* **Topic:** R and Basic Statistics Review

(1) Simple Linear Regression

response/dependent variable : y
explanatory/independent variable : x
(also called predictor, covariate, etc.)

$$y = \boxed{\beta_0} + \boxed{\beta_1} x \quad \text{(deterministic)}$$

β_0: intercept (value of y when $x=0$)

β_1: slope (controls the change in y for each unit change in x)

SLR Model: $Y = \beta_0 + \beta_1 x + \underbrace{\boxed{\varepsilon}}_{\text{error term}}$ ⇒ random (Y and ε are RV's)

without the error ε, y would be the equation of a straight line!

Obs: Notation for RV = capital letters (Y, ε, X, etc.)
For instance: $X = x$ ⇒ X is a RV
x is not: it's an observation ("data point")

Objective: estimate the parameters, while minimizing the error ε.

Parameters for SLR: β_0, β_1 and σ^2 (Error Variance)

⇒ We need to estimate these parameters. A commonly used approach is the Least Squares Method (although there are many other methods available)

* why LSEs? In a practical sense, LSEs make very efficient use of data (optimal estimates under the usual assumptions of linear models).

Least Square Estimation: minimize the Sum of Squared Errors (SSE) from the regression line

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left[(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) \right]^2 = \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2,$$ where n = # of data points.

↳ this is a calculus problem!

We want
$$\min_{\beta_0, \beta_1} SSE(\beta_0, \beta_1) \Leftrightarrow \min_{\beta_0, \beta_1} \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2 \Leftrightarrow \min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \hat{y})^2$$

How to solve this optimization problem?

Answer: Use First Order Condition (F.O.C)

Take partial derivatives with respect to β_0, β_1, set the equations to 0 and solve for the parameters.

- Let's start with $\hat{\beta}_0$

  * Partial derivative: $\dfrac{\partial SSE(\beta_0,\beta_1)}{\partial \beta_0} = \dfrac{\partial \left\{ \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \right\}}{\partial \beta_0} = \sum_{i=1}^{n} [2(y_i - \beta_0 - \beta_1 x_i)(-1)]$

  * Set equation to 0: $\sum_{i=1}^{n} [-2y_i + \beta_0 + \beta_1 x_i] = 0$

  * Solve for β_0: $\iff -\cancel{2}\sum_{i=1}^{n} y_i + \cancel{2}n\beta_0 + \cancel{2}\beta_1 \sum_{i=1}^{n} x_i = 0$

$$\iff n\beta_0 = \Sigma y_i - \beta_1 \Sigma x_i$$

$$\iff \boxed{\hat{\beta}_0 = \frac{\Sigma y_i - \hat{\beta}_1 \Sigma x_i}{n}}$$

check 2^{ND} order derivative (hint: it must be >0, so that we have a minimum)

- Now let's find $\hat{\beta}_1$

  * Partial derivative: $\dfrac{\partial SSE(\beta_0,\beta_1)}{\partial \beta_1} = \dfrac{\partial \left\{ \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \right\}}{\partial \beta_1} = \sum_{i=1}^{n} [2(y_i - \beta_0 - \beta_1 x_i)(-x_i)]$

  * Set equation to 0: $\sum_{i=1}^{n} [-2y_i x_i + 2\beta_0 x_i + 2\beta_1 x_i^2] = 0$

  * Solve for β_1: $\iff -\cancel{2}\sum_{i=1}^{n} x_i y_i + \cancel{2}\beta_0 \sum_{i=1}^{n} x_i + \cancel{2}\beta_1 \sum_{i=1}^{n} (x_i)^2 = 0$

$$\iff \beta_1 \sum_{i=1}^{n} (x_i)^2 = \sum_{i=1}^{n} x_i y_i - \left(\frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n} \right) \sum_{i=1}^{n} x_i$$

$$\iff \beta_1 \sum_{i=1}^{n} (x_i)^2 = \sum_{i=1}^{n} x_i y_i - \frac{\Sigma y_i \Sigma x_i}{n} + \frac{\beta_1}{n} \left(\sum_{i=1}^{n} x_i \right)^2$$ (multiply both sides by n)

$$\iff \beta_1 \left[n \sum_{i=1}^{n} (x_i)^2 - (\Sigma x_i)^2 \right] = n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i$$

$$\iff \boxed{\hat{\beta}_1 = \frac{n \Sigma x_i y_i - \Sigma y_i \Sigma x_i}{n \Sigma (x_i)^2 - (\Sigma x_i)^2} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}}$$

Notice that the 2 forms for $\hat{\beta}_1$ above are equivalent. The last one is achieved by doing

$$\frac{\frac{\cancel{n} \Sigma x_i y_i}{\cancel{n}} - \frac{\Sigma y_i \Sigma x_i}{n}}{\frac{\cancel{n} \Sigma (x_i)^2}{\cancel{n}} - \frac{(\Sigma x_i)^2}{n}} = \frac{\Sigma x_i y_i - n\bar{y}\bar{x}}{\Sigma (x_i)^2 - n\bar{x}^2} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

(Tip: replace Σx_i by $n\bar{x}$)

Mariana da Almeida Costa

- Expected value and Variance

- For $\hat{\beta}_1$

$$E[\hat{\beta}_1] = E\left[\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\right] = \frac{\sum(x_i-\bar{x})E(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{\sum(x_i-\bar{x})\left(E(y_i)-E(\bar{y})\right)}{\sum(x_i-\bar{x})^2} \quad (1)$$

where $E[y_i] = E[\beta_0 + \beta_1 x_i + \varepsilon_i] = \beta_0 + \beta_1 x_i + E[\varepsilon_i] = \beta_0 + \beta_1 x_i$

and $E[\bar{y}] = E[\beta_0 + \beta_1 \bar{x}] = \beta_0 + \beta_1 \bar{x}$

Replacing the above in (1)

$$E[\hat{\beta}_1] = \frac{\sum(x_i-\bar{x})\beta_1(x_i-\bar{x})}{\sum(x_i-\bar{x})^2} = \beta_1$$

$\Rightarrow$ so the LSE $\hat{\beta}_1$ is an unbiased estimator.

Now the variance

$$var(\hat{\beta}_1) = var\left[\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\right] = \frac{\sum(x_i-\bar{x})^2\left(var(y_i)-var(\bar{y})\right)}{\sum(x_i-\bar{x})^4} \quad (2)$$

where $var(y_i) = var(\beta_0 + \beta_1 x_i + \varepsilon_i) = 0 + var(\varepsilon_i) = \sigma^2$

$var(\bar{y}) = var(\beta_0 + \beta_1 \bar{x}) = 0$

Replacing the above in (2)

$$var(\hat{\beta}_1) = \frac{\sum(x_i-\bar{x})^2\sigma^2}{\sum(x_i-\bar{x})^4} = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$$

- for $\hat{\beta}_0$:

$$E[\hat{\beta}_0] = E\left[\frac{\sum y_i - \hat{\beta}_1\sum x_i}{n}\right] = \frac{\sum E[y_i] - \beta_1\sum E[x_i]}{n}$$

$$= \frac{\sum(\beta_0+\beta_1 x_i) - \beta_1\sum x_i}{n} = \frac{\sum\beta_0 + \cancel{\beta_1\sum x_i} - \cancel{\beta_1\sum x_i}}{n} = \frac{n\beta_0}{n} = \beta_0 \text{ (unbiased)}$$

$$var(\hat{\beta}_0) = var\left(\frac{\sum y_i - \beta_1\sum x_i}{n}\right) = \frac{\sum \overbrace{var(y_i)}^{\sigma^2} + \sum x_i^2\, var(\beta_1)}{n^2}$$

$$= \frac{n\sigma^2}{n^2} + \underbrace{\frac{\sum x_i^2}{n^2}}_{\bar{x}^2}\left(\frac{\sigma^2}{\sum(x_i-\bar{x})^2}\right) = \frac{1}{n}\sigma^2 + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\sigma^2 = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right]\sigma^2$$

Mariana da Almeida Costa

- Main Assumptions for Simple Linear Regression:
  (and how to check them in a statistical software)

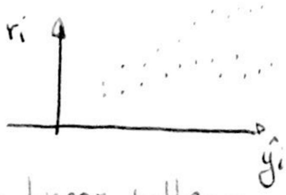
① Linearity (i.e. $E[Y|X=x] = \beta_0 + \beta_1 x$)

- check by doing a scatterplot
  (in R there are many functions, for example
  - → plot(x, y, main = "Title", xlab = "x axis", ylab = "y axis")
  - → add fit lines using abline(linearmodel, col = "red")
    where linearmodel = lm(y ~ x)
  - → use scatterplot function in the library car
    scatterplot(y ~ x | byVar, data = database,
    labels = row.names(database))
  - → scatterplot matrices: use pairs(y ~ $x_1 + x_2 + \dots + x_n$)
    or use scatterplot.matrix from car library.
  - → many other options (3D, high density data, etc.)

② Independence (uncorrelated data): residuals should be independent to prevent differences in the model for different subgroups, or in other words, the residuals ($\hat{\varepsilon}_i = y_i - \hat{y}_i$) should not be predictable (knowing one $\hat{y}_i$ does not tell you anything about the other predictions).

- check by making a Residual vs. fitted plot

→ No pattern
(uncorrelated data) ✓



→ linear pattern
(correlated) ✗

→ quadratic pattern
(correlated) ✗

(in R, the following code will plot 4 graphs to check assumptions for ②, ③ & ④)

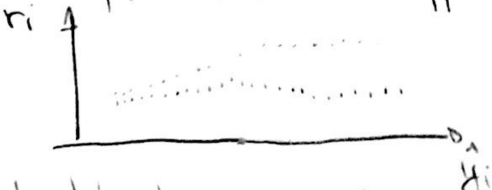
- dev.new() # opens a new window (optional)
- par(mfrow = c(2,2)) # drawing in 2 by 2 format
- plot(linearmodel)

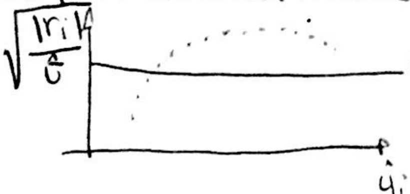
Obs: linear model = lm(y ~ x, data = data)

③ Constant variance ("homocedasticy"): variance of the errors is constant ($var(\varepsilon_i) = \sigma^2, \forall i$), and a violation of this assumption would cause the model to be more accurate (less σ^2) for some subgroups than others (larger σ^2), which in turn would make the estimates not efficient/consistent.

- check by making the Fitted vs Residuals plot (as in ②).

A violation of this assumption would appear as consistent differences in the spread of residuals as $\hat{y}_i$ increase/decrease, as in the figures below. Scale-location plots are also effective (standardized residuals vs. fitted).





→ horizontal line in scale-location plot equals 1. Residuals should be equally distributed around it.

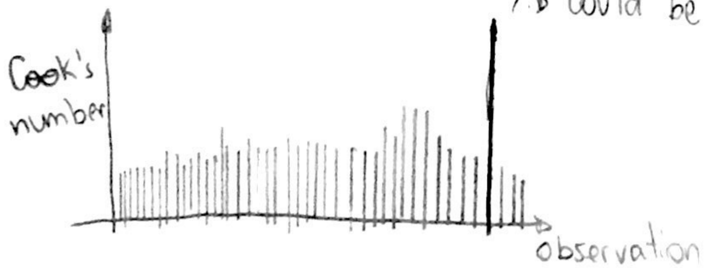
This plot should show no pattern (residuals randomly spread).

④ Normality ($\varepsilon \sim N(0, \sigma^2)$): required for inference purposes, that is, if it's desired to do some hypothesis testing or any confidence or prediction interval.

- check with Q-Q plot (also obtained from the previous formula) you should observe a straight line.

⑤ Outliers: unfortunately, linear regression is very sensitive to the presence of outliers, which may compromise the whole model. Identifying and treating them is, therefore, very important. The last of the 4 plots is the Cook's distance plot that shows how much influence a particular observation has on the y_i's.

→ could be an outlier (go check it!)



② Multiple Linear Regression

Model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$ (k covariates)

$$\underset{n\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \underset{n\times(k+1)}{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}} \underset{(k+1)\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}} + \underset{(n\times 1)}{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

In matrix form we can write as: $y = X\beta + \varepsilon$

- The idea is the same as Simple Linear Regression, however, since we're dealing with a lot of covariates, working in matrix form may be helpful. Let's re-do the LSE derivation in matrix form

Least Squares Estimation for MLR:

L1-norm $\|x\|_1 = \sum_i |x_i|$

L2-norm $\|x\|_2 = \sqrt{\sum_i x_i^2}$

Lp-norm: $\|x\|_p = \left[\sum_i |x_i|^p\right]^{1/p}$

$$\min_\beta \underbrace{\|Y - \hat{Y}\|_2^2}_{SSE} \rightarrow \|\ \|_2^2 \text{ is the L2-NORM (squared)}$$

$$\Leftrightarrow \min_\beta \|Y - X\hat{\beta}\|_2^2 = (Y - X\beta)^T(Y - X\beta) = Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta$$

$$= Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta$$

(notice that $Y^TX\beta \Leftrightarrow \beta^TX^TY$, $(Y^TX\beta)^T = \beta^TX^TY$ //)

F.O.C: $\dfrac{\partial SSE(\beta)}{\partial \beta} = 0 - 2X^TY + 2X^TX\beta$

$$\Leftrightarrow \cancel{2}X^TX\beta = \cancel{2}X^TY$$

$$\hat{\beta} = (X^TX)^{-1}X^TY \quad (LSE)$$

: assumes X^TX is invertible

$$\text{rank}(X^TX) = \begin{cases} k, & n > k \rightarrow \text{may be invert} \\ n, & n < k \rightarrow \text{not invertible} \end{cases}$$

- Expected value and variance

- $E[\hat{\beta}] = E[(X^TX)^{-1}X^Ty] = E[(X^TX)^{-1}X^T(X\beta + \varepsilon)] = E[(X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\varepsilon]$

$$= E[\beta] + (X^TX)^{-1}X^TE[\varepsilon] = \beta + 0 = \beta \text{ (unbiased!)}$$

- $\text{var}(\hat{\beta}) = \text{var}[(X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\varepsilon] = \underset{\text{constant}}{\text{var}(\beta)} + (X^TX)^{-1}X^T\text{var}(\varepsilon)X(X^TX)^{-1}$

$$= 0 + (X^TX)^{-1}X^T\sigma^2I_nX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$$ //

$\hat{\beta}$ has multinormal dist $\rightarrow \hat{\beta} \sim MN(\beta, \sigma^2(X^TX)^{-1})$

Mariana de Almeida Costa

• Simple example of MLR

A student was asked to find the weights of 2 balls, A and B using a scale with random measurement errors.
He found:

weight of ball A (measured alone) = 2 lbs
weight of ball B (measured alone) = 3 lbs
weight of balls A+B (measured simultaneously) = 4 lbs.

Use LSE to find estimates for the weights of A and B.

Solution:

$$\underbrace{\begin{matrix} 2 \\ 3 \\ 4 \end{matrix}}_{\text{observed values}} \begin{matrix} = \\ = \\ = \end{matrix} \underbrace{\begin{matrix} A \\ B \\ A+B \end{matrix}}_{\text{true values}} \begin{matrix} + \\ + \\ + \end{matrix} \underbrace{\begin{matrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{matrix}}_{\text{measurement error}}$$

Recall: the general model is

$$y_i = \beta_0 + \beta_A x_A + \beta_B x_B + \varepsilon_i$$

Two solutions:

① • find $\min_{\beta_A, \beta_B} SSE(\beta_A, \beta_B) = \min (2-\beta_A)^2 + (3-\beta_B)^2 + (4-(\beta_A+\beta_B))^2$

• $\frac{\partial SSE(\beta_A,\beta_B)}{\partial \beta_A} = (-2)(2-\beta_A) + (-2)(4-\beta_A-\beta_B) = 0$

$$2 - \beta_A + 4 - \beta_A - \beta_B = 0$$

$$\hat{\beta}_A = 3 - \frac{\beta_B}{2}$$

• $\frac{\partial SSE(\beta_A,\beta_B)}{\partial \beta_B} = (-2)(3-\beta_B) + (-2)(4-\beta_A-\beta_B) = 0$

$$3 - \beta_B + 4 - \beta_A - \beta_B = 0$$

$$\hat{\beta}_B = \frac{7}{2} - \frac{\beta_A}{2}$$

Plugging-in

$$\beta_A = 3 - \frac{1}{2}\left[\frac{7}{2} - \frac{\beta_A}{2}\right] = 3 - \frac{7}{4} + \frac{\beta_A}{4} \Leftrightarrow \frac{3}{4}\beta_A = \frac{12-7}{4} \Leftrightarrow \boxed{\hat{\beta}_A = 5/3}$$

$$\beta_B = \frac{7}{2} - \frac{1}{2}\left[\frac{5}{3}\right] = \frac{7}{2} - \frac{5}{6} = \frac{21-5}{6} = \frac{16}{6} = \frac{8}{3} \Leftrightarrow \boxed{\hat{\beta}_B = 8/3}$$

(2) Use Matrix Computation (this method is preferred by its ability to deal with big problems)

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{bmatrix} x_{1A} & x_{1B} \\ x_{2A} & x_{2B} \\ x_{3A} & x_{3B} \end{bmatrix} \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

which translates into

$$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

We know that $\hat{\beta} = (X^T X)^{-1} X^T Y$

Notice $X_{3\times 2} \Rightarrow X^T_{2\times 3} \Rightarrow X^T_{2\times 3} X_{3\times 2} \in \mathbb{R}^{2\times 2}$

Hence $(X^T X)^{-1} \in \mathbb{R}^{2\times 2}$

$(X^T X)^{-1}_{2\times 2} X^T_{2\times 3} \in \mathbb{R}^{2\times 3}$

$[(X^T X)^{-1} X^T]_{2\times 3} Y_{3\times 1} \in \mathbb{R}^{2\times 1}$

Hint: always check dimensions first!

$\Rightarrow$ Please check the matrix calculations in excel

Your final answer should match for both methods.

What's a good estimate for the error variance $\hat{\sigma}^2$?

$$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{\left(2-\frac{5}{3}\right)^2 + \left(3-\frac{8}{3}\right)^2 + \left(4-\frac{13}{3}\right)^2}{3-2} = \frac{1}{3}$$

($\hat{\sigma}^2$ is commonly called Mean Squared Error $E(MSE) = \sigma^2$)

$\hookrightarrow \beta_A, \beta_B$ (if we included an intercept, then $n-p-1$)

Therefore, std. dev $= \sqrt{1/3} = 0.5774$

CI's

90% CI for the weight of ball A

(Recall $var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$ so $var(\hat{\beta}_i) = (\hat{\sigma}^2 (X^T X)^{-1})_{ii}$)

$\hat{\beta}_A \pm t_{n-2, \frac{\alpha}{2}} \hat{\sigma} \sqrt{d_i}$ where d_i = diagonal element

$\hat{\sigma}\sqrt{d_i} = se(\hat{\beta}_A)$

$t_{0.05,1} = 6.31$ (in R use qt(0.95,1, lower.tail=T) default ($P(X \le x)$))

$\hookrightarrow \frac{5}{3} \pm (6.314)(0.5774)\sqrt{0.67} = 1.67 \pm 2.98$ or $[-1.31, 4.65]$

Question: does the interval above contain 0?

YES!

What does this indicate? β_A might be $0 \Rightarrow x_A$ has no explanatory power!

Remember, we're testing:

$H_0: \beta_A = 0$ (null hypothesis) $\Rightarrow$ interpretation: x_A has no explanatory power

$H_1: \beta_A \neq 0$ (two-sided alternative, two-tailed test)

Under the assumption of normality of the error terms ε_i, each y_i is a normally distributed RV.

then recall $\hat{\beta} \sim MN(\beta, \sigma^2(X^TX)^{-1})$

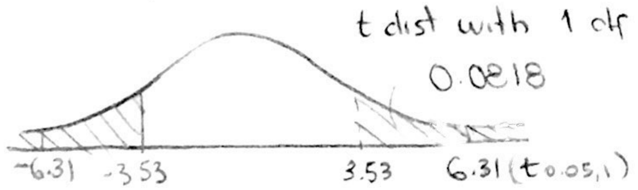
For testing H_0, the following statistic should be used

$$t = \frac{\hat{\beta}_A}{se(\hat{\beta}_A)}$$ which follows a t dist with $(n-p) = 1$ df.

In our case $t_{obs} = \dfrac{5/3}{(0.5774)(\sqrt{0.67})} = \dfrac{5/3}{0.4726} = 3.53$

To check whether is big or not, we compare with the value of a t RV having the same df, example: $t_{0.05,1} = 6.31$ ($\alpha = 0.1$ for 2-tailed)

Or... To get the p-value $= P(|t_{stats}| > |t_{obs}|) = P(t_{stats} > 3.53)$



$0.0818 * 2 = 0.175 \approx$ p-value!

2-tailed test

how to manually get in R?

One way: (1 - pt(3.53, 1)) * 2

So p-value $\approx 0.175 \Rightarrow$ we don't have enough evidence to reject H_0 at a confidence level of what? Pretty much anything < (1 - p-value)

In practice a very small p value (< 0.005) indicates that we observe something which rarely happens under H_0, suggesting we reject H_0 / H_0 is not true

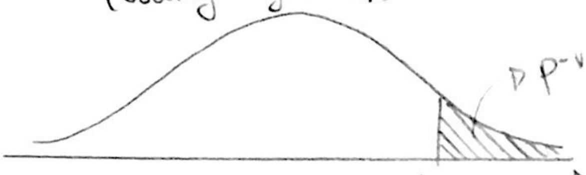
To get a small p-value, $|t_{obs}|$ must be larger than $t_{statistic}$

Understanding P-value

▷ P-value is just a "trigger" to decide whether you reject the null hypothesis or not.

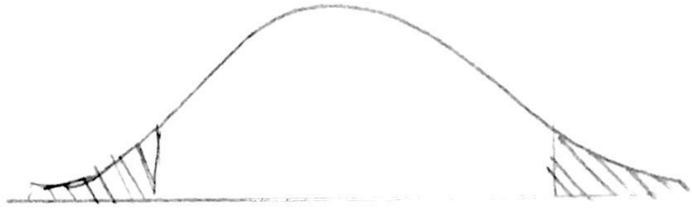
Before continuing, understand the difference between a one-tailed and two-tailed hypothesis test.

One-tailed t-test
(usually right tail)



t distribution

Two-tailed t-test



$H_0: \mu = 30$
$H_1: \mu > 30$ (right tail)
or $H_1: \mu < 30$ (left tail)

$H_0: \mu = 30$
$H_1 \quad \mu \neq 30$
(ie $\mu > 30$ or $\mu < 30$)

For a Confidence level of:

$\underline{C}$	$\underline{\alpha}$	($t_{\alpha,1}$) One-tailed t test		($t_{\alpha/2,1}$) Two-tailed t test	$\underline{df}$'s
0.90	0.10	3.078	notice the ≠? ▷	± 6.314	1
0.95	0.05	6.314 ◁		± 12.706	1
0.99	0.01	31.821		± 63.657	1

most t-tables will bring upper tail prob. of the dist.
(i.e. t s.t. $P(X > t) = 1-\alpha$)

So, you construct your hypothesis testing first. Then, you calculate, based on your data, the probability of obtaining a result more "extreme" than those observed in your data assuming Ho is true. This probability ($P(|t| > |t_{obs}|)$) is called the p-value. When it's very small, it means that the prob. of getting a result that confirms the null hypothesis is also very small, so you do not have enough evidence in the data to accept the null hypothesis (roughly speaking, you "reject" the null hypothesis based on your data).

In our previous example, p-value was 0.175.

Confidence Level	Accept/Reject Ho?
0.99	Accept
0.95	Accept
0.9	Accept
0.8	Reject

Notice that for a $\underline{C}$ of 0.99, I want strong evidence that Ho is <u>not</u> true, whereas for a $\underline{C}$ of 0.8, I'm willing to reject Ho if I have "weak" evidence that Ho is not true.