4/1/2023

Hypothesis Testing
One-Sample/Two-Sample

We start with our assumptions.
The assumptions that we make without any data (or based on previous data) is called the null hypothesis. Fallback position. The "safe" assumption. Notation is H_0.
The second hypothesis, the alternative hypothesis, is the claim that we are trying to prove based on the data. The notation is H_A, H_a, H_1.

The null hypothesis is going to be the information about the world that we assume is true, and base our probability calculations on that assumption. We are going to calculate the probability that the sample we collected could have been collected if the null assumption is, in fact, true.

If the probability we find is larger than the significance level that we set in advance, then we do not have enough evidence to establish the alternative is true. Fail to reject the null hypothesis – we fall back on our initial assumption. If the probability is smaller than the significance level, then the likelihood that we could get the data we have and the null hypothesis being true is very small, then we reject the null hypothesis and we accept the claims of the alternative (we do have strong evidence).

The significance level is designated as α, and it's related to the confidence interval. Typically the significance level is set at 0.05 or 5%, which is the complement of the standard confidence interval level of 95%. In some circumstances, there may be reasons to adjust the significance level. For the cases where the consequences of making an incorrect assertion of the alternative are not very severe, then we can raise the significance level to a higher number (say 10%). An example might be an education setting, where both methods of learning are good enough, but we are testing to see which is better. If the consequences of being wrong are very high, then you might lower the significance level to say 1%.

If problems do not say what the significance level is, assume it is 5%. But some problems will give other values.

Our tests typically involve means and proportions. The null hypothesis statement always includes an equal sign. The alternative will include an inequality or a not-equal-to sign.

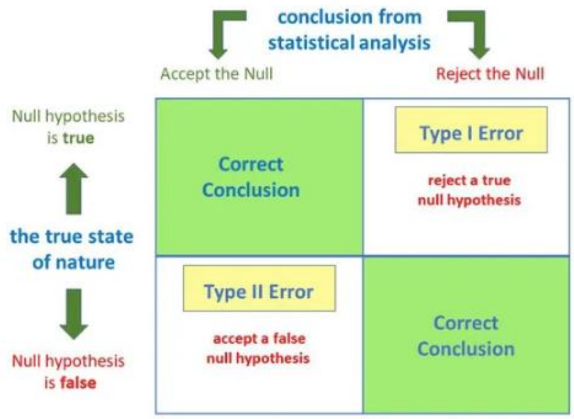
$$H_0: \mu = 500$$
$$H_A: \mu \neq 500$$

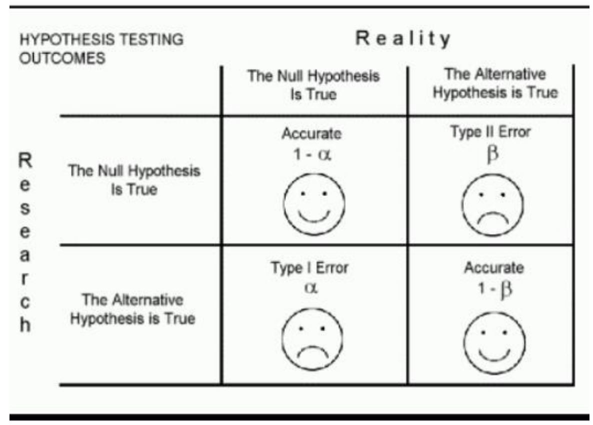
This is an example of a two-tailed test.

$$H_0: p = 0.5$$
$$H_a: p < 0.5$$

This is called a one-tailed test.

Because we are doing random sampling, there is a chance that despite our best efforts that we will come to the wrong conclusion. So we want to talk about errors.





Two kinds of errors:
Type I error: related to the significance level. This is the chance that the null is true, but my data says otherwise. (false positive)
Type II errors: the chance of a type II error is called β ($1-\beta$ is called the power of a test). The possibility that the null hypothesis is false, but we are unable to prove that it is false. (false negative)

We are going to use the p-value method rather than the rejection region method. Calculating a test-statistic and then converting that to a probability. Compare that p-value to the significance level.

Example:
Jeffrey, as an eight-year old, established a mean time of 16.43 seconds for swimming the 25-yard freestyle, with a standard deviation of 0.8 seconds. His dad, Frank, thought that Jeffrey could swim the 25-yard freestyle faster using goggles. Frank bought Jeffrey a new pair of expensive goggles and timed Jeffrey for 15 25-yard freestyle swims. For the 15 swims, Jeffrey's mean time was 16 seconds. Frank thought that the goggles helped Jeffrey to swim faster than the 16.43 seconds. Conduct a hypothesis test using a preset α = 0.05. Assume that the swim times for the 25-yard freestyle are normal.

$$H_0: \mu = 16.43$$
$$H_a: \mu < 16.43$$

We want to prove that the goggles decreased his time.
We want to conduct a t-test because even though we have the standard deviation of his previous times (population), we still only have a sample of 15, which is less than $30 \sim 40$.

The test-statistic is basically the standard score formula, but in the context of a sampling distribution.

$$t = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{16 - 16.43}{\frac{0.8}{\sqrt{15}}} = -2.0817 \ldots$$

Essentially, this is more than two standard deviations below the mean that you would expect. For the p-value method, we want to convert this to a probability.

If the probability is below the significance level (0.05), then we conclude the new data is unlikely to have come from the same distribution as the old data, and therefore the speeds are faster. If the probability is above 0.05, then there is not sufficient evidence to conclude that the goggles reduced the speed.

The p-value we found to the 0.028… < 0.05. We reject the null hypothesis, therefore, we have sufficient evidence to think that the alternative is true… in this context, the swim times are shorter.

Example.
A teacher believes that 85% of students in the class will want to go on a field trip to the local zoo. She performs a hypothesis test to determine if the percentage is the same or different from 85%. The teacher samples 50 students and 39 reply that they would want to go to the zoo. For the hypothesis test, use a 1% level of significance.

$$H_0: p = 85\%$$
$$H_a: p \neq 85\%$$

Sample proportion is $\frac{39}{50} = 0.78, 78\%$

The test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

$$z = \frac{0.78 - 0.85}{\sqrt{\frac{0.85(0.15)}{50}}} = -1.3862\ldots$$

The p-value we found is 0.1656… >0.01, we fail to reject the null. The alternative was that it was different than 85%, and we don't have enough evidence to think it is different.

Two sample tests. If we are comparing data from two samples.

The null hypothesis compares the two samples.

$$H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

Equivalently:

$$H_0: p_1 - p_2 = 0$$
$$H_a: p_1 - p_2 < 0$$

$$z = \frac{\left((\hat{p}_1 - \hat{p}_2) - 0\right)}{\sqrt{\frac{p_0(1-p_0)}{n_1 + n_2}}}$$

If p_0 is not given in the problem, assume the average of the two proportions:

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

In a means there are other options.

You can have a dependent or independent data.
Dependent data is also called paired data, the problem may say something like “matched”. We calculate the difference between the pairs, and then do a single sample test.

If the data is independent, then we calculate statistics on each sample separately. These can have different sample sizes. There is a pooled approach (which assumes that the standard deviation of both groups is the same), and there is a unpooled approach (which doesn’t).

Unpooled = heteroscedastic
Pooled = homoscedastic

Review for Exam #2
The exam begins with probability distributions (discrete), continuous, binomial, normal, confidence intervals.
Hypothesis testing is not this test.

Discrete probability expected values
Binomial probability question
Normal distribution: compare values in different distributions with a z-score, estimate with the Empirical rule, using normal distribution function
calculating probabilities with t-distribution
sampling distributions
confidence intervals
counting formulas