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 Solutions

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.

   > *Solution:* H_0: mean of amount of tomatoes produced by each plant $= 10lbs$, H_a: mean of amount of tomatoes produced by each plant $> 10lbs$.

   (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.

   > *Solution:* H_0: proportion of germs removed from a surface with proper usage $= 0.999$, H_a: proportion of germs removed from a surface with proper usage < 0.999.

   (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%.

   > *Solution:* H_0: proportion of people at Max's community centre that are left-handed $= 0.1$, H_a: proportion of people at Max's community centre that are left-handed $\neq 0.1$.

(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.

> *Solution:* H_0: proportion of successful attempted penalty shots $= 0.75$, H_a: proportion of successful attempted penalty shots < 0.75.

(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.

> *Solution:* H_0: average hours of usage ≥ 3, H_a: average hours of usage < 3.

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!)

   > *Solution:* Individual answers will vary, as the program utilises random generation. Below are examples of answers; solutions should, in general, be similar.

(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 $149cm$.

Solution: After generating our 1000 simulated sample means, we want to look for how many of those results are at least as extreme as our actual sample mean. In this case, when we assume that H_0, we see 0 instances where the sample mean is 149cm or less. Thus, we estimate the p-value to be $\frac{0}{1000} = 0$.
This means that, when H_0 is true, it is impossible or almost impossible to conduct a survey of 20 individuals and observe data as extreme as yours. Thus, considering our sample mean, it is impossible or almost impossible for H_0 to be true.

```
> r main.r
[1] "below is a table of 1000 simulated sample means:"

157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176
  2   2   4   5  19  34  66  99 125 136 151 123 101  63  41  18   6   4   0   0
177 178
  0   1
>
```

(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.

Solution: When we assume H_0, we see $1 + 10 + 11 + 38 + 75 + 100 = 235$ instances where the sample proportion is 0.45 or less. Thus, we estimate the p-value to be $\frac{235}{1000} = 0.235 = 23.5\%$.
This means that, when H_0 is true, it is very reasonable to conduct a survey of 20 individuals and observe data as extreme as yours. In fact, data at least as extreme as 0.45 will be observed around 23.5% of the time when H_0 is true, so this does not give us convincing evidence against H_0.

```
> r main.r
[1] "below is a table of 1000 simulated sample proportions:"

 0.2 0.25  0.3 0.35  0.4 0.45  0.5 0.55  0.6 0.65  0.7 0.75  0.8 0.85
   1   10   11   38   75  100  174  182  185  102   66   39   11    6
>
```

(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.

Solution: When we assume H_0, we see 0 instances where the sample mean is 3.4 or greater. Thus, we estimate the p-value to be $\frac{0}{1000} = 0$.
This means that, when H_0 is true, it would be impossible or almost impossible to have a sample mean as big as yours, leading us to believe that it is impossible or almost impossible for H_0 to be true.

```
> r main.r
[1] "below is a table of 1000 simulated sample means:"

2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9   3
  5  16 103 222 316 227  91  19   1
>
```

(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.

Solution: When we assume H_0, we see $1 + 20 = 21$ instances where the sample proportion is 0.85 or less. Thus, we estimate the p-value to be $\frac{21}{1000} = 0.021 = 2.1\%$.
This means that, when H_0 is true, it is pretty rare (but not impossible) to get a sample proportion of 0.85 or less. Thus, we have some reason to believe that H_0 may not be true.

```
> r main.r
[1] "below is a table of 1000 simulated sample proportions:"

 0.8 0.85  0.9 0.95    1
   1   20   73  293  613
>
```

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.

   > *Solution:* H_0: time that symptoms last = 5 days, H_a: time that symptoms last < 5.

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

   > *Solution:* H_0: time that symptoms last = 5 days, H_a: time that symptoms last < 5.

   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 $\alpha = 0.01$?

   > *Solution:* For a result to be statistically significant at $\alpha = 0.01$, it must have a p-value of less than 0.01; that is, it must be a result that has less than a 1% chance of being observed when H_0 is true.

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

```
> r main.r
[1] "below is a table of 1000 simulated sample means:"

4.3 4.4 4.5 4.6 4.7 4.8 4.9   5 5.1 5.2 5.3 5.4 5.5 5.6
  1   3   6  26  53 114 173 218 178 132  66  20   7   3
>
```

Solution: Assuming that H_0 is true, $1 + 3 + 6 + 26 = 36$ times out of 1000, we see sample means of 4.6 or less. This gives us a p-value of $\frac{36}{1000} = 0.036$, meaning that when H_0 is true, there is only a 3.6% chance of observing a sample mean at least as extreme as 4.6.

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

Solution: Assuming that H_0 is true, $1 + 3 = 4$ times out of 1000, we see sample means of 4.4 or less. This gives us a p-value of $\frac{4}{1000} = 0.004$, meaning that when H_0 is true, there is only a 0.4% chance of observing a sample mean at least as extreme as 4.6.

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

Solution: At $\alpha = 0.01$, the results for Treatment A are *not* statistically significant, because $0.036 > 0.01$. However, at $\alpha = 0.01$, the results for Treatment B *are* statistically significant, since $0.004 < 0.01$. Therefore, based on this significance test, Treatment B should be accepted for usage (but Treatment A should not)!

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

Solution: A Type I error would be concluding that the new water fountain decreased the frequency of students asking to leave class for water, when the frequency actually stayed the same. This would lead the school to approve plans to install another one under the incorrect premise that it could further decrease the frequency of students asking to leave class for water, which could mean extra unecessary spending without the intended benefit.
A Type II error would be concluding that the new water fountain did not decrease the frequency of students asking to leave class for water, when it actually did. This would lead to them cancelling the plans to install a new one, which could mean missing out on further intended benefits.

(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 $> 1000kg$

Solution: A Type I error would be concluding that the new watering system *does* increase crop yield when it actually does *not*. This would lead to them keeping the new system when there's no good reason to! Depending on the situation, there could be downsides—the new system might be more expensive to maintain, it can take money and time to uninstall/re-install systems, etc.
A Type II error would be concluding that the new watering system does *not* increase crop yield when it actually *does*. This would lead to them switching back to their old system and missing out on the increased yield they would've otherwise had.

(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

Solution: A Type I error would mean sending back the order when in actuality, the proportion of paper with imperfections over the entire order is less than or equal to 2%. As a consequence, the school loses the money that the manufacturer does not refund and also has to place, wait, and pay for a new order.

A Type II error would mean keeping the paper when in actuality, more than 2% of the paper in the order has imperfections. As a consequence, the school would have recieved a lower-quality product than they paid for, and teachers at the school may run into more inconveniences when using the paper than originally accounted for.