# Quantitative Methods (GEO 441)
## Hypothesis Testing
*From Zar, 1984*

Statistical procedures for addressing research questions involves formulating a concise statement of the hypothesis to be tested. The hypothesis to be tested is referred to as the *null hypothesis* (abbreviated $\boldsymbol{H_0}$) because it is a statement of *no difference.* Hypothesis testing starts with the assumption that the null hypothesis is true… that there is/are no difference(s).

Along with the null hypothesis we must also state an alternate hypothesis (abbreviated $\boldsymbol{H_a}$). The alternate hypothesis is a statement that a difference exists. If a null hypothesis is rejected, then we tentatively accept the alternate hypothesis and conclude that there is a difference.

Why is the null hypothesis the one that is tested? Think about it this way: we only have to find one instance in which the null hypothesis is not true (false) in order to be able to reject it. Conversely, we would have to continue to test the alternate hypothesis in order to be able to accept it. In other words we would have to test all possibilities since the alternate hypothesis can only be proven correct if *all* possible tests are performed.

The moral of the story: it is easier to prove a null hypothesis incorrect than to prove an alternate hypothesis correct.

Example Hypotheses:

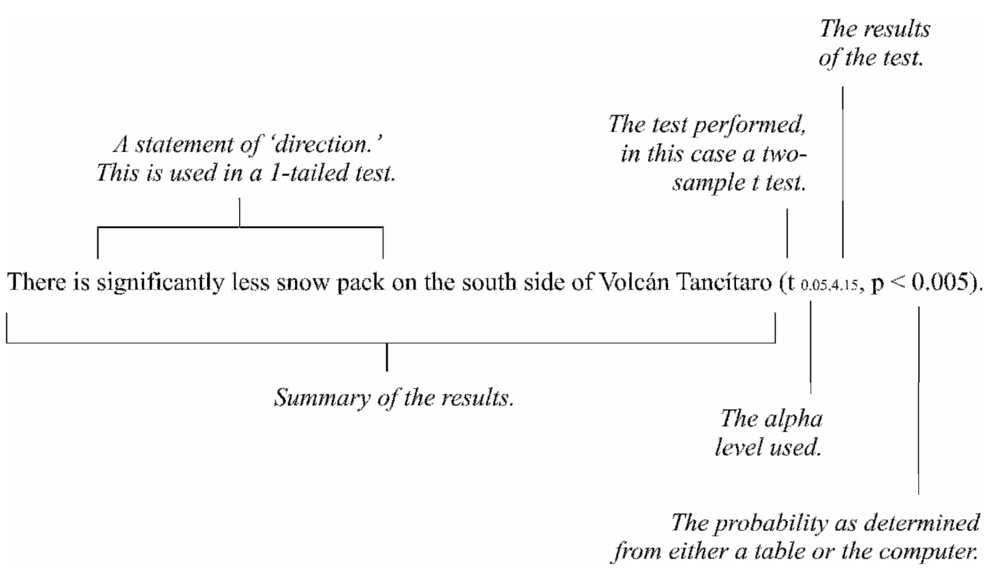
$H_0$ : There is no difference in eye color between the two groups.

$H_a$ : There is a difference in eye color between the two groups.

**IMPORTANT: In ALL cases, if the calculated value is GREATER than the critical value (from the table) then *reject* $H_0$. If the calculated value is LESS than the critical value then *accept* $H_0$.**

## Components of a Summary Statement

The results of any statistical test (e.g. one where you are testing a null hypothesis) must be stated in a concise summary statement. This statement should include a summary of the findings, the test that was performed, the alpha level used, the statistical results, and the probability. An example of a good summary statement is below:



Example of a Statistical Test and Reporting its Results
Comparing Housing Values between York and Lancaster, PA

Earlier research suggested that the average house value in York was lower than that in Lancaster. Data for the average house value for each block group in downtown York and Lancaster were gathered from the census. A two-sample t test was performed to determine whether housing values were lower in York than in Lancaster. Since we have *a priori* (prior) knowledge of the direction of the difference (e.g. housing values are *less* in York) we would use a one-tailed test.

$H_0$ : Housing values in York are not significantly less than those in Lancaster.
$H_a$ : Housing values in York are significantly less than those in Lancaster.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \quad where \quad s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \qquad s_p^2 = \frac{SS_1 + SS_2}{v_1 + v_2}$$

$\alpha = 0.05$
$n_1 = 7 \quad n_2 = 6$
$df = (n_1 + n_2 - 2) = (7 + 6 - 2) = 11 \qquad v_1 = 7 - 1 = 6 \qquad v_2 = 6 - 1 = 5$

**<u>Housing Value ($)</u>**

| York | Lancaster |
|---|---|
| 25368 | 49465 |
| 37045 | 37500 |
| 47500 | 53055 |
| 26785 | 48125 |
| 41493 | 45000 |
| 32864 | 52946 |
| 26140 | |

$$\bar{X}_{York} = \frac{25368+37045+47500+26785+41493+32864+26140}{7} 33885$$

$$\bar{X}_{Lancaster} = \frac{49465+37500+53055+48125+45000+52946}{6} = 47682$$

$$SS_{York} = (25368-33885)^2 + (37045-33885)^2 + (47500-33885)^2 + (26785-33885)^2 + (41493-33885)^2 + (32846-33885)^2 + (26140-33885)^2 = 437212244$$

$$SS_{Lancaster} = (49465-47682)^2 + (37500-47682)^2 + (53055-47682)^2 + (48125-47682)^2 + (45000-47682)^2 + (52946-47682)^2 = 2444393535$$

$$s_p^2 = \frac{437212244 + 2444393535}{6+5} = \frac{2881605779}{11} = 261964161.7$$

$$s_{X_1 - X_2} = \sqrt{\frac{261964161.7}{7} + \frac{261964161.7}{6}} = \sqrt{37423451.7 + 43660693.6} = \sqrt{81084145.3} = 9004.7$$

$$t = \frac{33885 - 47682}{9004.7} = -1.532 \qquad (\textit{ignore the sign})$$

$$t_{Critical} = 1.796 \qquad Since\ 1.532 < 1.796\ accept\ H_0$$

The housing values in York were not significantly less than the housing values in Lancaster ($t_{0.05,1.532}$, $0.10 > p > 0.05$).