Hypothesis Testing Terminology
STAT 420 FALL 2011

A hypothesis test is a common statistical procedure. A conjecture is made based on a model, data is obtained, and a conclusion is made using statistical techniques. Several standard tests are derived in Stat 430, but we introduce the concepts here. During the semester we can combine probability ideas with hypothesis testing ideas to provide examples of how probability is used by statisticians.

- **The Hypotheses:** There are always two hypotheses in a statistical test. The **null hypothesis** is the status quo, the "nothing's happening," or the "no difference" hypothesis. This is labeled as H_0. The **alternative hypothesis** is the claim that we are testing. It is labeled H_a, or sometimes as H_1.
- **The Decision:** We take some data, do some calculations, and come up with a decision. This is traditionally phrased in terms of the null hypothesis, so that we can either *reject* the null hypothesis, or *accept* the null hypothesis.
- **The Decision Rule:** Usually phrased as "we reject H_0 if ....."
- **Errors:** If there are two possible decisions, then there are two possible mistakes we can make. A **Type I Error** is rejecting the null hypothesis when it is really true. A **Type II Error** is accepting the null hypothesis when it is really false.
- **The Test Size:** The size of the test is the probability of making a Type I Error. This is represented by the Greek letter α.
- **Statistically Significant:** Results are called statistically significant if the null hypothesis is rejected.
- **p-value** The p-value is the probability of seeing our data or more extreme, given that the null hypothesis is true, where "more extreme" means "supports the alternative hypothesis more."
- **Power:** The power of a test is the probability that the null hypothesis is rejected, given that the alternative is true.

Example: Testing a die

Suppose a six-sided die is used in a gambling game, and you suspect the die is not "fair." For a fair die, the probability of attaining any of the six faces on a given roll is 1/6. Suppose you think that the die might be weighted and sixes will come up too often. You want to test $H_0 : p = 1/6$ versus $H_a : p > 1/6$ where p is the true probability of getting a six.

Note that the null hypothesis corresponds to the die being fair. The idea is that we want to have strong evidence against the null hypothesis before we will reject it and claim that the die is not fair. Otherwise, we accept the null hypothesis even if there is *some* evidence against it.

A Type I Error occurs if we reject H_0 and claim that the die is not fair, when the die really is fair. The consequences of a Type I Error might be to make a false accusation. A Type II Error occurs if we accept H_0 and decide that the die is fair, when it is not. The consequences of a Type II Error might be to let people get away with cheating.

Now we need to design the experiment, which can be as simple as rolling the die 30 times and counting the number of sixes. Our decision rule might be "reject H_0 if we get at least 8 sixes."

The test size will be the probability that we see at least 8 sixes when the die is really fair. Typically we want the test size to be small. If this is large we might want to change the decision rule.

Suppose we perform the experiment and we see that 10 out of the 30 rolls are sixes. The p-value would be the probability of at least 10 sixes (our data or more extreme) when the die is fair. The smaller the p-value is, the stronger the evidence against the null hypothesis.