Levels of Significance

Level of significance is denoted by Greek symbol alpha (α) which is the level of probability at which the null hypothesis can be rejected with confidence and the research hypothesis can be accepted with confidence. The 0.05 level of significance is found in the small areas of the tails of the distribution of mean differences. These are the areas that represent a distance of plus or minus 1.96 standard deviations from a mean difference of zero. If we obtain a *z* score that exceeds 1.96, it is called statistically significant at $\alpha=0.05$. If $\alpha=0.01$, the critical *z* score is 2.58.

Choosing Levels of Significance

Rejecting the null hypothesis when it is true is called Type I error, and its probability is symbolized by α. Retaining the null hypothesis when it is false is called Type II error and is represented by β. We predetermine our level of significance for a hypothesis test on the basis of which error is more costly or damaging, and therefore riskier.

Difference between P and α

P is the exact probability that the null hypothesis is true in light of the sample data; the alpha value is the threshold below which the probability is considered so small that we decide to reject the null hypothesis. We reject the null hypothesis if the P value is less than the alpha value and otherwise retain it.

Standard Error of Difference between Means

We do not have the exact knowledge of the standard deviation of the distribution of mean differences, so we can estimate it on the basis of the two samples we actually draw.

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}\right)\left(\frac{N_1 + N_2}{N_1 N_2}\right)}$$

Where,

$$s_1^2 = \frac{\sum X_1^2}{N_1} - \bar{X}_1^2$$

$$s_2^2 = \frac{\sum X_2^2}{N_2} - \bar{X}_2^2$$

Testing the Difference between Means

If we use the estimate for the standard error for the difference between means, we cannot use the *z* score but have to calculate the *t* statistic instead.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$$

Adjustment for Unequal Variances

The formula for estimating the standard error of the difference between means pools combines variances that are the same for the two groups, that is $\sigma_1^2 = \sigma_2^2$. In instances when either of the sample variances is more than double the other, the standard error is calculated without pooling by the following formula:

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}}$$

Comparing Dependent Samples
A variation of the two-mean comparison based on two independently drawn samples is the before-after, or panel, design, which is the case of a single sample measured at two different points in time, where $D = X_1 - X_2$. The standard deviation of the distribution of the before-after difference scores is given by s_D and the standard error of the difference between means is given by $s_{\bar{D}}$. The t statistic is measured using these.

$$s_D = \sqrt{\frac{\Sigma D^2}{N} - (\bar{X}_1 - \bar{X}_2)^2}$$

$$s_{\bar{D}} = \frac{s_D}{\sqrt{N-1}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{D}}}$$

Two-Sample Test of Proportions
The logic of testing the difference between two proportions is the same as for testing the difference between means. Our statistic (z) is the difference between sample statistics divided by the standard error of the difference and the standard error of the differences in proportions is given by $s_{P_1-P_2}$ and $P*$ is the combined sample proportion.

$$z = \frac{P_1 - P_2}{s_{P_1-P_2}}$$

$$s_{P_1-P_2} = P^*(1 - P^*)\left(\frac{N_1 + N_2}{N_1 N_2}\right)$$

$$P^* = \frac{N_1 P_1 + N_2 P_2}{N_1 + N_2}$$

Requirements for Testing the Difference between Means

1. Comparison between two means – z score and t ratio are employed
2. Interval data – z score and t ratio cannot be used for nominal data
3. Random sampling – this assumption is critical for the usage of z or t scores
4. A normal distribution – if the researcher has reason to believe that the normality assumption is false, the z or the t scores should not be used
5. Equal variances – the t ratio of independent samples assumes population variances are equal, even though sample variances may differ as a result of sampling

Summary (page 180)

In Chapter 6, we saw how a population mean or proportion can be estimated from the information we obtain from a single sample. In this chapter, we turned our attention to the task of testing hypotheses about differences between sample means or proportions (see figure 7.10). Studying one sample at a time, we previously focused on characteristics of the sampling distribution of means. As we saw in this chapter, there is also a probability distribution for comparing mean differences. As a result of sampling error, the sampling distribution of differences between means consists of a large number of differences between means, randomly selected, that approximates a normal curve whose mean (mean of differences between means) is zero.

With the aid of this sampling distribution and the standard error of the difference between means (our estimate of the standard deviation of the sampling distribution based on the two samples we actually draw), we were able to make a probability statement about the occurrence of a difference between means. We asked, "What is the probability that the sample mean difference we obtain in our study could have happened strictly on the basis of sampling error?" If the difference falls close to the center of the sampling distribution (that is, close to a mean difference of zero), we retain the null hypothesis. Our result is treated as merely a product of sampling error. If, however, our mean difference is so large that it falls a considerable distance from the sampling distribution's mean of zero, we instead reject the null hypothesis and accept the idea that we have found a true population difference. Our mean difference is too large to be explained by sampling error.

But at what point do we retain or reject the null hypothesis? To establish whether our obtained sample difference reflects a real population difference (that is, constitutes a statistically significant difference), we need a consistent cut=off point for deciding when to retain or reject the null hypothesis. For this purpose, it is conventional to use the .05 level of significance (also the stricter .01 level of significance). That is, we are willing to reject the null hypothesis if our obtained sample difference occurs by chance less than 5 times out of 100 (1 time per 100 at the .01 level). We determine whether or not the null hypothesis is plausible by calculating degrees of freedom and a significance test known as the *t* ratio. We then compare our obtained *t* ratio against the *t* ratios at varying degrees of freedom given in Table C in Appendix C. Using the same logic, there is also a test of significance for differences between proportions. In either case, the researcher must choose between a one or two-tailed test of significance, depending on whether he or she can anticipate in advance the direction of the difference.

Key Terms

Null hypothesis
Research hypothesis
Sampling distribution of differences
Between means
Level of significance
.05 level of significance
.01 level of significance

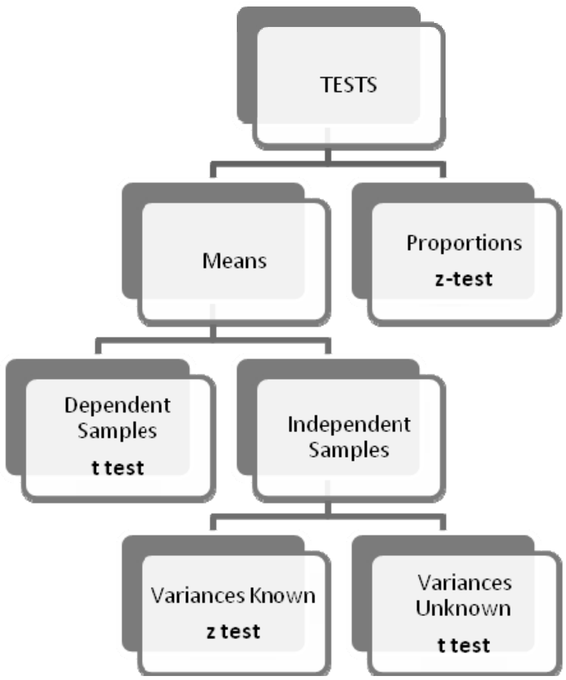
t ratio
Type I error
Type II error
Standard error of the difference between
Means
Statistically significant difference
Two-tailed test

Lecture Launchers and/or Discussion Topics

It is important to remember that the formulas often appear to be scary to students at this stage because the number of superscripts and subscripts is more than in previous chapters. It would be helpful to break down each formula into parts and then explain it step-wise. It would also help students to recognize broad steps of hypothesis testing first and then start using the appropriate formulas where necessary. This will create a "plan" which the student can follow for all hypothesis-testing scenarios.

Demonstrations and/or Activities

Use the textbook demonstrations or a demonstration of your own to show the students all the steps leading up to a result of hypothesis testing. It would help to show examples for options where retention of null hypothesis is required, and another example where the rejection of null hypothesis occurs.



OVERHEAD XIII
TESTING HYPOTHESES

From Chapter 7, the following overhead compares the .05 and .01 levels of significance and Type I and Type II errors. The features that you might point out are as follows:

- Based on its larger *z* score, the .01 level of significance is more difficult to achieve than the .05 level of significance.

- To reject the null hypothesis at the .05 level, you must be at least 1.96 standard deviations from the mean. To reject at the .01 level, you must be at least 2.58 standard deviations from the mean.

- You are more likely to commit a Type I error (rejecting a true null hypothesis) by using the .05 level of significance than by using the .01 level of significance.

- You are more likely to commit a Type II error (accepting a false null hypothesis) using the .01 level of significance than by using the .05 level of significance.