Statistics 10 Lecture 15 Hypothesis testing (Chapter 20)

1. Introduction to Hypothesis Testing (Chapter 20)

This is the basic idea in Chapter 20: we make assumptions about the unknown parameters, and then test to see if those assumptions could have led to the outcomes (statistics) we actually observe. We then use a probability calculation to express the strength of our conclusions, stated as a chance (probability) and not as a confidence interval (even though the parameter is unknown).

2. An Example from a recent article

Someone has recently reported that smoking marijuana may help sufferers of MS, they used statistics to help them reach a conclusion similar to this one:

"It involved 100 multiple sclerosis patients from around Britain. They received a capsule containing whole cannabis oil. Results were reported after 15 weeks of treatment. Fifty-seven percent of the patients taking the whole cannabis extract said their pain had eased compared with a prior research that 47 percent of MS patients will report some easing of pain after 15 weeks when using prescribed drugs or alternative medical treatments."

Hypothesis testing involves a series of steps

First, state a NULL HYPOTHESIS. The null hypothesis is a claim that observed results that differ from the population parameter of interest are due to chance alone. The null hypothesis is a statement about a parameter, e.g. the true population proportion is .47 or 47%. It is generally written H_0 and in this example, it would be written:

$$H_0 : p = .47$$

Second, state an ALTERNATIVE HYPOTHESIS. The alternative suggests that the observed results (sample outcomes) are due to more than just chance. It implies that the statement about the NULL is not correct (i.e. the proposed parameter is wrong) and any observed differences are real, not just luck. Usually, the ALTERNATIVE is what we're setting out to prove. The NULL is like a "straw man" that we set up to knock down. In this example, the alternative is written

$$H_0 : p > .47$$

because we believe that the patients on medicinal marijuana had greater pain relief than those who did not have medical marijuana. There are other alternatives possible (see page 376) , such as $p < .47$ or $p \neq .47$ but this one fits our example the best because it makes sense (it's consistent with evidence of interest).

The TEST PROCEDURE measures how different the observed results are from what we would expect to get if the null hypothesis were indeed true. When using the normal curve, the test is z,

where z = (observed value – hypothesized parameter)/(appropriate standard deviation)

All a Z test does is it tells you how many standard deviations away the observed value is from the hypothesized population parameter when the parameter is a value established by the NULL HYPOTHESIS. In this example:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{.57 - .47}{\sqrt{\frac{.47 * .53}{100}}} = \frac{.10}{.0499} = +2.00$$

This is called a "one proportion Z test" Notice the standard deviation that we use. It's established by the population parameter and the sample size. Recall that statistics come from samples and we work with samples as if they are

drawn from a population of samples. Under certain conditions (recall the CLT) the population of samples (the sampling distribution) will be normal and centered around the parameter.

The test just answers the question: are the sample results near or far from the parameter? We could also say… how likely is it to observe a 57% when we are expecting a 47%. The Z-test answers: a .57 is +2.00 Standard Deviations away from a .47. So how far is that? One checks the area under the curve that is as extreme or more extreme than the result of the Z-test, so the area at +2.00 and beyond (.0228 or 2.28%)

The result of the Z-test is called a **P-VALUE (a probability value).** This is the chance (probability) of getting results (Z) as or more extreme than what we got, IF the null hypothesis were true.

P-VALUE could also be called an "observed significance level" and it is simply the "tail" area associated with the Z score calculated using the sample information and parameter information.

p-values can be interpreted as "if-then" statements:

"If the null hypothesis were true, then there would be a p% chance of getting these kind of results from a single sample."

The less probable an outcome is, the stronger the evidence that we would reject the null in favor of the alternative.

3. Notes on Hypothesis Testing

A. Clearly identify the parameter and the outcome.

B. State the null hypothesis. This is what is being tested. A test of significance assesses the strength of evidence (outcomes) against the null hypothesis. Usually the null hypothesis is a statement of "no-effect" or "no difference" and it is ALWAYS a statement about the parameter.

C. The alternative hypothesis is the claim about the population that we are trying to find evidence in favor of. In our example today, you are seeking evidence that marijuana helps. The null hypothesis would say that the percentage experiencing an easing of pain will not be different from existing treatments (i.e. p=.47), the alternative would say the parameter is larger when marijuana is involved (i.e. p>.47). Note this is a ONE-SIDED alternative because you are only interested in deviations in one direction.

D. Choose an appropriate test statistic. In the above example, the parameter is the population proportion and the outcome is the sample proportion and the appropriate test-statistic is Z.

The significance test assesses the evidence by examining how far the sample statistic falls from the proposed null (parameter) using the Z score as the measure of distance (in Standard Deviations).

The answer the test gives is ultimately the probability of getting an outcome as extreme or MORE than you actually observed. So to test the outcome of the sample, you would ask a question like "what is the chance of getting a 57% or higher when we are expecting 47%?

E. The probability that results from the test is called a P-VALUE and that is the tail area associated with the resulting Z score. The smaller the p-value the stronger is the evidence against the null hypothesis. If instead you had gotten a sample proportion of .51 with the same standard deviation, there would not be enough evidence to make the claim that marijuana helps MS sufferers. (A Z of about .80 has about 21% of the area under the normal curve, so here, there was a 21% chance of getting a sample with 51% or more showing pain relief).