CHAPTER 18 Appendix

Comparing Two Means

Comparing two groups or two treatments is the "heart" of comparative statistics. The ideas in this chapter require two independent (separate) samples. If the two samples are related (paired), use the methods described in Chapter 17.

Some key points to think about when deciding whether the samples are independent:

1. Are the sample sizes different? If so, they are probably independent. In practice, however, individuals often drop out of a study and will not have an "after" measurement. In that case, the "before" measurements for those individuals should be deleted from further consideration of the treatment.
2. If you are given only summary statistics for two samples, they should be independent. With only this information, conducting inference on paired data is impossible.
3. If the data values in one column were scrambled in some way, would that rearrangement alter the data structure? That is, for the data to make sense, should you switch the values in both columns at the same time? If so, these data are paired.

Thinking critically about your data, how the data were collected and structured is more important than anything else.

Excel

Using Excel to compare two means requires raw data—not simply the sample means, standard deviations, and sample sizes. If you only have data summaries, you could calculate the test statistic "by hand" then use the Excel function T.DIST or T.DIST.2T to find a p-value for a test.

1. **Data Analysis ➔ t-test: Two Sample Assuming Unequal Variances**
   Alternatively (not recommended):
   ➔ t-test: Two Sample Assuming Equal Variances
2. Select or enter the range of cells for the first variable.
3. Select or enter the range of cells for the second variable.
4. Enter a hypothesized mean difference, usually 0.
5. **OK**

Excel does not automatically calculate the confidence interval for the difference in means.

For more help and examples, see the Excel Video Technology Manuals *Two Sample Mean Inference t—with data* and *Two Sample Mean Inference t—summarized data.*

This procedure requires that all response variable values be in one column and a "sample identifier" (factor) be stored in a second column.

1. **Analyze ➔ Fit Y by X**
2. Select the numeric response column and click **Y, Response** to enter this variable in the *y* role.
3. Select the categorical factor (sample identifier variable) and click **X, Factor** to enter this variable in the *x* role.

4. **OK**
5. Click the Red Triangle and select **t-Test** for a test assuming unequal variances or **Means/ANOVA/Pooled t** for a *t* test assuming equal variances (not recommended).

JMP does not calculate and display the confidence interval for the difference in means, but it does give the standard error of the difference and the degrees of freedom in its results.
For more help and examples, see the JMP Video Technology Manuals *Two Sample Mean Inference t—with data.*

Minitab

Stat ➔ Basic Statistics ➔ 2-Sample t
There are three ways to enter the data.

- If the data for the two samples are in two separate columns, choose **Each sample is in its own column**, and then select and click in the two columns of data.
- If the data for the two samples are in one column with a group label (sample ID) in a second column, choose **Both samples are in one column**, and then select and click in the samples column and sample IDs column.
- If you know the sample sizes as well as the values of the sample means and sample standard deviations, you can choose the **Summarized data** option and input those items in the boxes.

The default is not to assume equal population variances. If you wish to assume equal population variances (not recommended), select **Options**, then check the box for **Assume equal variances**.

If you wish to change the level of confidence (from the default of 95%), the hypothesized mean difference (from 0), or the form of the alternative hypothesis (the default is "not equal"), click **Options** and input your desired values. After changing the option, click **OK** to return to the main dialog.

Minitab displays a confidence interval that corresponds to the alternative hypothesis in addition to the test results.

For more help and examples, see the Minitab Video Technology Manuals *Two Sample Mean Inference t—with data*, *Two Sample Mean Inference t—summarized data*, and *Paired Samples Inference*.

Comparing Two Means: With Data

This requires all data to be in a single column, with a second column containing the group identifier.

1. **Analyze ➔ Compare Means ➔ Independent-Samples T-Test**
2. Enter the data variable name into the **Test Variables** section and the grouping variable into the **Grouping Variable** box.
3. Click **Define Groups** and supply the labels for "Sample 1" and "Sample 2."
4. The default is to provide the test statistic, *P*-value, and a 95% confidence interval. For a different confidence level, click **Options** and change to the level you want.
5. **OK**

Comparing Two Means: With Summary Statistics (requires Python Plug-in)

1. **Analyze ➔ Compare Means ➔ Summary Independent-Samples T Test**
2. Enter the summary statistics for each sample and the confidence level.
3. **OK**

For more help and examples, see the SPSS Video Technology Manuals *Two Sample Mean Inference t—with data*.

CRUNCH

1. **Statistics ➔ t ➔ Two-sample**
2. Select either **Columns** (for data in separate columns), **Grouped** (for data in one column, and a group ID in another column), or **Summarized** (to enter summary statistics).
3. Enter the columns or values necessary.
4. Select **Pooled Variance** to perform a test assuming equal population variances (not recommended).
5. Click either the **Hypothesis Test** tab or the **Confidence Interval** tab.
6. For a test, enter the mean difference under the null hypothesis (usually 0, the default) and the desired type of alternative hypothesis.
7. For a confidence interval, enter the desired confidence level.
8. **Calculate**

For more help and examples, see CrunchIt! Help Video *Two Sample Mean Inference*.

TI-83/-84

1. Enter data in L1 and L2, if you have them.
2. Press STAT ▶ ▶ to select **TEST**.
3. Select **4:2-SampTTest** or **0:2-SampTInt**.
4. Select **Data** (or **Stats** to enter summary statistics). Enter L1 and L2 (or your lists) as the data location, or type in the mean, standard deviation, and sample size for each sample.
5. For the test, enter the form of the alternative hypothesis; for the interval, enter the confidence level.
6. For the Pooled option, **No** isypical (Yes assumes equal variances).
7. **Calculate**

For more help and examples, see the TI Video Technology Manuals *Two Sample Mean Inference t—summarized data* and *Two Sample Mean Inference t—with data*.

The R command for both a test and a confidence interval is the same as for the paired test, but omitting the "paired=T" part.

```
>t.test(mydat$numvar1,mydat$numvar2,alternative=
 "two.sided", conf.level=0.95,data=mydat)
```

The other choices for the alternative parameter are "less" and "greater". Note the quotes are required.

For more help and examples, see the R Video Technology Manuals *Two Sample Mean Inference with data* and *Two Sample Mean Inference with summarized data*.