**Formal Hypothesis:**
**Making Decisions with a Single Sample**

This worksheet continues to build on the previous concepts of inferential statistics, only now, we’re going to give it a fancier name: “Formal Hypothesis Testing”. But don’t be alarmed! It’s still the same idea. We use formal hypothesis testing to decide if a sample result is significantly different enough (beyond the regular experimental variation) to disprove a population claim or expected value.

**Step 1:** Set up the null hypothesis. Recall that in inferential statistics we use a sample result to investigate an expected or claimed population parameter. We describe this parameter with the null hypothesis ($H_0$, pronounced “H-nought”). It states what you are going to believe about the population (the usual result) unless the sample gives you strongly contradictory evidence (an unusual result). It’s written using one of the two forms below:

- $H_0: \mu = 5$ (for a mean)
- $H_0: p = 0.82$ (for a proportion)

A null hypothesis must always have $\mu$ or $p$, an equal sign, and the claimed value for the parameter in it!

**Step 2:** Set up the alternative hypothesis, $H_1$, to represent the situation where there is strong evidence that the null hypothesis is not true. Two key things to remember about the alternative hypothesis: 1) the population parameter to the left of the sign and the number value to the right of the sign must exactly match the ones from the null hypothesis, and 2) the alternative hypothesis must use a $>$, $<$, or $\neq$ sign.

- $H_1: \mu > 5$
- $H_1: \mu < 5$
- $H_1: \mu \neq 5$
- $H_1: p > 0.82$
- $H_1: p < 0.82$
- $H_1: p \neq 0.82$

The alternative hypothesis represents the situation we would be concerned about from a business perspective – whether it’s producing a good that weighs too much, a service taking more time than is economical, or if either too little or too much volume/weight/proportion/etc will cost us money or customers. We pick the sign ($<$, $>$, or $\neq$) based on the scenario we’re interested in. The first two forms of the alternative hypothesis ($>$ and $<$ signs) are called one-tailed tests, because we only care about extreme sample results that are to one side of the population parameter (or in one tail of the sampling distribution). The third form of the alternative hypothesis ($\neq$) is called a two-tailed test, because extreme sample results on either side of the population parameter (in either tail of the sampling distribution) would provide evidence against the null hypothesis.
Example 1: A shipping company has been working to reduce the time it takes for a package to arrive. In the past, the average arrival time was 4 days. You take a random sample of 30 shipments and track how long it takes them to arrive. What are the null and alternative hypotheses?

Solution: We are looking at a population mean, $\mu$. The average arrival time in the past is the null hypothesis, $H_0: \mu = 4$ days. This is what we assume to be the case unless evidence indicates otherwise. We are interested in seeing if the average arrival time has been reduced, so the alternative hypothesis is $H_1: \mu < 4$ days.

Step 3: Determine the significance level, $\alpha$. For the example above, we only care about significantly low sample results. But how low is too low? Previously, we compared the probability of an observed sample result to a specified level of 0.05 to decide whether the result was usual or unusual. That was an example of a significance level. To explain what the significance level means, you must first know that there are two types of errors that can occur when performing a hypothesis test. These errors occur because we are using sample data (which usually reflects the population characteristics, but occasionally does not) to make a conclusion about the population.

A Type I error occurs when the null hypothesis is mistakenly rejected, even though it’s true. For example, if we were looking at the filling of soda bottles in a manufacturing plant and sampled 40 bottles, a Type I error would lead us to believe that the desired volume was not being produced and we should adjust the fill line, even though it is in fact producing the desired volume. The significance level represents the maximum allowable probability of a Type I error. In other words, if $\alpha = 0.05$ (or 5%), when the hypothesis test is repeated multiple times, the null hypothesis will be incorrectly rejected about 5% of the time.

A Type II error occurs when we mistakenly fail to reject the null hypothesis, even though it’s false. In the case of a soda manufacturing plant, this error would lead us to decide (based on an observed sample result) not to adjust the fill line even though the desired fill volume is not being produced. The probability of a Type II error is called beta, or $\beta$.

A sample will never provide perfect certainty that the population is what we think it is. Statisticians find a balance between these two types of errors – reducing the chances of one type of error (by setting $\alpha$ lower, for example, to reduce the Type I errors) will always increase the chance of the other type. All we can do is seek to minimize the type of error that is more critical for us.
Example 2: The Learning Centre at VCC wants to ensure that adequate computer access is provided for students. In the past, an average of 15 hours per week of use on LC computers was used for planning, but the tutors think that students are using the computers more often now. What are the null and alternative hypotheses? Explain the Type I and Type II errors and which party (Learning Centre or students) cares more about the types of errors.

Solution: H₀: μ = 15 hours per week; H₁: μ > 15 hours per week. The Type I error would mean we mistakenly believe that computer use has increased, when in reality, it has not. A Type II error would mean concluding that computer use has not increased, when in reality it has. The Learning Centre would be more concerned about Type I error, as extra computers would be purchased even though they are not needed. Students would care more about a Type II error as there would not be extra computers provided, even though they are needed, leading to insufficient access for homework, research, etc.

Step 4: Calculate the probability (p-value) of the sample result for a set of sample data. For a sample proportion, this is the same calculation as before. For a sample mean, we will have to calculate a t-score. This calculation is very similar to calculating a z-score, but it will be covered in a separate worksheet.

Step 5: Compare the probability of the sample statistic (the p-value) to the significance level (α) and make a decision to reject or fail to reject the null hypothesis. This is very similar to deciding about usual or unusual results. If we get a usual result (p-value > α), this means we don’t have enough evidence to disprove the null hypothesis. We fail to reject the null hypothesis. If we get an unusual result (p-value ≤ α), we would say there is strong evidence against the null hypothesis and we reject the null hypothesis. The smaller the p-value, the stronger the evidence it provides against the null hypothesis.

The method of calculating the p-value depends on the alternative hypothesis:

<table>
<thead>
<tr>
<th>When H₁ contains</th>
<th>Calculate probability as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;</td>
<td>1 – P(sample statistic ≤ observed sample result)</td>
</tr>
<tr>
<td>&lt;</td>
<td>P(sample statistic ≤ observed sample result)</td>
</tr>
<tr>
<td>≠</td>
<td>2 x P(tail area beyond the observed sample result)</td>
</tr>
</tbody>
</table>

Example 3: A poll surveyed 560 business owners in BC about the new HST and its effect on businesses. The poll reported that 78% thought the HST impact would be negative. Is there enough evidence to conclude that more than three-quarters of all British Columbian business owners think the HST will harm their business? Use α = 0.07.

Solution: Start by setting up null and alternative hypotheses. Sometimes it’s easier to identify the alternative hypothesis first. In this case we care about whether more than three-quarters (or 75%) of business owners think that the new HST will harm their business. In other words, H₁: p > 0.75. The null hypothesis must match, but use an equal sign. So H₀: p = 0.75. We are told α is 0.07.
The population of business owners would need to be at least $560/0.05 = 11,200$ people, which is reasonable. We must check that the sampling distribution of $\hat{p}$ is approximately normal:

$$np = 560(0.75) = 420 \geq 10$$
$$nq = 560(0.25) = 140 \geq 10$$

The conditions are met to use the sampling distribution of $\hat{p}$ to decide about the population. Now we calculate the probability of $\hat{p} >$ observed sample result, which means we will want to find $1 - p(\hat{p} \leq$ observed sample result).

$$P(\hat{p} \leq 0.78) = P \left( Z \leq \frac{0.78 - 0.75}{\sqrt{\frac{0.75(0.25)}{560}}} \right) = P(z \leq 1.64) = 0.9495$$

$$1 - 0.9495 = 0.0505$$

The probability of getting a survey result where the sample proportion is 0.78 or higher is 5.05%. Since this is a one-tailed test, we don’t have to multiply the probability by 2. Our p-value is less than $\alpha$ ($0.0505 < 0.07$). This is a significant (unusual) result, so we would reject the null hypothesis. The evidence supports the alternative hypothesis. We conclude that more than 75% of business owners believe the HST will have a negative impact on their business.

**Exercises**

1. Identify which of the following are incorrect pairs of null and alternate hypotheses and correct them.
   a. $H_0 = 7; H_1 > 7$
   b. $H_0: p = 0.65; H_1: p \geq 0.65$
   c. $H_0: \mu = 54; H_1: \mu < 54$
   d. $H_0: p < 0.30; H_1: p = 0.30$
   e. $H_0: \mu = 18.5; H_1: \mu \neq 18.5$
   f. $H_0: p = 0.50; H_1: p > 0$

2. Given the following probabilities and significance levels, decide whether to reject or fail to reject the null hypothesis.
   a. p-value: 0.05, $\alpha = 0.10$
   b. probability: 10%, $\alpha = 0.07$
   c. p-value: 0.02, $\alpha = 0.02$
   d. probability: 20%, $\alpha = 0.25$
3. A national poll in September 2010 found that 35% of post-secondary students had more $30,000 worth of loans upon graduation. The results were based on a random sample of 2,500 students. At the 5% significance level, does the survey support the notion that more than a third of post-secondary students graduate with more than $30,000 worth of loans?

4. A survey of the morning beverage market shows that the most common breakfast beverage of 83% of post-secondary college employees is coffee. A random sample of 400 VCC employees is taken and they are asked to name their most frequently consumed breakfast beverage. If 344 replied that coffee was their most common breakfast drink, does this provide evidence than a greater proportion of VCC employees drink coffee with breakfast as compared to post-secondary employees? Use $\alpha = 0.25$.

---

### Solutions

1. a) $H_0: \mu = 7$  $H_1: \mu > 7$
   b) $H_0: p = 0.65$  $H_1: p > 0.65$
   c) correct
   d) $H_0: p = 0.30$  $H_1: p < 0.30$
   e) correct
   f) $H_0: p = 0.50$  $H_1: p > 0.50$

2. a) reject $H_0$
   b) fail to reject $H_0$
   c) reject $H_0$
   d) reject $H_0$

3. $n = 2500$, $p = 1/3$, $q = 2/3$  $np$ and $nq$ are both $\geq 10$ so we can approximate with the normal distribution.

   $H_0: p = 0.3333...$
   $H_1: p > 0.3333...$
   $z = 1.77$, $p$-value $= 0.0384$

   Since $\alpha (0.05) > p$-value, this is a significant result and we reject the null hypothesis. The survey provides evidence that more than a third of post-secondary students graduate with more than $30,000$ worth of loans.

4. $H_0: p = 0.83$  $H_1: p > 0.83$  $z = 1.60$, $1 - p$-value $= 1 - 0.9452 = 0.0548$

   Since the $p$-value is $< \alpha$, the result is significant. We reject $H_0$, so there is sufficient evidence that the average proportion of VCC employees who drink coffee as a morning beverage is greater than the proportion of post-secondary employees.