Hypothesis Testing I: *Getting started*

Null and Alternative Hypotheses

- H_0: the *Null* Hypothesis (the hypothesis we are testing)
- H_1: the *Alternative* Hypothesis, the alternative to H_0
- Overwhelmingly, the Null Hypothesis we will be focusing on is:
  - **H_0: The true parameter value is 0**

Two Types of Error

- **Type I – False Rejection**:
  - Rejecting the Null, H_0, (and accepting the Alternative, H_1) when H_0 is true
- **Type II – False Acceptance**:
  - Accepting the Null, H_0, (and rejecting the Alternative, H_1) when H_0 is false

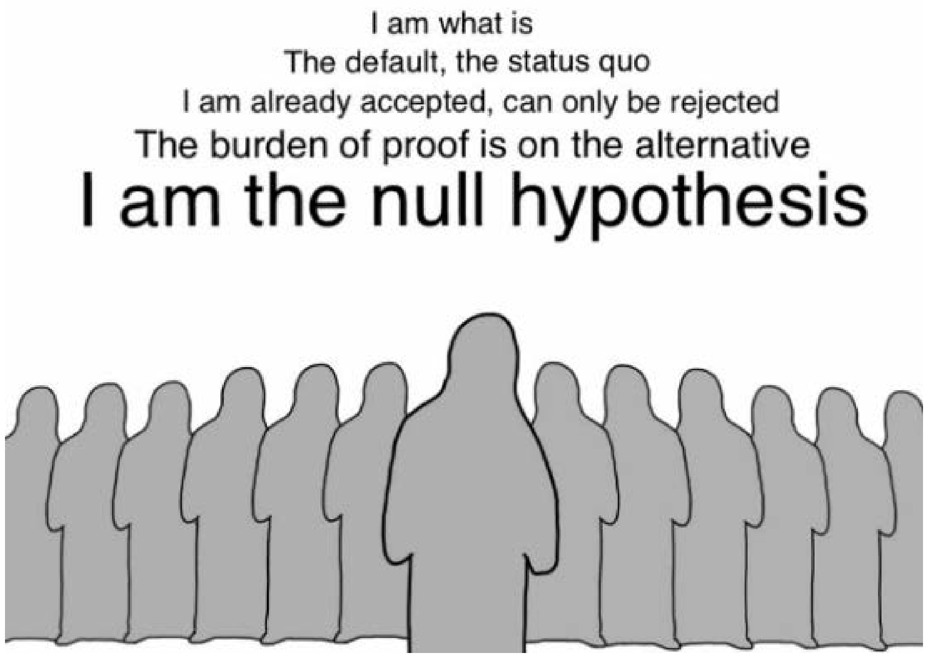
Focus on Type I Error: *False Rejection*

- We generally focus on Type I error and *protect* the Null hypothesis… only rejecting the Null hypothesis in the face of *overwhelming* evidence to the contrary… where, for example, the probability of being wrong (incorrectly rejecting the Null) is very small… $\leq$ 10%, … $\leq$ 5%, or … $\leq$ 1%, or… .
- So while mistakes may happen, their probability will be small small small.

Significance levels (α):

- We call these probabilities *significance levels*, and typically denote them with α. They are the maximum acceptable probability of a Type I error = P(Reject H_0 | H_0 is true)
- Where do significance levels come from?

  We make them up!



Estimating the Mean of the Distribution: *Reprise*

- ***Random sample***: $\{Y_1, Y_2, \ldots Y_n\}$ are an *iid* random sample from Y.
- ***Sample Mean* estimator of** μ: $\bar{Y} = \frac{1}{n}\sum Y_i$ is a ***BLUE*** estimator of μ.
- ***Normal distribution***: Assume $Y \sim N(\mu, \sigma^2)$, so $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ and $\frac{\bar{Y}-\mu}{\sigma / \sqrt{n}} = \frac{\bar{Y}-\mu}{sd(\bar{Y})} \sim N(0,1)$.
- ***Unknown variance***: If $\text{var}\left(Y\right) = \sigma^2$ is unknown then need to estimate $sd\left(\bar{Y}\right) = \frac{\sigma}{\sqrt{n}}$.
- ***Unbiased estimator***: $E\left(S_{YY}\right) = \sigma^2$, so use $se\left(\bar{Y}\right) = \frac{S_Y}{\sqrt{n}}$ (***standard error*** of $\bar{Y}$) to estimate $sd\left(\bar{Y}\right)$.
- ***t statistic***: $\frac{\bar{Y}-\mu}{se\left(\bar{Y}\right)} = \frac{\bar{Y}-\mu}{S_Y / \sqrt{n}} \sim t_{n-1}$ has a (Student's) t distribution with n-1 degrees of freedom

Testing the Null Hypothesis: $H_0: \ \mu = 0$

Run the Hypothesis Test: I

Step 1: Constrict the t-statistic, the cornerstone of inference: $t\ statistic = \dfrac{\bar{Y} - \mu}{S_Y / \sqrt{n}}$.

- Note that t statistics can be positive or negative.

Step 2: Evaluate the t statistic under the Null Hypothesis ($H_0 : \mu = 0$):

- Since $\mu = 0$, the t statistic, $t\ stat = \dfrac{\bar{Y}}{S_Y / \sqrt{n}}$, has a t distribution with n-1 degrees of freedom. As before, we express this as $\dfrac{\bar{Y}}{S_Y / \sqrt{n}} \sim t_{n-1}$.

- This is the ***t-stat*** under the Null Hypothesis ($H_0 : \mu = 0$). (Notice that I did not call it a ***t statistic***… if you see or hear the term ***t stat***, unless you know otherwise, it's reasonable to make the underlying assumption that the true parameter value is zero).

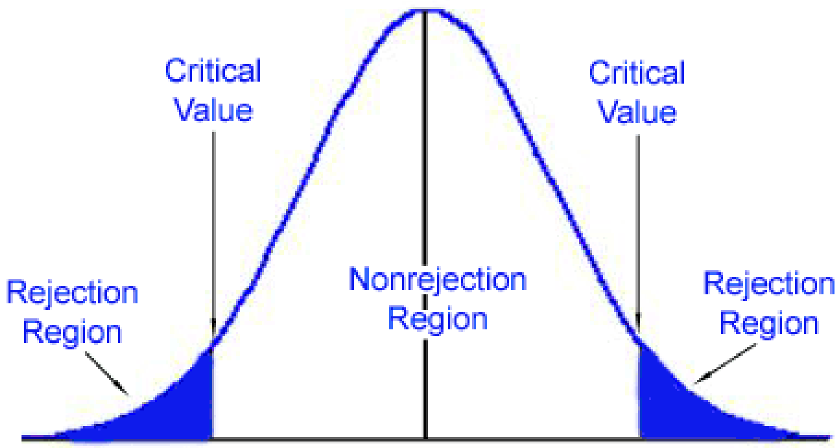
Run the Hypothesis Test: II

- Reject if $\left|\dfrac{\bar{Y}}{S_Y / \sqrt{n}}\right| > c$ … if the ***t stat*** is larger in magnitude than some *critical value* $c > 0$.

- So reject the Null Hypothesis $H_0 : \mu = 0$

  - if $\dfrac{\bar{Y}}{S_Y / \sqrt{n}} > c$ or if $\dfrac{\bar{Y}}{S_Y / \sqrt{n}} < -c$.

  - The rejection region consists of two *tails* of the t-distribution… which is why this is called a **two-tailed** test.



The Probability of a Type I Error

- The probability of Type I error is the probability of rejecting H_0 when it is true.
- For this test, that probability is $prob\left(\left|\frac{\bar{Y}}{S_Y / \sqrt{n}}\right| > c\right)$.
- But since $\frac{\bar{Y}}{S_Y / \sqrt{n}} \sim t_{n-1}$ under the Null Hypothesis,

  $prob\left(\left|\frac{\bar{Y}}{S_Y / \sqrt{n}}\right| > c\right) = prob\left(\left|t_{n-1}\right| > c\right)$ is just the probability that we're in the tails of a t distribution with n-1 degrees of freedom (below $-c$ or above $+c$)

Selecting the *Critical Values*

- Suppose that we want to select the critical value c so that the probability of falsely rejecting the Null hypothesis is $\alpha = 5\%$ (the *significance level* of the test). Then because the t-distribution is symmetric, we want to find c^* such that:

$$prob\left(\frac{\bar{Y}}{S_Y / \sqrt{n}} > c^*\right) = prob\left(t_{n-1} > c^*\right) = .025$$

(focusing just on the upper tail probability).

- For this critical value, c^*, we reject $H_0 : \mu = 0$ if

$$\left| t\ stat \right| > c^* \text{ or, equivalently, if } \left| \bar{Y} \right| > c^* \frac{S_Y}{\sqrt{n}}.$$

- With this test, we will falsely reject H_0 5% of the time… a low Type I error rate (a small risk that we rejected the Null Hypothesis when it was true).

Critical Values for the t and Standard Normal distributions

Degrees of Freedom	***Significance Levels (two-tailed test)***			
	20%	***10%***	***5%***	***1%***
5	1.48	2.02	2.57	4.03
10	1.37	1.81	2.23	3.17
15	1.34	1.75	2.13	2.95
20	1.33	1.72	2.09	2.85
25	1.32	1.71	2.06	2.79
30	1.31	1.7	2.04	2.75
35	1.31	1.69	2.03	2.72
40	1.3	1.68	2.02	2.7
45	1.3	1.68	2.01	2.69
50	1.3	1.68	2.01	2.68
N(0,1)	1.28	1.64	1.96	2.58

The Test & Statistical Significance

The Test: ***Reject the Null Hypothesis*** $H_0 : \mu = 0$ ***if*** …

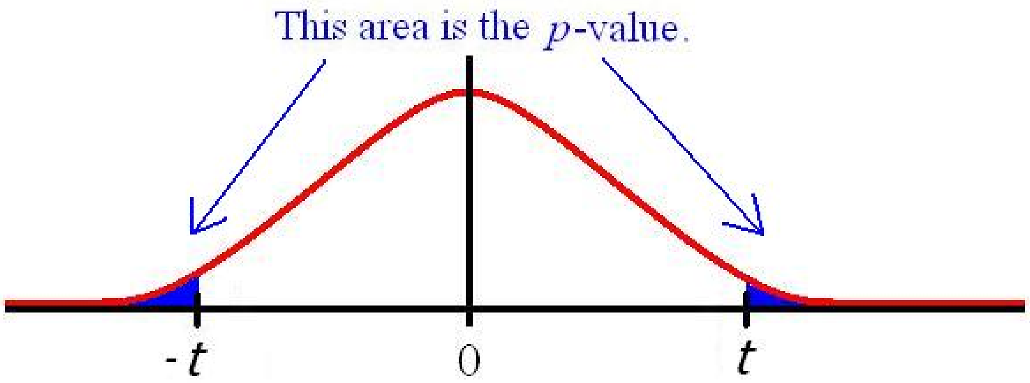
- the observed sample mean is at least c^* Standard Errors away from 0, or put differently,
- if the t-stat is larger in magnitude than c^*, where the particular value of c^* reflects the significance level of the test, α , and the degrees of freedom.

Statistical Significance

- If we can reject the Null Hypothesis $H_0 : \mu = 0$ at, say, the 5% significance level, then we say that the estimate is ***statistically significant*** at the 5% level (using a two-tailed test).
- It makes no sense to talk about statistical significance without referencing the significance level.
- Common significance levels are 10%, 5%, 1% and .1%.
- Where do they come from? ***We make them up!***
- And remember: Every estimate is statistically significant at some significance level! … but in some cases, that level is embarrassingly large!

Probability Values (p values): *The Easy Test*

- ***Probability values*** *(or p* values for short) are the maximum significance levels at which we can conduct a hypothesis test and fail to reject the Null Hypothesis…. Typically, p values are reported for two-tailed tests.
  - So if the p value is, say .02, then the null hypothesis can be rejected at significance levels above 2%, but not at smaller significance levels.
- More formally: Suppose we have a particular sample mean $\bar{y}$ and particular standard error $se = \frac{S_y}{\sqrt{n}}$. Then we can determine the p value as the probability of being that far or further away from 0 under the Null Hypothesis that $\mu = 0$: $prob\left(\left|t_{n-1}\right| > \left|t\ stat\ \right|\right) = p$.



Rejection Rule II: *p values and significance levels*

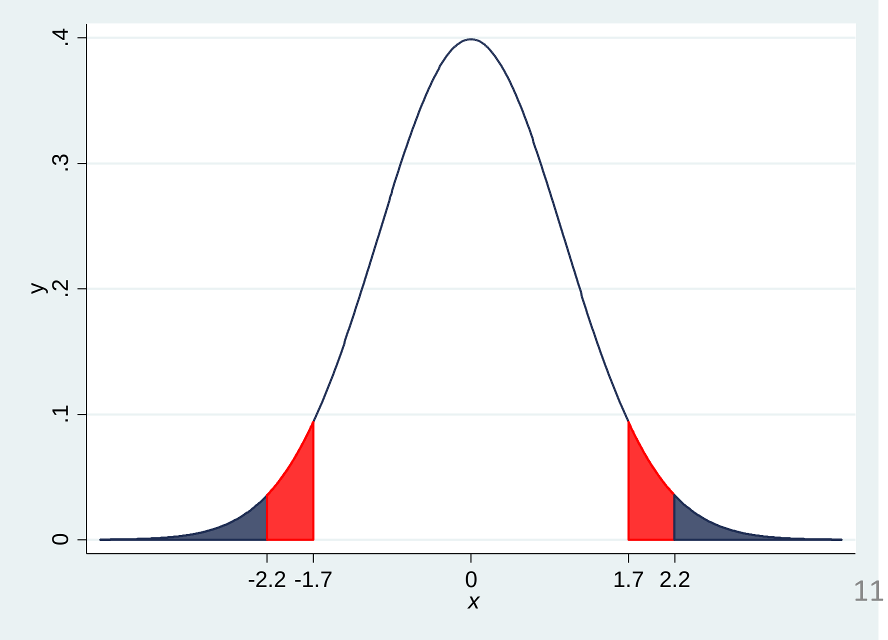
- We will reject the two-tailed Null Hypothesis that $\mu = 0$ at the significance level α if and only if $p < \alpha$ (the p-value for the given sample is smaller than the significance level for the test):
- From before, we reject the Null hypothesis if the t stat is larger than the critical value. So reject if:

$$prob\left(\left|\frac{\bar{Y}}{S_Y / \sqrt{n}}\right|\right) > c_\alpha \text{ where } c_\alpha \text{ is defined by } prob\left(\left|t_{n-1}\right| > c_\alpha\right) = \alpha \text{ .}$$

- But the t stat is larger than the critical value if and only if the p value is less than the significance level, so we can equivalently reject if $p < \alpha$.
- Since ***small p values ~ large t stats***, we reject if: $p < \alpha \Leftrightarrow \left|\frac{\bar{y}}{S_y / \sqrt{n}}\right| = \left|t\ stat\right| > c_\alpha$

Equivalence of Rejection Rules: *An Example*

- In this case $\alpha = .10,\ dofs = 30\ \ and\ \ c^* = 1.7$.
- Reject the Null Hypothesis that $\mu = 0$ if $|t\ stat| > 1.7$ or if $p\ value < .10$.
- Here, $|t\ stat| = 2.2 > 1.7$ and $p\ value < .10$, since the shaded region to the right of 2.2, $\frac{p\ value}{2}$, is less than the shaded region to the right of 1.7, $\frac{\alpha}{2}$. So: *Reject! Reject!*
- Accordingly: p values make hypothesis testing easier, since we don't need to determine critical values. If we want the Type I Error to be less than 10% in a two-tailed test, then we reject the null hypothesis only if the p value is below 10%. Done!
- You'll discover that Stata gives you the p-values in the regression output… making hypothesis testing and the determination of statistical significance a snap!



A Peek Ahead: p values & Statistical Significance

```
.   reg Brozek BMI if _n < 10

      Source |       SS           df       MS      Number of obs   =         9
-------------+----------------------------------   F(1, 7)         =      4.00
       Model |  178.534262         1  178.534262   Prob > F        =    0.0857
    Residual |  312.734627         7  44.6763753   R-squared       =    0.3634
-------------+----------------------------------   Adj R-squared   =    0.2725
       Total |  491.268889         8  61.4086111   Root MSE        =     6.684

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |   4.191929   2.096969     2.00   0.086    -.7666158    9.150474
       _cons |  -88.37095   52.05255    -1.70   0.133    -211.4557    34.71376
------------------------------------------------------------------------------
```