Lecture 32

So far we have considered one statistical inference technique: Confidence intervals. Today our discussion moves to another statistical inference technique known as hypothesis testing.

<u>Hypothesis Test</u>: a formal method to test a claim about the population parameter (of model) based on sample data

<u>Elements of a hypothesis test</u>

1. Null hypothesis H_0
2. Alternative hypothesis H_A
3. Test statistic
4. Pvalue or level of significance and rejection region.
5. Conclusion.

Hypotheses

the null hypothesis H_0 is a claim that we intially assume to be true.

the alternative hypothesis H_A is an assertion that is contradictory to H_0.

A hypothesis test is designed to assess the strength of evidence against H_0.

For our purposes the null hypothesis will usually be a statement of "no difference" or "no effect".

Note: Both H_0 and H_A should be stated in terms of the population parameter.

eg

$H_0: \mu = 110.4$ $\qquad$ $H_A: \mu \neq 110.4$

$H_0: p \geq .5$ $\qquad$ $H_A: p < .5$

$H_0: \sigma_1^2 \leq \sigma_2^2$ $\qquad$ $H_A: \sigma_1^2 > \sigma_2^2$

The test statistic

This is a quantity which we compute based upon the sample data. It is used to decide between H_0 and H_A. In otherwords we use it to decide which of the null or alternative hypotheses is more likely to be true given our observed data.

P value

The p-value is the probability, if we assume H_0 is true, that the test statistic would take the value observed or something more extreme. This means that the smaller the p-value the less likely it is that H_0 is true. In practise, the smaller the p-value the more evidence you have against H_0.

Some example P-values

P-value	Evidence against H_0
$< .1$	very weak evidence
$< .05$	weak evidence
$< .01$	evidence
$< .001$	strong evidence

$< .0005$ Very strong evidence.

Level of significance/rejection region

This is an alternative method of deciding between a null and alternative hypothesis. First we choose a value α, which we call the level of significance, that is the largest probability of falsely rejecting H_0 when it is true that we can tolerate. Typical values of α are .1, .05, .01.

The rejection region is the set of all possible values of the test statistic for which we reject H_0. For our purposes this rejection region is related to α (the significance level). Ie the probability of getting a test statistic in the rejection region assuming H_0 is true is α.

For the most part our focus will be on the P-value approach, but you should be aware of how to find a rejection region.

Note you can compare your P-value with α when you want to decide between H_0 and H_A. If

$$\text{p-value} \leq \alpha \implies \text{reject } H_0 \text{ in favor of } H_A$$

$$\text{P-value} > \alpha \implies \text{cannot reject } H_0.$$

Drawing a conclusion

You should state your conclusion in terms of whether or not you ~~[illegible]~~ reject or cannot reject H_0. Note that it is ok to say that H_0 is false and H_A is true as a conclusion of your test (ie "I reject H_0 and accept H_A"). It is not correct however to say the reverse ie H_0 is true and H_A is false. Why? because you assume already that H_0 is true so there is no way to prove it is true. Instead you say ("I cannot reject H_0")

Errors in hypothesis testing (ie why we choose α)

type I error – rejecting H_0 when it is really true

type II error – not rejecting H_0 when it is really false

We call

$$\alpha = P(\text{making a type I error})$$

$$\beta = P(\text{making a type II error})$$

Note ~~always~~ α depends only on a specific value of the parameter which is specified by H_0 whereas β varies ~~[illegible]~~ depends on which particular value of the parameter is specified by the alternative (there are typically many). ~~[illegible]~~

Also not α here is same as α above (level of significance).

Ideally we would like a test procedure which had zero possibility of either error. Unfortunately in the real world because of sampling variability this is impossible. Instead we try to make these probabilities as small as possible. Since α is easy to work out for a given null hypothesis we usually fix α to some level and then live with whatever value we get for β.