Intro to Hypothesis Testing - Lecture Notes

Confidence intervals allowed us to find ranges of reasonable values for parameters we were interested in. Hypothesis testing will let us make decisions about specific values of parameters or relationships between parameters.

There are 5 settings we will be seeing in depth:
1. π - population proportion
2. μ - population mean
3. $\mu_1 - \mu_2$ - difference in population means (example: compare average heights of men and women)
4. μ_d - population mean difference (for paired data) (example: compare average heights of fathers and sons)
5. $\pi_1 - \pi_2$ - difference in population proportions (example: compare proportion of students who study abroad at Amherst vs. Smith)

You will need to be able to distinguish these settings from each other. Notably, to distinguish proportion settings from mean settings, think about the question being asked. Are responses to the question of interest yes or no? If so, you are dealing with proportions. If the response is a numerical value, you are dealing with means. (At least for now).

In hypothesis testing, we use sample data to choose between two competing hypotheses. Think of it like a jury trial. There are two options: innocent and guilty. You assume innocence until shown guilty beyond a reasonable doubt. In hypothesis testing, there are 2 choices, the null hypothesis and the alternative hypothesis. You assume the null hypothesis is true until the alternative is shown beyond "chance".

Hypothesis - a claim about a population characteristic (parameter)
Null Hypothesis - the status quo - initially assumed true
Alternative Hypothesis - the researcher's proposal - what you hope to show

Main idea: Reject the null hypothesis in favor of the alternative only with convincing/significant evidence. We do NOT say that we accept the alternative, only that we have significant evidence to reject the null. This is because we could have made a mistake (see below).

Forms of Hypotheses:
Null Hypothesis: H_0 : population parameter = some hypothesized value
Alternative Hypothesis:
H_A : population parameter $\neq$ that same hypothesized value (two-sided) OR
H_A : population parameter $>$ that same hypothesized value (one-sided to the right) OR
H_A : population parameter $<$ that same hypothesized value (one-sided to the left)

The null hypothesis implicitly maintains values not mentioned in the alternative hypothesis. If $H_A : \mu > 25$, then the null hypothesis is written as $H_0 : \mu = 25$, but it could also be written as $H_0 : \mu \leq 25$.

Hypotheses must be written in terms of parameters, not statistics, and the hypothesized values must match and make sense! From the choices below, which are valid pairs of hypotheses?

1. $H_0 : \bar{x} = 30$ vs. $H_A : \bar{x} < 30$
2. $H_0 : \mu = 270$ vs. $H_A : \mu \neq 270$
3. $H_0 : \pi > .3$ vs. $H_A : \pi = .3$
4. $H_0 : \mu = 70$ vs. $H_A : \mu = 69$
5. $H_0 : \pi = 3$ vs. $H_A : \pi > 3$

It is very important to try to DEFINE your parameter of interest as clearly as possible. For example $\mu =$ population average height is not as informative as $\mu=$ population average height of immigrant workers in the 1940s in California. Writing a parameter definition can also help you catch mistakes in your choice of setting.

Try setting up some hypotheses on your own, and be sure to define your parameters:

1. For the now nearly infamous soda bottle example, the company claims that there are 20 oz. per bottle. You (a consumer) will sue if there is less than 20 oz. on average per bottle. What hypotheses should you test?

2. You are interested in whether a proposal related to environmental conservation will be voted into law by a majority vote at the next election. What hypotheses should you test?

3. What if the law had to be passed by a 2/3 majority in the state senate (assume you could obtain a random sample of state senators)? Then what hypotheses would you test?

Errors in hypothesis testing:

Just like you could have a mistake in a jury trial, you can make mistakes in hypothesis testing. Let's consider the mistakes in a jury trial and determine their statistical counterparts.

1. Innocent man is pronounced guilty. This would mean you rejected the null hypothesis when in fact, the null hypothesis was true. This is a Type I Error.
2. Guilty man is pronounced innocent. This would mean you did not reject the null hypothesis when in fact, the alternative was true. This is a Type II Error.

Both Type I and Type II errors have different consequences. Type I is usually considered more serious, and it is the value we have direct control over.

Related quantities:
Probability of a Type I Error = P(Type I Error) = α = significance level (you get to set this in advance of the test)
Probability of a Type II Error = P(Type II Error) = β
Power = 1-β

In any hypothesis test, you need to balance α and β. You can never get them both to be 0 with a sample. In practice, we choose the largest α that is tolerable - usually .01, .05 or .10. The relationship between α and β is not deterministic. That is to say, when α increases, β decreases (and power would increase), and vice versa, but you can't say how much because other factors affect the quantities.

Example of Errors and their Consequences:
You want to try a new cancer treatment for your patients. The current treatment has a remission rate of 40 percent. You want to see if the new treatment is more effective.
$H_0 : \pi = .4$ vs. $H_A : \pi > .4$ where π is the population proportion of cancer patients who enter remission after undergoing the new treatment.
A Type I error in this context would mean that you concluded that the new treatment is more effective than the current treatment, when in fact, it isn't more effective. As a consequence, you may assign patients to the new treatment when it is not any better than the current treatment.
A Type II error in this context would mean that you concluded that the new treatment is not more effective than the current treatment, when in fact, it is more effective. As a consequence, you may not swap patients to the new treatment which would give them a better chance at remission than the current treatment.

Example for you to try:
We want to explore the mean number of eggs laid by female fish of a certain species because we are concerned about the survival of the species. (My numbers are made up, but you can see the idea). If the average number of eggs laid is less than 25, the species will be considered endangered and precautions to help it survive may be implemented. What hypotheses should you test? (Define your parameter!) Interpret what Type I and Type II Errors mean in context of this example and give a consequence associated with each.

```
Steps
1. Define parameter.
2. Give null hypothesis.
3. Give alternative hypothesis.
4. Select α -- when want to accept Ha or how confident you need
5. Give test stat formula.
6. Check assumptions.
7. Compute test statistic.
8. Compute p-value.
9. State conclusion in context.
```