Hypothesis Testing Key terms

1. STATISTICAL HYPOTHESES:

A statistical hypothesis is a statement concerning the distribution of a random variable, e.g. the hypothesis that a coin is fair. This is equivalent to the statement "the probability of obtaining a head is 0.5" and completely defines the probabilities involved in tossing a coin.

1.1. Testing Hypotheses

Suppose we believe that a coin may be biased. One way to test whether we are correct is to toss the coin a large number of times (say 100) and count the number of heads. If we get approximately 50 heads, we must admit that the coin is probably fair, but if we get a lot more heads than tails we are probably safe if we conclude that it is biased. The decision procedure is a **statistical hypothesis test**.

1.2. Null and Alternative Hypotheses

The purpose of the test is to choose between the two hypotheses. One is called the **null** hypothesis (denoted H_0) and the other is the **alternative** hypothesis (denoted H_1). Generally, the null hypothesis is the one we are attempting to **disprove**. So we can write:

$H_0 : p = 0.5$ (null hypothesis)

$H_1 : p \neq 0.5$ (alternative hypothesis).

1.3. Decision Rule

We must now formulate a decision rule for deciding between two hypotheses. The decision rule is a **function** of the data. A typical decision rule is "accept H_0 (that the coin is fair) if we obtain between 40 and 60 heads".

1.4. Critical Region

The set of values of the test statistic that cause us to reject the null hypothesis is called the Critical Region. In the above example, the Critical Region includes 0-39 heads and 61-100 heads.

1.5. Two types of errors

It is possible to obtain 61 heads with a fair coin and it is possible to obtain 60 heads with a biased coin, so the above rule could lead to the wrong conclusion in two ways.

- The incorrect rejection of the null hypothesis (saying the coin is biased when it's really fair) is called a **Type I Error**.
- The incorrect acceptance of the null hypothesis (saying it is fair when it is really biased) is called a **Type II Error**.

1.6. Error Probabilities

The probabilities associated with **Type I and Type II Errors** are determined by the choice of Critical Region. The probability α of a Type I Error is called the **Significance Level** of the Hypothesis Test. The probability of a Type II Error is denoted by β. It is desirable to keep α and β as small as possible, but any decrease in α leads to an increase in β, unless the sample size (number of tosses) is increased. We usually select a value (such as 0.05 or 0.01) for α, the Significance Level, and adjust the Critical Region accordingly.

1.7. One-Tailed Hypothesis Tests

If we are willing to reject a null hypothesis because of extreme values on either side of the mean, *the* test is called a two-tailed test. Sometimes we are interested in rejecting just one side only. Given a null hypothesis

$$H_0 : \mu = \mu_0$$

a one-tailed test is required for an alternative hypothesis of the form $H_1 : \mu > \mu_0$ (as opposed to $H_0 : \mu = \mu_0$).

The need for a one-tailed test is usually implied by the context of the problem. The decision to use a one-tailed test must never be taken on the basis of the data itself.

1.8. p-value

The value of the significance level used for H_o to be rejected is largely dependent on the circumstances, and the consequences of rejecting H_o. An alternative approach is to quote the *p-value*. This is the probability of getting "**the result or a more extreme one**" if H_o true.