University of Regina
Department of Mathematics & Statistics
Final Examination
202110

Statistics 160
Introductory Statistics

Name: ______________________________ **Student Number:** ______________

Instructor: Dr M Kozdron

Read all of the following information before starting the exam.

This exam is due electronically through UR Courses by 5:00 pm (Regina Time) on Saturday, December 11, 2021. Some options include: you may print out this exam and write your solutions on it, or you may write your solutions on your own paper and photograph your solutions, or you may type up your solutions using a word processor.

Although not required, it would be preferred if you uploaded your solutions as a .pdf file.

Late submissions will not be accepted. Please refer to the "Deferrals" section on pages 43–44 of the 2021-2022 University of Regina Undergraduate Calendar.

Please read all instructions carefully, and check your answers. Show all work neatly and in order, and clearly indicate your final answers. Answers must be justified whenever possible in order to earn full credit. Answers to multiple choice questions do not need to be justified.

You may use standard notation; however, any new notations or abbreviations that you introduce must be clearly defined.

You may consult any notes or books or websites that you wish, provided that proper acknowledgements are included. For example, if you find a solution using Google, that must be cited. However, you may not discuss this exam with anyone except the course instructor before 5:00 pm on Saturday, December 11, 2021. This includes students, other professors, and colleagues, as well as in online chat groups or on social media.

*This exam has **14** problems and is worth a total of **140** points. For problems with multiple parts, each part is equally weighted.*

For each of the following four problems, choose ALL answers that are correct.
You do NOT need to justify your answer.

1. (*3 points*) The Central Limit Theorem says

(a) the sample mean has a normal distribution;

(b) the sample mean is approximately normal with mean equal to the population mean and standard deviation equal to the population standard deviation divided by the square root of the sample size, when the sample size is large;

(c) the t-distributions are nearly normal, and in the limit of high degrees of freedom are normal with mean zero and standard deviation one;

(d) the sample mean has a t-distribution with $n-1$ degrees of freedom;

(e) the sampling distribution for the sample mean tends to a normal with mean equal to the population mean and standard deviation equal to the (true) standard error of the mean.

2. (*3 points*) A random sample of size $n = 25$, taken from a normally distributed population, gives the sample mean $\bar{x} = 50$ and the sample standard deviation $s = 5.00$. Based on this information, and using a 95% confidence level,

(a) the point estimate of the population mean is 50;

(b) the margin of error is 1.96;

(c) the standard error of the sample mean is 1.00;

(d) the standard deviation of the sample mean is 5.00;

(e) the upper confidence limit is 51.96.

3. (*3 points*) Testing $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$

(a) gives a two-tailed rejection region, one on each side of zero;

(b) by way of a t-statistic is equivalent to asking if the t-based confidence interval contains zero, for an appropriate confidence level;

(c) cannot be done unless you know that (i) $\sigma_1 = \sigma_2$, and (ii) their common value is known;

(d) cannot be done unless you know that (i) $\sigma_1 = \sigma_2$, and (ii) their common value is unknown;

(e) is equivalent to asking if a certain sort of confidence interval does not contain zero.

4. (*3 points*) A sample mean and a sample median

(a) are usually close for symmetric distributions;

(b) are equally good measures of centre when the tails of a distribution are light;

(c) have the same sampling distribution;

(d) are usually far apart for skewed distributions;

(e) can help you determine if the distribution is skewed.

5. (*12 points*) During January 2019, Michael (a Canadian from Regina, SK), was escaping the snow and vacationing in Phoenix, Arizona. While watching the evening news, the American meteorologist would announce the day's high temperature in degrees Fahrenheit. Michael remembered to write down the temperature 18 times during his 31-day stay in Phoenix. Here is what he recorded:

$$74, 78, 82, 83, 78, 73, 84, 89, 91, 78, 77, 82, 69, 93, 76, 74, 80, 83.$$

(a) Construct a stem-and-leaf plot of this temperature data by completing the following table.

5	
6	
7	
8	
9	

(b) Determine the median temperature (in degrees Fahrenheit) for those days in which Michael remembered to record the temperature.

(c) In order to gloat with his friend's back home, Michael needed to convert his temperature observations into degrees Celsius. Remembering that

$$\text{fahrenheit} = \left(\frac{9}{5} \times \text{celsius}\right) + 32$$

determine, in *degrees Celsius*, the median of this temperature data.

6. (*12 points*) A national restaurant chain conducts a simple random sample of some of its customers. The card these customers are given to fill out asks their opinions about their meal, the service, the cleanliness, and so on. One question asks the customer to rate the quality of the service as poor, below average, average, above average, or outstanding. The following data represent the results obtained for the Regina branch of this restaurant chain.

Poor	Below average	Average	Above average	Outstanding
7	17	38	57	21

(a) Describe **(i)** the population of interest, and **(ii)** the sample. **(iii)** What is the variable of interest? Is it categorical or quantitative?

(b) Let p denote the true (but unknown) proportion of customers that would rate the sampled restaurant as either above average or outstanding. Give a point estimate $\hat{p}$ of p, and use $\hat{p}$ to construct an approximate 99% confidence interval for p.

(c) Can the results derived from the given data be extended to *all* other branches in this restaurant chain? Explain why or why not.

7. (*12 points*) Acid rain, caused by the reaction of certain air pollutants with rainwater, is a growing problem in Canada. Pure rain falling through clean air registers a pH value of 5.7. Note that pH is a measure of acidity: 0 is pure acid and 14 is pure alkaline. The smaller the pH, the more acidic the rain. The sample of $n = 49$ rainfalls in the Greater Toronto Area (GTA) produced pH readings with $\overline{x} = 5.5$ and $s = 0.7$. Do the data provide sufficient evidence to indicate that the mean pH for rainfalls in the GTA is **more acidic** than pure rainwater? Test with $\alpha = 0.05$.

(a) Choose the correct null and alternative hypotheses. *You do NOT need to justify your answer.*

- **(i)** $H_0 : \mu = 5.7$ against $H_a : \mu > 5.7$
- **(ii)** $H_0 : \mu = 5.7$ against $H_a : \mu \neq 5.7$
- **(iii)** $H_0 : \mu = 5.7$ against $H_a : \mu < 5.7$
- **(iv)** $H_0 : \mu \neq 5.7$ against $H_a : \mu = 5.7$
- **(v)** $H_0 : \mu < 5.7$ against $H_a : \mu > 5.7$

(b) Compute the test statistic.

(c) Determine the rejection region.

(d) Select the correct conclusion. *You do NOT need to justify your answer.*

- **(i)** H_0 is not rejected. There is sufficient evidence to indicate that the average pH for rainfalls in the GTA is more acidic than pure rainwater.
- **(ii)** H_0 is rejected. There is sufficient evidence to indicate that the average pH for rainfalls in the GTA is more acidic than pure rainwater.
- **(iii)** H_0 is not rejected. There is insufficient evidence to indicate that the average pH for rainfalls in the GTA is more acidic than pure rainwater.
- **(iv)** H_0 is rejected. There is insufficient evidence to indicate that the average pH for rainfalls in the GTA is more acidic than pure rainwater.

8. (*12 points*) Suppose that a word-association experiment is conducted using eight people and making a comparison of reaction times within each person; that is, each person is subjected to both stimuli in a random order. The reaction times (in seconds) for the experiment are as follows.

Person	Stimulus 1	Stimulus 2
1	2	5
2	1	2
3	1	4
4	2	1
5	1	3
6	2	3
7	3	4
8	2	3

Do the data present sufficient evidence to indicate a difference in mean reaction times for the two stimuli? Test with $\alpha = 0.05$. (Use $\mu_1 - \mu_2 = \mu_d$. Round your answers to three decimal places.)

(a) Choose the correct null and alternative hypotheses. *You do NOT need to justify your answer.*

- **(i)** $H_0 : \mu_d < 0$ against $H_a : \mu_d > 0$
- **(ii)** $H_0 : \mu_d = 0$ against $H_a : \mu_d \neq 0$
- **(iii)** $H_0 : \mu_d \neq 0$ against $H_a : \mu_d = 0$
- **(iv)** $H_0 : \mu_d = 0$ against $H_a : \mu_d < 0$
- **(v)** $H_0 : \mu_d = 0$ against $H_a : \mu_d > 0$

(b) Compute the test statistic.

(c) Determine the rejection region.

(d) Select the correct conclusion. *You do NOT need to justify your answer.*

- **(i)** H_0 is not rejected. There is sufficient evidence to indicate a significant difference between the stimuli.
- **(ii)** H_0 is rejected. There is insufficient evidence to indicate a significant difference between the stimuli.
- **(iii)** H_0 is rejected. There is sufficient evidence to indicate a significant difference between the stimuli.
- **(iv)** H_0 is not rejected. There is insufficient evidence to indicate a significant difference between the stimuli.

9. (*10 points*) Suppose that a certain species of tomato plant has been extensively studied. It has been found that the distribution of the diameters of these tomatoes is roughly bell-shaped with light-tails. In order to test whether a new growth hormone produces larger tomatoes, a botanist decides to conduct a hypothesis test. If μ_1 denotes the true mean of tomatoes with the growth hormone and μ_2 denotes untreated tomatoes, then her hypotheses are $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$. She randomly samples 6 tomatoes, applies the new growth hormone to 3 of them, and measures the following diameters (in cm).

Experimental Group	12.7	13.4	13.9
Control Group	9.7	10.4	13.1

(a) Explain why the two-sample t-test is the appropriate hypothesis test for the botanist to perform.

(b) Based on the two sample t-test as taught in Stat 160 this semester, is there sufficient evidence to reject H_0 in favour of H_A at the $\alpha = 0.05$ significance level?

10. (*16 points*) Four chemical plants, producing the same product and owned by the same company, discharge effluents into streams in the vicinity of their locations. To check on the extent of the pollution created by the effluents and to determine whether this varies from plant to plant, the company collected random samples of liquid waste, five specimens for each of the four plants. The data are shown in the table.

Plant	Polluting Effluents (100 g/L of waste)				
A	1.65	1.73	1.51	1.38	1.60
B	1.70	1.85	1.46	2.05	1.80
C	1.40	1.75	1.39	1.66	1.55
D	2.11	1.95	1.66	1.88	2.00

Do the data provide sufficient evidence to indicate a difference in the mean amounts of effluents discharged by the four plants? Test with $\alpha = 0.05$. (Round your answers to two decimal places.)

(a) Choose the correct null and alternative hypotheses.

(i) H_0: One or more pairs of population means differ, against
$H_a : \mu_A = \mu_B = \mu_C = \mu_D$

(ii) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$, against
$H_a : \mu_A < \mu_B < \mu_C < \mu_D$

(iii) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$, against
H_a: One or more pairs of population means differ

(iv) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$, against
H_a: None of the population means are the same

(v) $H_0 : \mu_A < \mu_B < \mu_C < \mu_D$, against
$H_a : \mu_A = \mu_B = \mu_C = \mu_D$

(b) Compute the test statistic.

(c) Determine the rejection region.

(d) Select the correct conclusion.

(i) H_0 is not rejected. There is sufficient evidence to indicate a difference in treatment means.

(ii) H_0 is rejected. There is sufficient evidence to indicate a difference in treatment means.

(iii) H_0 is not rejected. There is insufficient evidence to indicate a difference in treatment means.

(iv) H_0 is rejected. There is insufficient evidence to indicate a difference in treatment means.

11. (*16 points*) Suppose that a biologist wishes to see if there is a relationship between the heights of trees in a certain forest and their diameters. The biologist selects 10 trees at random from the forest and measures their heights and diameters. To ensure consistent measurements, the biologist records the diameters around the tree from a height of 1.4 metres above the ground.

Diameter (m)	26.0	24.1	11.5	12.8	19.3	16.4	18.0	14.9	11.2	13.9
Height (m)	79.6	97.8	66.8	85.6	48.5	25.3	58.2	43.0	70.7	32.9

Note: In case you are curious, there are some pretty simple ways to measure heights of trees. `https://bigtrees.forestry.ubc.ca/measuring-trees/height-measurements/`

(a) Determine the equation of the linear trend line (i.e., least squares regression line or best-fit line) for height (response variable) in terms of diameter (explanatory variable).

(b) Determine the sample correlation between height and diameter for trees in this study.

Do the data provide evidence to indicate that height and diameter are linearly related? To answer this question, consider testing $H_0 : \rho = 0$ against $H_a : \rho \neq 0$ at the $\alpha = 0.05$ level.

(c) Compute the test statistic.

(d) Select the correct conclusion. *You do NOT need to justify your answer.*

- **(i)** H_0 is not rejected. There is sufficient evidence to indicate that height and diameter are linearly related.
- **(ii)** H_0 is rejected. There is insufficient evidence to indicate that height and diamter are linearly related.
- **(iii)** H_0 is not rejected. There is insufficient evidence to indicate that height and diameter are linearly related.
- **(iv)** H_0 is rejected. There is sufficient evidence to indicate that height and diameter are linearly related.

12. (*10 points*) In a breeding experiment, white chickens with small combs were mated and produced 190 offspring of the type shown below.

TYPE	NUMBER OF OFFSPRING
White feathers, small comb	111
White feathers, large comb	37
Dark feathers, small comb	34
Dark feathers, large comb	8

Use an appropriate chi-square goodness-of-fit test at significance level $\alpha = 0.10$ to determine if these data are consistent with the expected ratios of 9:3:3:1 for the four types. Be sure to carefully state your null and alternative hypotheses, your test statistic, your rejection region, and your conclusion. For your conclusion, state whether or not you reject H_0, and state whether or not this data provides statistically significant evidence for the expected ratios.

13. (*10 points*) One explanation for the widespread incidence of the hereditary condition known as sickle-cell trait is that it confers some protection against malarial infection. In one investigation, 543 African children were checked for the trait and for malaria. The results are shown in the following table.

	Heavily infected (with malaria)	Noninfected or lightly infected (with malaria)	
Yes (has sickle-cell trait)	36	100	136
No (does not have sickle-cell trait)	152	255	407
	188	355	543

Do the data provide evidence in favour of the explanation? Answer this question using an appropriate chi-square test as taught in Stat 160 at significance level $\alpha = 0.05$. Be sure to carefully state your null and alternative hypotheses, your test statistic, your rejection region, and your conclusion.

14. (*18 points*) Answer each of the following with a short paragraph. Your answer should draw on as many appropriate ideas as possible that were discussed in Stat 160 this semester.

(a) A number of volunteers were assigned to one of two groups, one of which received daily doses of vitamin C, and one of which received a placebo (a “sugar” pill without any active ingredients). It was found that the rate of colds was lower in the vitamin C group than in the placebo group. It became evident, however, that many of the subjects in the vitamin C group correctly guessed that they were receiving vitamin C, rather than the placebo, because of the taste. Can it still be said that the difference in treatments is what caused the difference in cold rates?

(b) Ten marijuana users, aged 14 to 16, were drawn from patients enrolled in a drug abuse program and compared to nine drug-free volunteers of the same age group. Neuropsychological tests for short-term memory were given, and the marijuana group average was found to be significantly lower than the drug-free group average. The marijuana group was held drug-free for the next six weeks, at which time a similar test was given with essentially the same result. The researchers concluded that marijuana use caused adolescents to have short-term memory deficits that continue for at least six weeks after the last use of marijuana. Can these results be generalized to other 14– to 16–year olds?

(c) Michael is a pizza inspector for the *Saskatchewan Health Authority*. He has received numerous complaints about a certain pizzeria for failing to comply with its advertisements. The pizzeria claims, on the average, that 85% of its customers receive their orders within 15 minutes. To test their claim, Michael decides to construct a confidence interval. He interviews 100 customers and discovers that only 80% of them receive their order within 15 minutes. Is this sufficient evidence for Michael to refute the pizzeria’s claim at the 95% confidence level? Explain.

The End.