



Seminario de Tesis

Docente: Dr. Ing. Aland Bravo Vecorena

Escuela Profesional
Ingeniería Civil

Periodo académico: 2019-I

Semestre: I

Unidad: I

Generalidades de Herramientas para Tesis e Investigación

Ficheros:
CAD, csv, etc.



- Repositorio gratuito de ficheros en línea.

Github y Git

Ficheros:
Encuestas, etc.



- Formularios para Encuestas de Google.

Google
Forms

Ficheros:
PDF, png, etc.



- Indicadores y Métricas de Desempeño.

Google
Data Studio

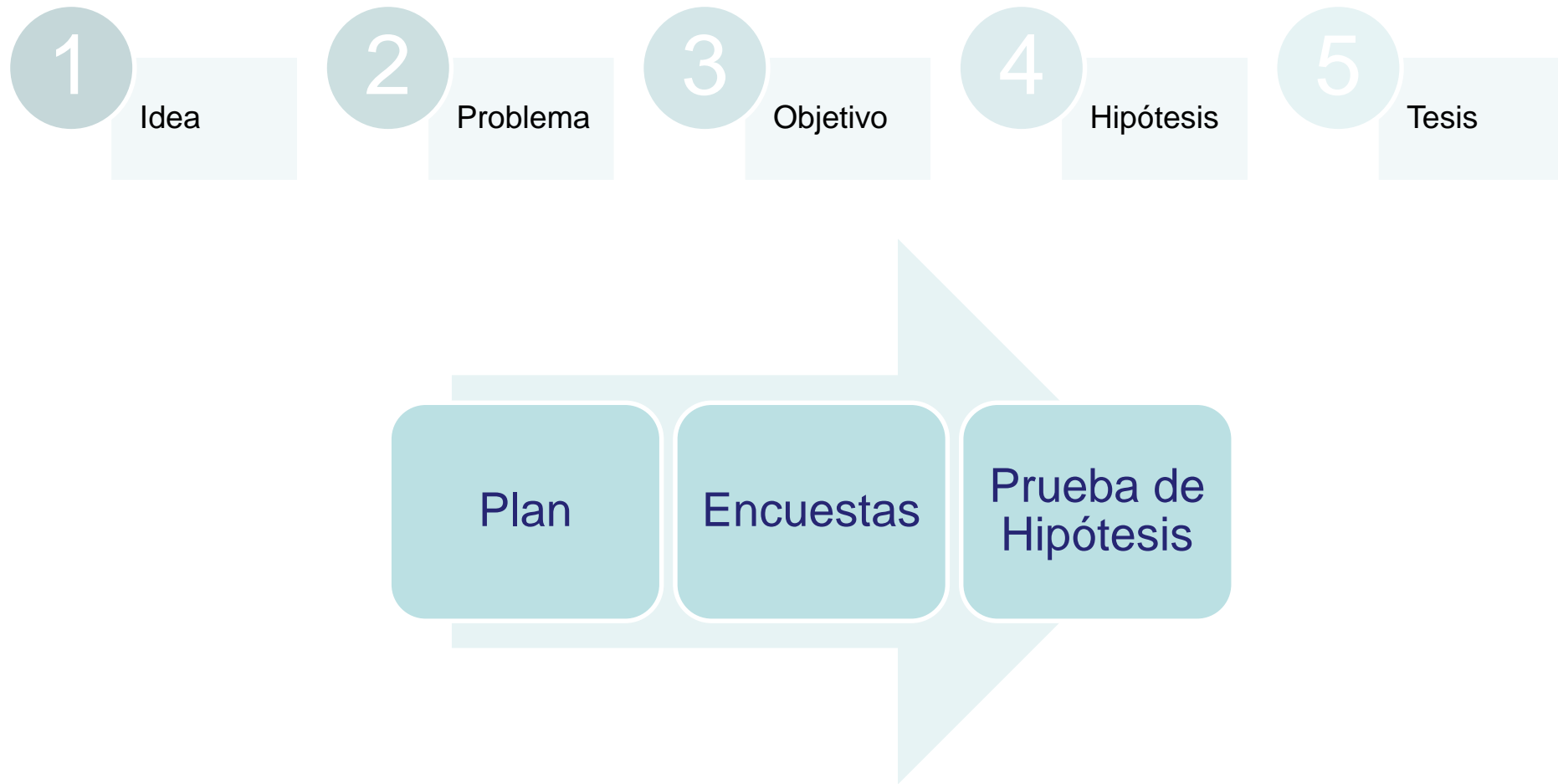
Ficheros:
html, etc.



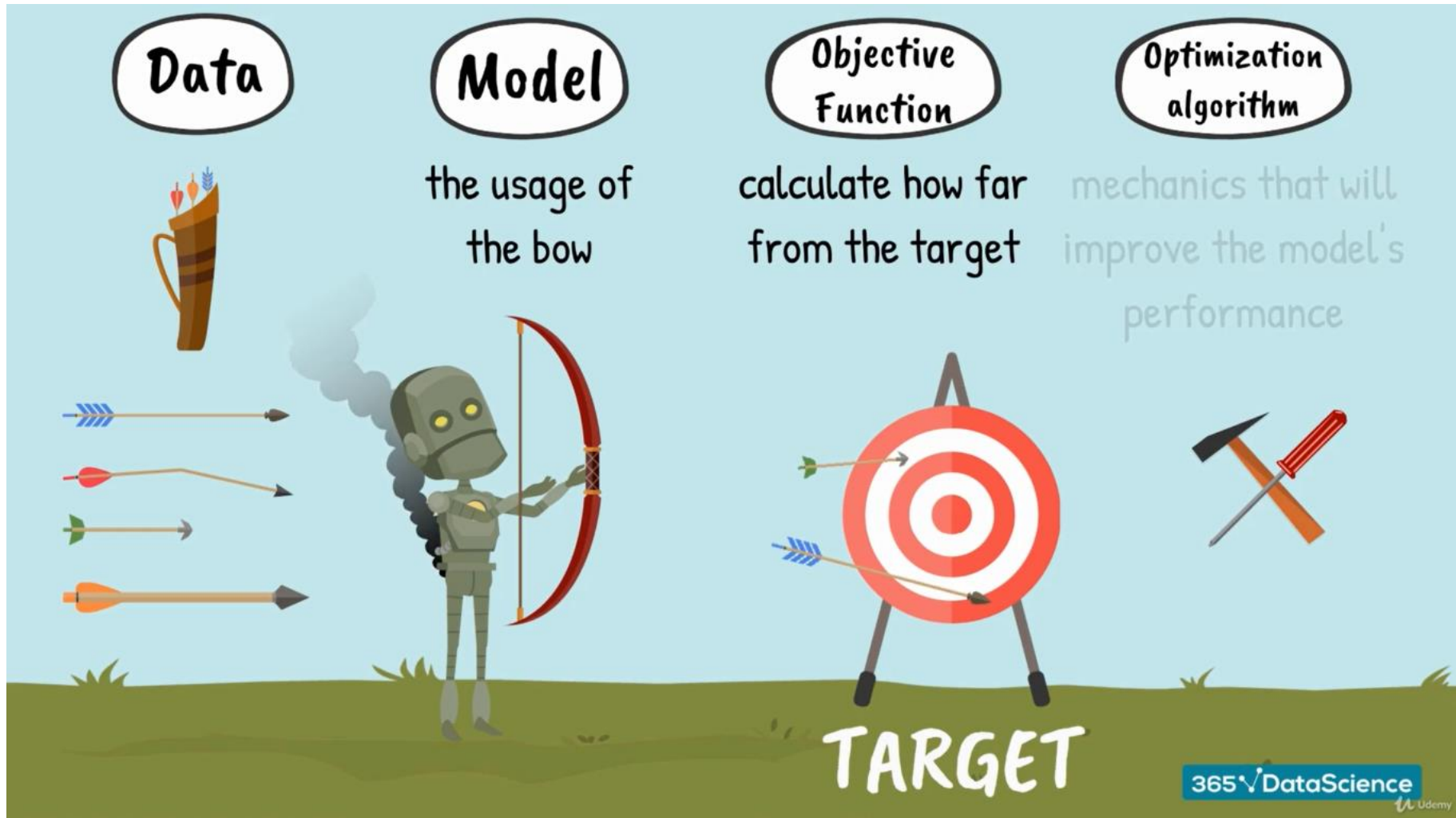
- Bitácora o agenda social.

Google
Blogger

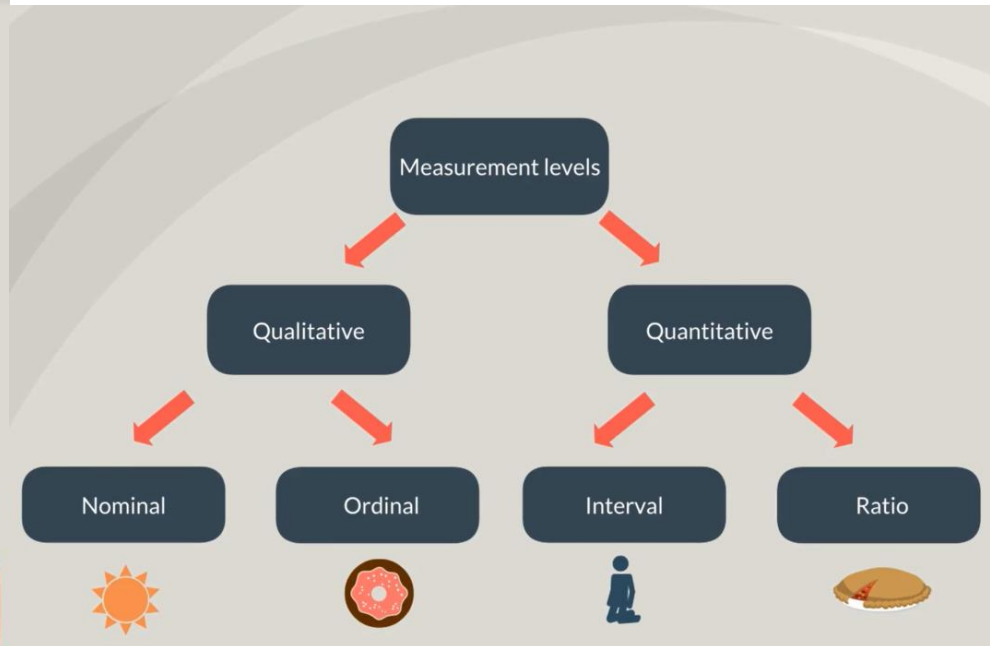
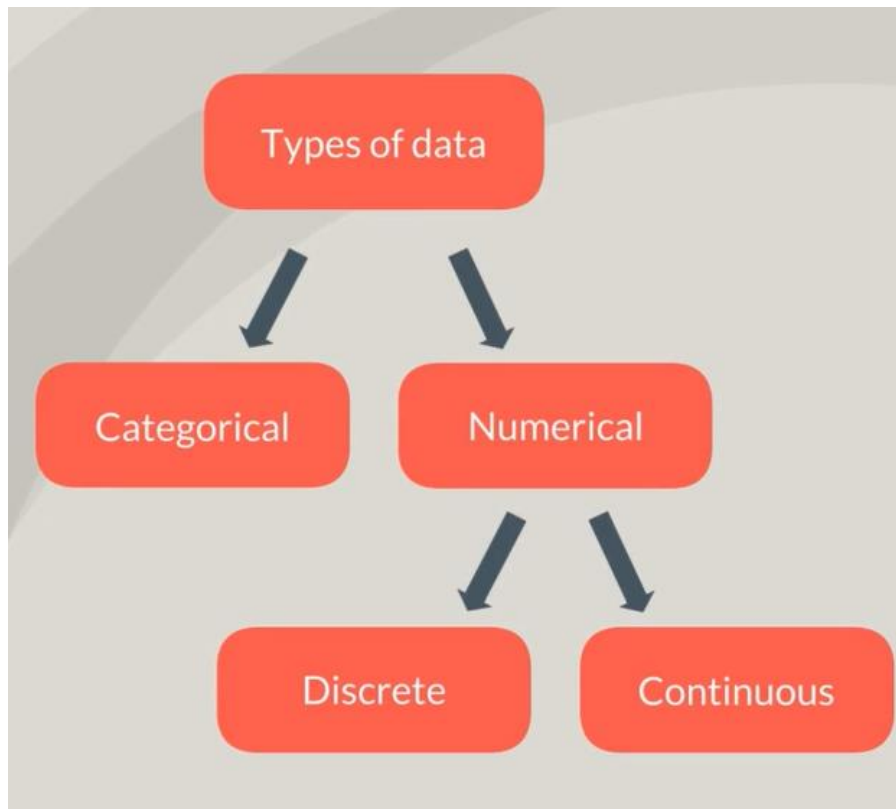
Validando Nuestra Tesis



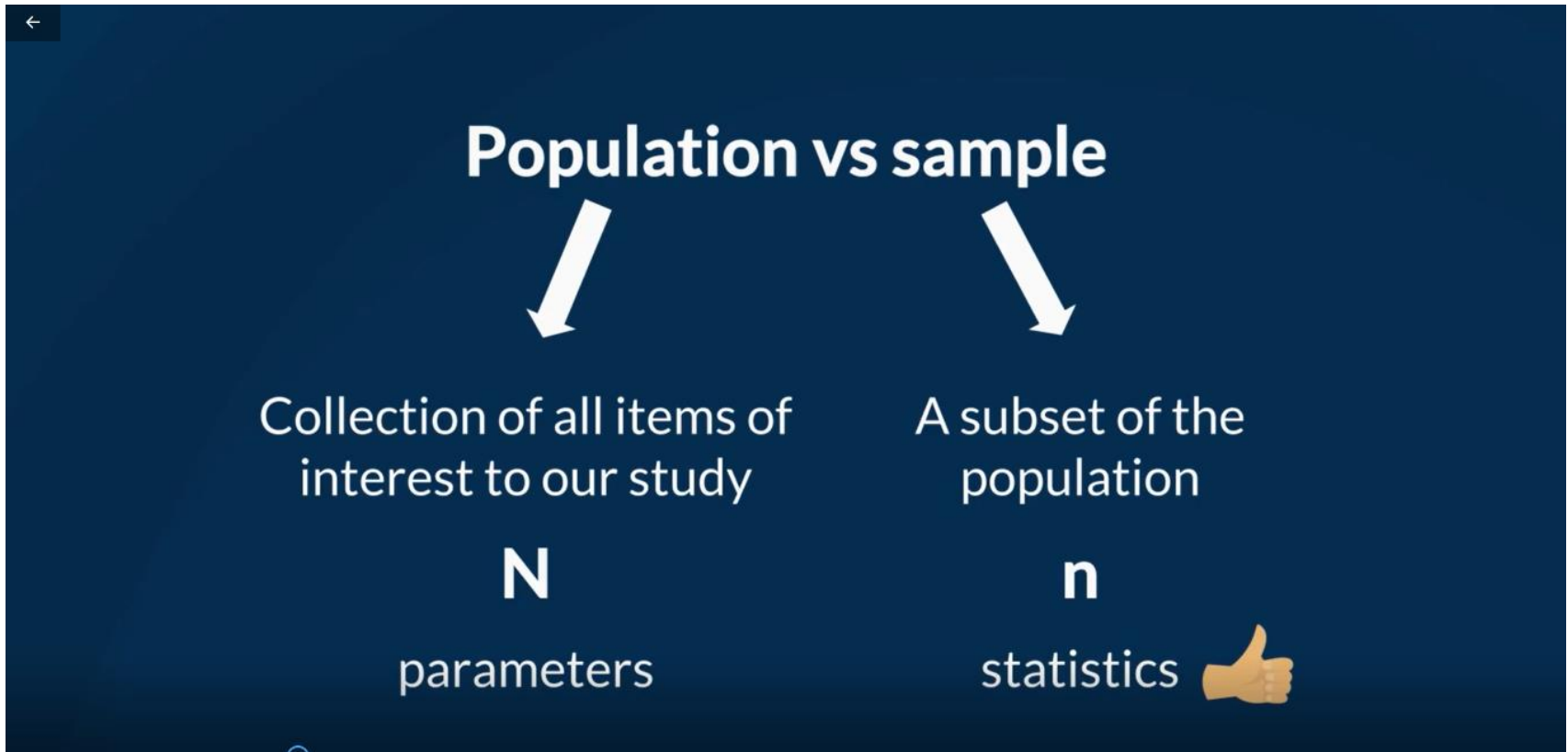
Evolución de la Ciencia de los Datos



Tipos y Niveles de Medida de los Datos



Población versus Muestra



La Media Muestral versus la Media Poblacional

Mean

Sample formula

$$\frac{\sum_{i=1}^n x_i}{n}$$

Population formula

$$\frac{\sum_{i=1}^N x_i}{N}$$

N is the size of the population,
n is the size of the sample

Varianza Poblacional versus Varianza Muestral

Variance

Population variance

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$



Sample variance

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$



Desviación Estándar Poblacional y Muestral

Standard deviation formulas

Population standard
deviation

$$\sigma = \sqrt{\sigma^2}$$

Sample standard
deviation

$$s = \sqrt{s^2}$$



Coeficiente de Variación Poblacional y Muestral

Coefficient of variation (CV)

/relative standard deviation/

standard deviation

mean

Population formula

$$c_v = \frac{\sigma}{\mu}$$

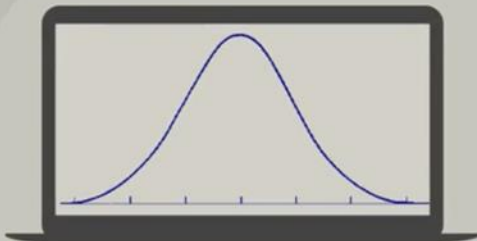
Sample formula

$$\hat{c}_v = \frac{S}{\bar{x}}$$

Distribuciones de la Función de Probabilidad

In statistics

Distribution = Probability distribution



Normal



Binomial



Uniform

Def: A distribution is a function that shows the possible values for a variable and how often they occur.

Probabilidad de Lanzar un Dado

Rolling a die

Outcome	Probability
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$
7	0



Probabilidad de Lanzar un Dado

Visual representation

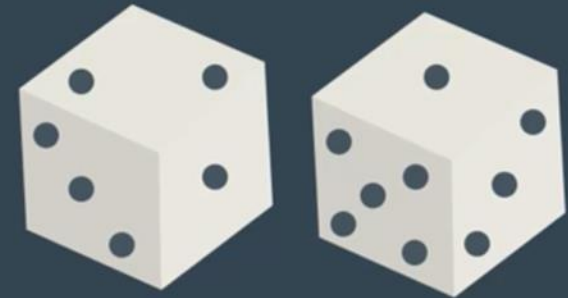
Uniform
distribution



Probabilidad de Lanzar Dos Dados

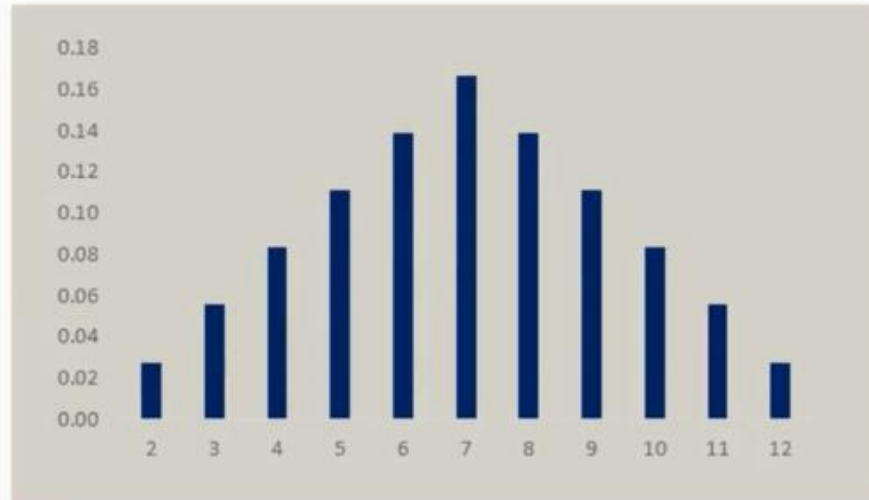
Rolling two dice

Outcome	Probability
2	0.03
3	0.06
4	0.08
5	0.11
6	0.14
7	0.17
8	0.14
9	0.11
10	0.08
11	0.06
12	0.03
All else	0

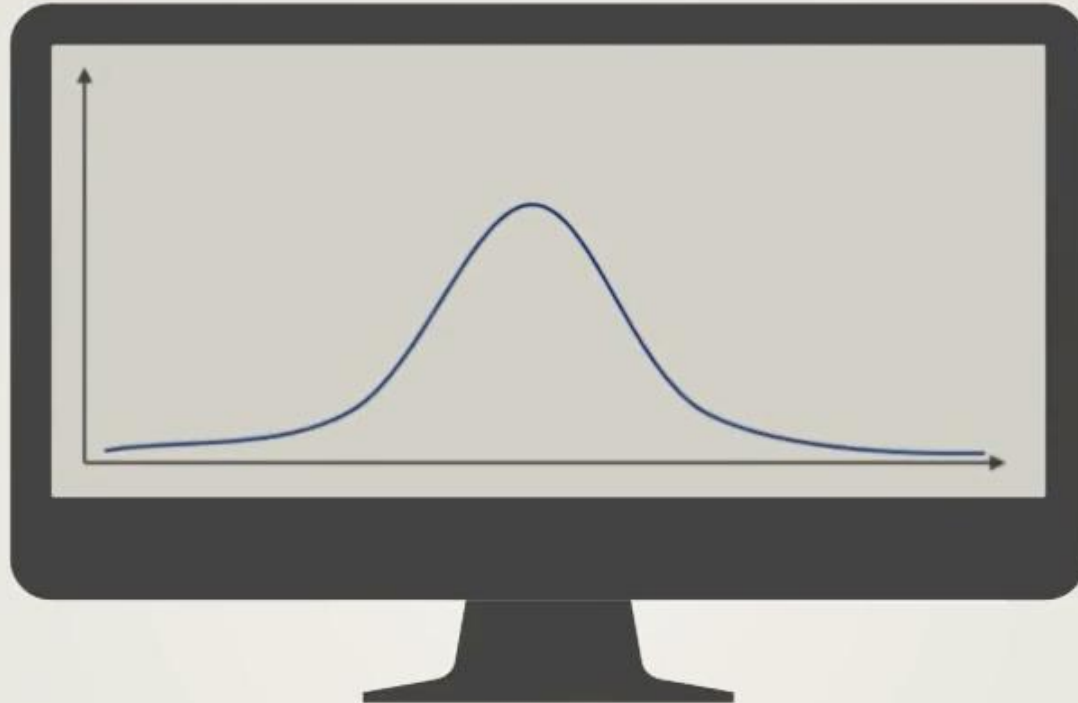


Probabilidad de Lanzar Dos Dados

Visual representation

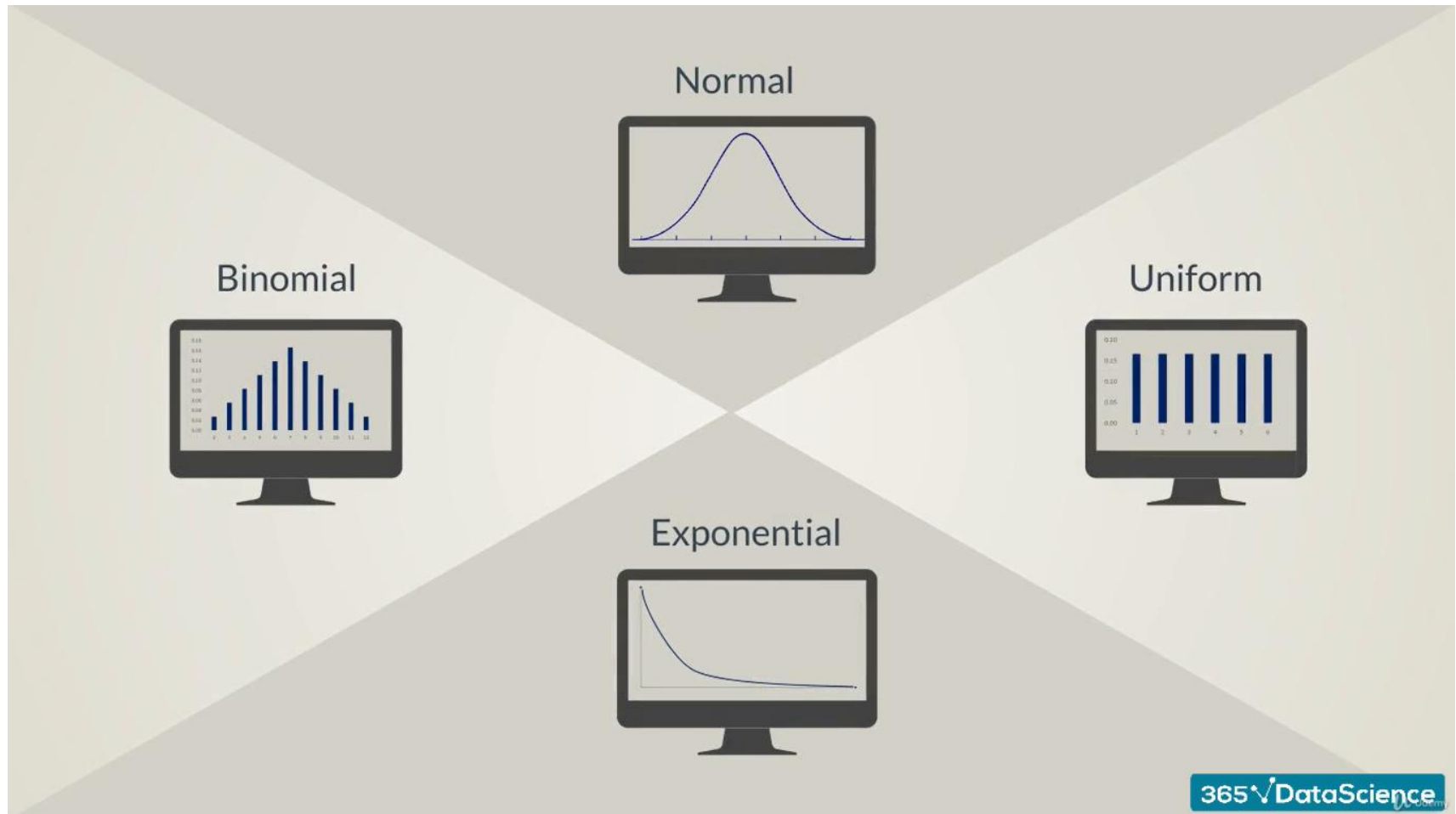


Distribución de la Probabilidad



The distribution of a dataset shows us the frequency at which possible values occur

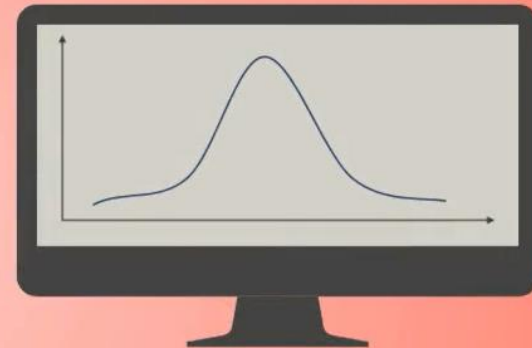
Distribuciones de Probabilidades



Distribución Normal versus T-Student

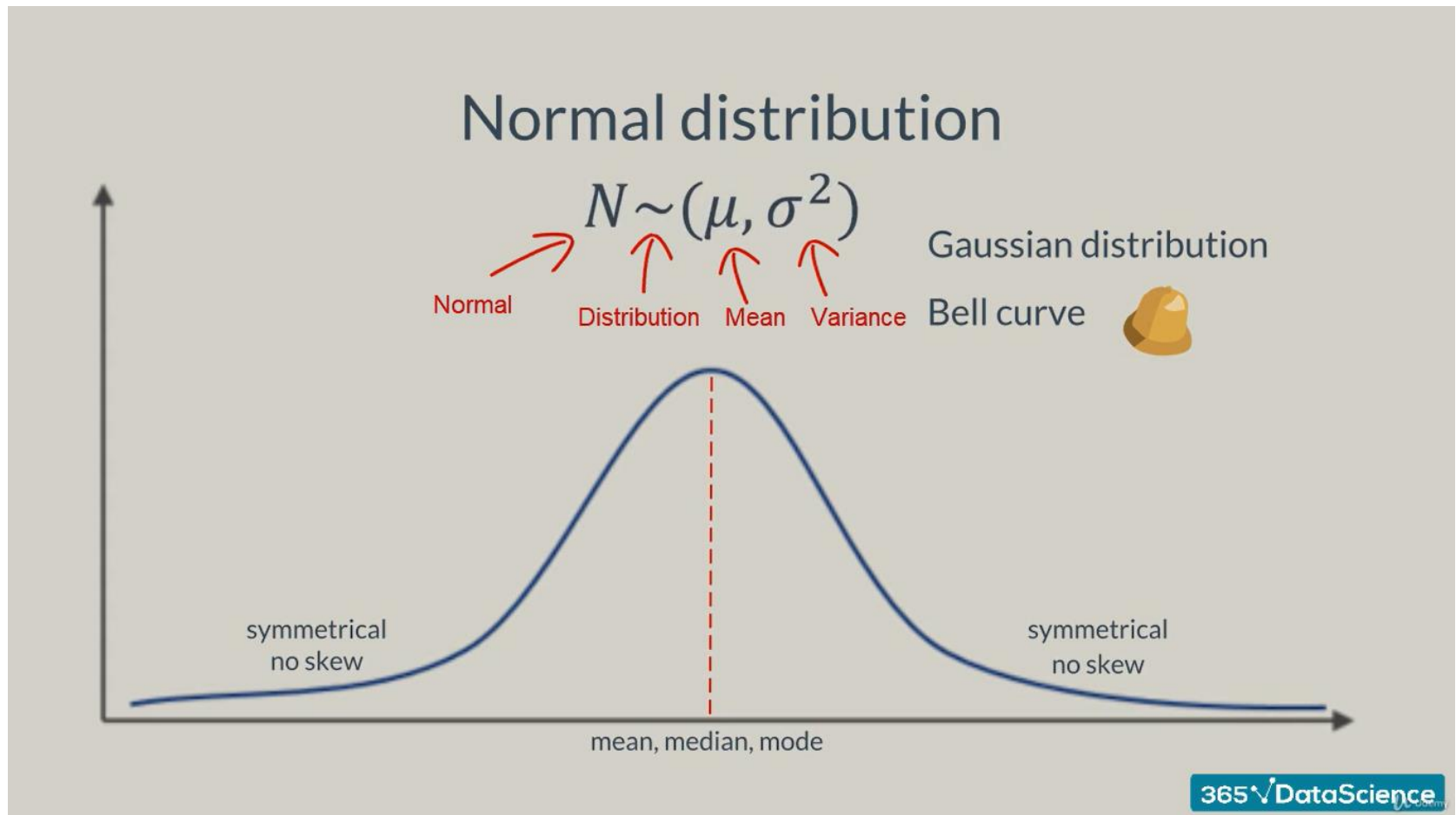


Normal

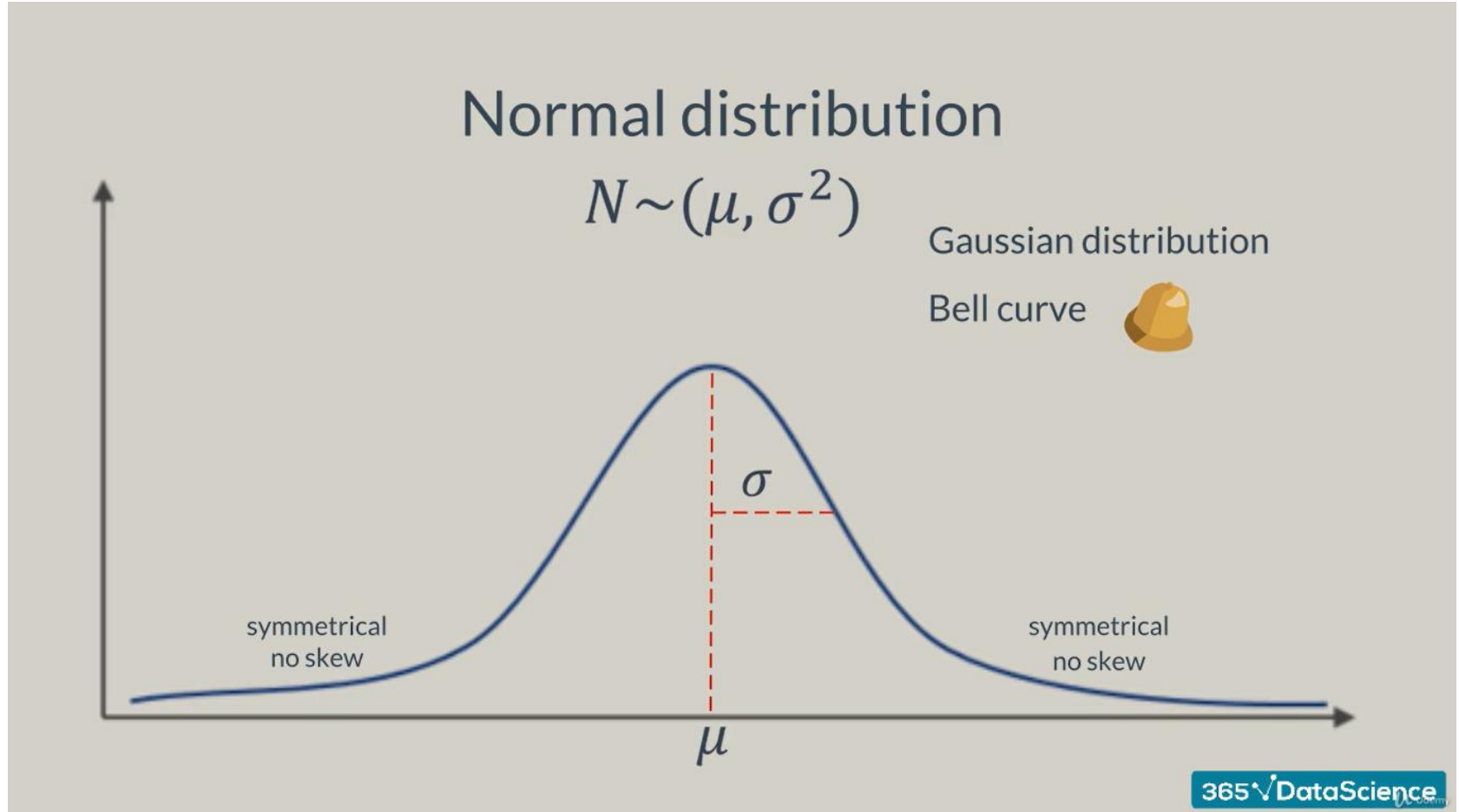


Student's T

Distribución Normal

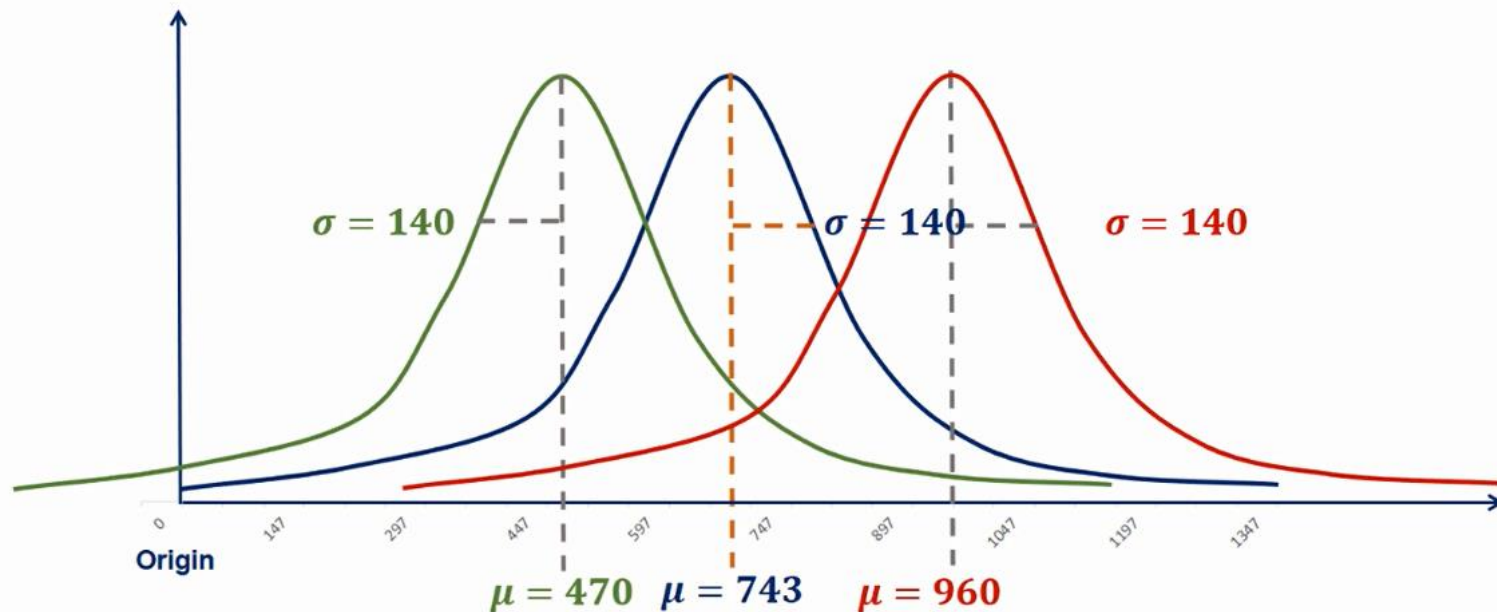


Distribución Normal



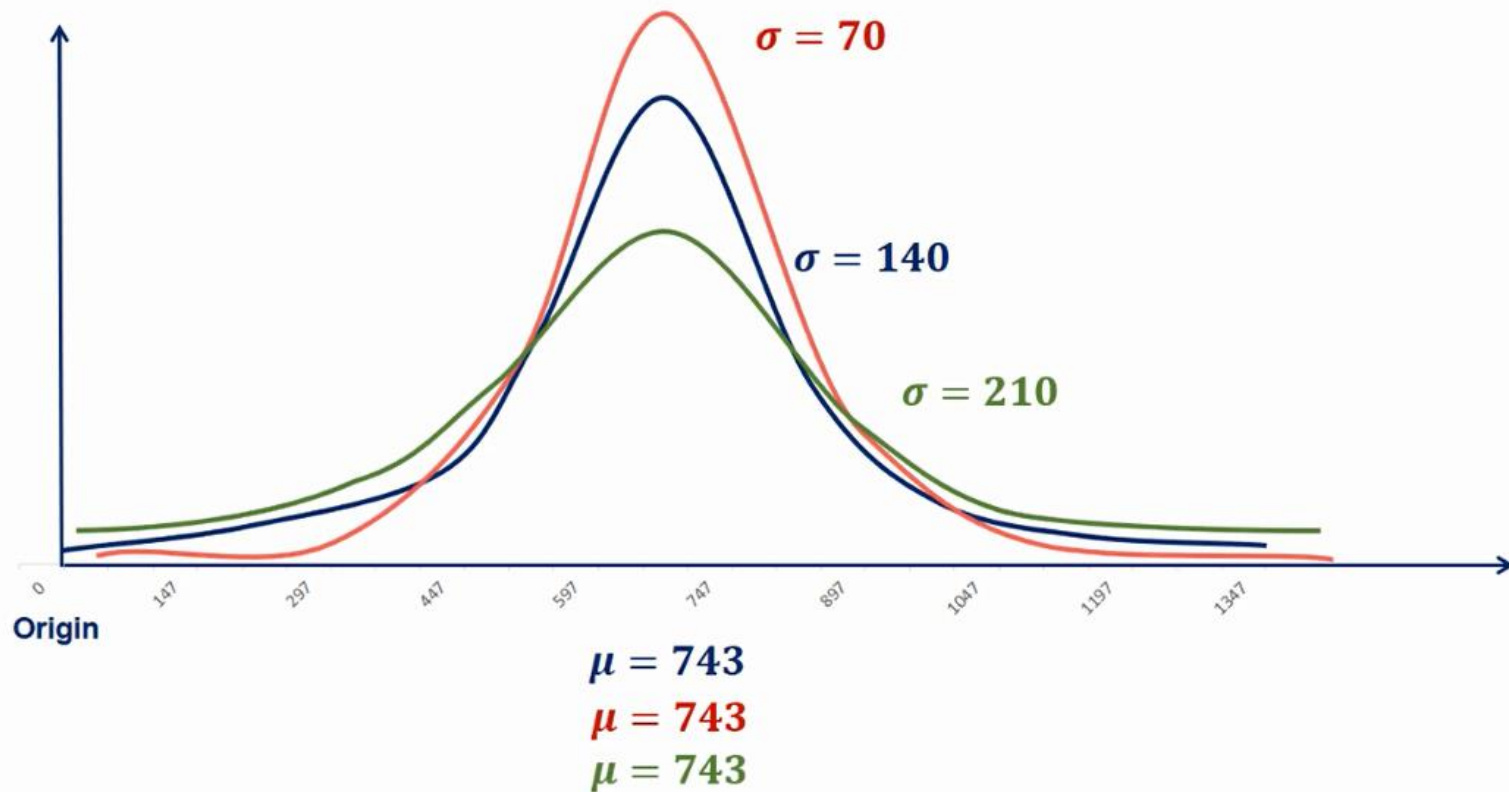
Detalles de una Distribución Normal

Normal distribution. Controlling for standard deviation



Detalles de una Distribución Normal

Normal distribution. Controlling for the mean



Normalización hacia la Distribución Estándar Normal

From Normal to Standard normal



Every normal distribution can be
'standardized'

Normalización hacia la Distribución Estándar Normal

$$N \sim (\mu, \sigma^2) \longrightarrow \text{Diagrama de una órbita atómica} \longrightarrow N \sim (0, 1)$$

$$\text{Standardized variable} = \frac{\text{Original variable} - \text{Mean}}{\text{Standard deviation}}$$

$$\text{z-score} = \frac{x - \mu}{\sigma}$$

If you take a dataset, subtract its mean from each data point and then calculate the mean once again, you will get 0.

Ejemplo de Normalización de Datos

Microsoft Excel ribbon: File, Home, Insert, Page Layout, Formulas, Data, Review, View, Power Pivot, Tell me what you want to do, Share.

Formulas: =O8

Standard normal distribution		Standardization	
Original dataset		Subtract mean	
1		-2	
2		-1	
3		-1	
4		0	
5		0	
6		0	
7		1	
8		1	
9		2	
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			

Mean 3
St. dev 1.22
N~(3,1.49)

Mean 0
St. dev 1.22
N~(0,1.49)

Mean 0.00
St. dev 1.00
N~(0,1)

x

$x - \mu$

$\frac{x - \mu}{\sigma}$

Teorema del Límite Central

The Central limit theorem

No matter the distribution



$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$$

The more samples you extract

The bigger the samples

$$N \sim \left(\mu, \frac{\sigma^2}{n} \right)$$

$$k \rightarrow \infty$$

$$n \rightarrow \infty$$



Teorema del Límite Central

How to find the standard error?

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k \quad N \sim \left(\mu, \underbrace{\frac{\sigma^2}{n}}_{\text{Variance}} \right)$$

$$\text{Standard deviation} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$



Estimadores de los Parámetros

Point estimators and estimates

Estimator

/how to estimate/

Parameter

/what to estimate/

Estimate

/concrete result/

\bar{x}

of

μ



52.22

s^2

of

σ^2



1724.93

Estimando Precios de un Restaurante

You visit 5% of the
restaurants



Average price

22.50£



20 £

25 £

Confidence interval: [20£, 25£]

5% chance that the population
parameter is outside the range

Nivel de Significancia y Nivel de Confianza

$\alpha \rightarrow$ Nivel de Significancia =

Tamaño de Región de Rechazo del Error Tipo I.

$1 - \alpha \rightarrow$ Nivel de Confianza = *Tamaño de Región de Aceptación del Error Tipo I.*



Tipos de Errores

- **Hipótesis Nula:** denotada como H_0 ; siempre especifica **un solo valor del parámetro** de la población.
- **Hipótesis Alternativa:** denotada como H_1 ; es la que **responde nuestra pregunta**, la que se establece en base a la evidencia que tenemos.
- **Error Tipo I:** Se comete cuando la **hipótesis nula es verdadera** y, después del contraste, **se rechaza**.
- **Error Tipo II:** Se comete cuando la **hipótesis nula es falsa** y, después del contraste **se acepta**.

	H_0 es VERDADERA	H_0 es FALSA
Rechazamos H_0	Error Tipo I $P(\text{Error Tipo I}) = \alpha$	Decisión Correcta
Aceptamos H_0	Decisión Correcta	Error Tipo II $P(\text{Error Tipo II}) = \beta$

Nivel de Significancia α y Nivel de Confianza ($1-\alpha$)

α	Confidence level
5%	95%
1%	99%



Estimación Puntual



$$\left[\text{Point estimate} - \text{reliability factor} * \text{standard error}, \text{Point estimate} + \text{reliability factor} * \text{standard error} \right]$$

Estimación Puntual



$$\left[\bar{x} - \text{reliability factor} * \frac{\sigma}{\sqrt{n}}, \bar{x} + \text{reliability factor} * \frac{\sigma}{\sqrt{n}} \right]$$

Intervalos de Confianza

Confidence intervals



Population variance known

Z



Population variance unknown

T

Intervalos de Confianza

Confidence intervals



$$N \sim (\mu, \sigma^2)$$

CLT

Population variance known

Distribución T-Student

Degrees of freedom (d.f.)

$$t_{n-1,\alpha} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

sample size: n
d.f.: n-1

Ejemplos con y sin Varianza de la Población

Microsoft Excel interface showing a dataset of salaries and calculated statistics.

Confidence intervals, t-score	
Data scientist salary	
Dataset	
\$ 78,000	Sample mean \$ 92,533
\$ 90,000	Sample standard deviation \$ 13,932
\$ 75,000	Standard error \$ 4,644
\$ 117,000	
\$ 105,000	
\$ 96,000	
\$ 89,500	
\$ 102,300	
\$ 80,000	

Population variance unknown

$$\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

Population variance known

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Excel tabs: Salaries, Mean, std, Confidence interval

Fórmulas de los Intervalos de Confianza

Confidence intervals formulas

Population variance known

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Population variance unknown

$$\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

Margin of error

$$\bar{x} \pm ME$$

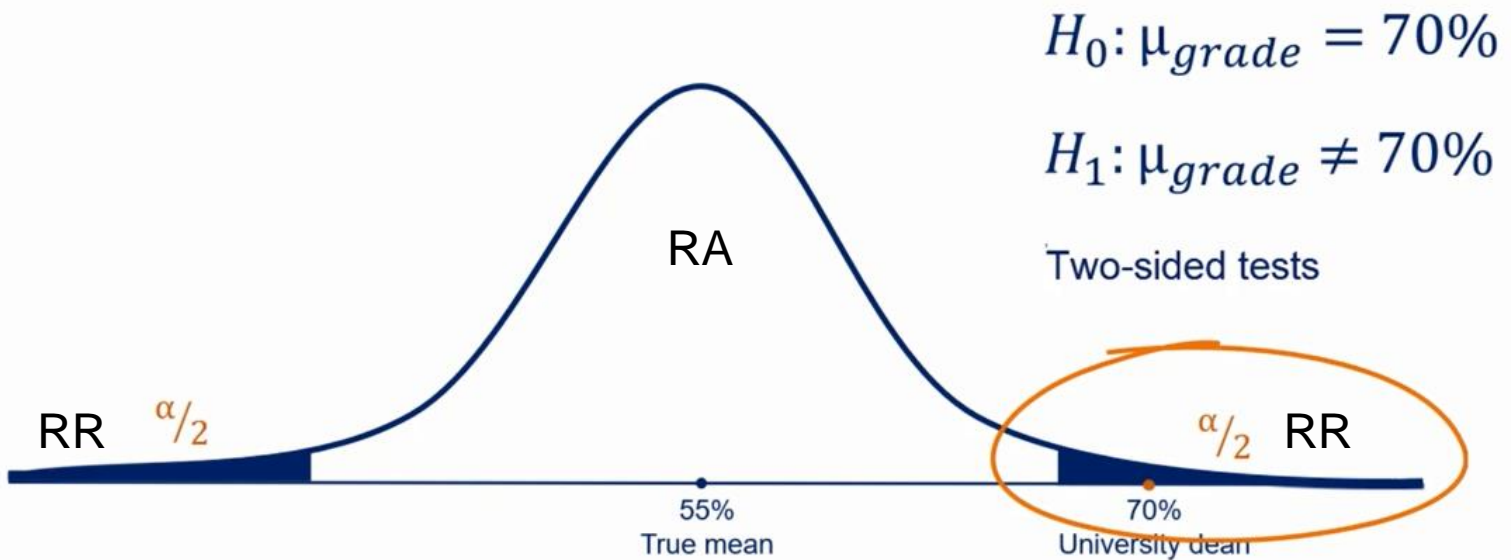
Aplicando la Distribución a las Hipótesis

Example

Hypotheses	Notation
Null hypothesis	H_0
Alternative hypothesis	H_1 or H_A

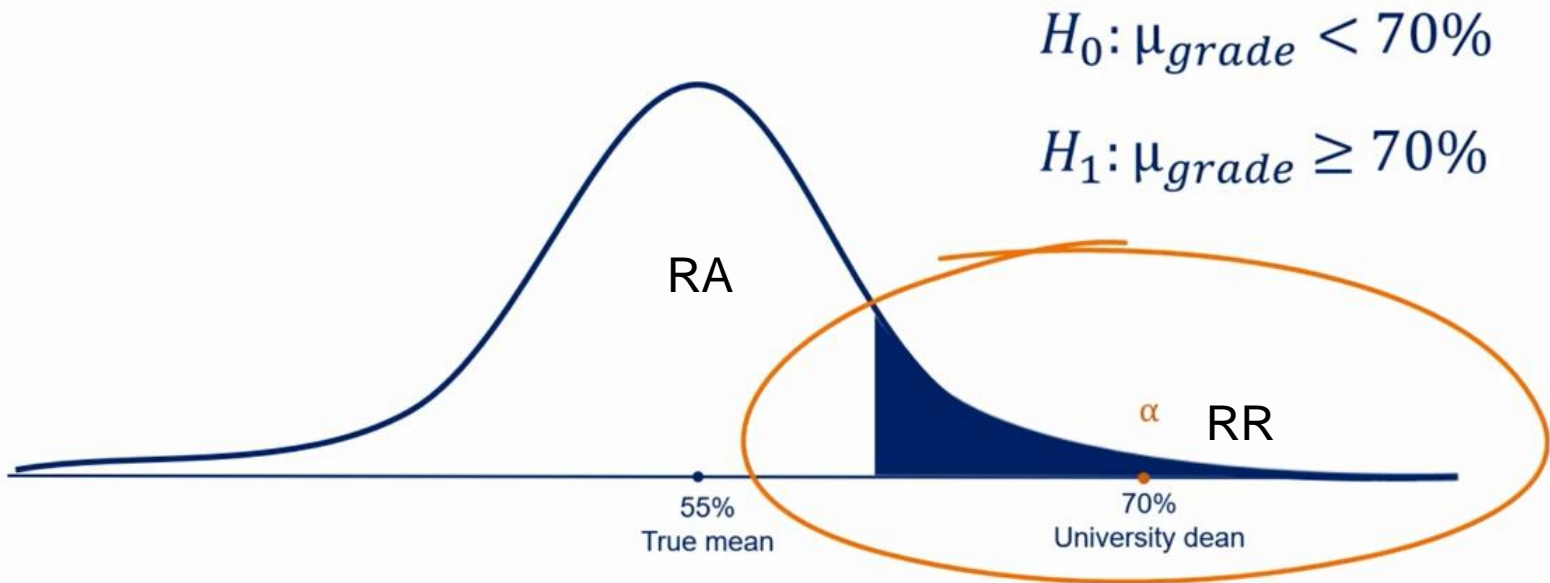
Aplicando la Distribución a las Hipótesis: Bilateral

Grades in a UK university



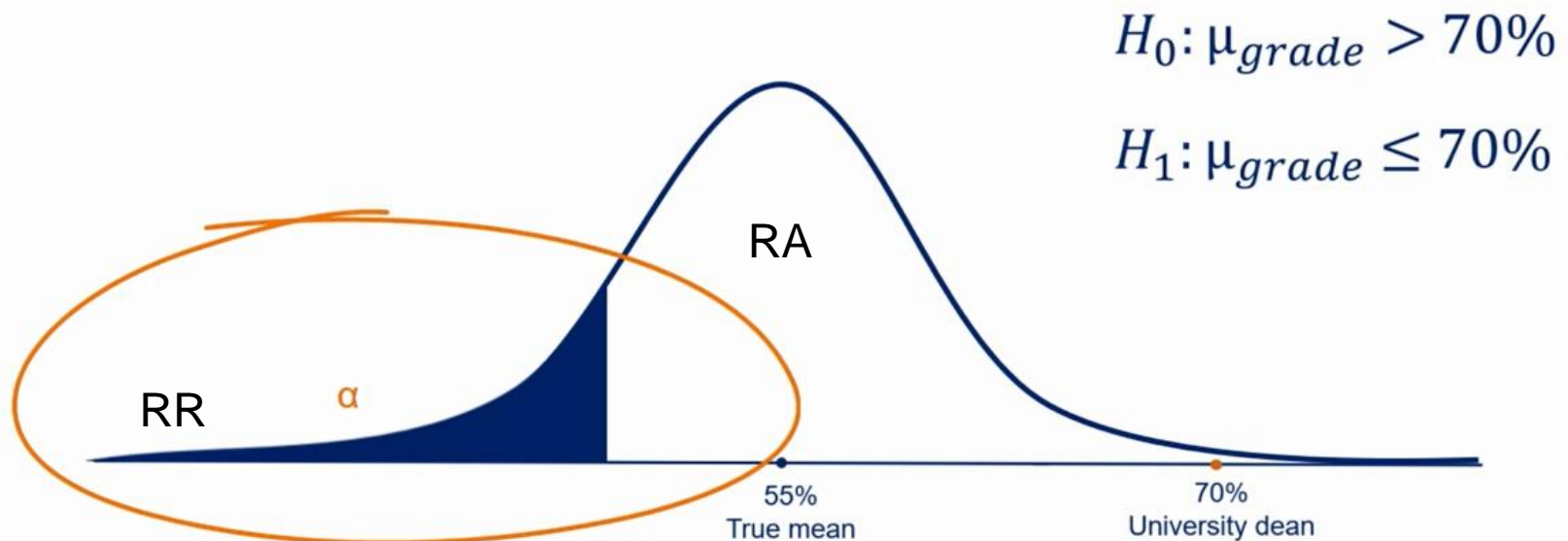
Distribución a las Hipótesis: Unilateral Derecha

Grades in a UK university



Distribución de las Hipótesis: Unilateral Izquierda

Grades in a UK university



Valor del estadístico p-value

p-value



p-value is the smallest level of significance at which we can still reject the null hypothesis, given the observed sample statistic

#mostcommonwaytotesthypotheses

Caso de Estudio: Prueba de Hipótesis

- 1) Un alumno de la EAPIC-UAP se ha planteado la siguiente hipótesis de trabajo:

“Motivación y Desempeño Laboral en los Trabajadores de Construcción Civil del Distrito de Huánuco”

Ho: La motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco.

H1: La motivación no influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco.

VARIABLES	DIMENSIONES	INDICADORES
VI: La Motivación en Obras Civiles	Intensidad	Esfuerzo
		Tareas
	Dirección	Orientación
		Toma de decisiones
	Persistencia	Continuidad
		Esfuerzo Tiempo
VD: Desempeño Laboral	La participación del trabajador	Trabajo en equipo
		Productividad
		Participación activa
		Actitud del trabajo
	Formación de desarrollo profesional	Crecimiento personal
		Mejor desempeño del puesto
		Autoestima
	Adecuación / Ambiente de trabajo	Conocimiento y habilidades
		Adaptabilidad

Caso de Estudio: Prueba de Hipótesis

Para efectuar la prueba de su hipótesis de trabajo, ha decidido utilizar como instrumento de medición un **Cuestionario**, que será realizado a los trabajadores de construcción civil, para ello ha utilizado:

II. INSTRUMENTO DE MEDICIÓN DE MOTIVACIÓN

	NUNCA	CASI NUNCA	A VECES	CASI SIEMPRE	SIEMPRE					
	1	2	3	4	5					
Nº	INTENSIDAD					1	2	3	4	5
Esfuerzo										
1	¿Considera Ud. que se esfuerza lo suficiente para realizar sus actividades?									
2	¿Ud. cumple con el horario de trabajo establecido por la empresa?									
Tareas										
3	¿Considera que las tareas que realiza son acorde a sus capacidades?									
4	¿Las tareas que realiza tienen buen resultado?									
DIRECCION										
Orientación										
5	¿Ud. Es guiado hacia la consecución de sus metas?									
6	¿Ud. Recibe orientación antes de realizar una actividad?									
Toma de decisiones										
7	¿Cuándo su jefe no está, asume el cargo en su área de trabajo?									
8	¿Cree que los jefes se relacionan con los demás trabajadores positivamente?									
PERSISTENCIA										
Continuidad										
9	¿Considera Ud. que se esfuerza continuamente en la realización de sus labores?									

Caso de Estudio: Prueba de Hipótesis

10	¿Cree Ud. que debería haber una capacitación continua para mejorar su productividad?						
Esfuerzo							
11	¿Considera Ud. que sus esfuerzos de superación deberían ser capacitadas?						
12	¿Considera que sus esfuerzos están acorde con su remuneración salarial?						
Tiempo							
13	¿Ud. ejecuta sus actividades en el menor tiempo establecido?						
14	¿Es recompensado por el sobre tiempo que pasa en sus labores?						

III. INSTRUMENTO DE MEDICIÓN DEL DESEMPEÑO LABORAL

		NUNCA	CASI NUNCA	A VECES	CASI SIEMPRE	SIEMPRE
		1	2	3	4	5
N°	LA PARTICIPACIÓN DEL TRABAJADOR	1	2	3	4	5
Trabajo en equipo						
1	¿Prefiere Ud. trabajar en equipo más que trabajar solo?					
2	¿Fomenta Ud. el trabajo en equipo?					
Productividad						
3	¿Ud. ha demostrado a los demás trabajadores la efectividad?					
4	¿Ud. produce igual cuando su pago no es muy bien remunerado?					
Participación activa						
5	¿Cómo trabajador participa con los demás compañeros?					
6	¿Ud. Soluciona problemas inmediatamente en su trabajo?					
Actitud del trabajo						
7	¿Cómo trabajador participa con los demás compañeros?					
8	¿Ud. Tiene las ganas de actuar e incentivar al grupo de trabajo?					

Caso de Estudio: Prueba de Hipótesis

	FORMACIÓN DE DESARROLLO PROFESIONAL					
Crecimiento personal						
5	¿Ud. Es guiado hacia la consecución de sus metas?					
6	¿Ud. Recibe orientación antes de realizar una actividad?					
Mejor desempeño del puesto						
7	¿Cuándo su jefe no está, asume el cargo en su área de trabajo?					
8	¿Cree que los jefes se relacionan con los demás trabajadores positivamente?					
Autoestima						
9	¿La empresa le proporciona, oportunidades de crecimiento profesional?					
10	¿La empresa me proporciona oportunidades de crecimiento económico?					
	ADECUACIÓN / AMBIENTE DE TRABAJO					
Conocimiento y habilidades						
11	¿Conoce cuáles son sus obligaciones y derechos dentro de la organización?					
12	¿Cree Ud. que tendría más habilidades si recibiría capacitación permanente?					
Adaptabilidad						
13	¿Disfruta trabajar en equipo más que trabajar solo?					
14	¿Le gusta influir a otras personas para que imiten su trabajo?					
Clima del trabajo						
15	¿Acostumbra construir relaciones estrechas con sus compañeros de trabajo?					
16	¿Prefiere trabajar con mucha libertad, sin supervisión directa o muy cercana?					

Caso de Estudio: Prueba de Hipótesis

Para definir la cantidad de encuestas a realizar ha utilizado el teorema de muestreo, lo cual le ha definido una muestra de 9 sindicatos detallados en 65 trabajadores a encuestar (a criterio del investigador):

Tamaño de Muestra						
Entradas:						
N=	9	Tamaño del Universo de Población de Sindicatos - Huánuco				
p=	0.5	Probabilidad de Ocurrencia de los Casos				
E=D=	0.05	Error de muestreo				
α =	0.05	Nivel de Significancia				
Salidas						
q=(1-p)=	0.5	$n = \frac{N (Z)^2 P Q}{(D)^2 (N-1) + (Z)^2 P Q}$				
n=	9					
$Z_{(1-\alpha/2)}$ =	1.959963985					

Caso de Estudio: Prueba de Hipótesis

Los resultados parciales de su Ficha de Cuestionario se muestra a continuación:

H_0 : La motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco.

H_1 : La motivación no influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco.

NUNCA	CASI NUNCA	A VECES	CASI SIEMPRE	SIEMPRE
1	2	3	4	5

$$\alpha = 5\%$$

$$n = 65$$

$$\bar{X} = 4.1$$

$$S = 0.19$$

$$H_0: \mu \geq 4$$

Population variance unknown

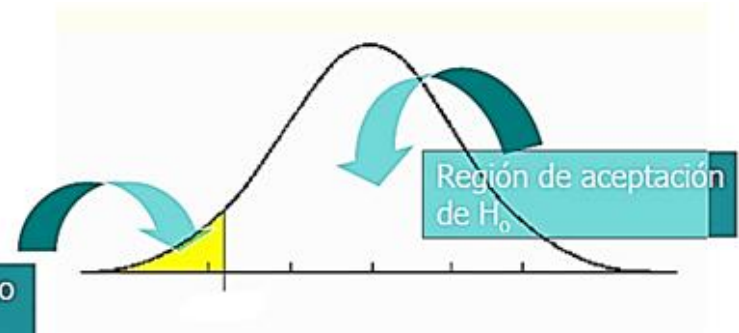
$$\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

$$t_{n-1, \alpha} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

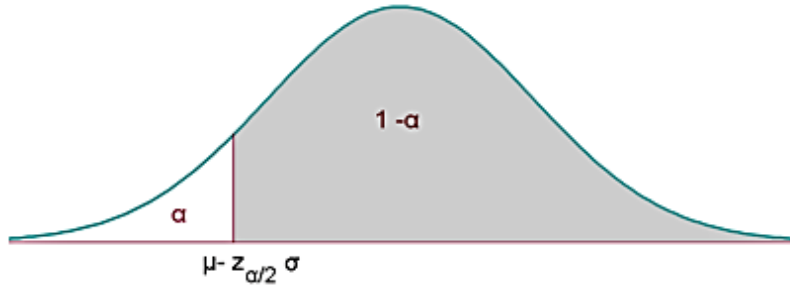
sample size: n
d.f.: n-1

Región de rechazo de $H_0 = \alpha$

Región de aceptación de H_0



Caso de Estudio: Prueba de Hipótesis



$1 - \alpha$	α	z_{α}
0.90	0.10	1.28
0.95	0.05	1.645
0.99	0.01	2.33

$1 - \alpha$	$\alpha/2$	$z_{\alpha/2}$
0.90	0.05	1.645
0.95	0.025	1.96
0.99	0.005	2.575

Distribución Normal

Pasos de una Prueba de Hipótesis:

1. Se plantea la **hipótesis nula** y la **alternativa**.
 2. Se selecciona el **nivel de significancia** y el **nivel de confianza**.
 3. Se identifica el **estadístico de prueba**.
 4. Se forma la **regla de decisión** y la **región de rechazo**.
 5. Se toma una **muestra**, **analiza** y se **decide**:
 - ❖ **Acepta H_0**
 - ❖ **O se Rechaza H_0 y se Acepta H_1 .**
- El **propósito** de la prueba de hipótesis **no es cuestionar el valor calculado del estadístico (muestral)**, sino hacer un juicio con respecto a la **diferencia entre estadístico de muestra y un valor planteado del parámetro.**

Caso de Estudio: Prueba de Hipótesis con Media Poblacional

- **Con VARIANZA CONOCIDA:** Se utiliza el siguiente estadístico de prueba:

$$\boxed{Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad (\geq 30) \quad \text{ó} \quad t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}; (n - 1)gl \quad (< 30)}$$

- **Con VARIANZA DESCONOCIDA:** Se utiliza el siguiente estadístico de prueba:

$$\boxed{Z = \frac{\bar{x} - \mu}{S / \sqrt{n}} \quad (\geq 30) \quad \text{ó} \quad t = \frac{\bar{x} - \mu}{S / \sqrt{n}}; (n - 1)gl \quad (< 30)}$$

Caso de Estudio: Prueba de Hipótesis con Media Poblacional

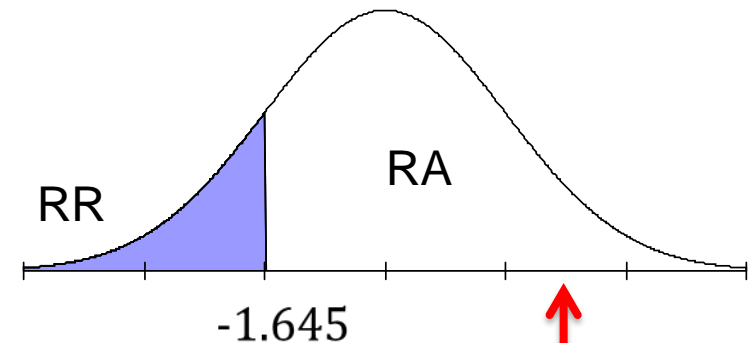
Como se desconoce la varianza de la población, y se tiene 65 muestras, entonces utilizaremos el estadístico de prueba Z:

$$H_0: \mu \geq 4$$

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{4.1 - 4}{0.19/\sqrt{65}} = \frac{0.1}{0.0235} = 4.2432$$

$$\alpha = 0.05$$

$$-Z_{\alpha} = -Z_{0.05} = -1.645$$



[Tabla Normal Z](#)

Concluimos: Como se Acepta la H_0 ; «Estamos 95% seguros, que la motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco»

4.2432

Caso de Estudio: Prueba de Hipótesis con Media Poblacional

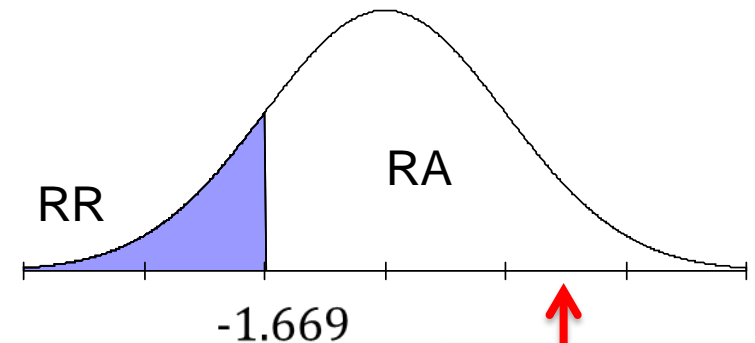
Como se desconoce la varianza de la población, y se tiene 65 muestras, entonces utilizaremos el estadístico de prueba T:

$$H_0: \mu \geq 4$$

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}} = \frac{4.1 - 4}{0.19 / \sqrt{65}} = \frac{0.1}{0.0235} = 4.2432$$

$$\alpha = 0.05 \quad n = 65 \quad gl = n - 1 = 64$$

$$-t_{\alpha; gl} = -t_{0.05; 64} = -1.669$$



[Tabla T Student](#)

4.2432

Concluimos: Como se Acepta la H_0 ; «Estamos 95% seguros, que la motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco»

Caso de Estudio: Prueba de Hipótesis

Los resultados parciales de su Ficha de Cuestionario se muestra a continuación:

H_0 : La motivación no influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco.

H_1 : La motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco.

NUNCA	CASI NUNCA	A VECES	CASI SIEMPRE	SIEMPRE
1	2	3	4	5

$$H_0: \mu \leq 3$$

$$H_1: \mu > 3$$

$$\alpha = 5\%$$

$$n = 65$$

$$\bar{X} = 4.1$$

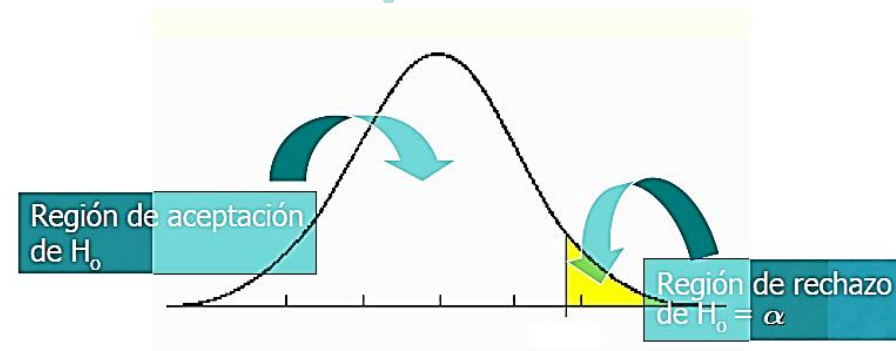
$$S = 0.19$$

Population variance unknown

$$\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

$$t_{n-1, \alpha} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

sample size: n
d.f.: n-1



Caso de Estudio: Prueba de Hipótesis con Media Poblacional

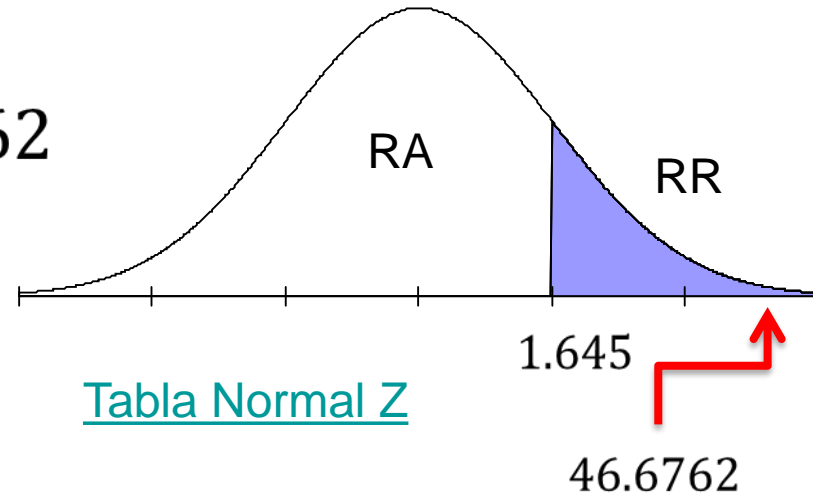
Como se desconoce la varianza de la población, y se tiene 65 muestras, entonces utilizaremos el estadístico de prueba Z:

$$H_0: \mu \leq 3$$

$$Z = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{4.1 - 3}{0.19 / \sqrt{65}} = 46.6762$$

$$\alpha = 0.05$$

$$Z_{\alpha} = Z_{0.05} = 1.645$$



Concluimos: Como se Rechaza la H_0 y se Acepta la H_1 ; «Estamos 95% seguros, que la motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco»

Caso de Estudio: Prueba de Hipótesis con Media Poblacional

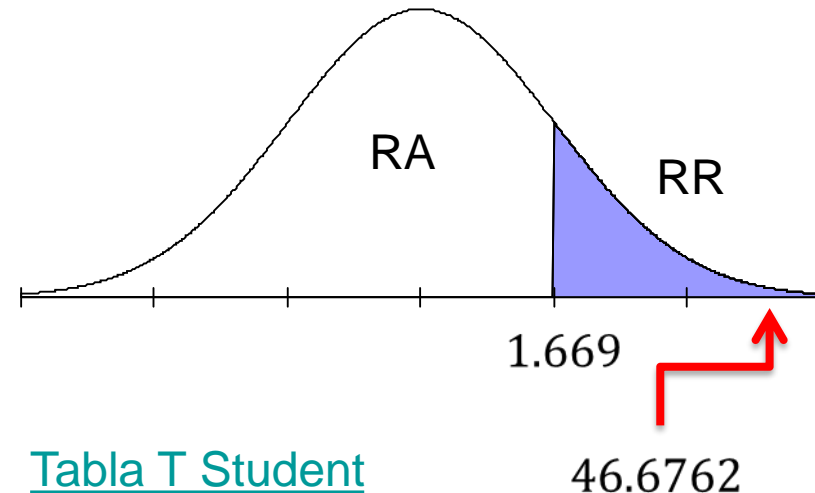
Como se desconoce la varianza de la población, y se tiene 65 muestras, entonces utilizaremos el estadístico de prueba T:

$$H_0: \mu \leq 3$$

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{4.1 - 3}{0.19 / \sqrt{65}} = 46.6762$$

$$\alpha = 0.05 \quad n = 65 \quad gl = n - 1 = 64$$

$$t_{\alpha; gl} = t_{0.05; 64} = 1.669$$



Concluimos: Como se Rechaza la H_0 y se Acepta la H_1 ; «Estamos 95% seguros, que la motivación influye positivamente en el desempeño laboral de los trabajadores de construcción civil del Distrito de Huánuco»

Tarea N° 01

- 1) Defina el tema de su tesis y el problema a solucionar.
- 2) Elabore una Tabla de Variables, Dimensiones e Indicadores.
- 3) Realice una **Ficha de Encuesta** sobre su propuesta de Tesis.
- 4) Simule datos para llenado de registros de encuestados en Excel, con la finalidad de Probar su Hipótesis utilizando la Distribución Normal y de T-Student (ambos métodos).



¡Gracias!