

# ANÁLISIS ESTADÍSTICO DE DATOS LINGÜÍSTICOS: ¿POR DÓNDE EMPEZAR?

Adrián Cabedo Nebot

## Table of contents

1	¿Quién era yo en 2006?	4
2	¿Quién soy yo en 2024?	5
3	Objetivos de la charla de hoy	6
4	Acercamientos comunes en investigación lingüística	6
5	¿En qué piensa muchas veces un alumno de lingüística sobre la estadística?	7
6	Reticencias ante el uso de datos	7
7	¿Qué puede ser también estadística en lingüística?	8
8	¿Qué puede ser también estadística en lingüística diacrónica?	8
9	¿Qué puede ser también visualización estadística en lingüística de corpus?	9
10	¿Qué es la estadística?	10
11	Utilidad	10
12	Por dónde empezar (bibliografía)	10
13	Estadística descriptiva vs inferencial	11
14	¿Qué es un dato?	11
15	¿De qué puede recogerse datos en lingüística?	12

<b>16 Ejemplo sobre un dato lingüístico: la palabra “perro”.</b>	<b>12</b>
<b>17 ¿Qué es una variable?</b>	<b>13</b>
<b>18 ¿Qué programas puedo usar para hacer estadística?</b>	<b>13</b>
<b>19 ¿Cómo recoger los datos?</b>	<b>13</b>
<b>20 Estructura de una base de datos</b>	<b>14</b>
<b>21 Bases de datos (Fonocortesía)</b>	<b>14</b>
<b>22 Exploración inicial de datos con Excel o Google Sheets</b>	<b>14</b>
<b>23 Opciones de transformación y visualización de datos</b>	<b>15</b>
<b>24 Correlaciones espúreas</b>	<b>15</b>
<b>25 ¿Qué es una población y una muestra?</b>	<b>16</b>
<b>26 Tamaño de la muestra</b>	<b>16</b>
<b>27 Hipótesis nula y alternativa</b>	<b>17</b>
<b>28 ¿Qué es un test de hipótesis?</b>	<b>17</b>
28.1 Ejemplos de hipótesis nula y alternativa . . . . .	17
28.2 Elección de pruebas estadísticas . . . . .	18
<b>29 Errores tipo I y tipo II</b>	<b>18</b>
<b>30 Paradoja de Simpson</b>	<b>18</b>
<b>31 Valor p</b>	<b>19</b>
<b>32 Distribución normal</b>	<b>19</b>
32.1 Regla de la normalidad . . . . .	20
32.2 Valor p y distribución normal . . . . .	20
<b>33 Ejemplos de valor P</b>	<b>21</b>
33.1 ¿Es la intensidad o volumen de la voz distinta en los enunciados corteses y descorteses? . . . . .	21
33.2 ¿Es la F0 o tono de la voz distinta en los enunciados corteses y descorteses? . . . . .	21
<b>34 Práctica online de pruebas estadísticas</b>	<b>22</b>
<b>35 Uso de un programa especializado (R)</b>	<b>22</b>

<b>36</b>	<b>Ejemplo de gráfico simple con R</b>	<b>23</b>
<b>37</b>	<b>¿Qué más puedo hacer con R?</b>	<b>24</b>
<b>38</b>	<b>¿Desde dónde instalo R y RStudio?</b>	<b>24</b>

Seminario diseñado para la [Asociación de Jóvenes Lingüistas \(AJL\)](#)

**Email:** [adrian.cabedo@uv.es](mailto:adrian.cabedo@uv.es)

**Celebrado:** 12 de junio de 2024

**Copyright y derechos:**

ANÁLISIS ESTADÍSTICO DE DATOS LINGÜÍSTICOS: ¿POR DÓNDE EMPEZAR? para la Asociación de Jóvenes Lingüistas (AJL) by Adrián Cabedo Nebot is licensed under CC BY 4.0



## 1 ¿Quién era yo en 2006?



Images created by an AI from OpenAI

- Estudiante de Filología Hispánica
- NO Interesado en datos
- NO Interesado en estadística
- NO Interesado en informática
- Me gusta leer literatura (Cortázar, Poe, Lovecraft, Muñoz Molina...) y escribir poesía.
- Me gusta la música (Queen, Alice Cooper, AC/DC...).
- Me gusta más la literatura que la lingüística.
- Tengo 25 años, me estoy haciendo mayor y necesito trabajo. Futuro: oposiciones.

## 2 ¿Quién soy yo en 2024?



Images created by an AI from OpenAI

- Doctor en Lengua Española (2009)
- Fui miembro de la AJL de 2006 a 2009.
- Profesor titular en la UV (desde 2019)
- Interesado en datos
- Interesado en estadística
- Interesado en informática
- Me sigue gustando leer, escribir y la música
- Tengo 17 años más y ya no soy “joven lingüista”.
- Voy a congresos y hago cosas con R.
- Escribo *papers*...

### 3 Objetivos de la charla de hoy

#### **i** Objetivos

- Introducir conceptos básicos de estadística.
- Introducir nociones de recogida de datos.
- Presentar herramientas para la exploración de datos.
- Mostrar ejemplos de aplicación de la estadística en lingüística.

### 4 Acercamientos comunes en investigación lingüística



Fuente de la imagen:

<https://www.rae.es/academico/ignacio-bosque-munoz>

- Erudición
- Hermenéutica discursiva
- Deducción
- Obra importante: coordinador de la Gramática...



Fuente de la imagen:

<https://www.rae.es/academico/guillermo-rojo>

- Erudición
- Big data
- Inducción
- Obra importante: coordinador del Corpes...

#### **i** ¿Qué decimos de estos caminos?

- Ambos acercamientos se necesitan mutuamente.
- No hay uno mejor que otro, son aproximaciones diferentes.
- Puede seguirse cada camino de manera independiente (y hacerlo muy bien, o hacerlo “menos bien”).
- El rigor científico es necesario en ambos casos.

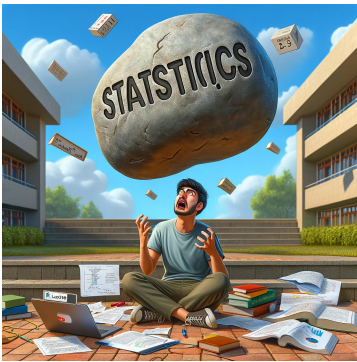
## 5 ¿En qué piensa muchas veces un alumno de lingüística sobre la estadística?

Fórmula de la chi-cuadrado:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Fórmula ANOVA:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$



Images created by an AI from OpenAI

## 6 Reticencias ante el uso de datos

### ! Reticencias ante el uso de datos

- Soy de letras.
- No sé de matemáticas.
- Todo lo de números es muy difícil y no entiendo nada.
- Los datos no sirven sin interpretación, es una manera de manipular al lector y hacer creer que un estudio es más “científico”.
- Se puede hacer en sincronía, pero no en diacronía.
- Yo no quiero programar ni usar software especializado. No se me da bien.

## 7 ¿Qué puede ser también estadística en lingüística?

Tabla 5. Distribución de funciones mitigadoras por sexo

<i>Funciones mitigadoras</i>	<i>Mujeres</i>	<i>Hombres</i>
Evitar o reducir compromiso dicho	118 (24,3 %)	121 (25,5 %)
Autoimagen	169 (34,8 %)	165 (34,8 %)
Reducir la repercusión de una generalización o efecto de lo dicho	64 (13,2 %)	63 (13,3 %)
Manifiestar y/o buscar acuerdo	26 (5,3 %)	30 (6,3 %)
Reparar o mitigar una amenaza al oyente	52 (10,7 %)	51 (10,8 %)
Evitar imponer el yo	36 (7,4 %)	25 (5,3 %)
Justificar desacuerdo o evitar un previsto desacuerdo	18 (3,7 %)	19 (4 %)

Fuente: Albelda Marco, Marta. «Variación sociolingüística de los mecanismos mitigadores: diferencias de uso en edad y sexo». *Cultura, lenguaje y representación: revista de estudios culturales de la Universitat Jaume I*, 2018, Vol. 19, pp. 7-29, <https://raco.cat/index.php/CLR/article/view/355769>.

## 8 ¿Qué puede ser también estadística en lingüística diacrónica?

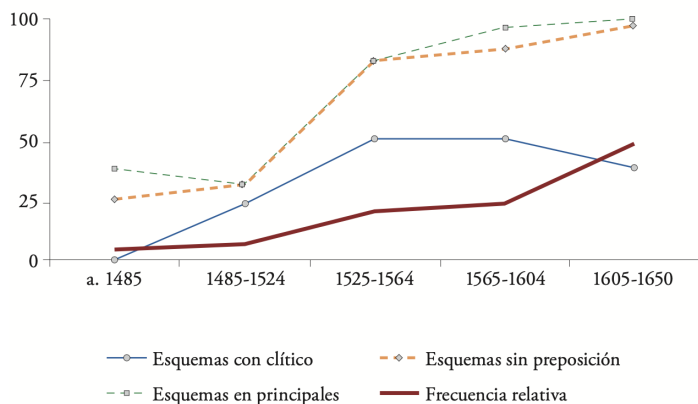


Figura 1. Valores porcentuales (para cada periodo) de varios parámetros combinados y frecuencia global relativa para los esquemas con infinitivo antepuesto a *tener*

Fuente: de Toledo y Huerta, Á. S. O. (2016). Enseñanzas del cambio fracasado: Trayectoria y estela de una perífrasis fugaz (infinitivo + tener). *Cuadernos de Lingüística de El Colegio de México*, 3(1), 119. <https://doi.org/10.24201/clecm.v3i1.28>



## 10 ¿Qué es la estadística?

La estadística es una disciplina que aborda la recogida, el estudio, la valoración, organización y presentación de datos.

## 11 Utilidad

- Superar el sesgo (a veces las cosas no son como creo que son).
- Comprobar la validez de las hipótesis. Por ejemplo, en esta sesión tengo la hipótesis de que hay más gente del centro norte de España que del sur o del este. La mayor parte de vosotros tenéis una media de 25 años, sois todos guapísimos o muy inteligentes...

## 12 Por dónde empezar (bibliografía)

- Uso de CHATGPT u otras IA (cuidado con “alucinaciones”).
- Gries, S. (2016). Quantitative corpus linguistics with R: A practical introduction. En *Quantitative Corpus Linguistics with R: A Practical Introduction*. <https://doi.org/10.4324/9781315746210>
- Gries, S. Th. (2021). *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter. <https://doi.org/10.1515/9783110718256>
- Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.
- Moore, D. S., & McCabe, G. P. (1999). *Introduction to the practice of statistics*. W.H. Freeman.
- Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners*. University of Adelaide. <https://learningstatisticswithr.com/>
- Cabedo Nebot, A. (2021). *Fundamentos de estadística con R para lingüistas*. Tirant Lo Blanch.

## 13 Estadística descriptiva vs inferencial

### **i** Estadística descriptiva vs inferencial

- **Estadística descriptiva:** descripción de los datos.
- **Estadística inferencial:** generalización de los resultados a partir de una muestra (esta es en la que piensa la gente cuando piensa en estadística).

## 14 ¿Qué es un dato?

“Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho” (DRAE 2024)

¿Qué puede ser ese **algo en concreto**?



Images created by an AI from OpenAI

### **i** Definición de dato

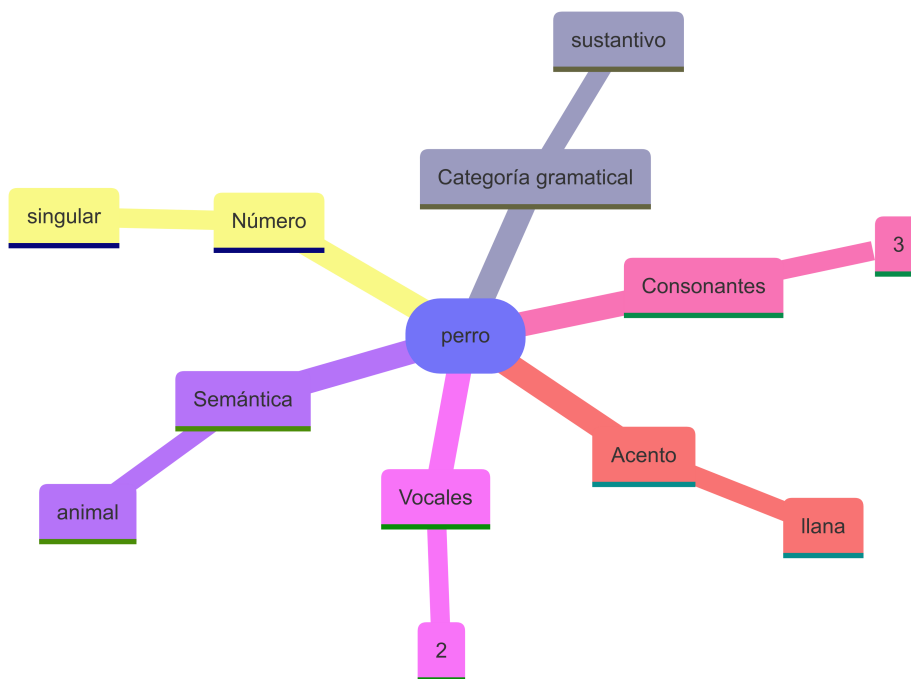
- Un dato o registro puede decirnos algo, por ejemplo, sobre un ser humano con sus características físicas o de comportamiento (orejas grandes y rojas, nariz gorda, simpático...).
- Cada rasgo o característica de ese “algo” es un elemento de variación (también conocido como categoría o **variante**) y se aglutina en lo que llamamos **variables**.

## 15 ¿De qué puede recogerse datos en lingüística?

💡 Tipo de elementos en lingüística sobre los que se pueden recoger datos

Un elemento lingüístico sobre el que recoger datos puede ser una palabra, una sílaba, un fonema (o alófonos correspondientes), un morfema, un sintagma, un enunciado, un párrafo, un texto, una posición discursiva, un acto de habla, un corpus, una obra, un autor, una lengua, una familia lingüística...

## 16 Ejemplo sobre un dato lingüístico: la palabra “perro”.



Trasladado a formato de tabla:

Palabra	Categoría gramatical	Número	Semántica	Vocales	Consonantes	Acento
perro	sustantivo	singular	animal	2	3	llana

## 17 ¿Qué es una variable?

Variable es una característica que puede cambiar y cuya variación es susceptible de medirse.

### Tipos de variables

- **Variables cualitativas:** no se pueden medir con una escala numérica. Por ejemplo, en lingüística, la categoría gramatical de una palabra.
- **Variables cuantitativas:** se pueden medir. Por ejemplo, en lingüística, la frecuencia de uso de una palabra, el valor de duración de una sílaba, el valor en Hz de un formante, la longitud en palabras de un enunciado, etc.

## 18 ¿Qué programas puedo usar para hacer estadística?

- Excel
- SPSS
- R
- ChatGPT (alucinaciones...)

## 19 ¿Cómo recoger los datos?

### Recogida de datos

Si menos de 500000 o 1000000 datos,

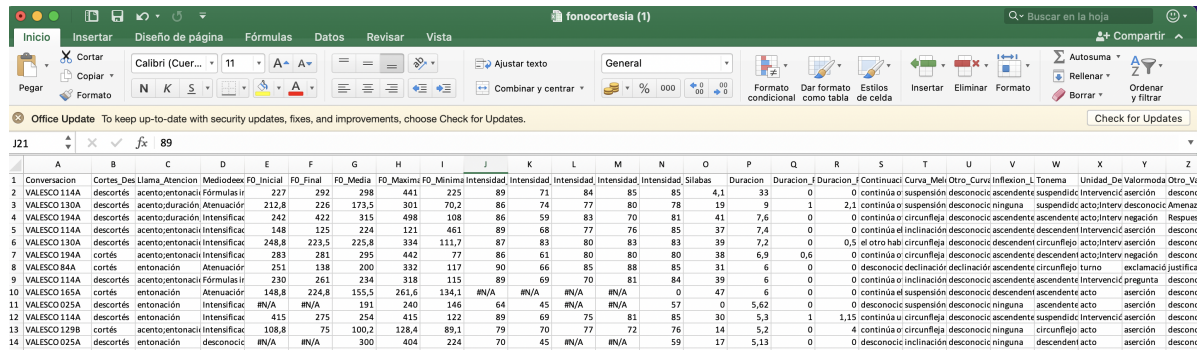
- Google Sheets (viene gratis con Google Drive)
- Microsoft Excel (de pago, pero más utilizado)

### Precaución

Si más de 1000000 datos,

- PostgreSQL (gratis)
- MySQL (gratis)
- SQLServer (de pago)

## 20 Estructura de una base de datos



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Conversacion	Cortes_Des	Llama_Atencion	Mediodesx	FO_Inicial	FO_Final	FO_Media	FO_Maximi	FO_Minima	Intensidad	Intensidad	Intensidad	Intensidad	Silabas	Duracion	Duracion	Duracion	F	Continuaci	Curva_Mel	Otro_Curv	Inflexion	L_Tonema	Unidad_De	Valormoda	Otro_V
2	VALESCO 114A	descortés	acento;entonaci	Fórmulas Ir	227	292	298	441	225	89	71	84	85	85	4,1	33	0	0	continúa o suspensión	desconoc	ninguna	suspendid	acto;interv	desconoc	Amenaz	
3	VALESCO 130A	descortés	acento;duración	Atenuaciór	212,8	226	173,5	301	70,2	86	74	77	80	78	19	9	1	2,1	continúa o suspensión	desconoc	ninguna	suspendid	acto;interv	desconoc	Amenaz	
4	VALESCO 194A	descortés	acento;duración	Intensificac	242	422	315	498	108	86	59	83	70	81	41	7,6	0	0	continúa o circunfleja	desconoc	ascendente	ascendente	acto;interv	negación	Respuer	
5	VALESCO 114A	descortés	acento;entonaci	Intensificac	148	125	224	121	461	89	68	77	76	85	37	7,4	0	0	continúa el inclinación	desconoc	ascendente	descendente	Intervenci	aserción	descon	
6	VALESCO 130A	descortés	acento;entonaci	Intensificac	248,8	223,5	225,8	334	111,7	87	83	80	83	83	39	7,2	0	0,5	el otro habi	circunfleja	desconoc	descendent	circunflejo	acto;Interv	aserción	descon
7	VALESCO 194A	descortés	acento;entonaci	Intensificac	283	281	295	442	77	86	61	80	80	80	38	6,9	0,6	0	continúa o circunfleja	desconoc	ascendente	descendente	acto;Interv	negación	descon	
8	VALESCO 84A	descortés	entonación	Atenuaciór	251	138	200	332	117	90	66	85	88	85	31	6	0	0	desconoc	declinaciór	declinaciór	ascendente	circunflejo	turno	exclamaciór	justifica
9	VALESCO 114A	descortés	acento;entonaci	Fórmulas Ir	230	251	234	318	115	89	69	70	81	84	59	6	0	0	continúa o inclinación	desconoc	ascendente	ascendente	Intervenci	pregunta	descon	
10	VALESCO 165A	descortés	entonación	Atenuaciór	148,8	224,8	155,5	261,6	134,1	#N/A	#N/A	#N/A	#N/A	0	47	6	0	0	continúa el suspensión	desconoc	descendent	ascendente	acto	aserción	descon	
11	VALESCO 025A	descortés	entonación	Intensificac	#N/A	#N/A	191	240	146	64	45	#N/A	#N/A	57	0	5,62	0	0	desconoc	suspensión	desconoc	ninguna	ascendente	acto	aserción	descon
12	VALESCO 114A	descortés	entonación	Intensificac	415	275	254	415	122	89	69	75	81	85	30	5,3	1	1,15	continúa u	circunfleja	desconoc	ascendente	suspendid	Intervenci	aserción	descon
13	VALESCO 129B	descortés	acento;entonaci	Intensificac	108,8	75	100,2	128,4	89,1	79	70	77	72	76	14	5,2	0	4	continúa o circunfleja	desconoc	ninguna	circunflejo	acto	aserción	descon	descon
14	VALESCO 025A	descortés	entonación	desconoc	#N/A	#N/A	300	404	224	70	45	#N/A	#N/A	59	17	5,13	0	0	desconoc	inclinación	desconoc	ninguna	descendent	acto	aserción	descon

Ejemplo: base de datos del proyecto Fonocortesia, desarrollado en 2013 y dirigido por el catedrático Antonio Hidalgo Navarro

## 21 Bases de datos (Fonocortesia)

Enlace recomendado: [https://docs.google.com/spreadsheets/d/1wJ970x3PmcV9GQ7ZvjoBmdE\\_1NgPS7uvWp8xEIhhMA8/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1wJ970x3PmcV9GQ7ZvjoBmdE_1NgPS7uvWp8xEIhhMA8/edit?usp=sharing)

Google DataStudio: <https://lookerstudio.google.com/reporting/57f127e6-8e3b-4d63-8fc7-7e0226431e9c/page/9OMAB?s=iJDWulQ3Ees>

Oralstats: [https://adrin-cabedo.shinyapps.io/oralstats\\_v\\_1\\_3/](https://adrin-cabedo.shinyapps.io/oralstats_v_1_3/)

## 22 Exploración inicial de datos con Excel o Google Sheets

### **i** Sobre tablas dinámicas

Una tabla dinámica es una herramienta avanzada para calcular, resumir y analizar datos que le permite ver comparaciones, patrones y tendencias en ellos. Las tablas dinámicas funcionan de forma un poco distinta dependiendo de la plataforma que use para ejecutar Excel.

(cita directa extraída de: <https://support.microsoft.com/es-es/office/crear-una-tabla-din%C3%A1mica-para-analizar-datos-de-una-hoja-de-c%C3%A1culo-a9a84538-bfe9-40a9-a8e9-f99134456576>)

## 23 Opciones de transformación y visualización de datos

### Lenguaje de programación

- R / RStudio
- Python

### Soluciones de software

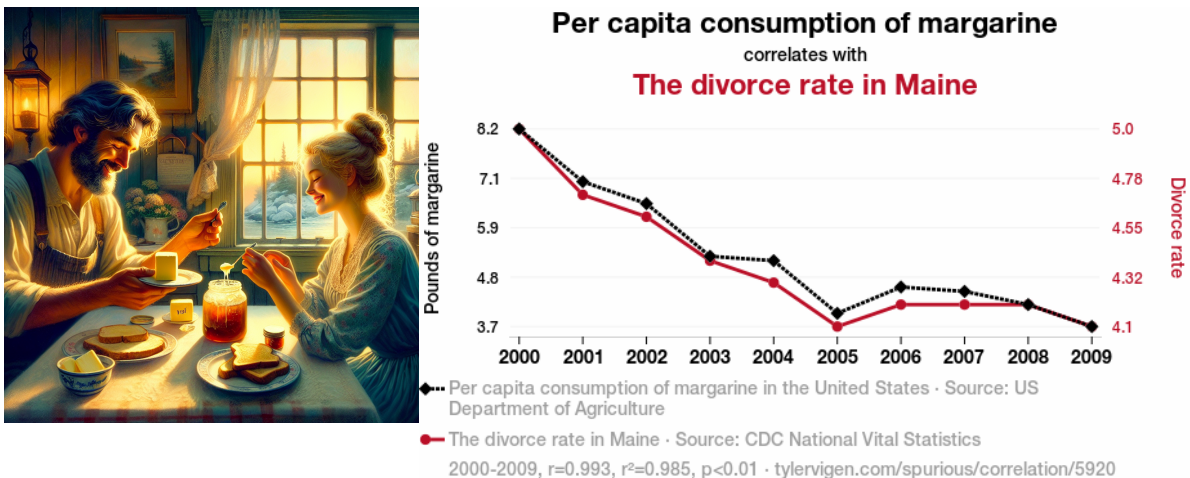
- Tableau
- Microsoft Power BI
- Voyant tools
- Google Data Studio

## 24 Correlaciones espúreas

La estadística puede llevar a conclusiones erróneas si no se tiene en cuenta el contexto y si no aplicamos un análisis crítico, procedente de nuestra experiencia y nuestro bagaje como investigadores (lecturas realizadas, experimentos previos...).

El ejemplo de más abajo lo conocí a partir de la lectura de Levshina (2015); ejemplifica perfectamente las relaciones “espúreas” entre dos elementos que parecen estar emparentados estadísticamente, pero que no lo están en realidad (a no ser que creamos en el mundo de la magia y la brujería)

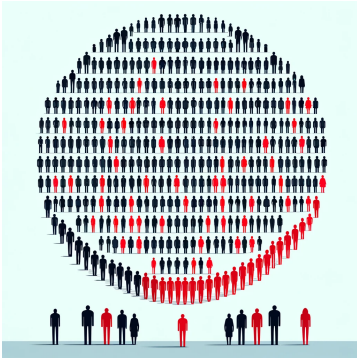
[https://tylervigen.com/spurious/correlation/5920\\_per-capita-consumption-of-margarine\\_correlates-with\\_the-divorce-rate-in-maine](https://tylervigen.com/spurious/correlation/5920_per-capita-consumption-of-margarine_correlates-with_the-divorce-rate-in-maine)



## 25 ¿Qué es una población y una muestra?

### **i** Población y muestra

- Población es el conjunto de todos los elementos que cumplen una serie de características comunes.
- Muestra es un subconjunto de la población.



Images created by an AI from OpenAI

## 26 Tamaño de la muestra

<https://www.calculator.net/>

### **i** Calcular tamaño de la muestra

- **Error muestral:** es la diferencia entre el valor real y el valor estimado.
- **Nivel de confianza:** es la probabilidad de que el valor real esté dentro del intervalo de confianza.

### **💡** Ejemplo

- En una población de 100000 personas, si queremos un error muestral del 5% y un nivel de confianza del 95%, necesitamos una muestra de 384 personas.

### **⚠️** Precaución

**Problema:** si quiero estudiar un fenómeno discursivo (la entonación de oraciones interrogativas, por ejemplo), debo asegurarme de que la muestra sea representativa de la

población, que no solo es el conjunto de hablantes, sino también el propio hablante en sí.  
**Casuística:** a veces solo tenemos la opción de recoger un único ejemplo para un hablante.  
Hay que intentar ampliarlos

## 27 Hipótesis nula y alternativa

### **i** Hipótesis nula y alternativa

- La **hipótesis nula** ( $H_0$ ) es una afirmación que se supone verdadera hasta que se demuestre lo contrario.
- La **hipótesis alternativa** ( $H_1$ ) es una afirmación que se supone falsa hasta que se demuestre lo contrario.

La nuestra suele ser la  $H_1$ .

## 28 ¿Qué es un test de hipótesis?

### **i** Test de hipótesis

Un test de hipótesis es una técnica estadística que se utiliza para tomar decisiones sobre la hipótesis nula.

Ejemplos de test de hipótesis: t-test, ANOVA, chi-cuadrado, test de Wilcoxon, test de Mann-Whitney, test de Kruskal-Wallis, test de Friedman...

### 28.1 Ejemplos de hipótesis nula y alternativa

#### **i** Ejemplos con variables numéricas

- $H_0$ : La media de la intensidad de la voz en los enunciados corteses y descorteses es la misma.
- $H_1$ : La media de la intensidad de la voz en los enunciados corteses y descorteses es distinta.

### **i** Ejemplo con variables nominales

- H0: Los enunciados corteses y descorteses (variable *cortesía*) no se relacionan con otros valores pragmáticos como la ironía, el humor, la intensificación o la atenuación (variable *valor pragmático*)
- H1: Los enunciados corteses y descorteses (variable *cortesía*) se relacionan con otros valores pragmáticos como la ironía, el humor, la intensificación o la atenuación (variable *valor pragmático*)

## 28.2 Elección de pruebas estadísticas

Hay muchas pruebas estadísticas, pero no todas son adecuadas para todos los casos. El investigador debe elegir la prueba adecuada para su estudio, así como la mejor manera para visualizar la información.

Tests sugeridos y abordados por Stefan Gries: [https://www.stgries.info/research/sflwr/sflwr3\\_navigator.png](https://www.stgries.info/research/sflwr/sflwr3_navigator.png)

<https://statkat.com/statistical-technique-selection/tool-for-selecting-a-statistical-technique.php>

## 29 Errores tipo I y tipo II

Los errores tipo I y tipo II son dos tipos de errores que se pueden cometer al realizar un test de hipótesis. Por eso operamos siempre con un nivel de confianza (en síntesis, existe una probabilidad de error).

### **!** Errores tipo I y tipo II

- Un **error tipo I** es la conclusión de que una hipótesis nula es falsa cuando en realidad es verdadera.
- Un **error tipo II** es la conclusión de que una hipótesis nula es verdadera cuando en realidad es falsa.

## 30 Paradoja de Simpson

Es una paradoja en la probabilidad y estadística en la que una tendencia que aparece en varios grupos de datos desaparece o se invierte cuando se añaden nuevas informaciones o variables.

### **i** Ejemplo de la paradoja de Simpson

Ninguna disciplina está libre de esta paradoja o posibilidad de error. Nuestra disciplina tampoco lo está. Por ejemplo, en un estudio sobre la variación de la entonación en el español de Valencia, si no se tiene en cuenta el sexo del hablante o la edad, se puede llegar a conclusiones erróneas.

## 31 Valor p



Images created by an AI from OpenAI

El valor p es una medida de la probabilidad de que un resultado estadístico (o uno más extremo) se produzca bajo la hipótesis nula.

### **i** Regla del valor p

- Menor a 0.05: rechazamos la hipótesis nula.
- Superior a 0.05: no rechazamos la hipótesis nula.

**Normalmente, la  $H_0$  es la hipótesis distinta de la que queremos demostrar.**

### **!** Importante

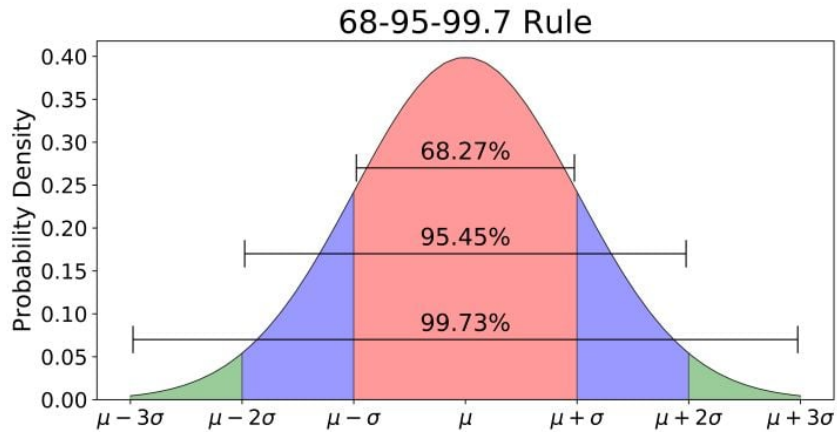
Un valor p en la prueba estadística inferior a 0.05 es lo que buscamos.

## 32 Distribución normal

Una distribución normal es una distribución de probabilidad que es simétrica alrededor de la media, la mediana y la moda. La curva de campana de una distribución normal es simétrica.

### 32.1 Regla de la normalidad

En una distribución normal, el 68% de los datos se encuentran a una desviación estándar de la media, el 95% a dos desviaciones estándar y el 99.7% a tres desviaciones estándar.



Fuente de la imagen: <https://builtin.com/data-science/empirical-rule>

### 32.2 Valor p y distribución normal

#### **i** Ejemplo

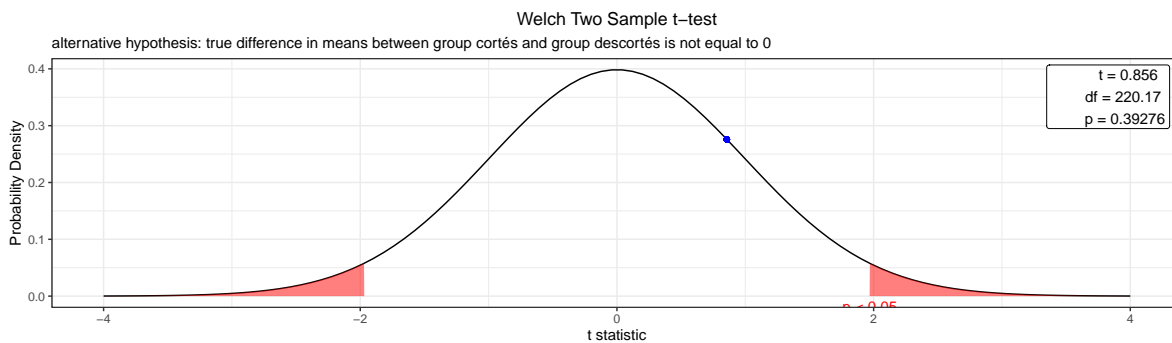
- Sobre una distribución normal estandarizada, el valor p es la probabilidad de que un valor igual o más extremo que el observado se produzca por azar.
- Hay que tener en cuenta que el valor p no es la probabilidad de que la hipótesis nula sea cierta, sino la probabilidad de que los datos observados se produzcan si la hipótesis nula es cierta.

## 33 Ejemplos de valor P

### 33.1 ¿Es la intensidad o volumen de la voz distinta en los enunciados corteses y descorteses?

#### **i** Ejemplo 1

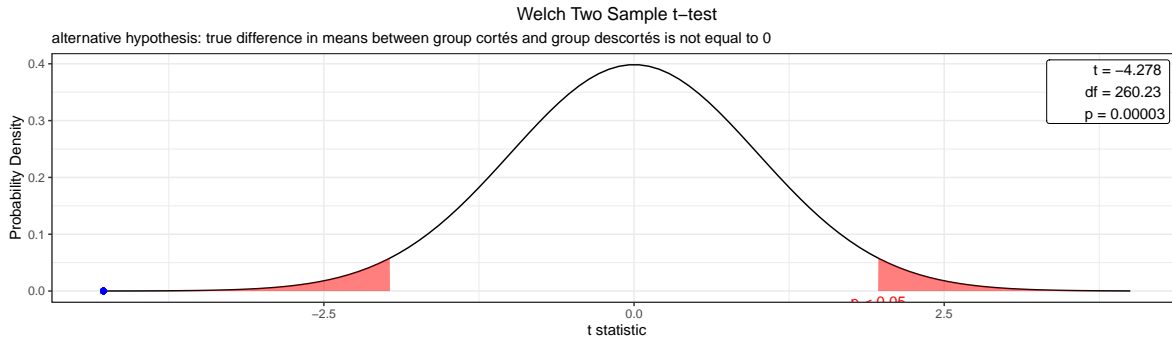
- Enunciados corteses: 77.61 decibelios de media
- Enunciados descorteses: 75.93 decibelios de media
- Valor p: 0.39
- Conclusión de investigación: los enunciados corteses y descorteses no tienen un volumen distinto.



### 33.2 ¿Es la F0 o tono de la voz distinta en los enunciados corteses y descorteses?

#### **i** Ejemplo 2

- Enunciados corteses: 198.8 Hz de media
- Enunciados descorteses: 230.04 Hz de media
- Valor p: 0.000003
- Conclusión de investigación: los enunciados corteses y descorteses tienen valores tonales distintos (mayores en la expresión de la descortesía).



#### ⚠ Precaución

Puede existir una variable oculta: el tono de la voz es distinto en hombres y mujeres. Si hay más hombres o mujeres en un grupo que en otro, la variable oculta puede estar influyendo en el resultado.

## 34 Práctica online de pruebas estadísticas

En la siguiente página, sin necesidad de instalar ningún software, pueden realizarse pruebas estadísticas básicas. Es útil si solo se requiere el dato de relación y no un gráfico.

<http://www.vassarstats.net/>

## 35 Uso de un programa especializado (R)

Recomiendo utilizar el programa gratuito R junto con Rstudio para realizar estadísticas, gráficos y análisis de datos.





### Funcionalidades de R

- **Análisis Estadístico:** Desde análisis descriptivos básicos hasta modelos estadísticos avanzados y pruebas de hipótesis.
- **Visualización de Datos:** Creación de gráficos y mapas detallados para explorar y presentar datos de manera efectiva.
- **Manipulación de Datos:** Transformación, limpieza y preparación de datos para análisis mediante paquetes como dplyr y tidyr.
- **Modelado Predictivo:** Desarrollo de modelos de machine learning, incluyendo regresión, clasificación y clustering.
- **Generación de Informes:** Automatización de informes y creación de documentos reproducibles con R Markdown.
- **Interfaz de Programación:** Desarrollo de aplicaciones interactivas y dashboards usando Shiny para presentaciones dinámicas de datos.

## 36 Ejemplo de gráfico simple con R

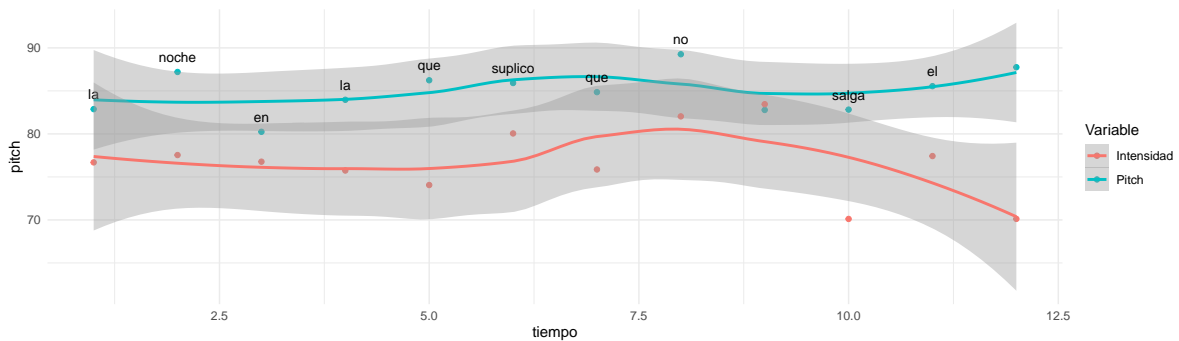


Figura extraída con GGplot2. Curva melódica y de intensidad del enunciado *la noche en la que no salga el sol*

## 37 ¿Qué más puedo hacer con R?

### Más funcionalidades

- Escribir documentos científicos mediante Rmarkdown o Quarto (este mismo documento ha sido escrito en R).
- Exportar tus documentos a varios formatos: PDF, Word o Powerpoint.
- Modificación de las plantillas. Ejemplo: la plantilla de la revista *Normas* de la UV se realiza con R.
- Programar scripts que realicen funciones de manera automática. Por ejemplo, abre todos los archivos de una carpeta e impórtalos.
- Crear aplicaciones web para consultar datos. Ej.: [Oralstats](#).

## 38 ¿Desde dónde instalo R y RStudio?

- [R](#)
- [RStudio](#)
- Puedes usar también una versión gratuita online (con limitaciones de uso, pero suficiente para este curso y para un uso puntual): <https://posit.cloud/>