Chapter 11

Differences Between Two Populations

This chapter focuses on the comparison of two populations. This type of comparison is used to answer questions such as:

- On average, do females and males have the same body temperature?
- Do people who take one aspirin tablet every morning have lower risk of heart attack?
- Is one toothpaste generally superior to another in fighting cavities?
- On average, do graduate students indicate liking statistics more than undergraduate students?

To answer these questions, a sample is drawn from each population (e.g., males and females) and the means (or proportions or medians) of the samples are compared. If the difference between the samples is greater than expected as a result of sampling error, we conclude that the populations are distinct. As with the hypothesis testing in Chapter 10, procedures differ depending upon whether the population standard deviations are known. In this book, we consider the more likely instance, when the population standard deviations are unknown.

11.1 COMPARISON OF TWO INDEPENDENT MEANS

In comparing population means, we rarely know the population standard deviations of the two distributions. They may, however, be estimated to determine the standard error of the difference of the means. Testing procedures are slightly different based on whether or not it can be assumed that the two standard deviations are equal. SPSS performs the test for both types of conditions. This section discusses the more common situation in which the standard deviations are treated as equal.

The "bodytemp.sav" data file contains information on body temperature and pulse rate for 130 male and female adults. Suppose we wish to determine whether males and females differ, on average, in normal body temperature. The null and alternative hypotheses are H_0: $\mu_{female} = \mu_{male}$ and H_1: $\mu_{female} \neq \mu_{male}$. We shall test the hypothesis at the 5% level.

We can now use SPSS to conduct the test as follows. After opening the data file:

1. Click on **Analyze** from the menu bar.
2. Click on **Compare Means** from the pull-down menu.
3. Click on **Independent Samples T-Test** from the pull-down menu to open the Independent-Samples T Test dialog box (Fig. 11.1).
4. Click on and move the "temp" variable the Test Variable(s) box using the **upper right arrow button**.
5. Click on and move the "sex" variable to the Grouping Variable box using the **lower right arrow button**.
6. Notice that two question marks appear in parentheses after the variable "sex." This signifies that you need to indicate the two values of the class variable for which you wish to calculate mean differences. To do so, click on **Define Groups** to open the Define Groups dialog box (see Fig. 11.2).
7. In our example, females are coded 0 and males are coded 1. Therefore, enter these numbers in the Group 1 and Group 2 boxes. (The cut point option is used if there are more than two values of the grouping variable.)
8. Click on **Continue** to close the dialog box.
9. Click on **OK** to run the procedure.

The output is displayed in Figure 11.3. The upper portion of the listing displays summary information (*n*'s, means, standard deviations, and standard errors) for each of the samples. In this case, females had an average of 98.394°F and males had an average temperature of 98.105°F. The difference is 98.394 – 98.105 = .289°F. The figure is listed in the lower table in the "Mean Difference" column.

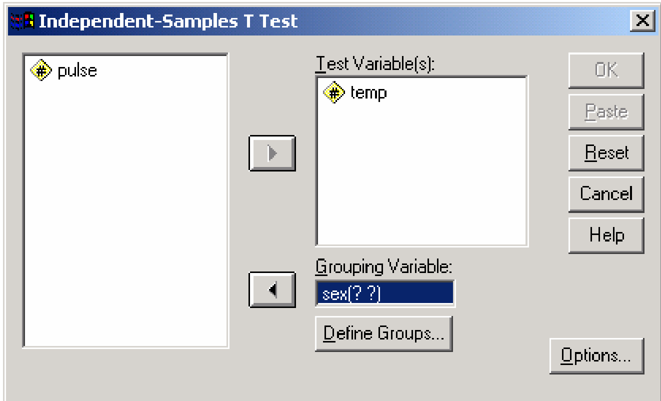

Figure 11.1 Independent-Samples T Test Dialog Box

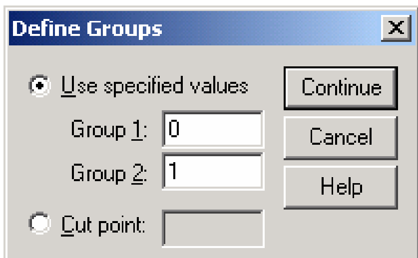

Figure 11.2 Define Groups Dialog Box

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
body temperature (degrees Fahrenheit)	female	65	98.394	.7435	.0922
	male	65	98.105	.6988	.0867

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
body temperature (degrees Fahrenheit)	Equal variances assumed	.061	.805	2.285	128	.024	.289	.1266	.0388	.5396
	Equal variances not assumed			2.285	127.510	.024	.289	.1266	.0388	.5396

Figure 11.3 T Test for Independent Samples

This table also displays two different t-statistics, one based on the assumption of equal variances, the other assuming unequal variances. We will only consider the equal variances case. The test statistic is $t = 2.285$.

The t-statistic is compared with significance points from the t-distribution with 130 – 2 = 128 degrees of freedom. This is done by SPSS, resulting in the printed P value (in the Sig. (2-tailed) column). Since $P = .024$ is less than .05, the null hypothesis is rejected, and we conclude that, on average, females do not have a higher body temperature than do males.

The output also includes a 95% confidence interval for the mean difference. That is, the difference between average temperature for men and women is between 0.04 and 0.54 degrees with 95% confidence. Because 0 (representing no difference in average temperature between women and men) is not in the range, the results of the significance test are confirmed.

One-Tailed Tests

The SPSS procedure for conducting a one-tailed test is the same as that for a two-tailed test; it differs only in how the P value is used. Because the reported P value is for a two-tailed test, we must compare $P/2$ to α, and also verify that the sample means differ in the direction supported by the alternative hypothesis. In the example, .012 < .05, and the sample mean for males is less than the sample mean for females. Thus, we would reject H_0: $\mu_{female} \leq \mu_{male}$ in favor of H_1: $\mu_{female} > \mu_{male}$ at the 5% level of significance.

Chapter Exercises

11.1 Use the "enroll.sav" data file and SPSS to test whether there is a significant difference in the racial disproportion index between districts with high and low percentages of students paying full price for lunch, as follows:

a. Recode the "pct_lnch" variable into a dichotomous variable, with 51% as the split point. (That is, values less than 51% will constitute the "low" group, and all other values the "high" group.)

b. Would you perform a one- or two-tailed test? Why?

c. Based on your response to part (b), state the null and alternative hypothesis.

d. Use SPSS to conduct the test. State the value of the test statistic and the P value. (Assume equal variances.) Is H_0 rejected if $\alpha = .05$? If $\alpha = .01$?

11.2 Use the "football.sav" data file to explore the relationship between the type of team (home or away) winning games and number of points by which the game is won.

a. State the null and alternative hypotheses for testing whether games that are won by the home team are won by, on average, more points than games that are won by the visiting team.

b. Assuming equal variances, use SPSS to conduct the test. What are your conclusions using $\alpha = .01$? Give the P value for the test.

11.3 The "cars.sav" data file contains information on cars collected from a university parking lot. The information collected was: age, color and owner (faculty/staff or student).

a. State the null and alternative hypotheses for testing whether students drive cars that are, on average, the same age as faculty/staff.

b. Assuming equal variances, use SPSS to conduct the test. What are your conclusions using $\alpha = .05$? Give the P value for the test.