

10 Más temas y más métodos empiricos en las ciencias del lenguaje

Métodos empíricos 2

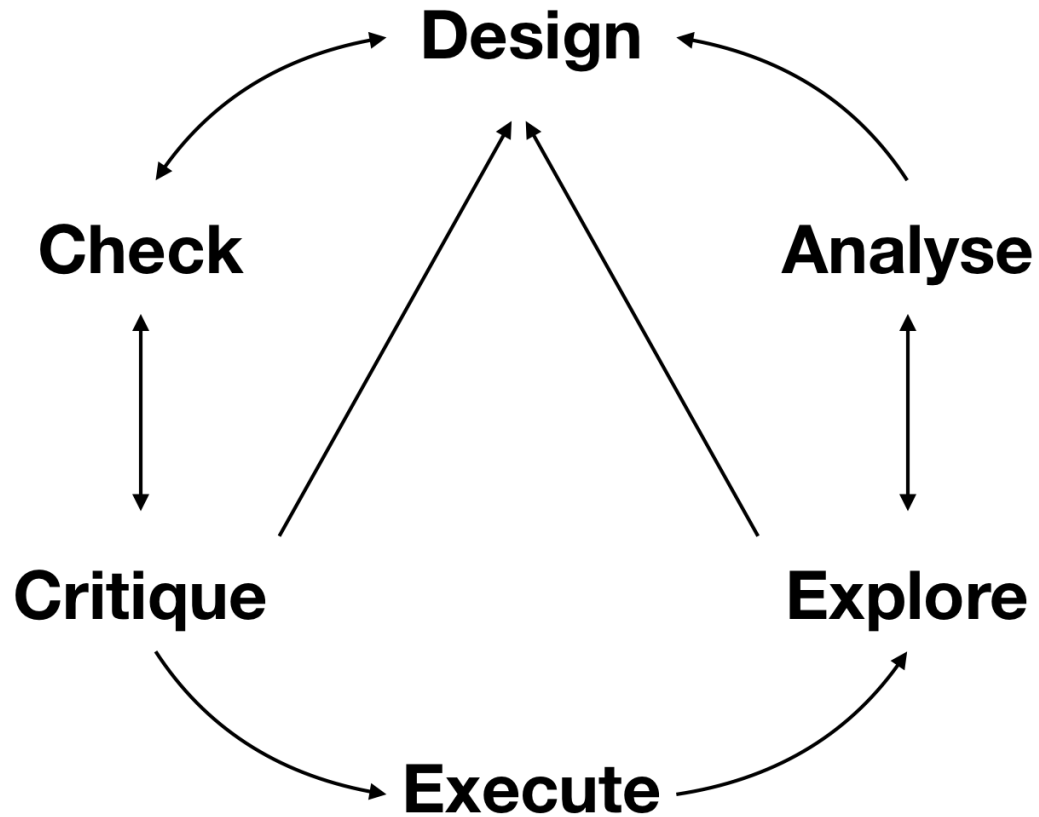
14/06/2022

Hoy

- Recapitulación
- Más allá: Métodos empiricos inferenciales
- Más allá: Ciencias del lenguaje empiricas

Recapitulación

Ciclo de análisis



Análisis inferencial vs. descriptivo

- Inferencia de propiedades (más allá de la muestra)
- Predicción
- Comparación
- Causa-efecto

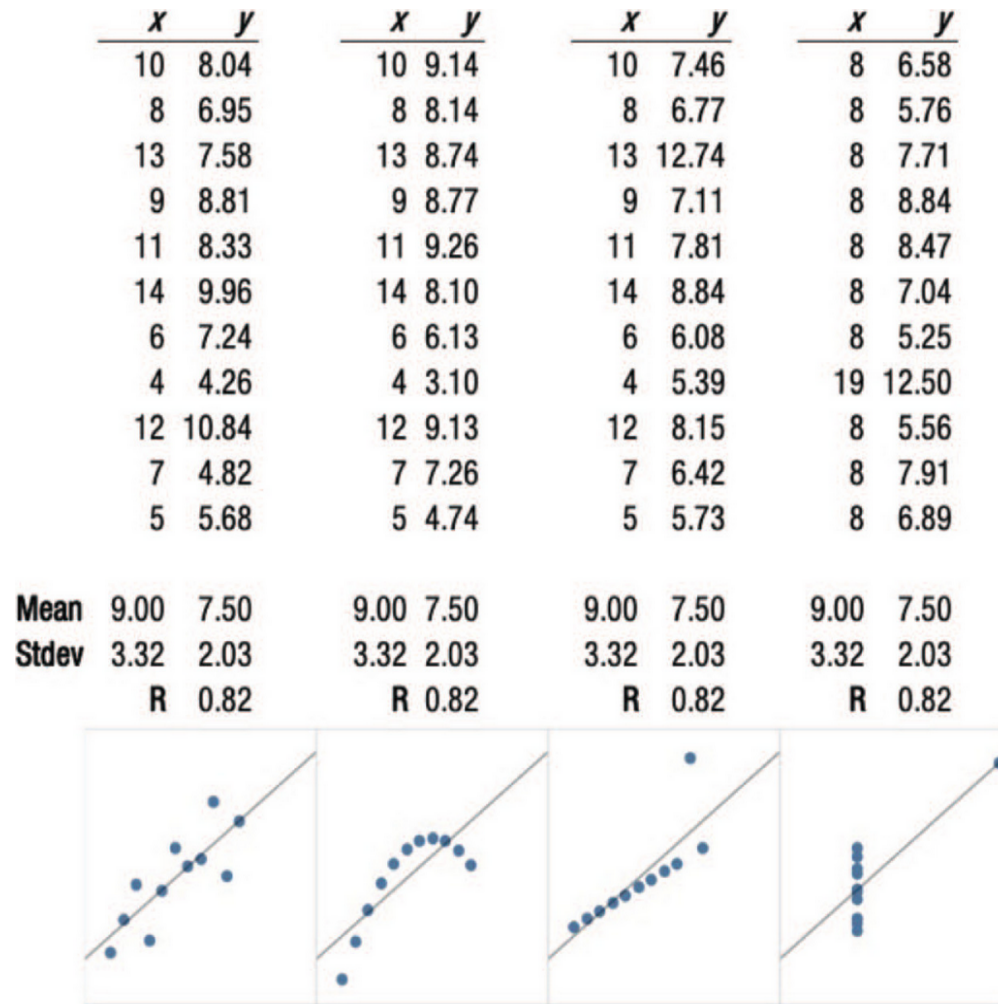


Fig. 1 de Franconeri et al. 2021 [The Science of Visual Data Communication: What Works](#)

Replicabilidad

Que se puedan obtener resultados consistentes con los mismo datos de entrada; pasos computacionales; métodos; código; y condiciones de análisis

Reproducibilidad

Que se puedan obtener resultados consistentes en diferentes análisis que buscan responder la misma pregunta, cada cual con sus propios datos

Diseño de análisis (componentes)

- Pregunta(s) del análisis
- Plan de diseño
- Plan de muestreo (sampling plan)
- Especificación de variables
- Plan de análisis

Terminología: Tipos de variables

- **Nominales**
- **Ordinales**
- **Binarias**
- **Booleanas**
- **Métricas**

Terminología: Tipos de muestras

Muestra completa: toda la población de interés

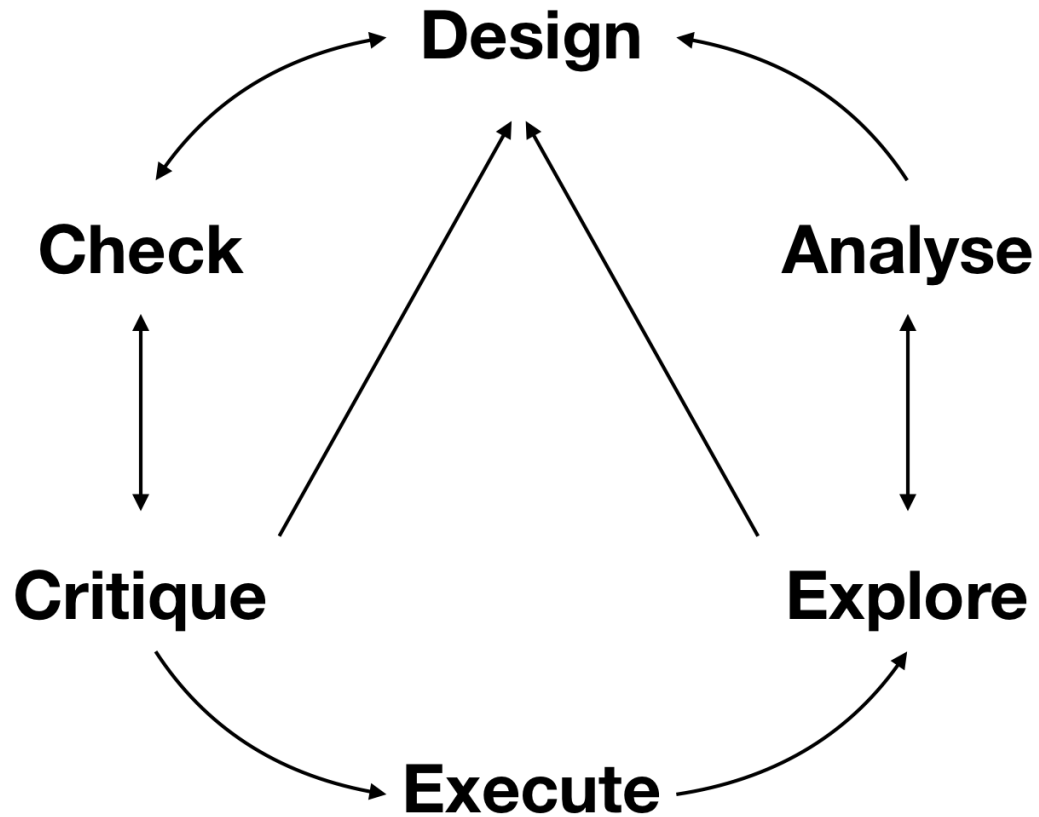
Muestra representativa/sin sesgo: tomada de la muestra completa con un método que no depende de la muestra que se está tomando

Muestra no representativa/con sesgo: los datos son influenciados por el método de toma

Terminología: Distribuciones

- Gaussiana/Normal
- Poisson
- Bernoulli (Binomial)

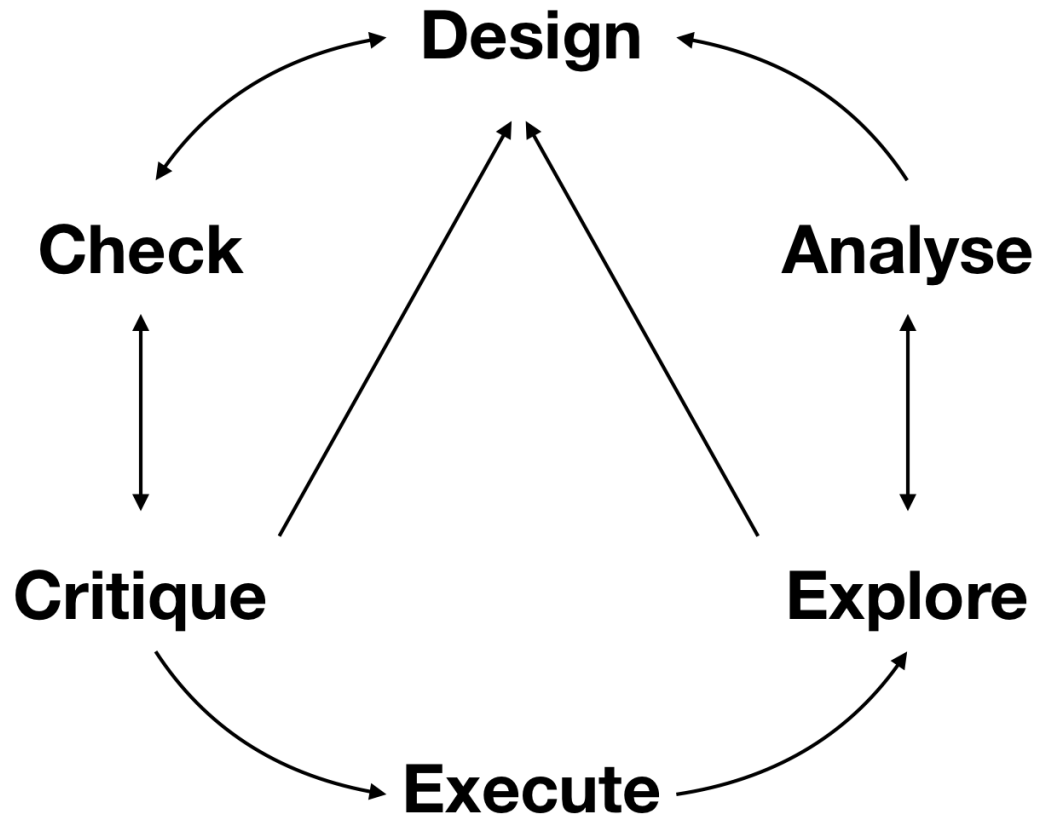
Ciclo de análisis



Control y crítica

- Estudios piloto
- Simulaciones

Ciclo de análisis



Exploración

- (Manipulación)
- Estadística descriptiva
- Visualización

Análisis

- Regresión lineal generalizada con uno o más predictores
 - Normal/Gaussiana
 - Poisson
 - Bernoulli
- Visualización
- Análisis de corpus

Fenómenos

- Tono
- Gestos
- Ambigüedad temporal
- Resolución de pronombres
- Leyes de Zipf (laboratorio y gran escala)
- ...

Kahoot!

www.kahoot.it

Más allá: Métodos empíricos inferenciales

Paramétricos

- Modelos lineales (hierárquicos)
 - k-means
 - ...
-

No-paramétricos

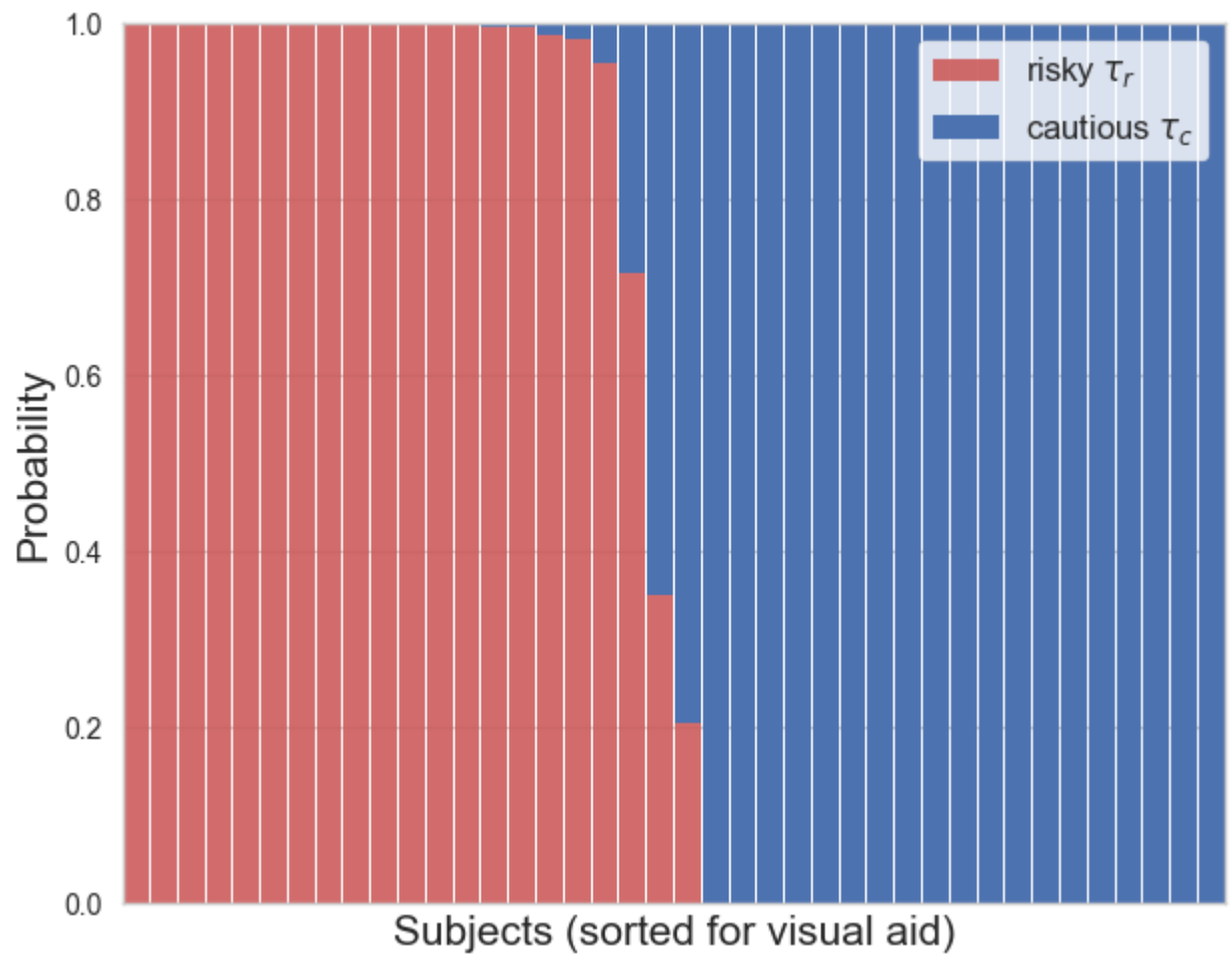
- Modelos generalizados aditivos
- Redes neuronales (en realidad no, pero en términos prácticos sí)
- ...

Modelo hierarquico para Kanwal et al. 2017

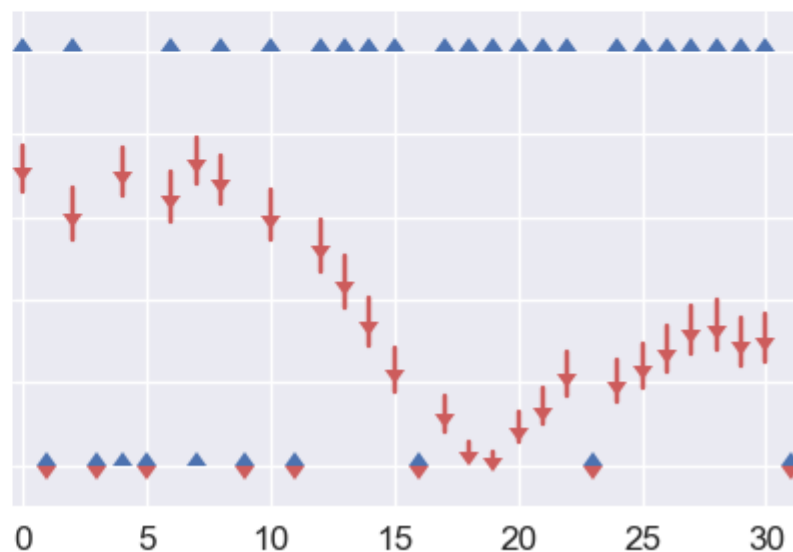
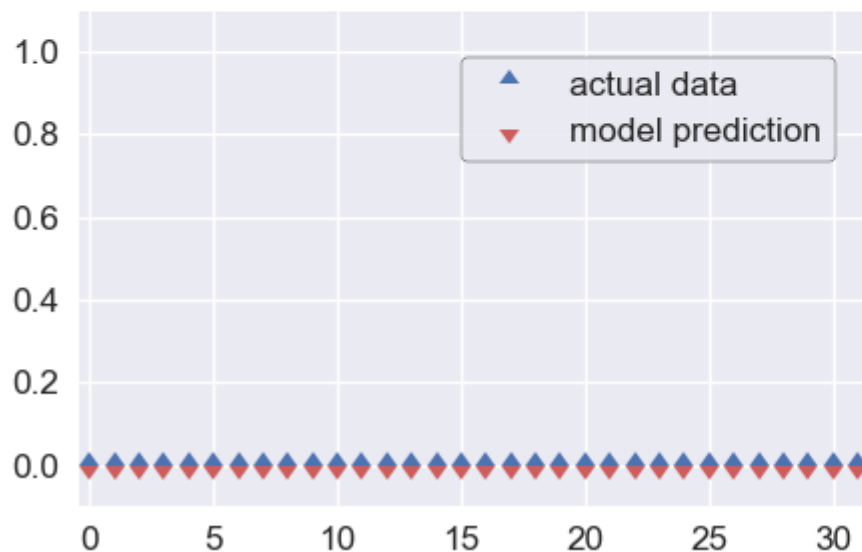
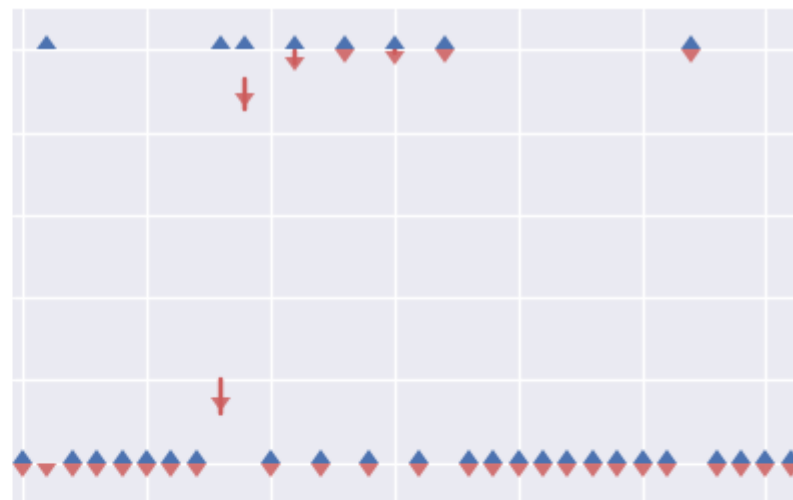
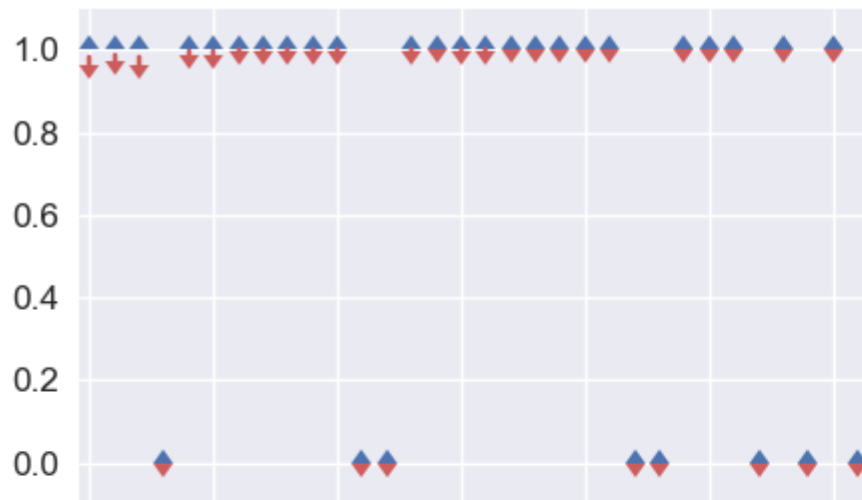
Hay dos tipos "latentes" de hablantes:

1. Arriesgados: Piensan que el interlocutor piensa que el significado frecuente es más esperado
2. Cuidadosos: Tienen incertidumbre sobre cual es el significado esperado

Hablantes usan el mensaje (que consideran) más probable a ser entendido, minimizando longitud de mensaje



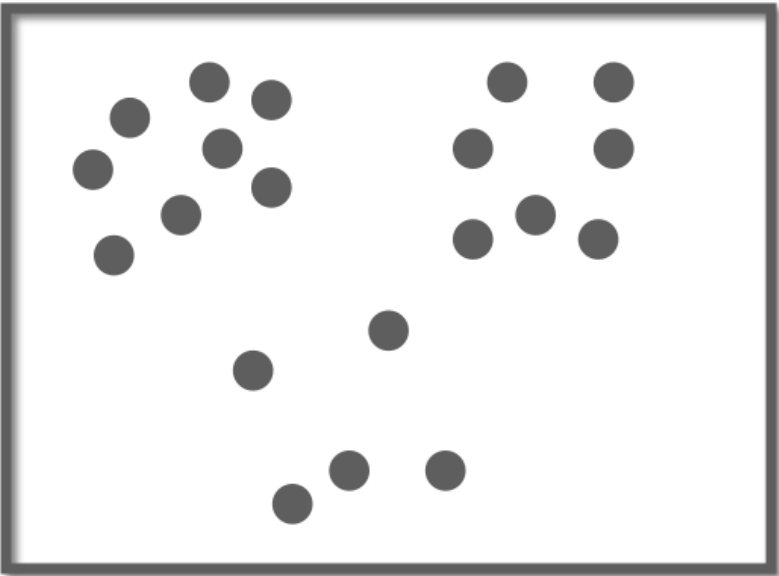
Probability of ambiguous message



Trial

K-means

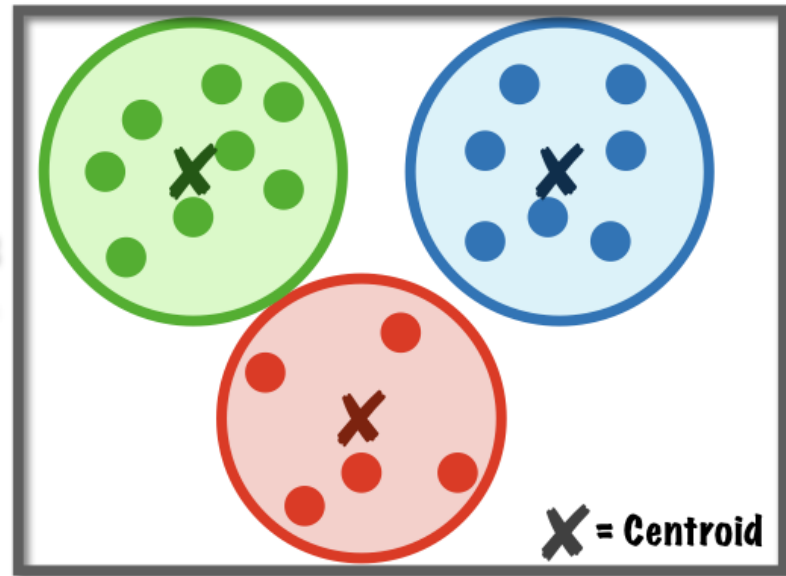
Unlabelled Data



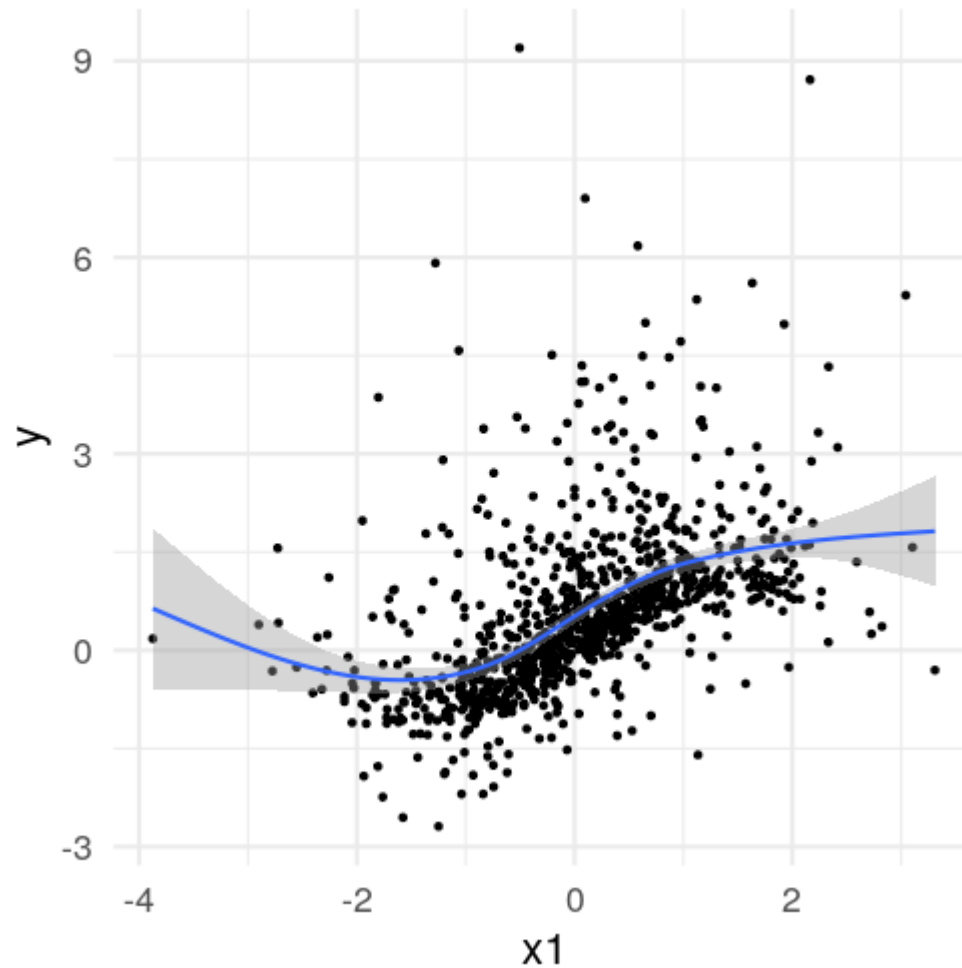
K-means



Labelled Clusters



Generalized Additive Models

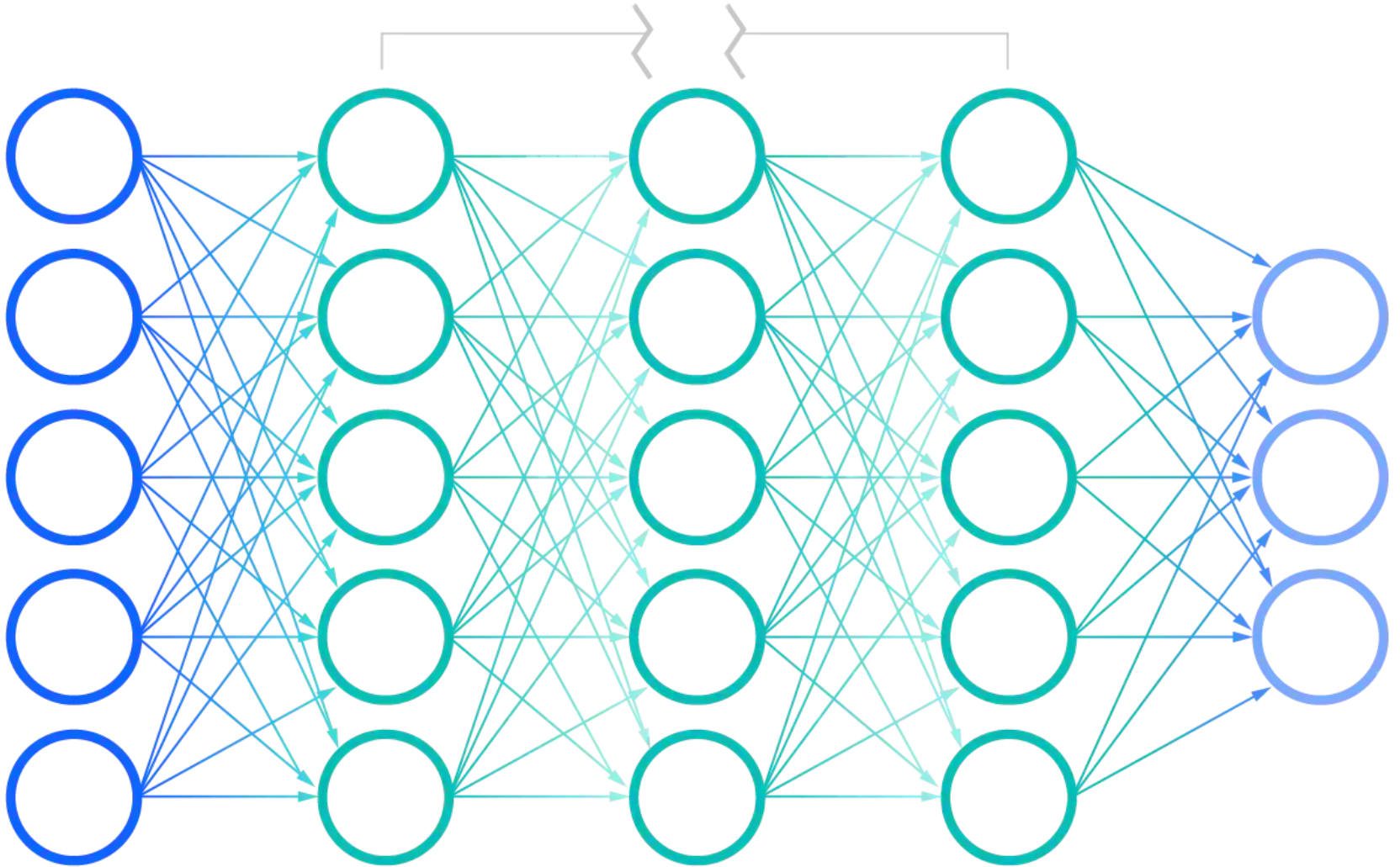


Deep neural network

Input layer

Multiple hidden layers

Output layer



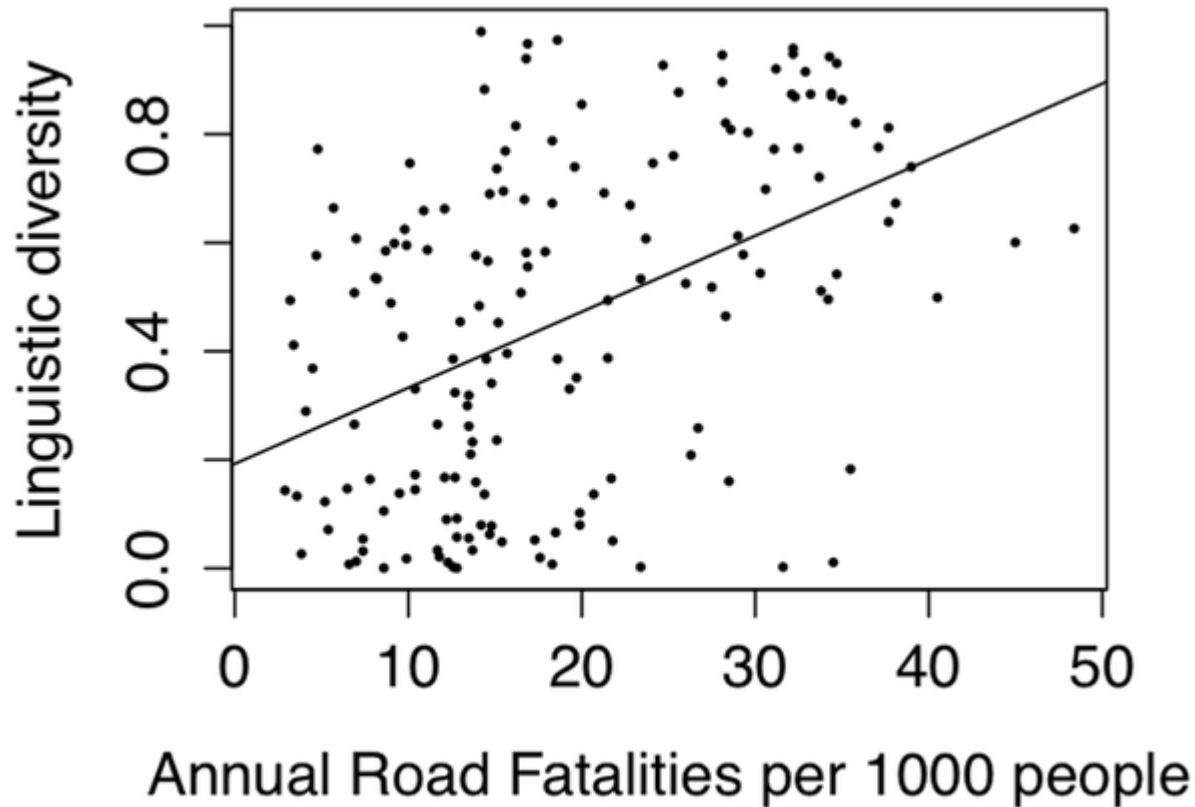
Todos estos métodos siguen el mismo proceso:

- Define objetivo ("función de pérdida")
- "Aprende" de los datos
- Encuentra el parámetro, o combinación de parámetros, que minimizan la pérdida (maximizan el objetivo)

Todos estos métodos siguen el mismo proceso:

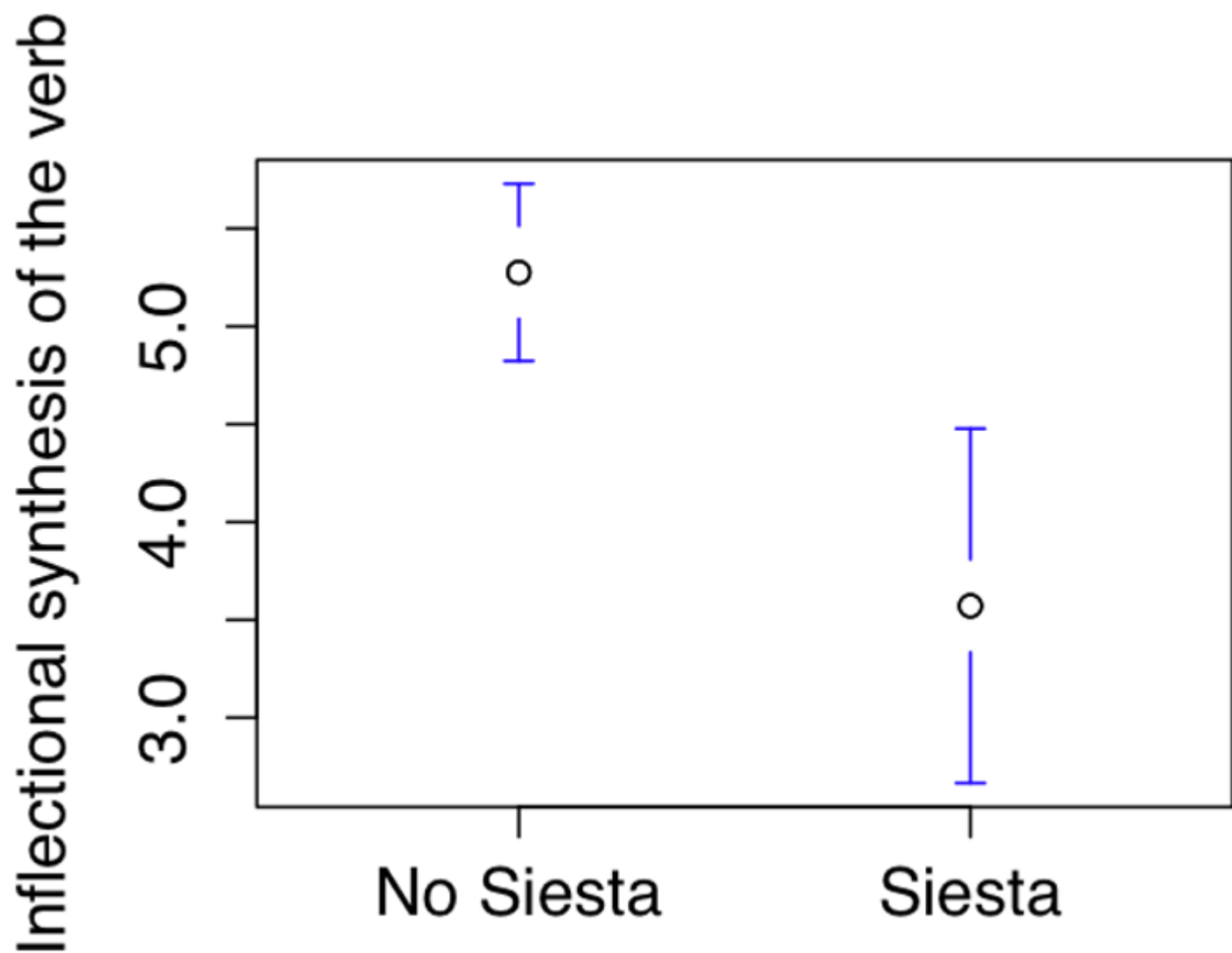
- Define objetivo ("función de pérdida")
- "Aprende" de los **datos**
 - Calidad de datos
 - Razón por qué una o más variables podrían solucionar el problema
- Encuentra el parámetro, o combinación de parámetros, que minimizan la pérdida (maximizan el objetivo)

-
- **Human in the Loop: Active learning and annotation for human-centered AI**
 - **Linguistic Diversity and Traffic Accidents: Lessons from Statistical Studies of Cultural Traits**



Linguistic Diversity and Traffic Accidents: Lessons from Statistical Studies of Cultural Traits

y = Linguistic diversity index: probabilidad que dos personas de una población tengan la misma lengua nativa



y = Promedio de categorías gramaticales que puede tomar el verbo

Más allá: Ciencias del lenguaje empiricas

De momento: Saliendo lentamente de la crisis de replicabilidad

"Nuevas" áreas de investigación/aplicación:

- Traducción asistida
- Tipología computacional
- Human in the loop QA
- PLN
- Metodologías de aprendizaje asistidas / automatizadas
- ...

Y ahora a dónde?

Preguntas?

Coda

- Avaldo (20/06)
- Informe final: 28/06

Gracias!

<https://brochhagen.github.io/>

thomas.brochhagen@upf.edu

52.631