 81% of the variation in a car's selling price can be accounted for by its regression on the age of a car. This is the definition of r^2 , so we know $r^2 = 0.81$. The correlation is therefore

$$r = -\sqrt{r^2} = -\sqrt{0.81} = -0.90$$

Note that we take the negative square root of r^2 , since there is a negative association between the two variables.

- (c) The residual for this car is

$$y_i - \hat{y}_i = 15500 - (18600 - 1700(5)) = 5400$$

3. (a) $S = \{GGG, GGC, GSG, GSC, SGG, SGC, SSG, SSC, CGG, CGC, CSG, CSC\}$
- (b) The probability distribution of X is as follows:

x	0	1	2	3
$P(X = x)$	0.336	0.452	0.188	0.024

4. The Central Limit Theorem states that, regardless of the population distribution of a variable X , when the sample size is high (we use $n \geq 30$), the sampling distribution of $\bar{X}$ is approximately normal.

5. (a) (25.87, 30.93)
- (b) If we took repeated samples of 27 trees and calculated the interval in a similar manner, 98% of all such intervals would contain the true mean height of all trees in the forest.
- (c) $t = -1.57$, $0.05 < \text{P-value} < 0.10$, fail to reject H_0 .
- (d) If the true mean height of all trees in the forest was 30 feet, the probability of observing a sample mean at least as low as 28.4 feet would be between 0.05 and 0.10.
- (e) No

Sample Final 2

Part B Solutions

1. (a)
 - i. The factors in an experiment are the explanatory variables. There are two factors in this experiment – brand of toothpaste and type of toothbrush.
 - ii. Brand of toothpaste has two factor levels – Crest and Colgate.
Type of toothbrush has two factor levels – regular and electric.
 - iii. There are $2 \times 2 = 4$ treatments – Crest/regular, Crest/electric, Colgate/regular and Colgate/electric.
 - iv. The response variable is the health of the patients' teeth.
 - v. The blocking variable is age (one block consisting of the children and the other block consisting of the adults).
- (b) Control is achieved by the comparison of the four treatment groups in each block. (Note that blocking is also another form of control. In this case, by using a randomized block design, we eliminate the possibility of age becoming a lurking variable.)
- (c) Yes. Since this was a properly designed experiment using the principles of randomization, replication and control, we can conclude that the treatment was likely the cause of the observed difference in the response.

2. (a) $S = \{WWW, WWB, WMW, WMB, PWW, PWB, PMW, PMB\}$

- (b) The probability distribution of X is as follows:

x	0	1	2	3
$P(X = x)$	0.024	0.188	0.452	0.336

3. (a) 5.235
- (b) 0.1587
- (c) 0.0793
- (d) exact

4. (a) (2.91, 3.33)
- (b) If we took repeated samples of 20 students at the university and calculated the interval in a similar manner, 95% of such intervals would contain the true mean GPA of all students at the university.
- (c) $z = 1.14$, P-value = 0.2542, fail to reject H_0 .
- (d) If the true mean GPA of all students at the university was 3.00, the probability of observing a sample mean at least as extreme as 3.12 would be 0.2542.
- (e) Yes. Fail to reject H_0 .

Sample Final 3

Part B Solutions

1. (a) The fraction of the variation in test score that can be accounted for by its regression on number of drinks consumed is, by definition, r^2 . Here, $r^2 = (-0.869)^2 = 0.755$.

- (b) The slope of the least squares regression line is

$$b_1 = r \frac{s_y}{s_x} = -0.869 \left(\frac{25.07}{2.61} \right) = -8.347$$

The intercept of the least squares regression line is

$$b_0 = \bar{y} - b_1 \bar{x} = 64 - (-8.347)(4) = 97.388$$

The equation of the least squares regression line is $\hat{y} = 97.388 - 8.347x$.

- (c) The predicted test score for a subject who consumes 5 drinks is

$$\hat{y} = 97.388 - 8.347(5) = 55.653$$

- (d) The predicted test score for a subject who consumes 11 drinks is

$$\hat{y} = 97.388 - 8.347(11) = 5.571$$

- (e) The prediction for the subject who consumed 5 drinks is more reliable, since the value $x = 11$ falls outside the range of x -values in our sample. Predicting the test score for the subject who consumed 11 drinks is extrapolating.

2. (a) 0.37

(b) A and B are **not** mutually exclusive.

B and C are **not** mutually exclusive.

A and C **are** mutually exclusive.

(c) A and B are **not** independent.

A and C are **not** independent.

B and C are **not** independent.

(d) 0.46

3. (a) (933.9, 1066.1)

(b) If we took repeated samples of size 30 and calculated the interval in a similar manner, 93% of all such intervals would contain the true mean size of all two-bedroom apartments in Winnipeg.

4. $t = 2.35$, $0.02 < \text{P-value} < 0.04$, reject H_0 .

5. (a) (0.1671, 0.2329)

(b) $z = -2.31$, $\text{P-value} = 0.0104$, reject H_0 .

(c) If the true proportion of Canadians who supported Donald Trump was 0.25, the probability of observing a sample proportion at least as low as 0.20 would be 0.0104.

Sample Final 4

Part B Solutions

1. (a) $S = \{GG, GY, GR, YG, YY, YR, RG, RY, RR\}$

(b) 0.43

(c) 0.65

(d) The probability distribution of X is as follows:

x	0	1	2
$P(X = x)$	0.42	0.46	0.12

2. (a) A binomial experiment is one for which the following four conditions are satisfied:

- There is a fixed number of trials n .
- There are only two possible outcomes of interest for each trial – a success or a failure.
- The outcome of one trial is independent of any other.
- The probability of success p is constant from trial to trial.

(b) 0.3085

(c) 0.2080

3. (a) A lurking variable is one that explains the relationship between variables in a study, but is not included in the study itself.

(b) A simple random sample of size n is one that is chosen in such a way that all possible groups of n individuals have the same chance of being chosen.

(c) A statistic is a number that describes a sample. Examples of statistics include the sample mean $\bar{x}$, the sample standard deviation s and a sample proportion $\hat{p}$. A parameter is a number that describes an entire population. Examples of parameters include the population mean μ , the population standard deviation σ and a population proportion p .

- (d) Two events that are mutually exclusive have no outcomes in common (i.e., they cannot both occur at the same time). Two events that are independent can both occur at the same time, but the outcome of one does not affect the outcome of the other.
 - (e) Statistical significance is a difference so large that it would rarely be observed by chance alone. We say that the results of a hypothesis test are statistically significant when the P-value $\leq \alpha$ and we reject H_0 .
4. (a) From a histogram of the data, we see the distribution of class sizes for this sample is skewed to the right.
- (b) Although it does not appear that the population distribution of class sizes is normally distributed, the Central Limit Theorem tells us that since the sample size is high (we use $n \geq 30$), the sampling distribution of the sample mean $\bar{X}$ is approximately normal.
- (c) (25.03, 42.43)
- (d) If we took repeated samples of 30 classes and calculated the interval in a similar manner, 99% of such intervals would contain the true mean class size of all third-year classes at the university.
- (e) $z = -1.86$, P-value = 0.0628, fail to reject H_0 .
5. (a) $t = -2.63$, $0.01 < \text{P-value} < 0.02$, fail to reject H_0 .
Note that $n = 8$ (and not 16), as n is the number of pairs.
- (b) If the service time in nightclubs was the same on average for males and females, the probability of observing a sample mean difference at least as low as -27 would be between 0.01 and 0.02.