UNIVERSITY OF
WATERLOO

Faculty of Mathematics
Waterloo, Ontario N2L 3G1

Centre for Education in
Mathematics and Computing

Grade 7/8 Math Circles

March 3, 2021

Significance Testing - Problem Set

1. State the null hypothesis and alternative hypothesis in each case. *(hint: make use of* $=, \neq, <, \leq, >,$ *and* $\geq$*)*

   (a) Paul likes growing tomato plants. On average, his plants each produce 10 pounds of tomatoes over the season. He read online that a higher-quality soil could improve his yield, so he sets an experiment where half of his plants are in the new soil, and half of his plants are in the old soil.

   (b) A company guarantees that, with their product, 99.9% of germs are removed from a surface with proper usage. A researcher conducts an experiment testing this claim, suspicious that the company is making their product seem better than it is.

   (c) Studies suggest that approximately 10% of people are left-handed. Max decides to do a survey at their community centre about handedness, curious if the proportion of left-handed people there is also 10%.

   (d) Avery claims that they make three-quarters of all the penalty shots they attempt. You think that they might be exaggerting, so you set up an experiment to test this.

   (e) Sam purchases a pack of sports tape that claims "at least three hours of usage." Skeptical, she decides to test this claim over the next few weeks.

2. You're interested in comparing your classmates to the general population of Canada, so you ask 20 students at your school to be part of a survey. Before collecting their answers, you hypothesize the differences you suspect you'll see when compared to the average Canadian. For each question, use the linked repl program to generate data to help you estimate a p-value for the stated hypotheses. *(Hint: take a look at Example 2 and Question 2 from the lesson!)*

Interpret what the p-value means in relation to your sample statistic. What conclusion would you draw about result's statistical significance at significance level $\alpha = 0.05$?

(As always, if the link isn't working for you, copy-paste into your browser!)

(a) `https://repl.it/@cemc/a-p-values#main.r`

H_0: the mean height of students at your school $= 167cm$.
H_a: the average height of students at your school $< 167cm$.

From your survey of 20 students, you have a sample mean of $149cm$.

(b) `https://repl.it/@cemc/b-p-values#main.r`

H_0: the proportion of students at your school who wear glasses $= 0.55$.
H_a: the proportion of students at your school who wear glasses < 0.55.

From your survey of 20 students, you have a sample proportion of 0.45.

(c) `https://repl.it/@cemc/c-p-values#main.r`

H_0: the mean family size of students at your school $= 2.6$.
H_a: the mean family size of students at your school > 2.6.

From your survey of 20 students, you have a sample mean of 3.4.

(d) `https://repl.it/@cemc/d-p-values#main.r`

H_0: the proportion of students at your school who spoke English as their first language $= 0.975$.
H_a: the proportion of students at your school who spoke English as their first language < 0.975.

From your survey of 20 students, you have a sample proportion of 0.85.

3. A clinical trial is testing two potential treatments for effectiveness at decreasing the time that symptoms last in a person experiencing illness. Currently, without treatment, symptoms in patients tend to last an average of 5 days. A trial with 100 participants will be conducted with each treatment to test for statistically significant decreases in the time spent experiencing symptoms.

We'll call the two treatments "Treatment A" and "Treatment B".

(a) State null and alternative hypotheses for Treatment A.

(b) State null and alternative hypotheses for Treatment B.

The trials are conducted, and sample results are recorded. For clinical trials, results must be statistically significant at $\alpha = 0.01$ to be accepted. To estimate p-values, use the linked repl program to simulate 1000 trials assuming the null hypothesis: `https://repl.it/@cemc/trial-p-values#main.r`.

(c) What does it mean for a result to be statistically significant at $\alpha = 0.01$?

(d) Treatment A had a sample mean of 4.6 days. Estimate and interpret a p-value for Treatment A.

(e) Treatment B had a sample mean of 4.4 days. Estimate and interpret a p-value for Treatment B.

(f) State your conclusion about each Treatment. Should either treatment be accepted for usage?

4. As noted in the lesson, drawing conclusions from p-values comes with an inherent risk of error. In particular, when carrying out significance tests, there are two types of potential errors:

- **Type I error**—rejecting H_0 when H_0 is true
- **Type II error**—failing to reject H_0 when H_0 is false (and H_a is true)

Example: We can identify what a Type I and Type II error would look like for the scenario in Question 1 (a) above.

A Type I error would be rejecting H_0 when it is true. In this context, that would be concluding that the higher-quality soil improved Paul's tomato yield, when in reality the population mean did not increase. This would happen if the sample data that Paul collected had, by chance, a significantly higher sample mean that caused him to conclude that it was unlikely for H_0 to be true. A potential consequence would be Paul spending more money than he has to each season to grow his tomatoes.

A Type II error would be failing to reject H_0 when H_0 is false (and H_a is true). In this context, that would mean concluding that the higher-quality soil did not improve Paul's tomato yield, when in reality the population mean did increase. This would happen if the sample data Paul collected had, by chance, a sample mean that was too low to conclude that H_0 was unlikely to be true. A consequence would be Paul deciding not to change over to the better soil and missing out on lots of potential tomatoes!

As you can see, neither of these options are ideal, but these things do happen! When generating data with the repl programs, you can see that some statistically significant sample statistic values *are* still possible when H_0 is true. Depending on the situation though, one of the two types of errors might be more preferable, and so we can choose a significance level at the beginning of the process accordingly.

Identify what a Type I vs. a Type II error would look like in each scenario, and a potential consequence of each.

(a) A school wants to assess if a newly-installed water fountain decreases how frequently students ask to leave class for a drink of water. If it does, they'll approve the pending plans to install another one.

- H_0: average # of times students ask to leave class for water $= 4$
- H_a: average # of times students ask to leave class for water < 4

(b) A farmer has recently changed their watering system and wants to track if there are significant increases to their crop yield. If there is no significant increase, then they will switch back to their old system.

- H_0: average crop yield $= 1000kg$
- H_a: average crop yield $> 1000kg$

(c) It is acceptable for up to 2% of the paper from an order by a school to have imperfections. Upon receiving the paper, the school will inspect a random sample of packs. If they conclude that the proportion of paper with imperfections is greater than 2%, then they can send the order back for a partial refund.

- H_0: proportion of paper with imperfections $= 0.02$
- H_a: proportion of paper with imperfections > 0.02