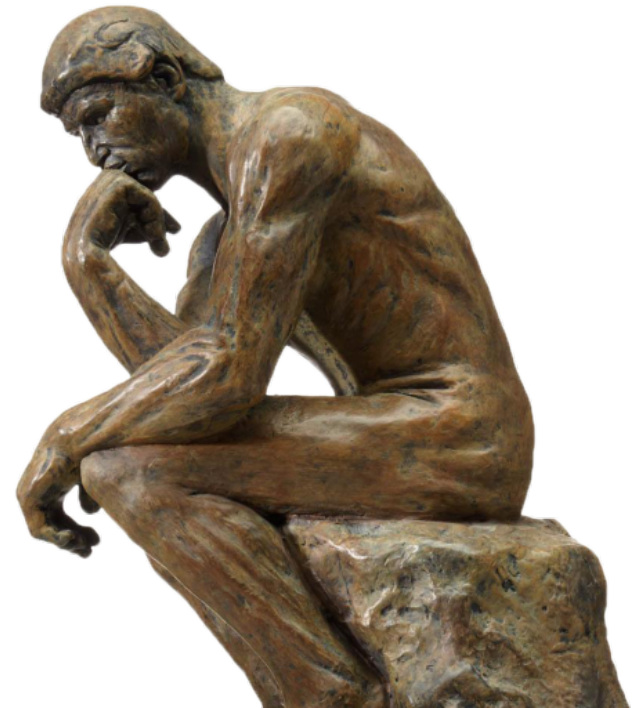


Introducción al análisis cuantitativo de datos lingüísticos

Ezequiel Koile (MPI SHH) y Carolina Gattei (IFIBA – CONICET)

Análisis cuantitativo en lingüística

- ▶ Tradicionalmente: análisis del lenguaje desde punto de vista **introspectivo**.
 - Diferencias **cualitativas** entre distintas lenguas o distintas estructuras lingüísticas en una misma lengua.



Análisis cuantitativo en lingüística

- ▶ Actualmente: tendencia hacia el análisis del lenguaje como **set de datos** que no proviene de la introspección del lingüista.
 - Tendencia hacia la búsqueda de diferencias **cuantitativas** entre distintas lenguas o distintas estructuras en una misma lengua para la formulación de hipótesis que apoyen teorías.





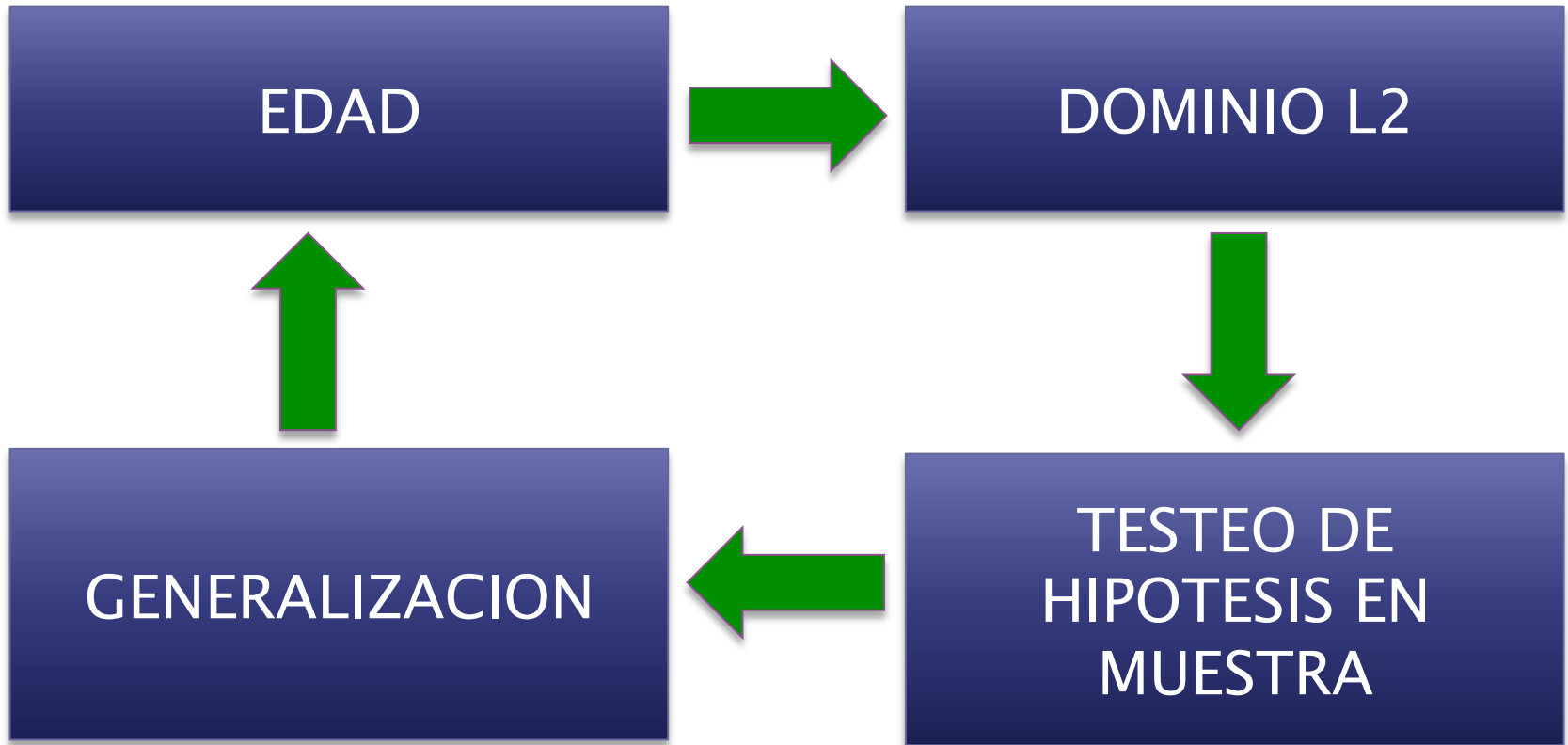
TESTEO DE HIPOTESIS

Testeo de Hipótesis

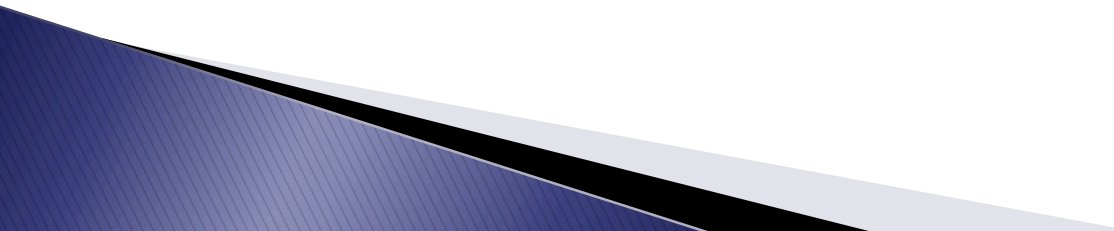
- ▶ “La gente mayor se olvida las segundas lenguas fácilmente.”
- ▶ “La gente mayor se olvida sus segundas lenguas más rápidamente que la gente de mediana edad.”
- ▶ “La gente mayor se olvida las palabras en su primera segunda lengua más rápido que la gente de edad media e igual nivel educativo.”

Testeo de Hipótesis

- ▶ H_0 = “No existen diferencias en el desempeño en L2 de la población xxx que dependa de la edad.”
- ▶ H_1 = “El desempeño en L2 en la población xxx es mejor cuando tienen edad joven o media. La gente de edad más avanzada se olvida más palabras que la gente de edad joven o media.”
- ▶ H_2 = “El desempeño en L2 de la población xxx es mejor cuando tienen edad más avanzada. La gente de edad más avanzada se olvida menos palabras que la gente de edad joven o media.”



Problema de la 'Generalizabilidad'

- ▶ ¿Hasta qué punto se pueden extender los resultados de un estudio a una población más grande que la muestra estudiada?
 - ▶ Características particulares de los individuos.
 - ▶ Necesidad de generalizar para poder explicar diferencias individuales.
- 

Problema de la Generalizabilidad

- ▶ Muestra vs. Población

μ : Se refiere a la **población** como tal.

n : Simboliza el número de elementos en una distribución de la **muestra**.

- ▶ Análisis de datos de una **muestra** permiten hacer inferencias sobre una **población** → **estadística inferencial**.

Muestra debe ser lo más representativa posible de la población que se quiere estudiar.

Problema de la generalizabilidad

- ▶ Selección de la muestra (N)
- ▶ Idealmente: al azar.
- ▶ Otra posibilidad: Agregar más variables de control (estudiantes de L2 de la universidad xxx, de edad xxx, de la carrera de xxx). Desventaja: problemas en la generalización de los resultados.
- ▶ Tamaño: depende del tamaño del efecto a buscar. Efectos grandes, una muestra pequeña es suficiente → *potencia estadística* del estudio

Testeo experimental vs. No experimental

- ▶ Hendriks (2002): adquisición de pedidos en estudiantes de inglés holandeses. Impacto de:
 - relaciones de poder;
 - distancia social; y
 - escenario de conversaciónen el uso de estrategias de cortesía en pedidos.

Testeo experimental vs. No experimental

The living room

You were in your room upstairs doing your maths homework, but you were not able to do the sums. You need some help. You go down to the living room where your dad is watching a documentary on television. What do you say to your dad?

.....

The supermarket

You are standing in line at the checkout with a shopping trolley full of groceries. You are late for an important meeting. There is one man in front of you. What do you say to the man in front of you?

.....



Estudios Longitudinales vs. Transversales

- ▶ Estudios longitudinales útiles para saber cómo influye una variable en el tiempo. P.ej.: adquisición de segunda lengua en situación de inmersión vs. clases.
- ▶ Problemas: TIEMPO y DINERO
- ▶ Posible solución: ESTUDIO TRANSVERSAL → Estudio de distintos grupos en etapas diferentes en un momento dado.

Estudios de Grupo vs. Estudios de Caso

- ▶ El tipo de objeto de estudio va a determinar el tipo de estudio más conveniente.
 - Pacientes lesionados en el cerebro
 - Estudiantes de escuela primaria de nivel x con nivel socio-económico y.

Análisis Cuantitativo vs. Cualitativo

- ▶ **Problema epistemológico 1:** ¿cuán confiable es el análisis cualitativo de datos?
- ▶ Posible solución: análisis intersubjetivo → varios ‘jueces’ testean la confianza de los datos (‘Inter-reliability tests’).
- ▶ **Problema epistemológico 2:** ¿sabemos qué están haciendo los participantes para resolver una tarea? Mayores TRs pueden significar mayor costo cognitivo, o uso de tiempo en poner en práctica una estrategia para resolver la tarea.
- ▶ Posible solución: usar tareas que marquen una o más líneas de base, usar distractores.

Testeo in-situ vs. Testeo en Laboratorio

▶ Laboratorio:

- **Pros:** ambiente controlado, reducción de variables que pueden influir resultados, posibilidad de muestra más pequeña.
- **Cons:** conclusiones reduccionistas de un fenómeno, falta de validez 'ecológica'.

▶ In-situ:

- **Pros:** validez ecológica, estudio del problema de estudio con todas las variables que podrían estar influyendo.
- **Cons:** falta de control de variables, artefactos que pueden esconder los resultados.

Testeo Estadístico: ¿Para qué?

“An independent-samples t-test showed that the mean difference between the groups turned out to be significant at $p < 0.05$, $t(89) = 2.4$.”



Nivel	Femenino	Masculino
A1	86	76
B1	90	90
C1	78	70