Hypothesis Tests

A **hypothesis test** starts with two competing claims or ideas, called hypotheses. The null hypothesis (H_0) represents a claim that we want to test; it is represented by a single value (e.g. $H_0 : p = 0.5$). Sometimes people think of the null hypothesis as a "straw man" - it is a hypothesis we would like to find evidence against. The alternative hypothesis (H_A) represents an alternative claim under consideration. Often it is represented by a range of values (e.g. $H_A : p \neq 0.5$).

Interpretation

In the following, we will discuss how to perform hypothesis tests with (a) confidence intervals and (b) using p-values. In both cases, we will "fail to reject the null hypothesis" if there is not enough evidence in favor of the alternative.

Hypothesis Tests with Confidence Intervals

One way to conduct a hypothesis test is using a confidence interval. The steps are:

1. Set up your null and alternative hypotheses (based on a particular situation).
   - Mean
     - Population mean (μ) is different than hypothesized mean (μ_0)

       $H_0 : \mu = \mu_0$

       $H_A : \mu \neq \mu_0$
     - Population mean (μ) is greater than hypothesized mean (μ_0)

       $H_0 : \mu = \mu_0$

       $H_A : \mu > \mu_0$
     - Population mean (μ) is less than hypothesized mean (μ_0)

       $H_0 : \mu = \mu_0$

       $H_A : \mu < \mu_0$
   - Population proportion (p) is different than hypothesized proportion (p_0)
     - Option 1:

       $H_0 : p = p_0$

       $H_A : p \neq p_0$
     - Population proportion (p) is greater than hypothesized proportion (p_0)

       $H_0 : p = p_0$

       $H_A : p > p_0$
     - Population proportion (p) is less than hypothesized proportion (p_0)

       $H_0 : p = p_0$

       $H_A : p < p_0$

2. Identify your sample statistic.
3. Check if it is appropriate to assume approximate normality (consider parent distribution/sufficiently large sample size).
   - *Note: When we are conducting a hypothesis test, we check binomial conditions using p_0 rather than p (unknown) or $\hat{p}$, like we did when we were estimating p but NOT conducting a hypothesis test.*
4. Determine what confidence level will be used for your confidence interval. If no level is specified, assume a 95% confidence interval.
5. Build a confidence interval for your sample statistic.
6. Assess whether your hypothesized value, μ_0 or p_0, is in the confidence interval. If it is, then there is not enough evidence to reject H_0. If it is not in the interval, then there is sufficient evidence to reject H_0 in favor of H_A.

Hypothesis Tests with p-values

The **p-value** is the probability of observing data at least as favorable to the alternative hypothesis as our current data, if the null hypothesis is true. We will use either the sample mean $\bar{x}$ or the sample proportion $\hat{p}$ to compute the p-value and evaluate the hypotheses. *You should know that while p-values are still widely used, they are controversial and there is a move towards using confidence intervals, for example, for hypothesis testing, rather than p-values.* The American Statistical Association issued a statement on p-values. Here is a brief summary: `https://www.amstat.org/asa/News/ASA-P-Value-Statement-Viewed-150000-Times.aspx`

Steps to hypothesis tests with p-values:

1. State the hypotheses as before.
2. Calculate the test statistic (this is a Z-score in this context)
   - $z = \frac{\text{point estimate} - \text{hypothesized value}}{SE}$
   - Test for mean: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
   - Test for proportion: $z = \frac{\hat{p}-p_0}{\sqrt{p(1-p)/n}}$
   - Check conditions: is a normality assumption appropriate?
   - Use your test statistic to calculate a p-value:
     - Alternative is $\neq$
       * $2 \times P(Z > z)$=`2*pnorm(q=`z`, lower.tail=FALSE)`
     - Alternative is $>$
       * $P(Z > z)$=`pnorm(q=`z`, lower.tail=FALSE)`
     - Alternative is $<$
       * $P(Z < z)$=`pnorm(q=`z`, lower.tail=TRUE)`

- If the probability is less than α, then we reject the null hypothesis in favor of the alternative. Otherwise we fail to reject the null hypothesis. Always interpret p-values in the context of the problem at hand.

Decision Errors

Hypothesis tests are not infallible. Sometimes we will make the wrong the decision about rejecting or failing to reject the null hypothesis. There are two types of errors we can make. Because statisticians are incredibly creative, they are helpfully (not really) termed "Type 1" and "Type 2".

- A **Type 1 Error** is rejecting the null hypothesis when H_0 is actually true. This corresponds to the **significance level**, α, which is the proportion of Type 1 errors we are willing to make. Most commonly, $\alpha = 0.05$; this is equivalent to building a 95% confidence interval.
- A **Type 2 Error** is failing to reject the null hypothesis when the alternative is actually true. There is always a trade-off between Type 1 and Type 2 errors. If we make α too small, then we will make more Type 2 errors, for example. You should be aware of the trade off - fewer Type 1 errors corresponds to more Type 2 errors and vice versa.

		Test conclusion	
		do not reject H_0	reject H_0 in favor of H_A
Truth	H_0 true	okay	**Type 1 Error**
	H_A true	**Type 2 Error**	okay

Practice Problems

1. According to official census figures, 8% of couples living together are not married (in the United States). A researcher took a random sample of 400 couples and found that 9.5% of them are not married.

   (a) Carry out a hypothesis test using a confidence interval. Be sure to state your conclusions in the context of the problem.

   (b) Carry out a hypothesis test using a test statistic and p-value. Be sure to state your conclusions in the context of the problem.

   (c) What do you notice about the conclusions in (b) and (c)? Do they agree? Are you surprised?

2. The board of a major credit card company requires that the mean wait time for customers for service calls is at most 3.00 minutes. To make sure that the mean wait time is not exceeding the requirement, an assignment manager tracks the wait times of 45 randomly selected calls. The mean wait time was calculated to be 3.4 minutes. Assume the population standard deviation is 1.45 minutes.

   (a) Is there sufficient evidence to say that the mean wait time is longer than 3.00 minutes with a 95% level of confidence? Use a test statistic and p-value to perform the hypothesis test.

   (b) Is there sufficient evidence to say that the mean wait time is longer than 3.00 minutes with a 98% level of confidence? Use a test statistic and p-value to perform the hypothesis test.

3. The population standard deviation for waiting times to be seated at a restaurant is know to be 10 minutes. An expensive restaurant claims that the average waiting time for dinner is approximately 1 hour, but we suspect that this claim is inflated to make the restaurant appear more exclusive and successful. A random sample of 30 customers yielded a sample average waiting time of 50 minutes.

   (a) Test the assertion that the restaurant's claim is too high (i.e., the mean waiting time is actually less than one hour)?

   (b) The original data (individual waiting times) is not normally distributed. What theorem allows us to do the calculations in part (a)?

4. In a survey of 1273 adults, 52% said it is not morally wrong to change the genetic makeup of human cells. We are interested in constructing a test for following statement: "The majority of adults do not think it is morally wrong to change the genetic makeup of human cells"?

   (a) Conduct this test using a confidence interval.

   (b) Conduct this test using a test statistic and p-value.