## Objectives:

- To develop an intuitive understanding of the reasoning process used in ***tests of significance***
- To become familiar with the formal structure of tests of significance and to learn to translate appropriate questions into that structure.
- To discover how to perform calculations relevant to a test of significance concerning a population proportion.
- To acquire the abilities to interpret and explain the results of tests of significance.
- To explore how the statistical significance of a sample result is related to the sample size of the study.

***Statistically significant:*** A result that is different enough from the expected that it is doubtful that it occurred by chance. For example, if a coin cam up heads 70% of the time in 100 flips then you may be suspicious that the coin is bias, this result would be statistically significant – since we doubt this many heads would occur by chance.

***Test of significance***: A test performed on a set of data to check to see if the difference between the expected and what occurred is statistically significant.

***Null Hypothesis:*** What we expect, if there is no bias or influence by some other source, about a set of data, in the above example we expect heads to occur 50% of the time. Therefore when flipping a coin 100 times we expect the number of heads to occur to be 50. Symbolically we say

$$H_0 : \pi = \pi_o$$

θ is the population proportion of interest, $\theta_o$ is the population proportion we expect. The null hypothesis is the statement of "*no effect*" or "*no difference"* . The significance test is designed to assess the strength of evidence against the *null hypothesis.*

***Alternative Hypothesis:*** What we expect or hope to be true. There is some influence by a source and it makes a difference (statistically significant

Can take one of the three following forms:

(a) $H_a : \pi < \pi_o$ or (b) $H_a : \pi > \pi_o$ or (c) $H_a : \pi \neq \pi_o$

Again, θ is the population proportion of interest. The alternative hypothesis states what researchers hope or suspect about the data in question. There is some influence that makes the results different than the normally expected outcome ($\theta_o$)
The first two forms shown are ***One-sided*** the third one is ***two-sided***.

***Test statistic:*** A value computed to standardize the difference between the null hypothesis and the we obtained (p-hat). This is used to assess the evidence against the null hypothesis.

$$z = \frac{\hat{p} - \pi_o}{\sqrt{\frac{\pi_o(1-\pi_o)}{n}}}$$

For the null hypothesis, we assume the null hypothesis is the true (that is $\theta_o$ is the parameter)

***p-value:*** the probability of obtaining a value at least as extreme as the test statistic's value assuming that the Null Hypotheses ($\theta_o$) is true. "Extreme" means in the direction of the alterative hypothesis. Therefore the p-value takes one of three forms corresponding with the appropriate $H_a$

| |
|---|
| $(a)\ \Pr(Z \le z)$ area below the z - score |
| $(b)\ \Pr(Z \ge z)$ area above the z - score |
| $(c)\ 2\Pr(Z \ge \lvert z \rvert)$ area more extreme than the z - score in both directions |

We judge the strength of evidence that the data provides against the null hypothesis by examining the p-value. Smaller the p-values suggest stronger evidence against the null hypothesis (the less likely a value that extreme occurred by chance alone.) and thus stronger evidence favoring the alternative hypothesis.

Typically we use the following criteria to judge the strength of evidence:

| | |
|---|---|
| p-value > .1 | Little or no evidence against $H_o$ |
| .05 < p-value ≤ .10 | Some evidence against $H_o$ |
| .01 < p-value ≤ .05 | Moderate evidence against $H_o$ |
| .001 < p-value ≤ .01 | Strong evidence against $H_o$ |
| p-value ≤ .001 | Very strong evidence against $H_o$ |

***Significance level:*** denoted as α, is an arbitrary value set before any research begins that gives the researcher a "*cut-off"* point for the p-value. The smaller the significance level the more evidence you need to reject the null hypothesis.

***Typical values for α include:***

***α = .10, α = .05 and α = .01***

(In education you can see an α = .2, because it is very difficult to get good data. it isn't "life critical" to be wrong, and researchers are trying to obtain a result and an α of .1 will hardly give results that support the $H_a$, which the point of their research.)

***Test Decision:*** The decision we make on the $H_o$ based on the p-value obtained.

If the p-value is less than or equal to the pre-chosen α the test decision is to ***reject $H_o$.***
If the p-value is at least the pre-chosen α then the test decision is to ***fail to reject $H_o$.***

***IMPORTANT: Not rejecting $H_o$ does not mean that we accept $H_o$ or that we affirm its truth, it only means that we do not have enough evidence to contradict it.***

*Symbolically*

*If p-value ≤ α then reject $H_o$*
*If p-value > α then fail to reject $H_o$*

It is also common to state that the result is ***statistically significant at the α level*** if our p-value is less than or equal the predetermined α

***Technical Conditions: Similar to what we have seen for proportions previously***

- ***Data are a SRS of the population in question***
- ***CLT rule of thumb $n(1-\pi_o) \ge 10$ and $n(\pi_o) \ge 10$***