Prueba de hipótesis

Análisis estadístico utilizando R

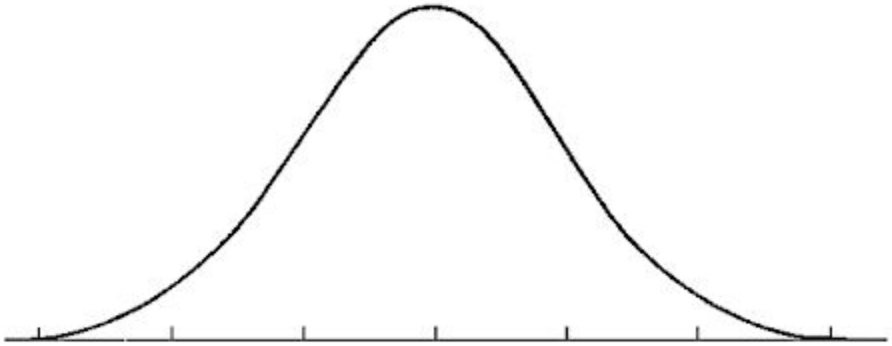
UNQ UNTreF CONICET

Pablo Etchemendy

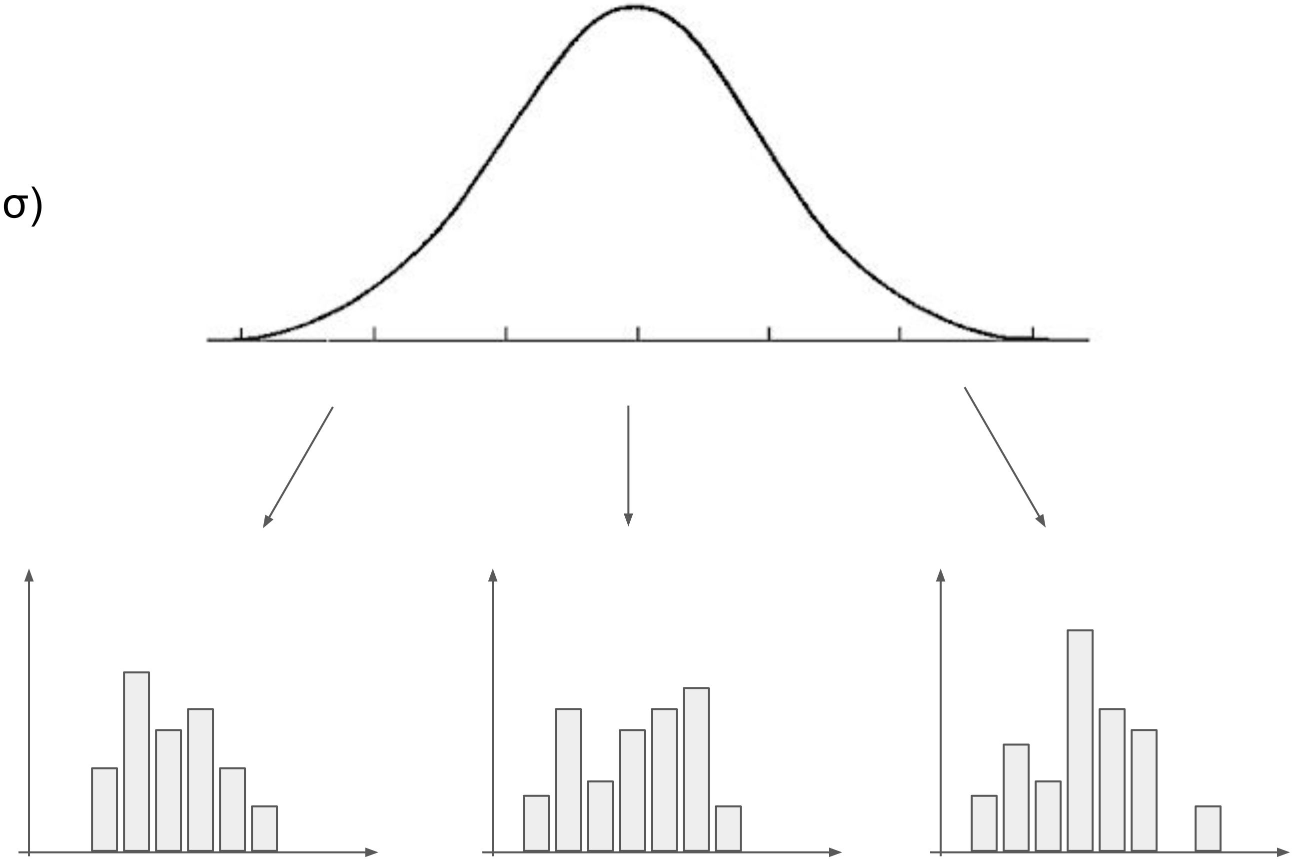
Ignacio Spiousas

Agosto 2021

$\mathcal{N}(\mu, \sigma)$



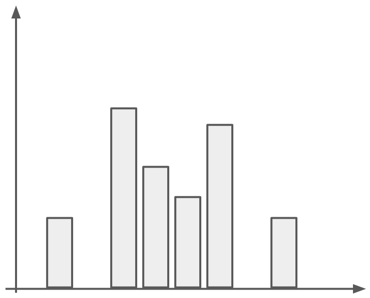
$\mathcal{N}(\mu, \sigma)$



Muestras de
tamaño n

Problema

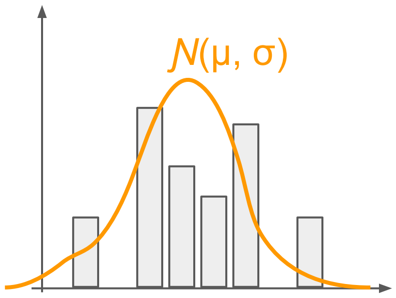
Me dan esta muestra



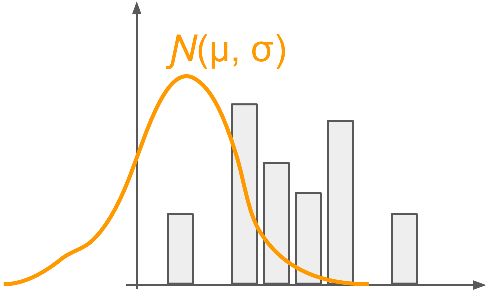
Quiero saber si fue generada por la distribución

$\mathcal{N}(\mu, \sigma)$

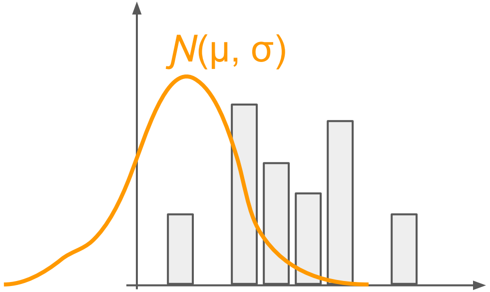
Puedo comparar “a ojo”



Puedo comparar “a ojo”



Puedo comparar “a ojo”



Pero es mejor si puedo dar una respuesta cuantitativa

Un camino posible

- Tomar una distribución de media conocida μ
- Generar muchísimas muestras de tamaño n

Un camino posible

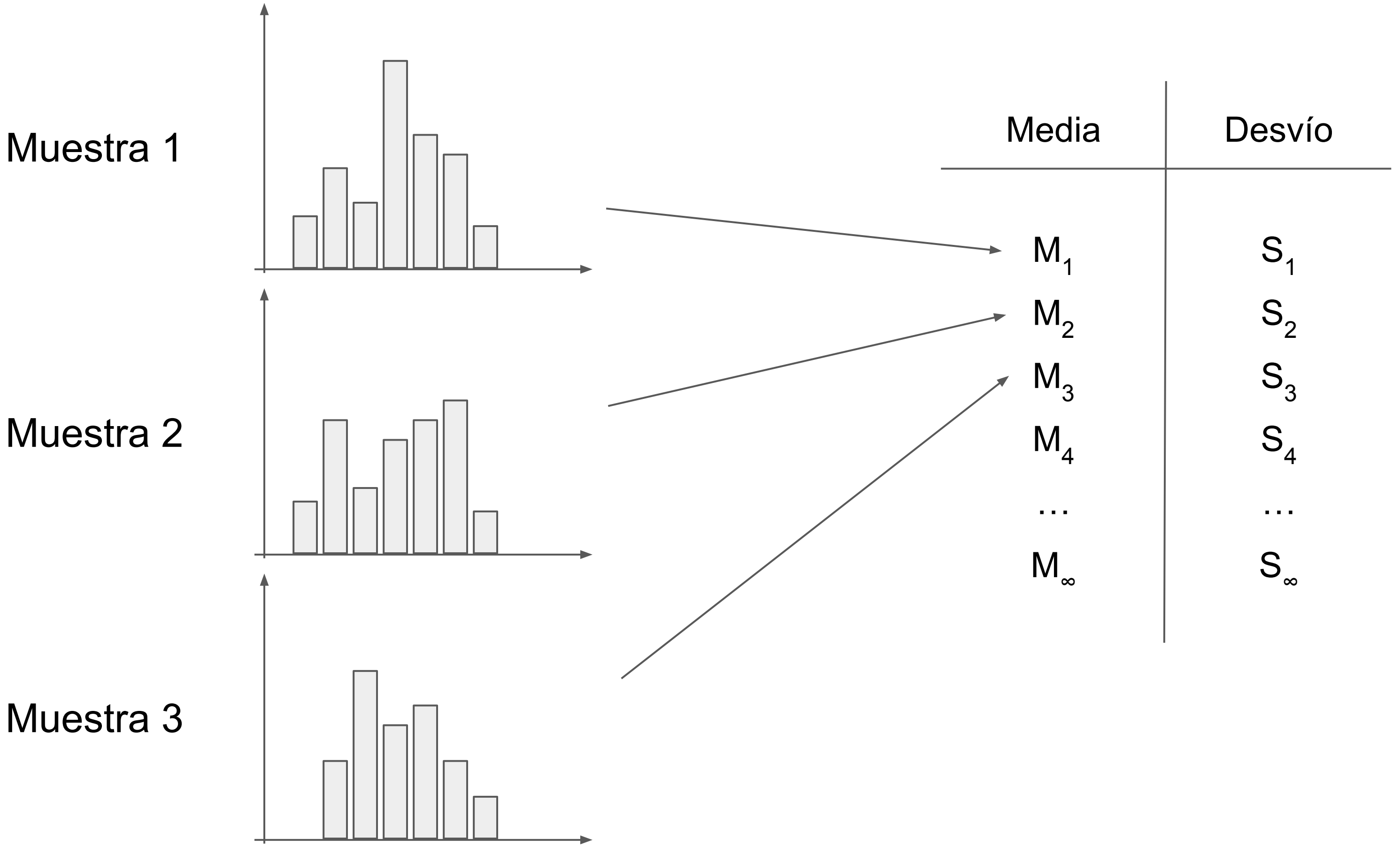
- Tomar una distribución de media conocida μ
- Generar muchísimas muestras de tamaño N
- Para cada una, calcular su media y su desvío estándar muestral

Un camino posible

- Tomar una distribución de media conocida μ
- Generar muchísimas muestras de tamaño N
- Para cada una, calcular su media y su desvío estándar muestral
- Obtener la distribución de media y desvío muestrales

Un camino posible

- Tomar una distribución de media conocida μ
- Generar muchísimas muestras de tamaño N
- Para cada una, calcular su media y su desvío estándar muestral
- Obtener la distribución de media y desvío muestrales
- De ese modo, podemos conocer la probabilidad de que la media muestral tenga un diferencia de cierto tamaño respecto a la media de la distribución


Muestra 1
Muestra 2
Muestra 3
Media
Desvío
M_1
M_2
M_3
M_4
…
M_∞
S_1
S_2
S_3
S_4
…
S_∞

Para cada muestra, voy a calcular

$$t = \frac{M - \mu}{S/\sqrt{n}}$$

¿Por qué usamos el cociente entre M - μ y S?

Recordemos que no conocemos la distribución.

Para saber si el desvío de M respecto a μ es grande, usamos al desvío estándar como referencia

Como no conocemos σ, usamos la mejor estimación que tenemos (S)

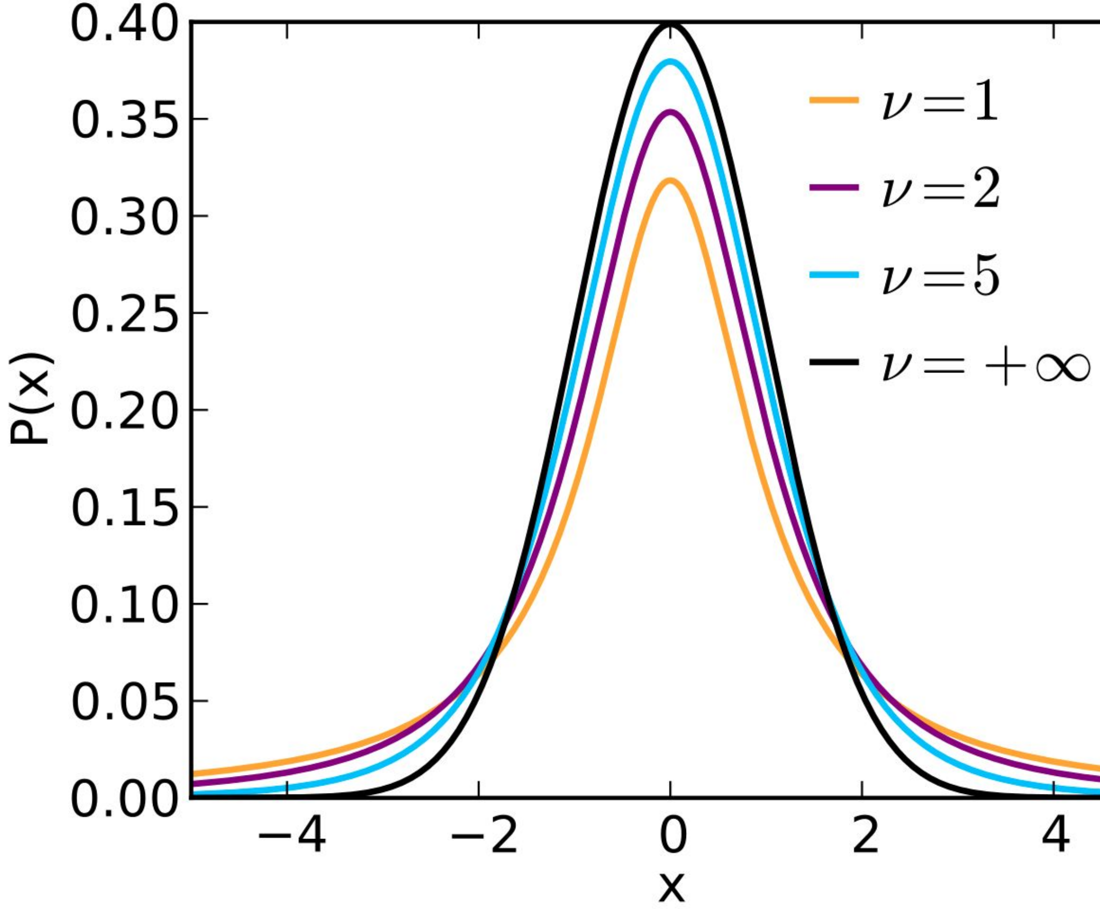
¿Cómo es la distribución de t?

Es en realidad una familia de distribuciones
(ya que depende del tamaño de cada muestra)

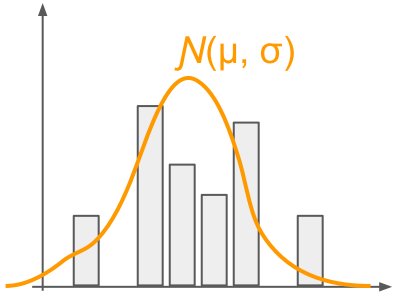
Se la representa como:

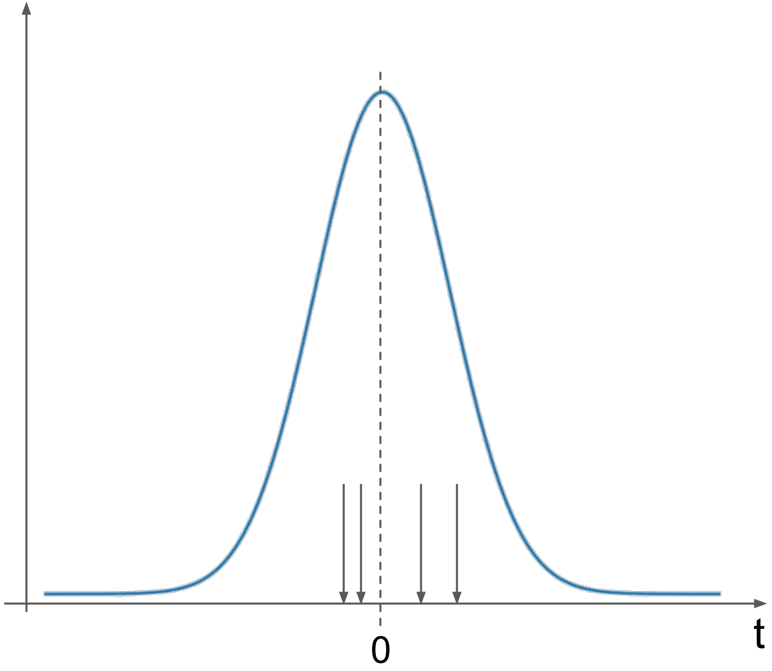
t_v

donde $v = n - 1$
(son los grados de libertad)

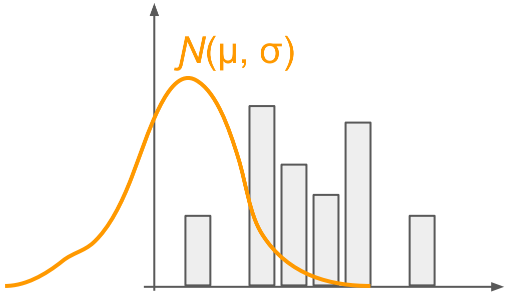

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu = 1$
$\nu = 2$
$\nu = 5$
$\nu = +\infty$

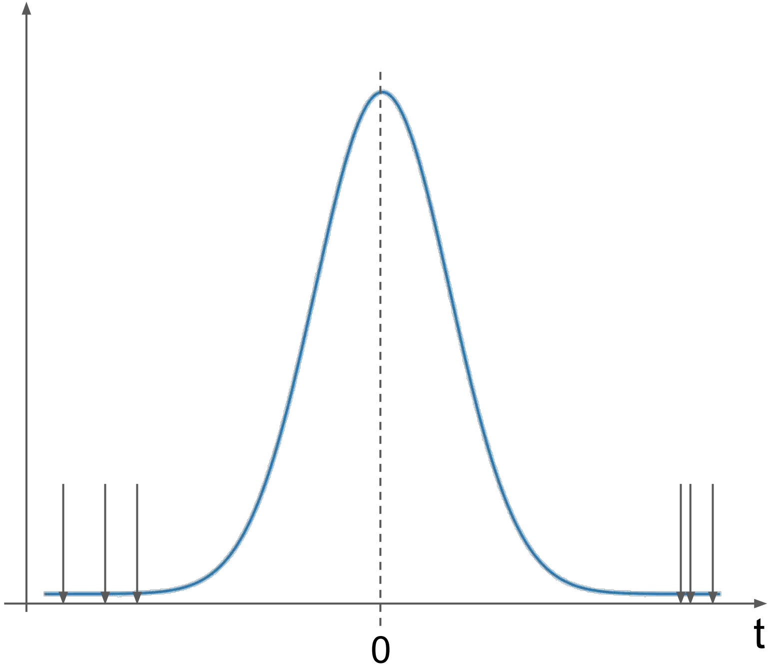
M y μ similares → t cercano a 0





M y μ diferentes → t lejano de 0





¿Qué probabilidad tengo de obtener un valor de t grande?

Definamos “grande”...

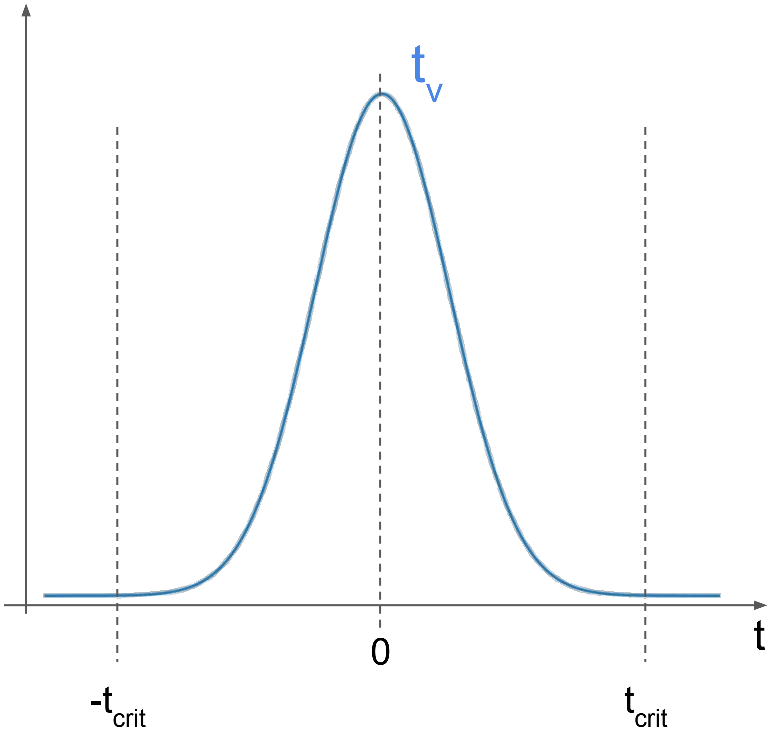
Partiendo de *N*(μ, σ), es posible obtener muestras que tengan

- M - μ grande
- S pequeño

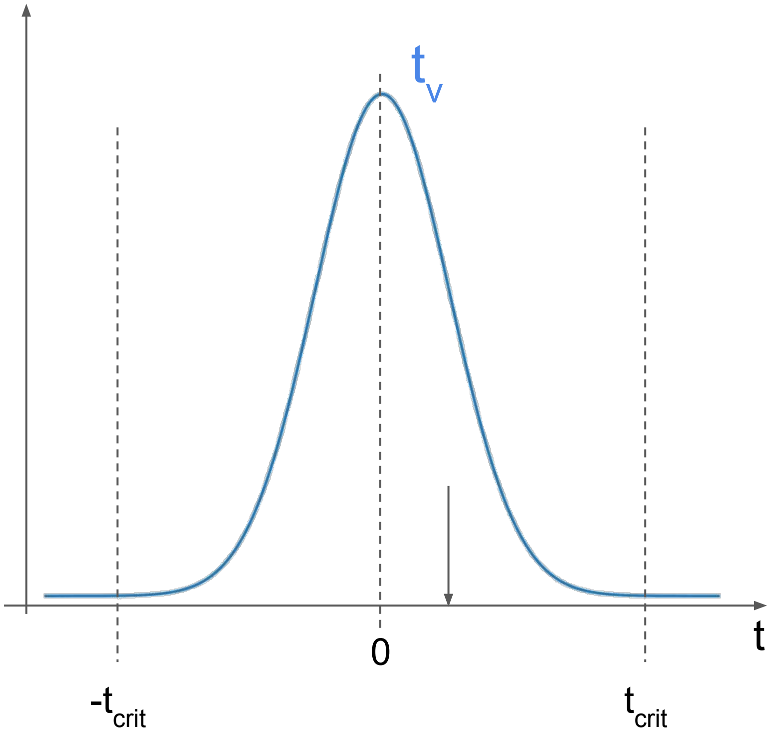
El t asociado será lejano a 0 y por lo tanto, serán identificadas como **no generadas** por *N*(μ, σ)

Este es un error que cometeremos, inevitablemente, un % de veces

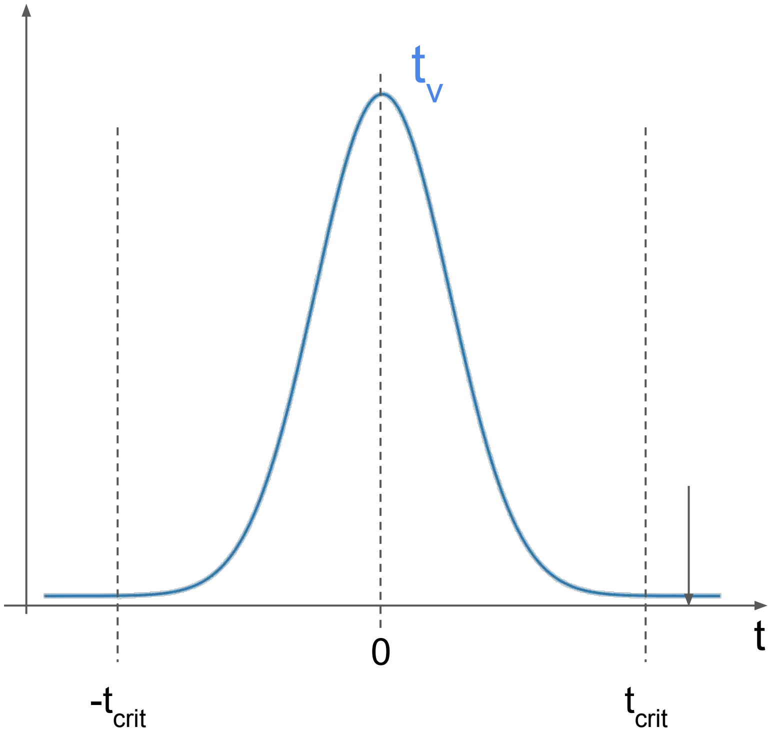
Definimos un valor de t crítico:



Si t está comprendido dentro de (-tcrit, tcrit), consideraremos plausible que la muestra provenga de $\mathcal{N}(\mu, \sigma)$



En el caso contrario, consideraremos no plausible que la muestra provenga de $\mathcal{N}(\mu, \sigma)$



Proviene de $\mathcal{N}(\mu, \sigma)$

t **cercano** a 0

a distancia **menor** que t_{crit}

Aceptamos que
proviene de $\mathcal{N}(\mu, \sigma)$

Proviene de $\mathcal{N}(\mu, \sigma)$

t cercano a 0

a distancia **menor** que t_{crit}

Aceptamos que proviene de $\mathcal{N}(\mu, \sigma)$

t lejano a 0

a distancia **mayor** que t_{crit}

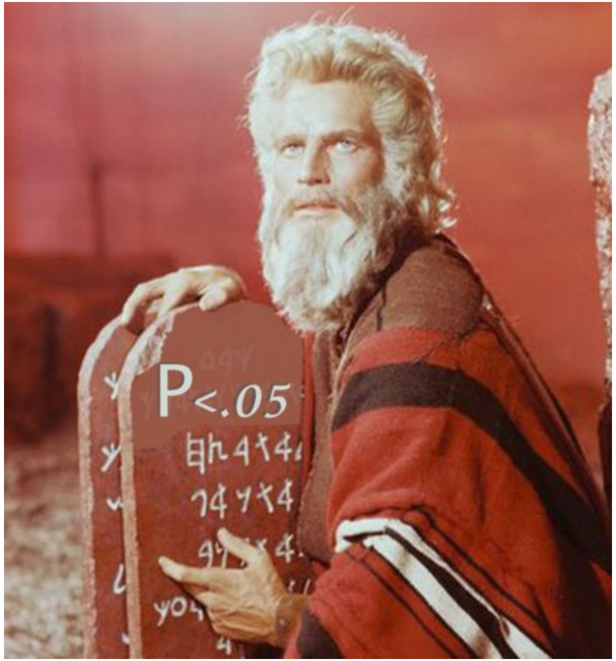
Rechazamos que proviene de $\mathcal{N}(\mu, \sigma)$

FALSO POSITIVO

Para definir t_{crit}, debemos elegir el % de falsos positivos que estamos dispuestos a aceptar

Comúnmente se lo representa con la letra α

En general, se emplea $\alpha = 5\%$



RA Fisher (1926)

"...If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty or one in a hundred. **Personally, the writer prefers to set a low standard of significance at the 5 per cent point**, and ignore entirely all results which fails to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment **rarely** fails to give this level of significance..."

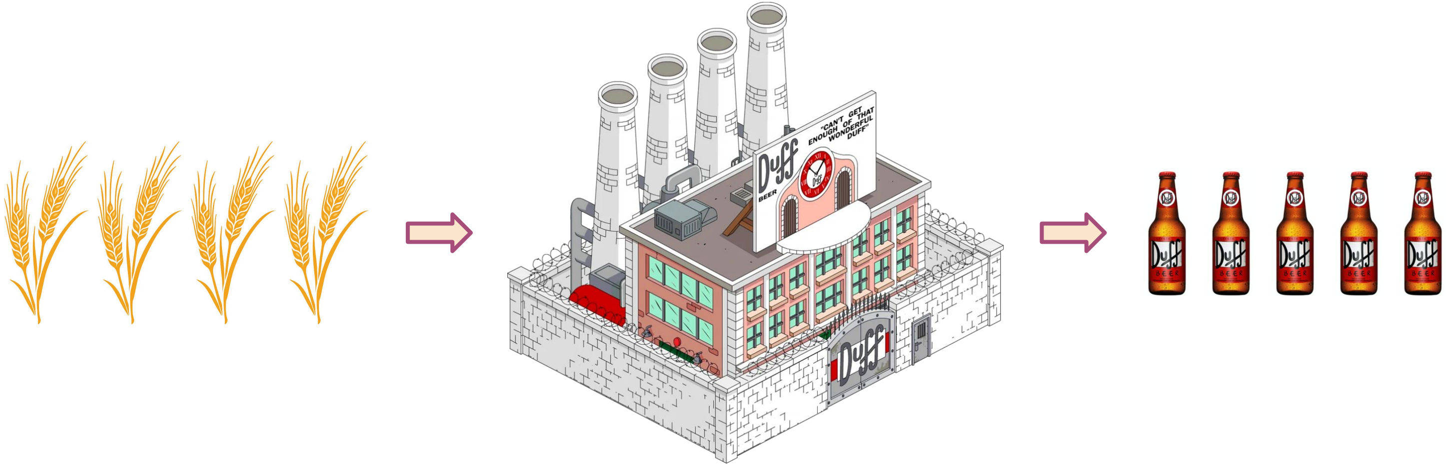
Fisher RA. The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain 1926; 33:503-513.

Nos queda ver

- ¿Cómo se calcula t_{crit} a partir de α?
- ¿Cuál es la probabilidad de obtener mi muestra a partir de $\mathcal{N}(\mu, \sigma)$
- ¿Qué ocurre si mi muestra es generada por una distribución diferente? Por ejemplo, $\mathcal{N}(\mu', \sigma)$

PARTE 2

¿Por qué hacemos esto?


"CAN'T GET ENOUGH OF THAT WONDERFUL DUFF"
Duff
BEER

Tengo variabilidad en la calidad de la materia prima

Y por lo tanto, también en la calidad del producto

Tengo un lote bueno

Tengo un lote malo

Quiero obtener una
muestra pequeña

Y decidir si es representativa de
un lote bueno o malo

M - µ

Lote de origen →

$t < t_{crit}$

FALSO NEGATIVO

$t > t_{crit}$

FALSO POSITIVO

W. S. Gosset
(aka Student)

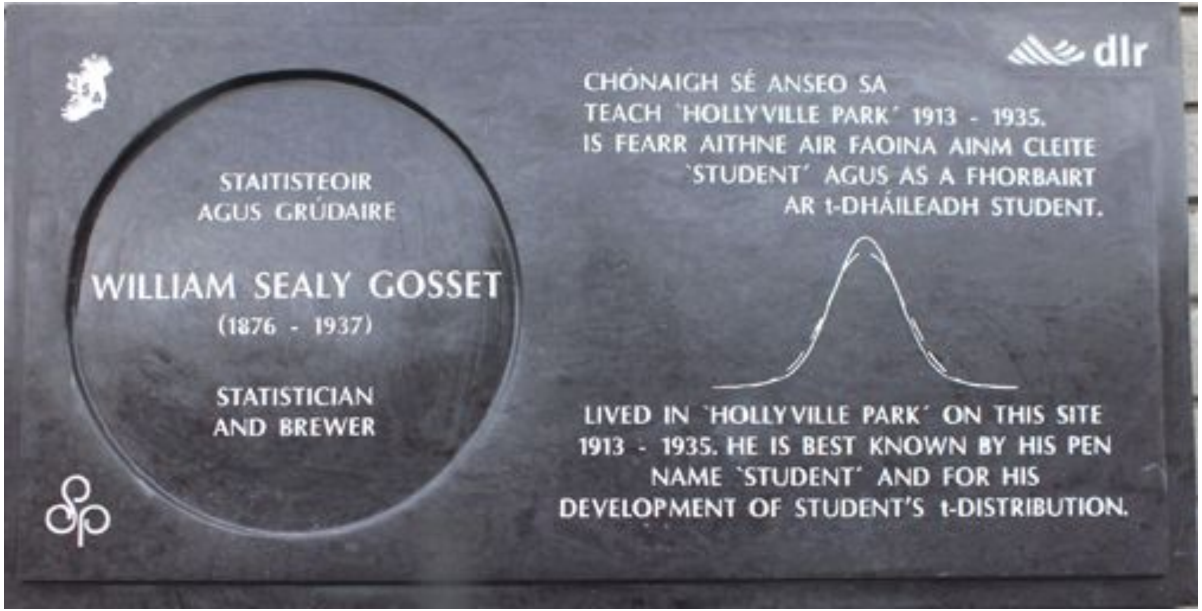
VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.


Chemist & Statistician
WILLIAM SEALY
GOSSET
1876-1937
Chief Brewer
Student 't' test


dlr
STAITISTEOIR
AGUS GRÚDAIRE
WILLIAM SEALY GOSSET
(1876 - 1937)
STATISTICIAN
AND BREWER
CHÓNAIGH SÉ ANSEO SA
TEACH `HOLLYVILLE PARK' 1913 - 1935.
IS FEARR AITHNE AIR FAOINA AINM CLEITE
`STUDENT' AGUS AS A FHORBAIRT
AR t-DHÁILEADH STUDENT.
LIVED IN `HOLLYVILLE PARK' ON THIS SITE
1913 - 1935. HE IS BEST KNOWN BY HIS PEN
NAME `STUDENT' AND FOR HIS
DEVELOPMENT OF STUDENT'S t-DISTRIBUTION.

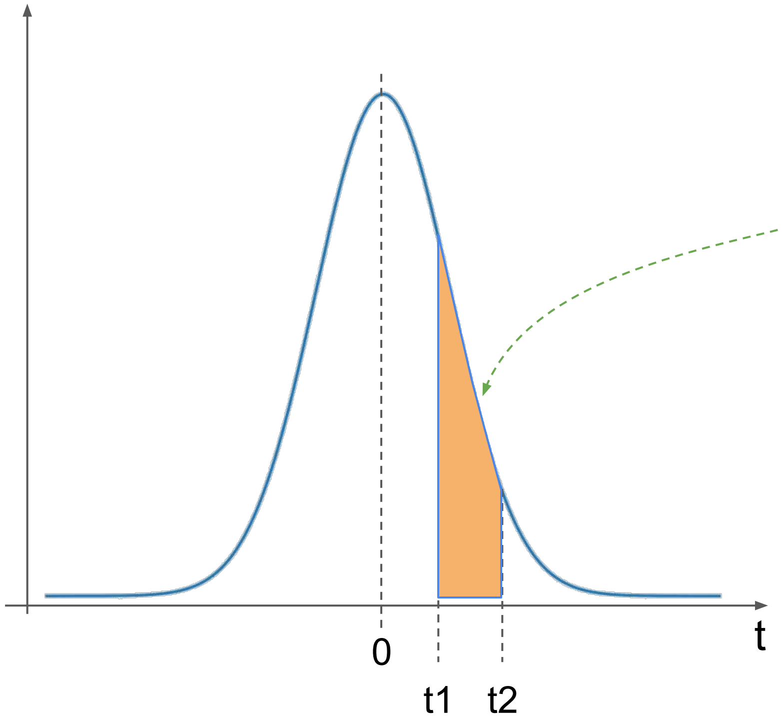
PARTE 3

¿Cómo se obtiene t_{crit}?

Cálculo de t_{crit} a partir de α

- α no tiene por qué ser 5% (o 0.05)
- Para el cálculo, vamos a usar la densidad de probabilidad de la distribución t
- Y también su función de probabilidad acumulada

¿Para qué sirve la función densidad?

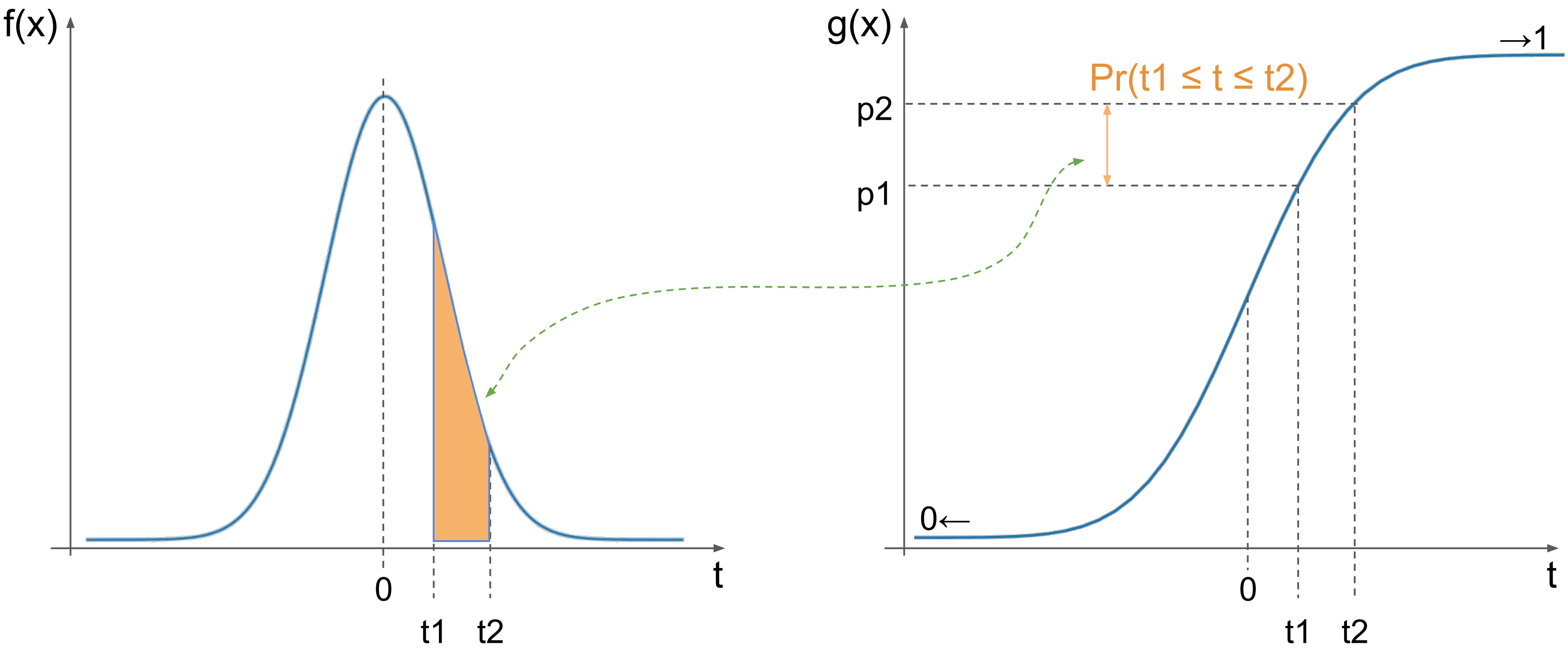


¿Cuál es la probabilidad de obtener un valor en el intervalo [t1, t2]

Pr(t1 ≤ t ≤ t2)

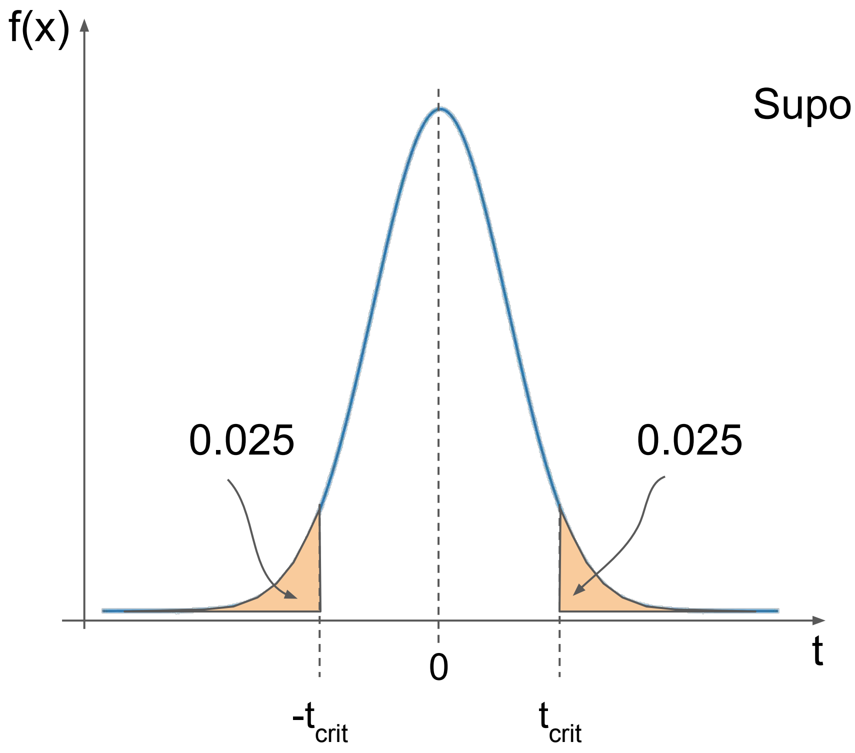
Es el área bajo la curva

Y para esto nos sirve la probabilidad acumulada

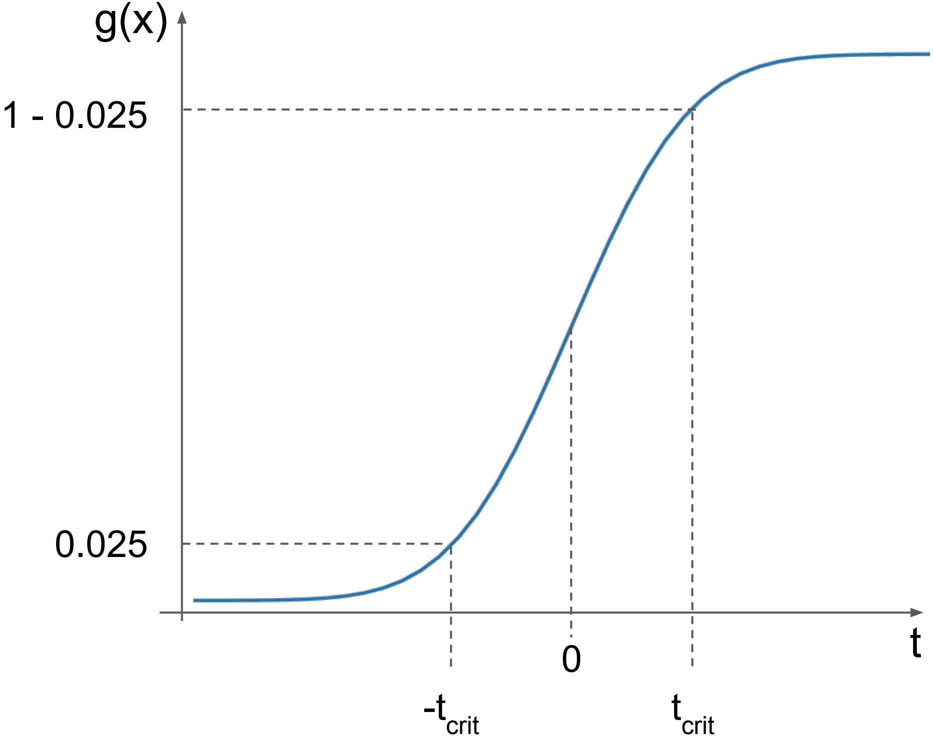


$$
\begin{aligned}
Pr(t1 \le t \le t2) &= p2 - p1 \\
&= g(t2) - g(t1)
\end{aligned}
$$

Apliquemos esto a nuestro caso



Supongamos α = 0.05



$g(t_{crit}) = 0.975$

$g(-t_{crit}) = 0.025$

Generalizando

$g(t_{crit}) = 1 - \alpha/2$

$g(-t_{crit}) = \alpha/2$

Despejemos

¿Cuánto vale t_{crit}?

$$g(t_{crit}) = 1 - \alpha/2 \longrightarrow t_{crit} = h(1 - \alpha/2)$$

$$g(-t_{crit}) = \alpha/2 \longrightarrow t_{crit} = -h(\alpha/2)$$

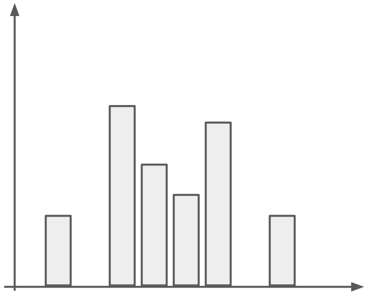
h es la función de probabilidad acumulada **inversa**

PARTE 4

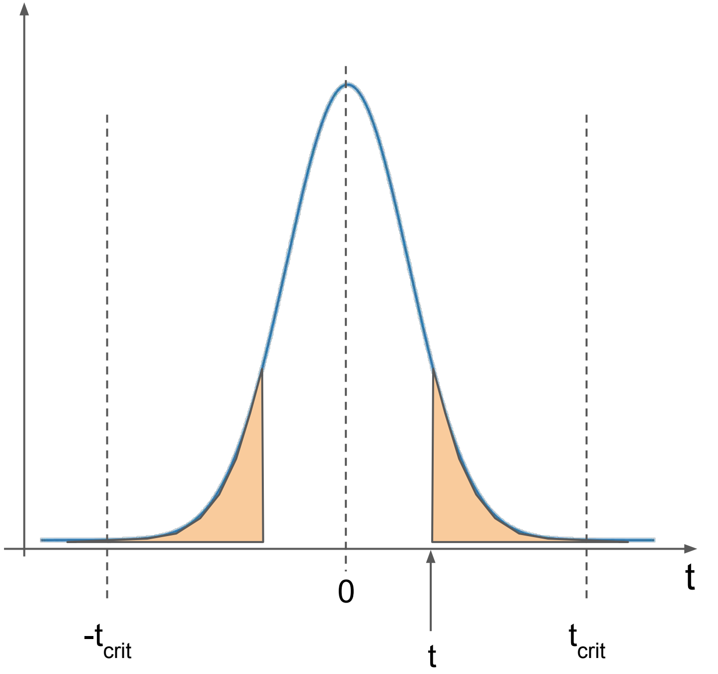
P-values

¿Qué tan probables son nuestros datos?

Suponiendo que provienen de $N(\mu, \sigma)$



Calculamos t



Probabilidad de obtener un t igual o más extremo a ambos lados

La probabilidad de obtener datos tan o más extremos que los nuestros...

...es el famoso p-value (o p-valor)

$$p = 2 \times g(-t)$$

El p-value:

- **NO es** la probabilidad de que los datos provengan (o no provengan) de $\mathcal{N}(\mu, \sigma)$

El p-value es una afirmación sobre los datos en relación a una posible explicación de su origen

No nos dice nada acerca de la explicación

El p-value:

- **NO indica** el tamaño o relevancia de la diferencia observada

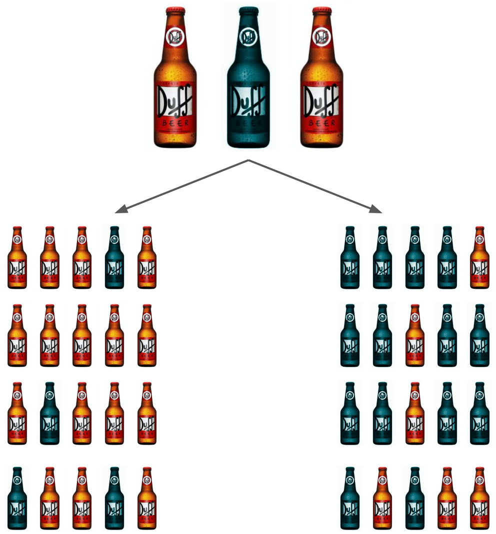
Esto se debe a que dependen del tamaño de la muestra

Para esto, es necesario una medida que sea independiente de n

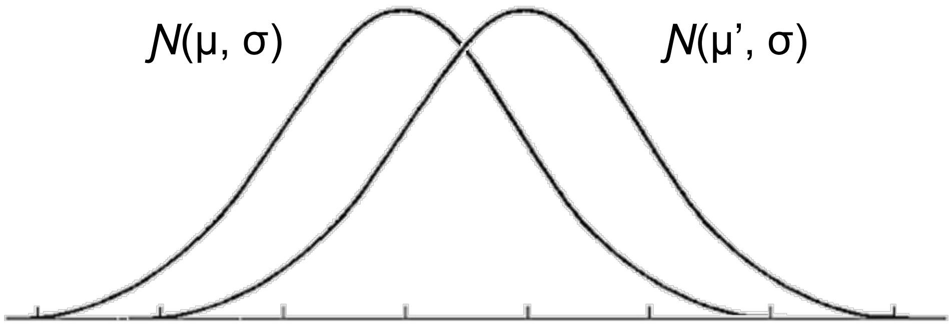
PARTE 5

Potencia estadística

Dada una muestra...



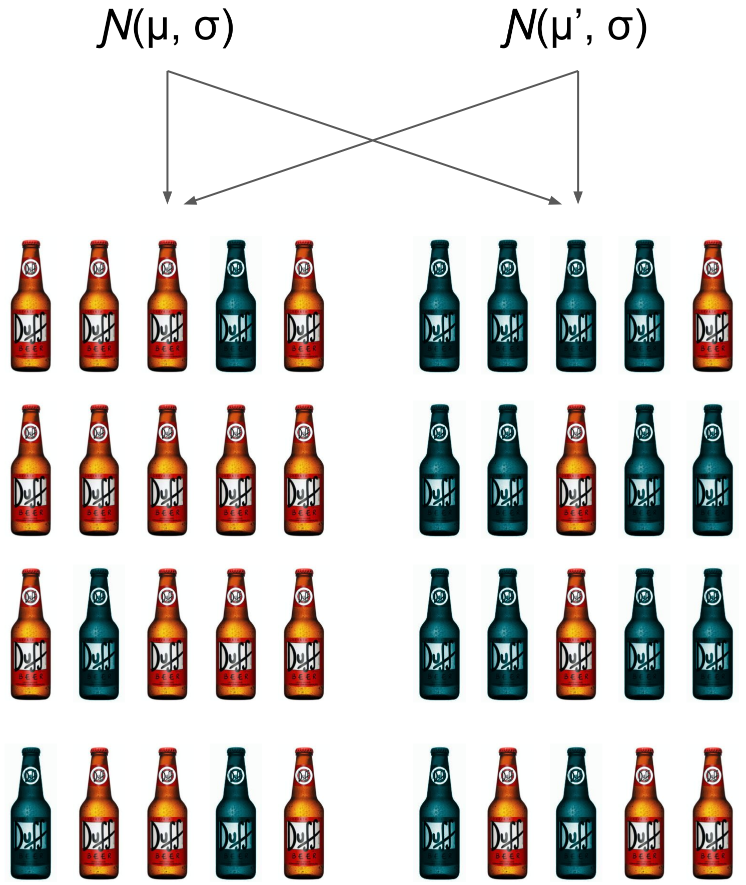
¿Qué seguridad tengo sobre mi decisión?



Hipótesis <u>nula</u> (H_0): El lote proviene de $N(\mu, \sigma)$

Hipótesis <u>alternativa</u> (H_1): El lote proviene de $N(\mu', \sigma)$

μ' es un valor relevante en función de nuestros fines


$\mathcal{N}(\mu, \sigma)$
$\mathcal{N}(\mu', \sigma)$

1. Obtengo mi muestra
2. Calculo la estadística t usando μ como referencia
3. Calculo el p-value asociado y comparo con α

¿Cuál es la probabilidad de obtener $p < \alpha$?

- Si los datos fueron generados por $\mathcal{N}(\mu, \sigma)$, es α (falso positivo)
- ¿Si en cambio fueron generados por $\mathcal{N}(\mu', \sigma)$?

  Depende de μ, μ’, σ y el tamaño de muestra

Potencia estadística

- Se define como la probabilidad de obtener $p < \alpha$ cuando los datos fueron generados por la distribución alternativa.
- Obtener $p < \alpha$ en un experimento con *baja* potencia **NO** es garantía de *replicabilidad* del resultado en un estudio futuro.
- Mayor tamaño de muestra → mayor potencia
- La probabilidad complementaria (falso negativo) se suele llamar β.
    - La potencia es $1 - \beta$

¿Cómo ajustamos la potencia?

- Conociendo:
  - La diferencia entre μ y μ'
  - La variabilidad σ
- Calculamos el tamaño de muestra n para que la potencia $(1 - \beta)$ tenga un valor “aceptable”

¿Cómo se calcula?

- Mediante simulaciones
- Mediante expresiones analíticas (si existen)

¿Con qué parámetros?

- Es necesario conocer μ y μ' y σ
    - μ y μ': informados por la teoría
    - σ: teoría, experimentos previos, o experimento piloto

PARTE 6

Tamaño del efecto

¿Qué es el tamaño del efecto?

(aka, *effect size*)

- Muchas veces se utiliza (erróneamente) al p-value como indicación de la relevancia de una diferencia.
- Aparte del problema de inconsistencia entre la definición del p-value y esta interpretación, el p-value, pensado como distribución estadística, depende del tamaño de la muestra.
- Sin embargo, la diferencia entre μ y μ' no depende del tamaño de la muestra

¿Qué es el tamaño del efecto?

(aka, *effect size*)

- Se define una medida que sea independiente del tamaño muestral.
- En el ejemplo que estamos viendo:

$$d = (M - \mu) / S \quad \text{(basada en la muestra)}$$

$$d = (\mu' - \mu) / S \quad \text{(teórico)}$$

- Conocido como d de Cohen (por Jacob Cohen).
- Existen otros “effect size” apropiados para diferentes tipos de estudios y pruebas.

Front. Psychol., 26 November 2013 | https://doi.org/10.3389/fpsyg.2013.00863

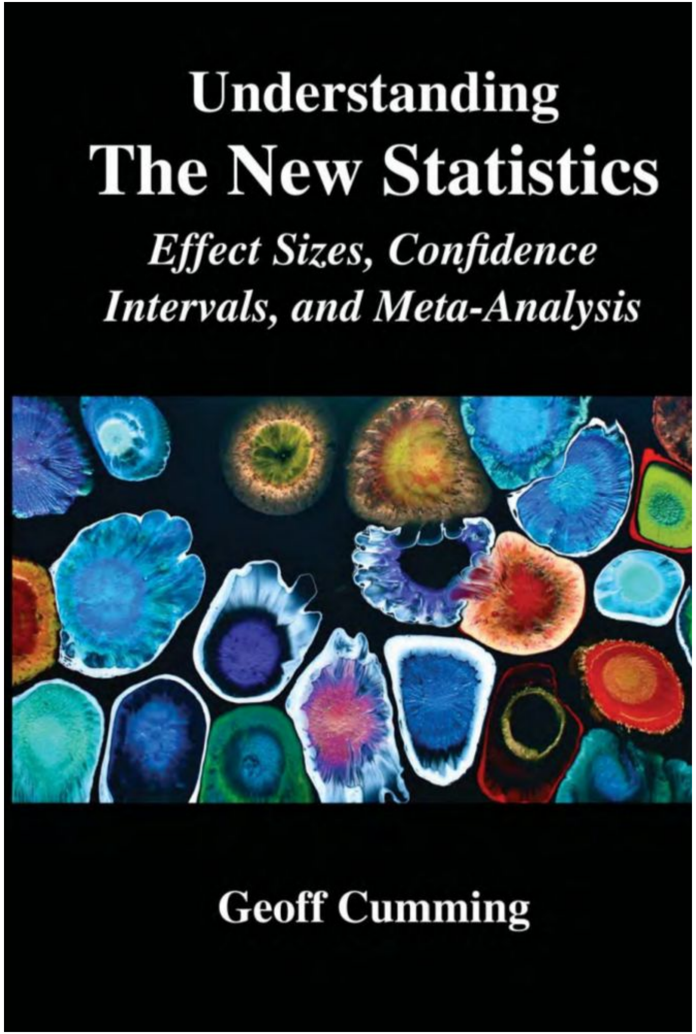
Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs

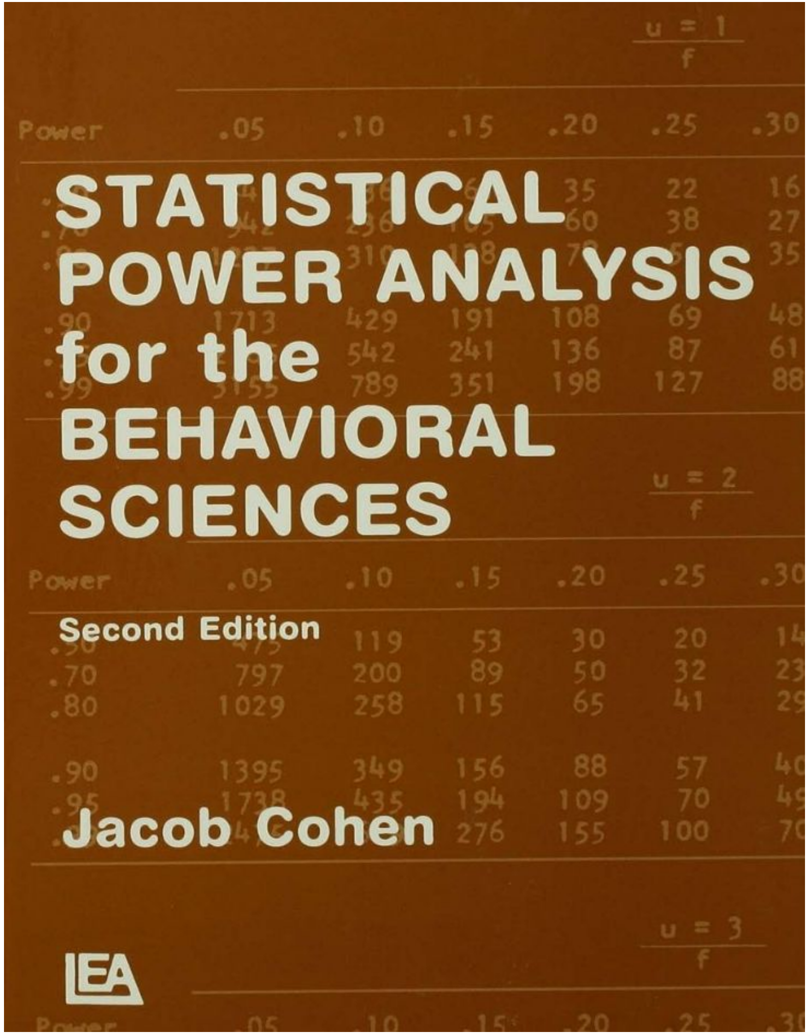
Daniël Lakens*

Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, Netherlands

Effect sizes are the most important outcome of empirical studies. Most articles on effect sizes highlight their importance to communicate the practical significance of results. For scientists themselves, effect sizes are most useful because they facilitate cumulative science. Effect sizes can be used to determine the sample size for follow-up studies, or examining effects across studies. This article aims to provide a practical primer on how to calculate and report effect sizes for *t*-tests and ANOVA's such that effect sizes can be used in a-priori power analyses and meta-analyses. Whereas many articles about effect sizes focus on between-subjects designs and address within-subjects designs only briefly, I provide a detailed overview of the similarities and differences between within- and between-subjects designs. I suggest that some research questions in experimental psychology examine inherently intra-individual effects, which makes effect sizes that incorporate the correlation between measures the best summary of the results. Finally, a supplementary spreadsheet is provided to make it as easy as possible for researchers to incorporate effect size calculations into their workflow.

https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00863/full


Understanding
The New Statistics
Effect Sizes, Confidence
Intervals, and Meta-Analysis
Geoff Cumming


STATISTICAL
POWER ANALYSIS
for the
BEHAVIORAL
SCIENCES
Second Edition
Jacob Cohen
LEA