Stat 6623 – Statistical Methods with SAS Programming

<u>More on hypothesis testing</u>: adapted from "Intro Stats" by Richard DeVeaux and Paul Velleman (Addison Wesley).

Hypothesis testing is very much like a court trial. Suppose the defendant has been accused of robbery. We wonder whether he is guilty or not. In U.S. law, the null hypothesis is that the defendant is innocent. Instructions to juries are explicit about this. In the trial the prosecutor gathers and presents evidence. This evidence takes the form of facts that seem to contradict the presumption of innocence. ("If the defendant were innocent, wouldn't it be remarkable that the police found him at the scene of the crime with a big bag of money in his hand, a mask on his face, and a getaway car parked outside?")

The next step is to judge the evidence. The jury considers the evidence in light of the presumption of innocence and judges whether the evidence against him would be plausible if the defendant were in fact innocent. We ask, "Could these data plausibly have happened by chance if the null hypothesis were true?" If they were very unlikely to have occurred, then the evidence raises reasonable doubt in our minds about the null hypothesis.

If the evidence is not strong enough to reject the defendant's presumption of innocense, what verdict does the jury return? They say "not guilty." They do not say the defendant is innocent. All they say is they have not seen sufficient evidence to convict (reject innocence).

The jury's null hypothesis is H_o: innocent defendant. If the evidence is too unlikely given this assumption, the jury rejects the null hypothesis and finds the defendant guilty. If there is insufficient evidence to convict the defendant, the jury does not decide that H_o is true and declare him innocent. Rather, juries can only fail to reject the null hypothesis and declare the defendant "not guilty."

In the same way, if the data are not particularly unlikely under the assumption that the null hypothesis is true, then the most we can do is to "fail to reject" the null hypothesis. We never declare the null hypothesis to be true, because we simply do not know whether it's true or not. Sometimes in this case we say the null hypothesis has been retained.

More on p-value: adapted from "The Basic Practice of Statistics" by David Moore (Freeman).

Sometimes we demand a specific degree of evidence in order to reject the null hypothesis. We can compare the p-value with a fixed value that we regard as decisive. This amounts to announcing in advance how much evidence against H_o we will insist on.

When there is a decisive value of p-value, it is called the significance level and we write it as α. For example, when α is set at .05, we require evidence against H_o so strong that it would happen no more than 5% of the time when H_o is true. When we set α = .01, we are insisting on stronger evidence against H_o – evidence so strong that it would appear only 1% of the time if H_o is in fact true.

If the p-value is as small or smaller than α, we say that the data are statistically significant at level α. "Statistically significant" does not mean important – it simply means "not likely to happen by chance."

How small a p-value is convincing evidence against H_o? The answer depends mainly on two circumstances.

- How plausible is H_o? If H_o represents an assumption that the people you must convince have believed for years, strong evidence (small p-value) will be needed to persuade them.

- What are the consequences of rejecting H_o? If rejecting H_o in favor of H_a means making an expensive change, you need strong evidence that making the expensive change will be beneficial.

There is no sharp border between "significant" and "insignificant," only increasingly strong evidence as the p-value decreases. It makes no sense to treat p-value $\leq$ .05 as a universal rule for what is significant!