Introducción al análisis cuantitativo de datos lingüísticos

Bloque 2.1:
Comparación de 2 grupos: t-test

Ezequiel Koile (MPI-SSH)
Carolina Gattei (IFIBA - CONICET)

Motivación

- Tenemos mediciones para dos grupos, y queremos saber si las diferencias son significativas
  - ¿L@s estudiantes de Puan son más alt@s que l@s de Exactas?
  - ¿Las palabras de mayor frecuencia de uso se codifican diferente en nuestro sistema cognitivo?
  - ¿Los cultivos del campo A son de mayor calidad que los del campo B?
  - ¿La gente reacciona de manera positiva a este nuevo medicamento?
- Entramos aquí en el dominio de la **estadística inferencial**
  - Esta "habla" de la población entera: no de una simple muestra: los resultados pueden extrapolarse más allá de nuestra muestra.
  - Dará lugar a resultados *replicables* por otros investigadores que utilicen otras muestras

Procedimiento

- PRIMERO: ¡visualizar los datos!
  - Graficar: histograma, boxplot, QQ-plot,…
- Luego, elegir el test adecuado
- Finalmente, analizar los resultados

Tipos de tests

- Paramétrico / No paramétrico:
  - ¿Qué tipo de variable? ¿Qué tipo de distribución?
- Independiente / Dependiente (no apareado) / (apareado):
  - ¿Las observaciones de ambos grupos están relacionadas?
- Una cola / Dos colas:
  - ¿La hipótesis es direccional?

t-test de Student

- William Sealy Gosset, estadístico de Oxford
- Contratado por la cervecería Guiness (entre otros estadísticos y bioquímicos) para mejorar los procesos industriales en la planta.

- Desarrolla el *t*-test como una forma de monitoriar la calidad (rendimiento) de la cebada usada en la producción de cerveza stout.
- Por razones de confidencialidad, los investigadores debían publicar sus resultados con seudónimos

- Por esto, en su paper de 1908, el Guille Gosset firmó como “Student”.

t–test de Student

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

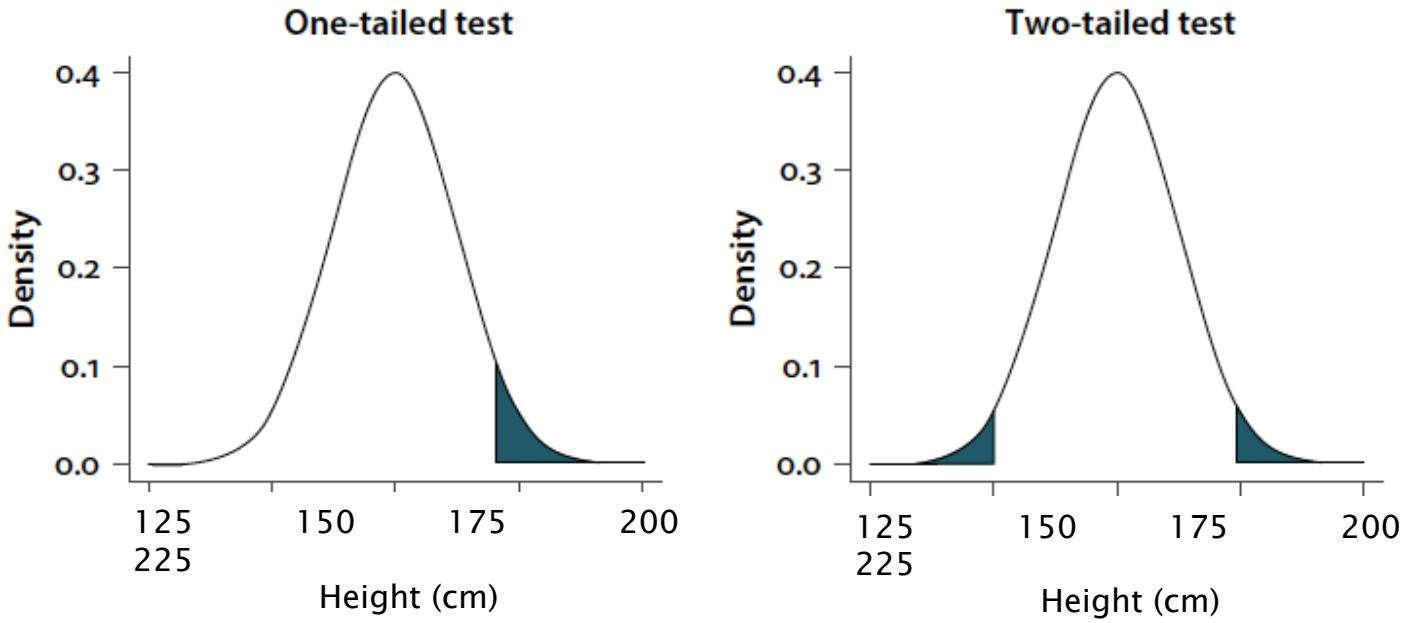
If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution

¿En qué sabores viene?

- Es un test paramétrico (sirve para variables numéricas)
- Puede ser *independiente* (*no apareado*) o *dependiente* (*apareado*):
- Puede ser de *dos colas (hipótesis no direccional)* o de *una cola (hipótesis direccional)*

Test de hipótesis

- Tests estadísticos de una y dos colas
  - Dependen de si la hipótesis es direccional o no (respectivamente)
  - ¡La distinción es relevante! El límite de significancia (digamos 0.05 o 5%) va de un lado o se divide en dos!



Test de hipótesis

- ¿Direccional o no direccional? (Levshina 2015)

A. H_0 (the null hypothesis): There is no difference in the number of lexemes that denote snow in Eskimo and Yucatec Maya.
   H_1 (the alternative hypothesis): There are more lexemes that denote snow in Eskimo than in Yucatec Maya.
B. H_0 (the null hypothesis): there is no relationship between the frequency of a word and how fast it is recognized in a lexical decision task.
   H_1 (the alternative hypothesis): the more frequent a word, the faster it is recognized in a lexical decision task.
C. H_0 (the null hypothesis): there is no difference in the relative frequencies of metaphoric expressions used by men and women when they speak about sex.
   H_1 (the alternative hypothesis): there is a difference in the relative frequencies of metaphoric expressions used by men and women when they speak about sex.

¿Cómo funciona el t-test?

- Definición y ejemplos
  ← ¡Pizarrón!

¿Cómo funciona el t-test?

$$t = \sqrt{\frac{n}{2}} \cdot \frac{\overline{x_A} - \overline{x_B}}{SD}$$

$$(n_1 = n_2)$$

$$(SD_1 = SD_2)$$

$$df = 2n - 2$$

```
t.test()
```

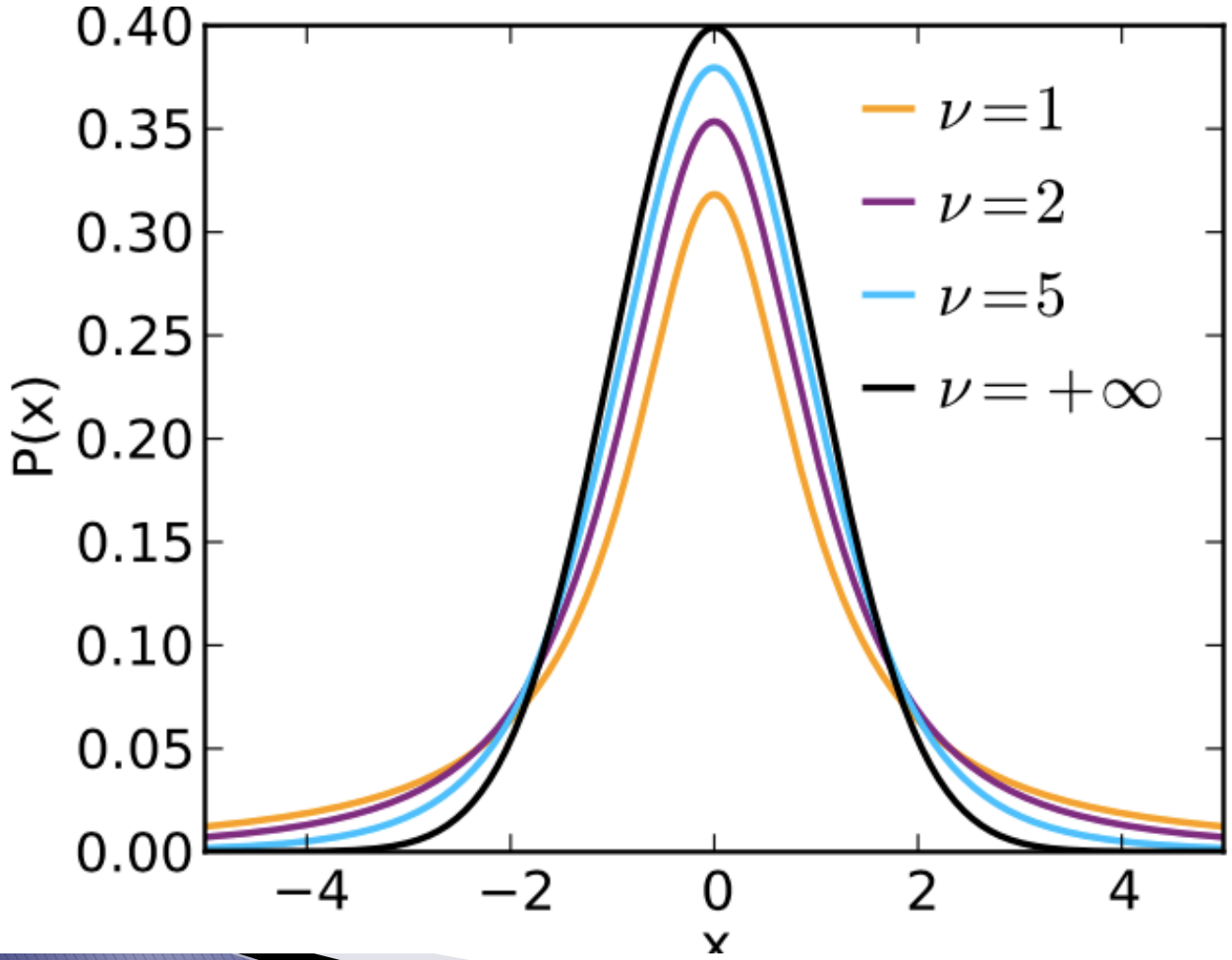
```
t.test(x, y, alternative = c("two.sided", "less", "greater"),
paired = FALSE, ...)
```

Analyze this

t Table

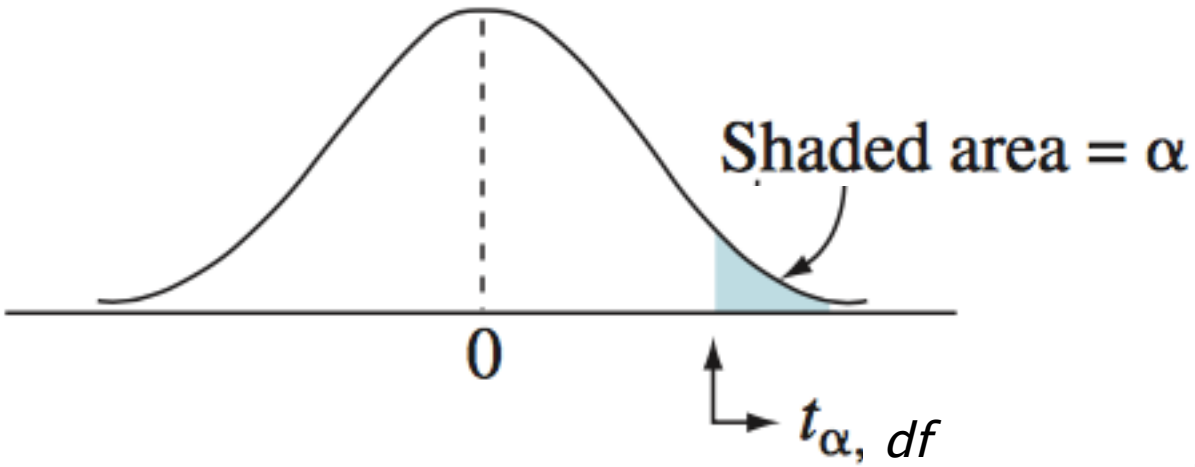
cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

La distribución *t* de Student



Intervalos de confianza

- Intervalo de confianza del 95%
  - *Si repitiéramos el estudio y el análisis una gran cantidad de veces, entonces el 95% de los intervalos calculados contendrá el valor "verdadero"*



t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Tamaño del efecto (effect size)

- Cohen's *d*
  - `cohen.d {effsize}`

$$t = \sqrt{\frac{n}{2}} \cdot \frac{\overline{x_A} - \overline{x_B}}{SD}$$

$$d = \frac{\overline{x_A} - \overline{x_B}}{SD}$$

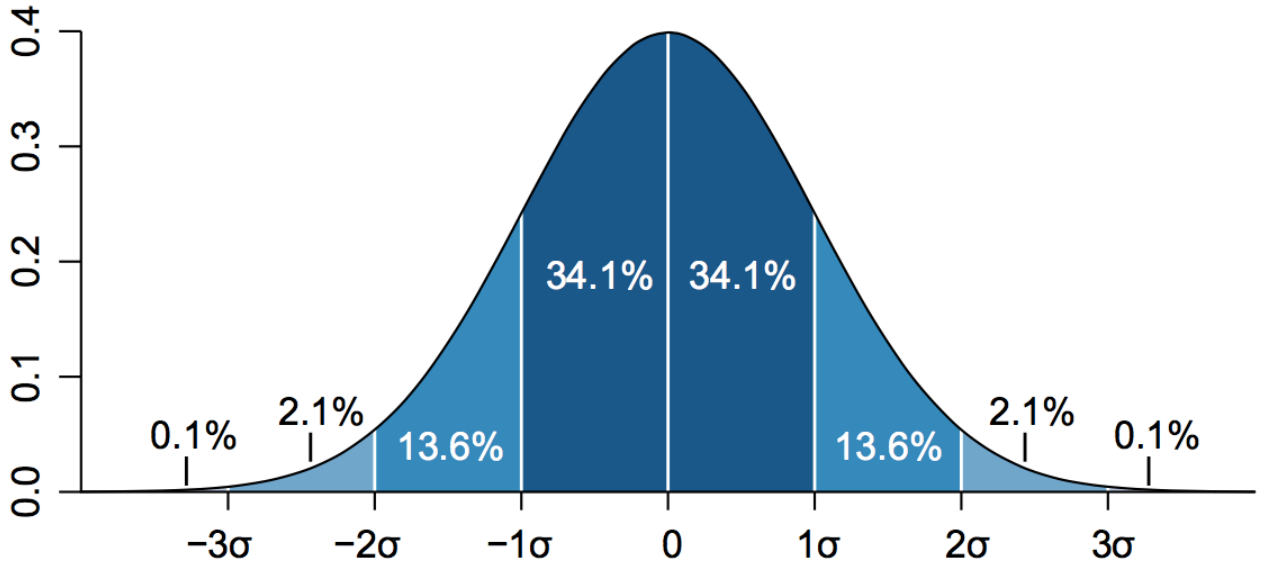
$$(n_1 = n_2)$$

$$(SD_1 = SD_2)$$

Effect size	***d***
Very small	0.01
Small	0.20
Medium	0.50
Large	0.80
Very large	1.20
Huge	2.0

La distribución normal

- Distribución normal
- Distribución gaussiana
- Campana de Gauss

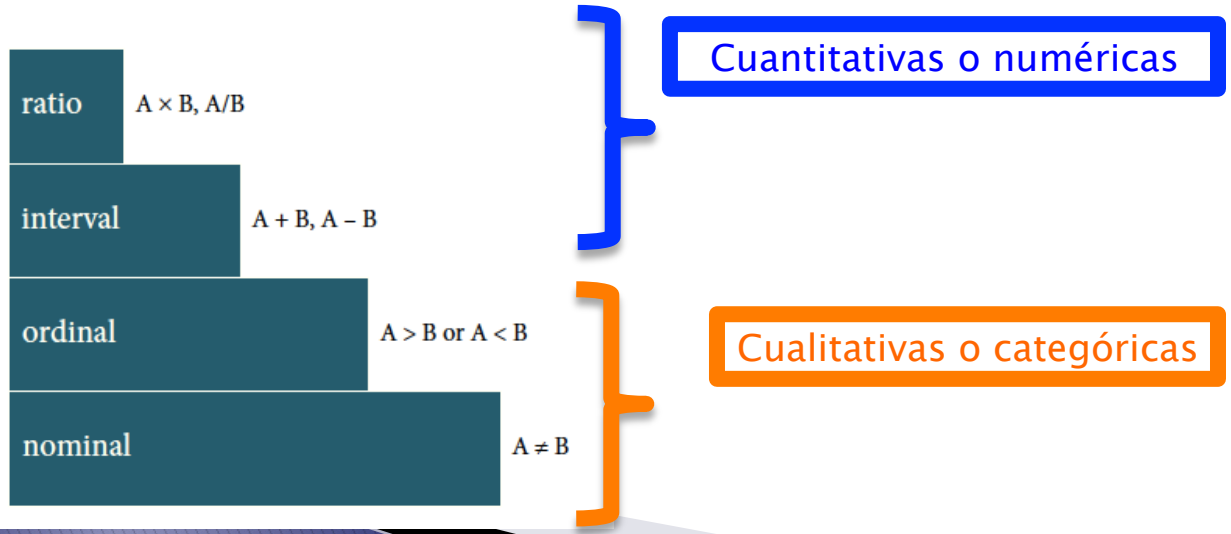


Test no paramétricos: *U*–test

- *U*–test de Mann–Whitney–Wilcoxon
- Usado para variables ordinales (no numércias) o datos distribuidos de manera no normal.

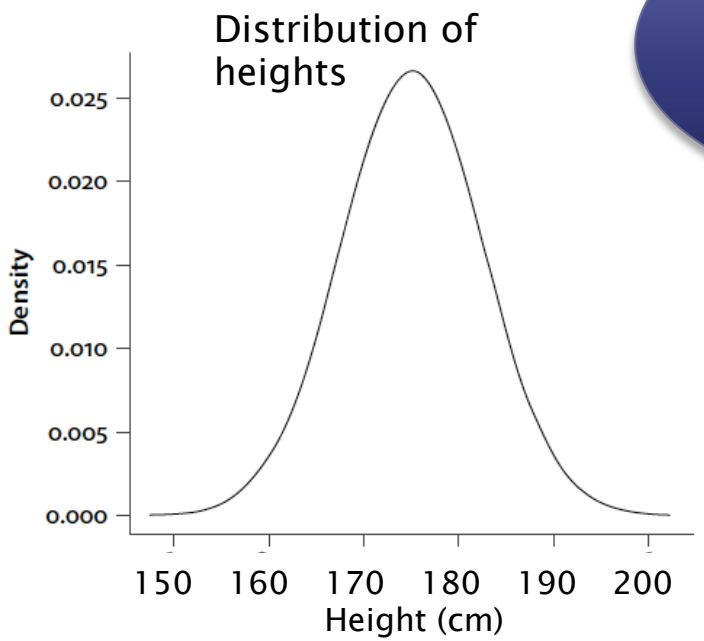
¿Cuándo usar cada test?

Requisitos	*t* – test	*U* – test
Muestras tomadas aleatoriamente de la población	Sí	Sí
Observaciones independientes	Sí	Sí
Variables al menos _____	intervalo	ordinales
Distribución normal subyacente (o n > 30)	Sí	No
Las varianzas de ambos grupos son similares	Sí	No



Test de hipótesis

- Distribución de alturas puaners:

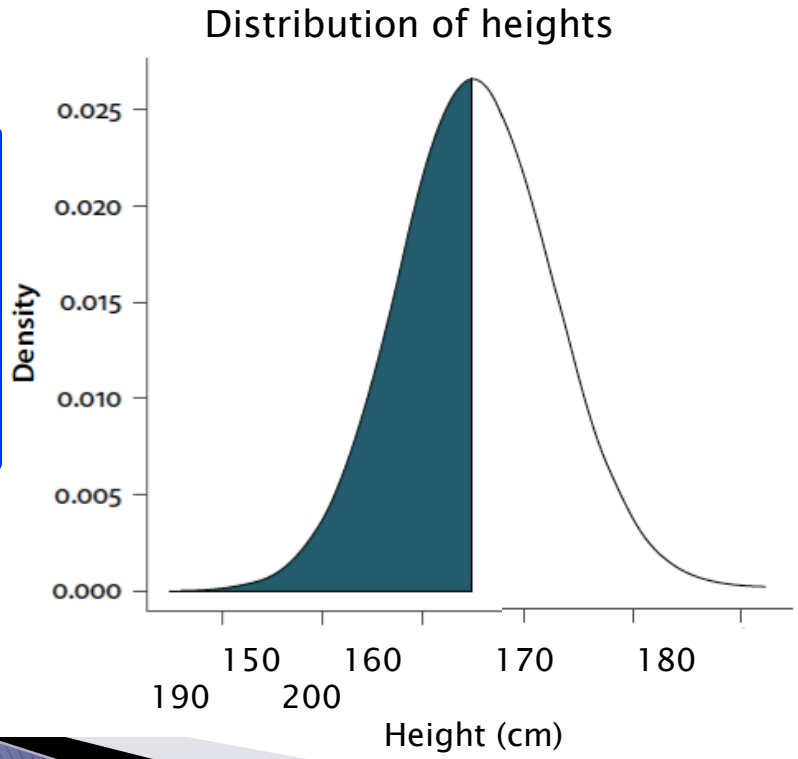


Soy una distribución normal ☺

Test de hipótesis

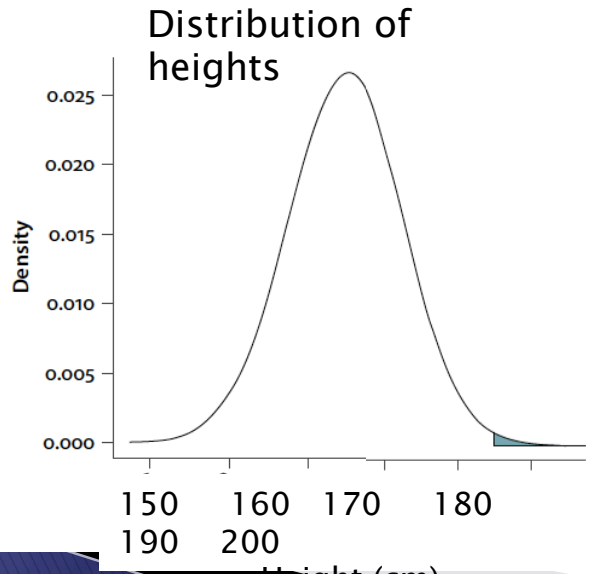
- La probabilidad de una altura por debajo de 1.75cm es 50%

50% del área total



Test de hipótesis

- Conozco a alguien de Puan que mide 1.95cm. La probabilidad de obtener este valor, *dado que la hipótesis (distribución de alturas) es correcta*, **es 0.023, o 2.3%**



2.3% del área total

Test de hipótesis

- Este es un *valor extremo*
- Si asumo que esta distribución es correcta, la probabilidad de obtener 1,95*m* o un valor más extremo por casualidad es 2.3%
- Este es el *p*–value

p– value: probabilidad, *dada la hipótesis nula,* de obtener el valor reportado por casualidad.

Test de hipótesis

- Lo que **no es** el *p*-value:
  - La probabilidad de que la hipótesis nula sea verdadera
  - La probabilidad de que la hipótesis alternativa sea falsa
  - La probabilidad de que los efectos observados se deban exclusivamente al azar
  - El tamaño o importancia del efecto observado
  - Cualquier otra definición que se te ocurra o veas en internet

Test de hipótesis

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Article | Talk

Misunderstandings of *p*-values

From Wikipedia, the free encyclopedia

Misunderstandings of *p*-values are an important problem in scientific research and scientific education. *P*-values are often used or interpreted incorrectly.[1] F
p-value to a significance level will yield one of two results: either the null hypothesis is rejected (which however does not imply that the null hypothesis is *false*),
hypothesis is *true*). From a Fisherian statistical testing approach to statistical inferences, a low *p*-value means *either* that the null hypothesis is true and a highly

Contents [hide]

Common misunderstandings of *p*-values [edit]

The following list corrects several common misconceptions regarding *p*-values: [1][2][3]

1. **The *p*-value is *not* the probability that the null hypothesis is true, or the probability that the alternative hypothesis is false.**[1] A *p*-value can indi
hypothesis). Specifically, the *p*-value can be taken as the prior probability of an observed effect given that the null hypothesis is true--which should not b
prosecutor's fallacy). In fact, frequentist statistics does not attach probabilities to hypotheses.
2. **The *p*-value is *not* the probability that the observed effects were produced by random chance alone.**[1] The *p*-value is computed under the assun
the relation of the data to that hypothesis, not a statement about the hypothesis itself.[1]
3. **The 0.05 significance level is merely a convention.**[2] The 0.05 significance level (alpha level) is often used as the boundary between a statistically si
scientific reason to consider results on opposite sides of the 0.05 threshold as qualitatively different.[2][4]
4. **The *p*-value does not indicate the size or importance of the observed effect.**[1] A small *p*-value can be observed for an effect that is not meaningful
significant *p*-value (see effect size).

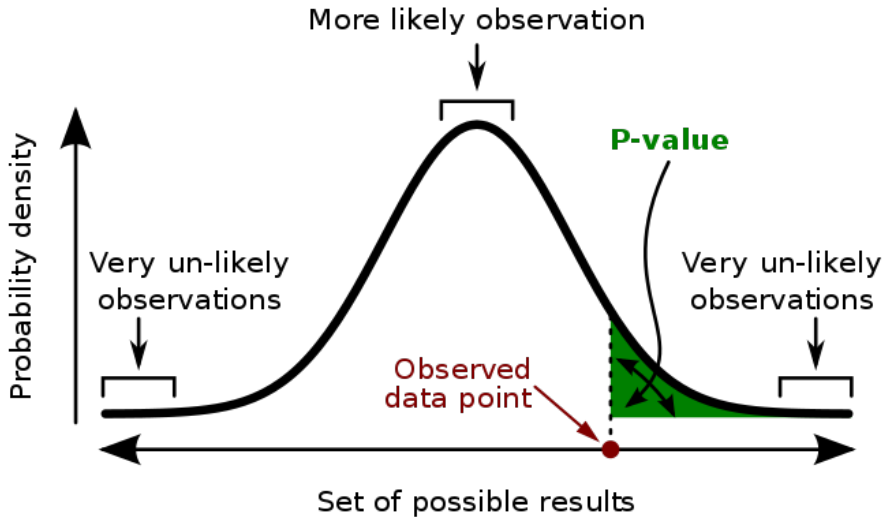
Test de hipótesis

Important:

Pr (observation | hypothesis) ≠ Pr (hypothesis | observation)

The probability of observing a result given that some hypothesis is true is *not equivalent* to the probability that a hypothesis is true given that some result has been observed.

Using the p-value as a "score" is committing an egregious logical error: **the transposed conditional fallacy.**



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Author: User: Rep[illegible]etilto @ Wikipedia ¬ User:Chen-Pan Liao @ Wikipedi

Test de hipótesis

- ¿Y si me equivoqué?

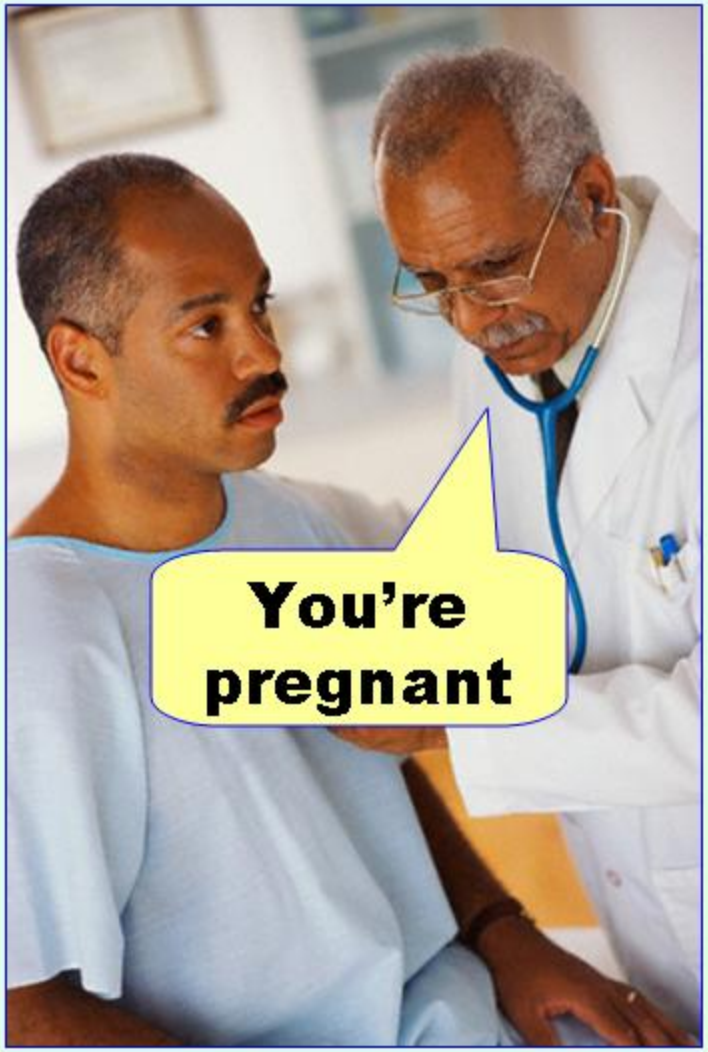
Error Tipo I

	Rechazo H_0	Conservo H_0
H_0 verdadera	Falso positivo	☺
H_0 falsa	☺	Falso negativo

Error Tipo II

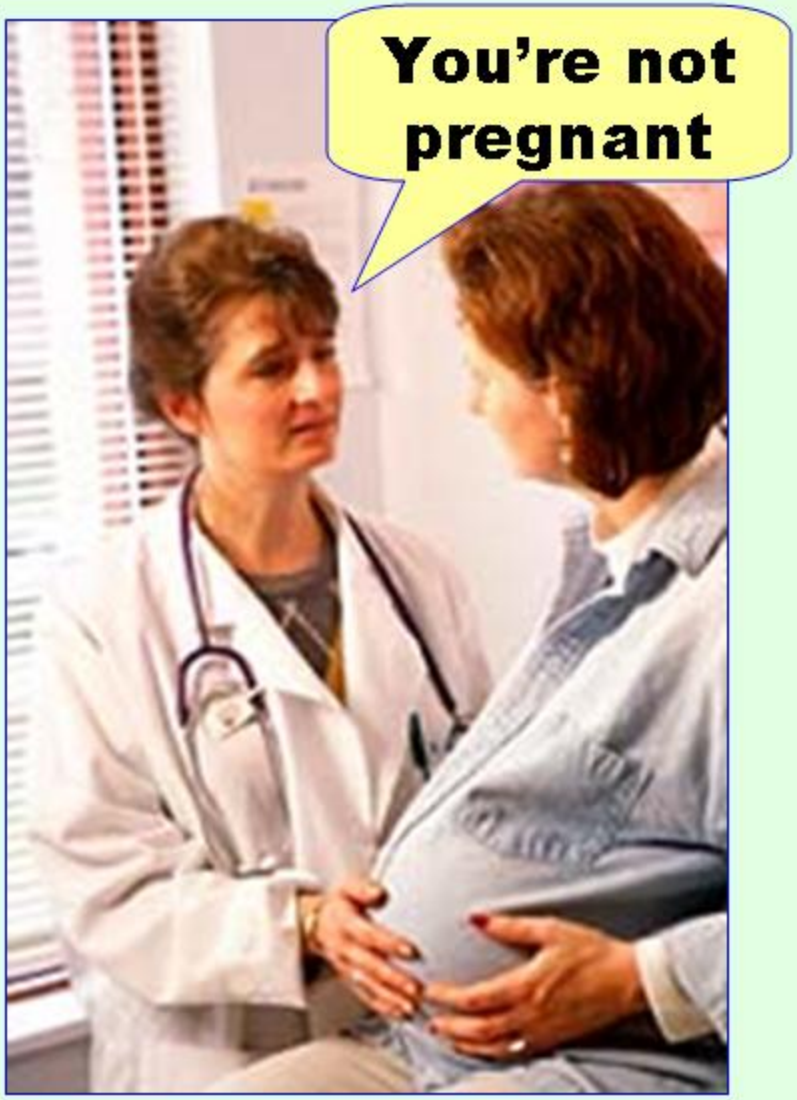
Type I error

(false positive)



Type II error

(false negative)



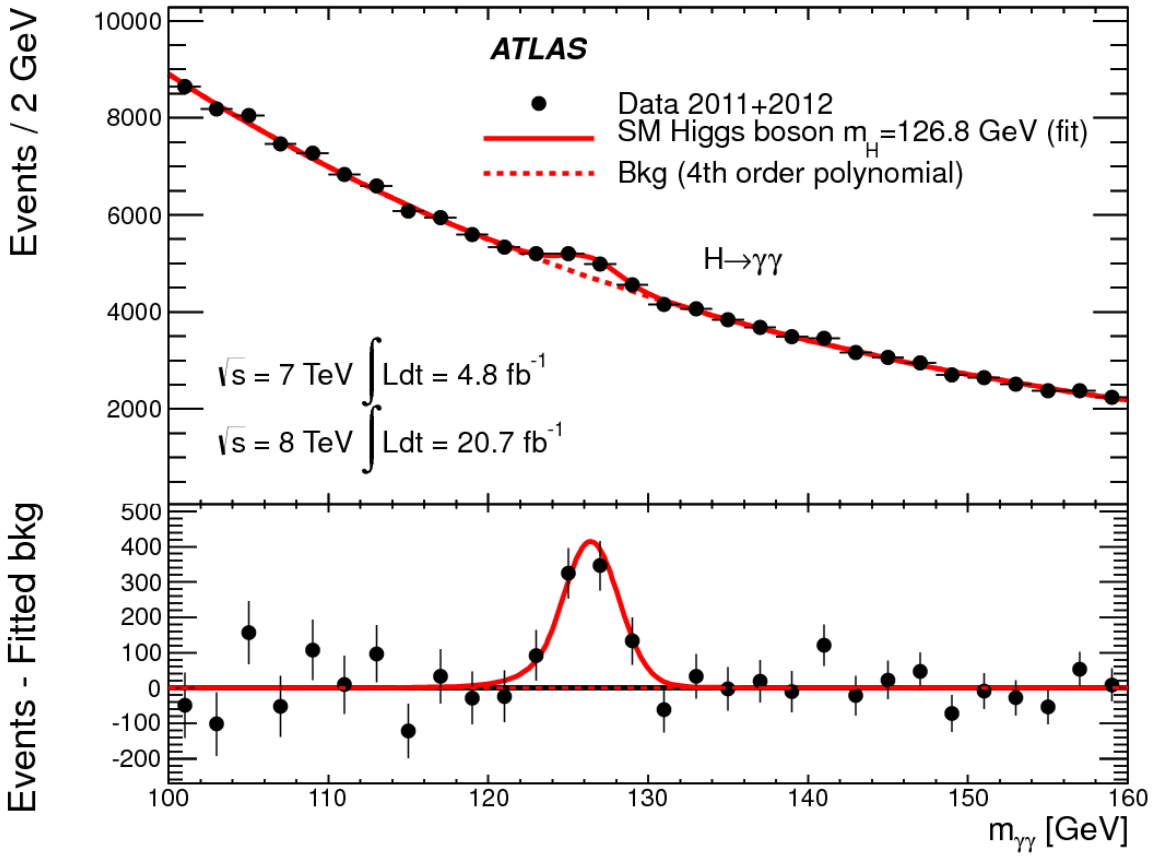
Test de hipótesis

- ***Nivel de significancia:*** es un umbral arbitrario que definimos para los *p-values*
- Si el *p-value* en nuestra observación/test es menor que nivel del significancia, rechazamos la hipótesis nula. Si es mayor, no podemos rechazarla.
- Un nivel de significancia menor (más estricto), disminuirá las chances de cometer un error tipo I, pero incrementa las chances de cometer un error tipo II.
- ¿Qué tipo de error preferirían cometer?

Test de hipótesis

- Qué significa un nivel de significancia “aceptable” depende fuertemente del campo

(*e.g.* en física de partículas, $p = 0.0000001$ (5 desviaciones estándar))



Test de hipótesis

- Qué significa un nivel de significancia “aceptable” depende fuertemente del campo

(*e.g.* en física de partículas, $p = 0.0000001$ (5 desviaciones estándar))