5 Steps in Testing an Hypothesis

1. Verify that assumptions are met
   (random sample, normal distri., level of measurement)
2. State research and null hypotheses and alpha
3. Select sampling distribution and test statistic to be used (Z or t statistic)
   (use Z if have population standard deviation, use t if have only sample SD)
4. Compute test statistic
5. Make a decision and interpret results

Computing the t statistic.

The t statistic is used in two different cases. It is calculated differently depending on which case you are interested in. These are:

1. Comparing a group to a whole population and

2. Comparing two groups to one another

Comparing a group to a population using the t statistic. The formula for the t statistic when <u>comparing a group to the population</u>

$$t = \frac{\overline{Y} - u_y}{\frac{S_y}{\sqrt{N}}} \quad \textbf{or} \quad \frac{\text{Sample Statistic – Population Parameter}}{\frac{\text{Sample SD}}{\sqrt{N}}}$$

The t formula is identical to the Z formula except the t formula uses the sample SD when calculating the standard error and the Z formula uses the population standard deviation when calculating the SE.

Steps for interpreting the t statistic

Unfortunately, locating where the t statistic falls on the normal curve is not as easy as when using the Z statistic (that is, unlike Z values, you cannot use the normal curve when using t values).

Once the t statistic is calculated it is compared to the t value needed to reject the null hypothesis. The t value needed can be found on a t distribution table (page 484-5 in textbook) and will vary depending on whether the researcher has chosen an alpha of .05, .01, etc.

How to use the t distribution table to determine significance

(1) Determine the **degrees of freedom** your sample provides (this is typically: N-1) and then locate the DF on the t-distribution table (table is on page 484-5) .

(2) Find on the table: the **alpha** which you selected at the start of the statistical analysis (an alpha of .05 and a two-tailed test are typically used by researchers)

(3) **Find the intersecting point** where the DF and the alpha cross. At the intersecting point you will find the t value needed to reject the null hypothesis.

---

t distribution table

Table 13.2 **Values of the *t* Distribution**

	Level of Significance for One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
10	1.372	1.812	2.228	2.764	3.169	4.587
15	1.341	1.753	2.131	2.602	2.947	4.073
20	1.325	1.725	2.086	2.528	2.845	3.850
25	1.316	1.708	2.060	2.485	2.787	3.725
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Source: Abridged from R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Table 111. Copyright © R. A. Fisher and F. Yates 1963. Reprinted by permission of Pearson Education Limited.

(4) **Compare** the t value calculated from the data to the t value identified on the t distribution table. If the calculated t value is larger than the t value found in the table, then the null hypothesis can be rejected and the difference between the group and the population can be considered statistically significant (but not necessarily "substantively" significant).

In-Class Exercise:

The average number of caviar eggs laid in a season by a single sturgeon fish is roughly 500,000. We have a sample of 80 sturgeon fish from lake Michigan that lay on average roughly 550,000 eggs in a season with a standard deviation of 200,000. We want to know whether our fish are different from the average. Begin by calculating the t statistic.

$$t = \frac{\bar{Y} - u_y}{\frac{S_y}{\sqrt{N}}} \quad \textbf{or} \quad \frac{\text{Group Mean – Population Mean}}{\frac{\text{Sample SD}}{\sqrt{N}}}$$

Why are we calculating the t statistic instead of the Z statistic?

What alpha (level of confidence) would you like to use and why?

The population mean is 500,000 and the sample mean is 550,000 with a sample standard deviation of 200,000 and sample size of 80.

$$\frac{550{,}000\text{-}500{,}000}{\frac{200{,}000}{\sqrt{80}}} = \frac{50{,}000}{22{,}371} = 2.24$$

Finding the t statistic in the t distribution table

Our **degrees of freedom** for this example is N – 1 or 79 and our t statistic is 2.24 (the larger the t statistic the more likely it will be significant).

On page 484-5 of your book we can find the **t distribution table**. It displays the **degrees of freedom** for 60 and for 120. Since ours is 79 it is less than 120. Therefore, to be conservative we will use 60 DF.

We won't assume a **one-tailed test** since there is no existing knowledge to support the hypothesis that sturgeon fish in Lake Michigan lay more eggs than the average sturgeon fish.

t distribution table

Table 13.2 **Values of the *t* Distribution**

	Level of Significance for One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
10	1.372	1.812	2.228	2.764	3.169	4.587
15	1.341	1.753	2.131	2.602	2.947	4.073
20	1.325	1.725	2.086	2.528	2.845	3.850
25	1.316	1.708	2.060	2.485	2.787	3.725
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Source: Abridged from R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Table 111.

Determining Statistical Significance

Since our t statistic is 2.24 we can conclude statistical significance at the .05 level.

Would our findings be significant if we had chosen an alpha of .01?

Comparing the Sample Statistics of Two Groups

(Presented above is a comparison of a group's sample statistic to a population parameter)

Example for comparing two groups:

Comparing the mean salary of new sociology professors (group 1) to the mean salary for new engineering professors (group 2). Previously we were comparing the group statistic (such as the mean salary of sociology professors) to the population parameter (such as the mean salary of the whole U.S. population).

Steps for comparing the sample statistics of two groups are the same as that for comparing a sample statistic to the population parameter with three exceptions:

(1) the formula for calculating the t statistic is different

(2) calculating the degrees of freedom is different, and

(3) must now determine whether the two groups have equal or unequal variances. (If the Levene's test is significant then their variances are unequal)

Formula for calculating the t statistic for comparing two groups:

(You will not be required to calculate this comparison because the formula for determining the Standard Error of the Differences Between the Means is complex. We will use the computer to do the comparison and then we will determine whether the null hypothesis can be rejected.)

$$t = \frac{\text{Mean of } 1^{st} \text{ Group} - \text{Mean of } 2^{nd} \text{ Group}}{\text{Standard Error of the Differences Between the Means}}$$

Step 4: Calculation of the t Statistic

$$t = \frac{\text{Mean of } 1^{st} \text{ Group} - \text{Mean of } 2^{nd} \text{ Group}}{\textbf{Standard Error} \text{ of the Differences Between the Means}}$$

$$\textbf{SE} = \sqrt{\frac{(N_1-1)\, SD^2 y_1 - (N_2-1)\, SD^2 y_2}{(N_1 + N_2) - 2}} \sqrt{\frac{N_1+N_2}{N_1 N_2}}$$

The sample findings were:

Males: mean = grade 4.39; SD=1.70; N=380

Females: mean = grade 4.61; SD=1.62; N=394

IMPORTANT You will not be asked to calculate the t statistic for comparing two groups. However, you are expected to use the computer to calculate the t statistic.

Example: Comparing Two Sample Means (male and female job burnout)

In SPSS:

1. Analyze
2. compare means
3. independent sample t test
4. move V101 (sex) to "group variable" box
5. click "define group"
6. in group 1 put "1" (female) and in group 2 put "2" (male)
7. click continue
8. move "burnout", V102 (age), V100 (education), V114 (# residents assigned) to "test variables" box and then click Okay

凩訝

(see you later)