PROBLEM-SOLVING STRATEGY FOR STATISTICS

Studying Statistics involves solving complex problems. Adapting George Pólya's problem-solving method (Pólya, 1957), this handout provides an organizational strategy that can help with multi-step problem-solving questions. By focusing on the "why" and the "how" behind difficult problems, you can build your problem-solving strengths and deepen your awareness and knowledge of Statistics. The examples included below are from a second-year Statistics course.

The Problem-Solving Strategy consists of three steps: **Step 1:** Plan & Prepare, **Step 2:** Organize & Solve, and **Step 3:** Cross Reference.

Step 1: Plan & Prepare

Questions for developing **metacognition,** defined as awareness or analysis of one's own learning or thinking processes (Merriam-Webster, 2023):

Do I have the knowledge needed to complete the question?

- **Yes:** How am I going to solve this problem? What first step is needed?
- **No:** What will I do to learn the content? Which section of course content do I need to return to for review? What information will help me understand the question?

Important Considerations:

Language in Statistics

In Statistics, hedging language is used to describe a well-reasoned answer. The use of this language conveys uncertainty around the answer, creating space for a margin of error. **STRATEGY:** Create a list of hedging terms to consult while studying can provide support for including this type of language required for statistical writing (for example, "possibility", "maybe", "perhaps", "probable", "certain", "can", "might")

Best Solution vs. Correct Answer

There are often multiple ways to get the correct answer. Marks are not always given just for reaching the correct answer; instead, they are often given based on the methods and processes you use to arrive at that answer. **STRATEGY**: note the methods or processes used in lecture content and what is recommended for these questions. What has your professor mentioned or emphasized in class?

What does the **best solution** mean to you? What are the factors that might make this the best solution?

What are the differences between another possible solution and the solution given by the professor?

Step 2: Organize & Solve

1. Confront the problem and spend time processing the details of the question to understand the problem.
2. Identify or define the problem and explore all the options. There may be many ways of answering a question. However, the best method may not be the most direct. In other words, the best method may be one that involves a deeper analysis and understanding of the problem.
3. Solve the problem. Execute the plan. Remember to show ALL workings.
    - a. **What** was done in each step?
    - b. **How** was it done and what formulas or solutions were followed?
    - c. **Why** was it done?
4. Review; Look back and reflect
    - a. **What** traps do I need to watch out for?
5. Extend. Are there other questions where this method can be applied

Step 3: Check & Cross Reference

Cross-reference your answer to the answer solution.

Carefully assess the professor's answer solution, do the steps match?

If there are differences...

- At what part of the question does your answer not align with the professors?
- **STRATEGY:** Access what kind of error(s) you may have made.
    - **Mathematical Error**- a numbers-based error
    - **Application Error**- applied the incorrect concept
    - **Knowledge Error**- did not learn the concepts
    - **Time Management Error**- caused by rushing and missing information

Return to the question, think about why your method may not have been the one included in the answer solution. Does the question outline or provide clues for a particular method?

Return to your course notes, review information for alternative methods or best practices. Are there similar problems available for review?

By following the steps outlined above, you can enhance your understanding and approach to complex statistical questions.

Let's look at three practical examples of Step 2: Organize & Solve. For each example, we have included responses from a student and an instructor to provide additional perspectives to deepen your understanding of the problem-solving strategy.

Question 1 a

Question 1 (9 points) in a study, out of a sample of 350 Toronto residents, 70 favoured increasing the speed limit on all parts of highway 401 from $100^{km}/_h$ to $110^{km}/_h$. In another sample of 180 people living in suburbs (outside of Toronto), 75 favoured the increase. Can it be concluded that the sentiment for increasing the speed limit is different for the two groups at the population level?

(a) (1 point) Formulate the appropriate null and alternative hypothesis.

Since we are dealing with a variable measured as a yes or no answer (i.e. a Bernoulli), we have Binomial data and therefore our hypotheses are referring to the proportion of individuals in each group in favour of the change:

$$H_0 : p_1 = p_2, \quad \text{versus} \quad H_A : p_1 \neq p_2$$

or equivalently

$$H_0 : p_1 - p_2 = 0, \quad \text{versus} \quad H_A : p_1 - p_2 \neq 0,$$

where p_1 refers to the proportion of all Toronto residents in favour, and p_2 is the proportion of all residents of suburbs in favour of the change.

Example of a Student's Response

Q1 a) Let P_1 denote the proportion of Toronto residents preferring speed limit increases and P_2 denote the proportion of non-Toronto residents preferring speed limit increase

$H_0 = P_1 = P_2$

$H_1 = P_1 \neq P_2$

Instructor's Comments

Deduct 0.5 if only one of below is present, otherwise if more than one present, deduct the full mark:

- wrote hypothesis as means instead of proportions
- wrote the alternative as anything other than ≠
- only included one group in the hypothesis
- wrote the hypothesis using notation for estimators instead of parameters

Decision Steps	
What was done in each step?	• First, we had to determine what parameter(s) we plan to test (e.g., proportions, means) • Then a null hypothesis was written out based on the context of the problem, followed by an alternative hypothesis for the same parameters.
How was it done and what formulas or solutions were followed?	• First, we had to recognize that we are dealing with proportions as our parameter. • We are told that, out of a certain number of people (sample), some number of them (a count) answered yes when asked a question (only possible answers were yes or no). By taking the ratio of yeses to the total surveyed, this is a proportion. • Then we had to realize we had two groups of people who were surveyed, and we have how many of each group responded yes. So, we have two proportions. • After we must notice from the question that we want to compare the two proportions from each group, so we are working with a two-sample hypothesis. • Next, we need to ask ourselves what direction are we testing? Do we want to know if Toronto people are more in favour than suburban residents, less in favour, or simply if the proportions are different. This tells us how we will relate the two proportions in the alternative hypothesis. • Finally, we write our hypotheses involving the unknown population values. • The null hypothesis is the claim of no difference, or that the proportions are equal, and the alternative is the claim we hope our data can support, that the proportions are different.
Why was it done?	Anytime we want to run a hypothesis test, we need to explicitly write out what the "status quo" (null) is and what we hope our data can let us conclude (alternative) because this lets us know what conclusions we can make with our test and how to compute the correct measure that helps us make this conclusion.

Decision Steps (continued)	

What traps do I need to watch out for?	• Writing hypotheses using notation that refers to the sample rather than the population
	  • An easy mistake but it makes no sense to write statistical hypotheses about the sample because we can see everything about our sample and thus have no need to test anything
	• Writing hypotheses where the testable claim is written as the null hypothesis
	  • The null hypothesis is always the status quo. Statistical tests are set up to investigate deviations from this, not to confirm if it is true
	• Not recognizing exactly what format our claim takes (greater than, less than, or not equal)
	  • Another easy mistake. Consider reading the question to look for key words that suggest a direction (at least, at most, larger, smaller, different, the same, etc.)

Are there other questions where this method can be applied?	• This general process of deciding how to write the hypotheses can be used to help formulate hypotheses for any number of other tests, including for means or even variances. • The questions used to highlight the important information needed to formulate the right conclusion can be used in many calculation problems where it's necessary to know what information pertains to what element of the calculation.

Question 1 b

(b) (4 points) By selecting an appropriate method, conduct a test of your hypothesis in (a). Show all your steps including the test statistic, distribution used, and parameters involved. Also write down any assumption that you have made.

The appropriate test to conduct is the two-sample test for proportions. This requires me to assume that the samples of individuals in each group are independent (i.e. no relationship between members of the sample in Toronto or in suburbs.). The sample proportions are:

$$\hat{p}_1 = \frac{70}{350} = 0.2, \quad \hat{p}_2 = \frac{75}{180} = 0.42.$$

Since both population proportions in favour are unknown to us, it is appropriate to use the pooled estimate of the standard error in the calculation of our test statistic:

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{70+75}{350+180} = 0.27$$

$$\begin{aligned} Z &= \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}} \\ &= \frac{(0.2 - 0.42) - 0}{\sqrt{0.27(0.73)(1/350 - 1/180)}} \\ &= -5.32 \sim N(0,1) \end{aligned}$$

where we have used the fact that the samples are both large and so the Central Limit Theorem applies and we can use the standard Normal distribution for our test statistic. Since we are looking to conduct a two-tailed test, our p-value will be

$$P(|Z| \geq |-5.32|) \leq 0.0002$$

Example of a Student's Response

b) Use two-sample t test.

test statistic $\hat{P}_1 = \frac{70}{350} = \frac{1}{5}$

$\hat{P}_2 = \frac{75}{180} = \frac{5}{12}$

$$T = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{350} + \frac{P_2(1-\hat{P}_2)}{180}}} = -5.0963 \approx -5.10$$

Under null hypothesis, the test statistic is assumed to follow a normal distribution

The p-value is $P(z \leq -5.10) \times 2 \approx 0.00$

Instructor's Comments

Each of the following earns 1 mark. If a student has not included one of the below or has included it incorrectly, deduct the corresponding mark. If a student makes a calculation error, then you should assess their remaining work assuming that this incorrect value is right and not deducting further marks unless an additional error occurs:

- assumptions of independence
- calculations of the test statistic, with correct values (work must be shown to earn this)
- noting that CLT is appropriate here and noting the distribution explicitly
- final p-value for the correct tail region is provided (note, tail region should be assessed in relation to their answer in (a), so if they noted a one-sided hypothesis they should perform that here)

Decision Steps	
What was done in each step?	• Assumptions were reported first • Estimates of the two proportions were computed • A decision regarding how the variance of the test statistic was to be computed was justified • This pooled proportion was computed to be used in calculating the variance of the test statistic • The test statistic was computed using above details • A justification was provided for the use of the standard Normal as the null distribution • A final p-value was computed using a standard Normal table of critical values.
Q. How was it done and what formulas or solutions were followed?	• Assumptions needed are determined by the type of hypothesis test being conducted (in this case a two-sample test of proportions). This would be noted directly in the lectures. • Estimates of the two proportions are computed by dividing the number of people in favour by the total number of respondents in each group (x/n). • The decision regarding how to calculate the variance in the test statistic is based on whether the samples are large. If not, an alternative variance would be computed, but in this case, we have in each group that the proportion of yes's and no's are greater than 10% and so we conclude that we should pool the groups. • We compute this pooled value by $\frac{n_1\widehat{p_1}+n_2\widehat{p_2}}{n_1+n_1}$ and plugging in the values for each term. • Using the estimated proportions of each group, and the pooled proportion we computed previously, we compute the test statistic which takes the difference in sample proportions and divides by the standard error. The standard error choice depends on whether we had large enough samples to start with. • We again use the fact that the samples are large meaning that it's reasonable to assume that the null distribution follows a standard Normal, rather than a T distribution. • Lastly, the p-value is computed using the null distribution from above and the value of the test statistic computed above. This also requires using the hypotheses from (a) as this tells us what tail probability we are looking for our p-value to represent. In our case, our hypothesis tells us we want two-sided tails so we must find the probability of being above the test statistic value and the probability of being smaller than the negative version of the test statistic, or alternatively take the probability of being larger than the test statistic and double it.

Decision Steps (continued)	
Why was it done?	• It's necessary to note what assumptions you have to make for any statistical procedure because if they do not hold, then the conclusions drawn from this procedure are not reliable and may not be true. By being explicit about it, you are noting your awareness of the limitations of the statistical procedure. • To test a hypothesis regarding any population parameter, we need to compute an estimate of this parameter based on our sample to have a guess for this unknown value. It's also needed for computing the test statistic as we must compare the distance between this value and the hypothesized value of the parameter to how much variation we expect to naturally occur from sample to sample. • Since we have multiple options for how to measure the variability in the sampling distribution, we need to justify the decision for which option we choose so that it is clear to a reader why the test was conducted the way it was. • We need this pooled proportion to find an estimate of the variability in the sampling distribution. As noted above, this variability is used to understand whether the sample values are reasonable under the condition that the hypothesized value is true or whether they are so extreme that it leads us to believe that the hypothesized value may not be true. • To formally compare our sample value to variation we expect to see in our sample values if the null were true, we compute the test statistic. This gives us a unit-less measure which we can better use for making a conclusion about the hypothesized value of the parameter. • Since we have different options for the null distribution, we need to formally state why we pick one over the other as this choice impacts how we make a conclusion about the parameter. • We finally compute a p-value because it is a measure that tells us whether it is reasonable that our data happened to arise from a population in which the true difference in proportion in favour was 0.
What traps do I need to watch out for?	• Nearly all hypothesis tests at this level require an assumption of independence, but it would be important to not mistake this for any sort of paired-sample situation. • This step is usually pretty well done by most, but being sure that you are transcribing the right values from the question is important. • It could be quite easy to not check whether the sample sizes are large enough and either jump to using the pooled values when it's not appropriate or possibly using the alternative option when it would be more accurate to use the pooled option. • The trickiest part of this step is making sure we use the right numbers from the question in the right spot. So all the information for group 1 should be consistently referring to whatever you have labelled as group 1. If group 1 is Toronto residents, then all numbers for group 1 should be for Toronto residents.

Decision Steps (continued)

What traps do I need to watch out for? (continued)	• This step will depend heavily on the previous ones and making the right choices. But since there are a lot of values to work with, computation mistakes are common. Breaking it down into smaller pieces and writing out middle steps helps. • Not recognizing that again the large samples help us by letting us apply the Central limit theorem. • Most commonly, the wrong p-value is computed (usually a one-sided option), or the wrong value is located on the table.
Are there other questions where this method can be applied?	The steps involved in this problem can be transferred to many other questions asking to test a hypothesis, as we always want to list assumptions, check properties of our sample to know when we can utilize asymptotic results that make calculations easier and making sure that the p-value obtained aligns with the hypothesis at the start.

Question 1 c

(c) (2 points) Interpret your p-value in the context of this problem and report your findings.

A p-value measures the strength of the evidence against the null hypothesis, and as such is interpreted as the probability of observing evidence at least as strong if not stronger than our current data. In our case, the p-value is very small, indicating that we have strong evidence against the null hypothesis that no difference exists between the proportion of Toronto residents and suburban residents in favour of increasing the speed limit.

Example of a Student's Response	Instructor's Comments
c) Since p-value is smaller than 0.05, so we reject Ho, which means that the sentiment for increasing the speed limit is different for two groups.	1 point for a correct reporting of the findings (i.e. small p-value implies reject the null) and 1 point for appropriate contextualisation.

Decision Steps

What was done in each step?	• A generic statement is provided regarding the meaning of p-values generally. • A statement that refers to what we have calculated previously as a p-value. • An interpretation in context of what that allows us to conclude about the proportion of residents in favour of the change.

Decision Steps (continued)	
How was it done and what formulas or solutions were followed?	• The definition of a p-value was something that would have been covered in class and so we just have to take this generic meaning and re-write it specifically using the information in the question, such as explicitly saying what is the claim being tested. • The conclusion in our findings is based on understanding that small values are strong evidence against the claim of the status quo, which would also have been emphasized from class.
Why was it done?	• Adding a generic definition of the p-value shows that you know the formal definition, but it also helps to orient one's thoughts regarding how this relates to our specific p-value computed earlier and what this lets us conclude. • The question asks us to frame the conclusion in context so that it is easiest to understand what this number tells us about the problem at hand. It's a skill that is needed in most scientific fields.
What traps do I need to watch out for?	• Missing the notion that the p-value is a measure of strength of our evidence against the null hypothesis, not in favour of it. • Not interpreting in context, especially when the questions explicitly asks for it. • Thinking that a high probability means we support the null (we only fail to reject, not support) • Not matching your conclusion correctly to the hypotheses, e.g. saying we reject the null but that means we conclude there is no difference in proportions.
Are there other questions where this method can be applied?	• Similarly to other parts of the questions, this applies to any time a hypothesis test is run and a p-value is computed as a means of concluding how strong the evidence is.

Question 1 d

(d) (2 points) You are told that the 95% confidence interval for the difference in two population proportions (Suburbs - Toronto) is (0.129, 0.304). Does this support the claim that the true difference in proportion is 20%? Explain briefly (no calculation needed).

A confidence interval represents a range of plausible values for the difference in proportions of our two groups. Since 20% is contained in the interval provided, this supports the conclusion as it is one of the plausible values for this true difference.

Example of a Student's Response	Instructor's Comments
d) Yes, since 0.2 is included in the 95% confidence interval and it's close to the mid point of the interval $\frac{0.129+0.304}{2} = 0.2165$ this supports the claim.	1 point for conclusion of supporting claim or not, and 1 point for explanation that involves the meaning of the confidence interval.

Decision Steps	
What was done in each step?	• Like above, a general definition of what a confidence interval means is provided. • Then a note about where the claimed value falls relative to the interval provided is given and a final conclusion of whether there is evidence to reject the claim or not.
How was it done and what formulas or solutions were followed?	• The general definition is one that would be found in the lecture slides and this is simply rewriting it. • The connection between testing and confidence is used here by reflecting on the fact that to reject a claim the confidence interval must not contain the null value.
Why was it done?	• The general definition is again written partly to act as an addition to the justification, but also as a setup or way to orient one's thoughts to make the appropriate conclusion. • We want to take the information provided and make it as clear as possible, so quoting the values provided gives a stronger justification than simply saying that because of the first sentence the claim is supported.
What traps do I need to watch out for?	• Questions that ask to explain or justify are being used to assess your understanding of the topic. So you want to provide as much information that showcases your understanding as possible. However it is tempting to simply answer the question portion and not provide the explanation. • Confidence intervals are also commonly misinterpreted so you would want to be careful that you explain it correctly. • Since it is common to ask if 0 is contained in the confidence interval (since hypothesis tests often test a parameter value of 0), be sure to read carefully what value is being tested in this situation (note it's not the same from the rest of the question).

References

Pólya, G. (1957). *How to solve it: a new aspect of mathematical method (2nd ed.).* Doubleday.

Merriam-Webster. (n.d.). Metacognition. In Merriam-Webster.com dictionary. Retrieved December 11, 2023, from https://www.merriam-webster.com/dictionary/metacognition

Developed by Learning Strategists Julia Andrews and Rahul Bhat from the Centre for Learning Strategy Support and Professors Nathalie Moon and Katherine Daignault from the Department of Statistical Sciences.

For more learning strategies, visit the Centre for Learning Strategy Support learningstrategy.utoronto.ca.