

# Modelos lineales y modelos lineales generalizados

Rolando Gonzales Martinez, PhD

Fellow postdoctoral Marie  
Skłodowska-Curie

Universidad de Groningen  
(Países Bajos)

Investigador (researcher)

Iniciativa de Pobreza y Desarrollo  
Humano de la Universidad de  
Oxford (UK)

## En esta sesión:

1. Conocernos como clase
2. Revisar el contenido de la clase
3. Discutir el software y hardware de la clase
4. Discutir los objetivos de aprendizaje y el método de aprendizaje
5. Evaluación de diagnóstico

# Bellringer/ice breaker

Doctor en Negocios internacionales por la Universitetet i Agder (Noruega), Máster en Estadística Aplicada por la Universidad de Alcalá (España). **Cursando UTQ (Netherlands)**

**Actualmente:** Fellow postdoctoral Marie Skłodowska-Curie (Universidad de Groningen), Investigador (researcher) Iniciativa de Pobreza y Desarrollo Humano (Universidad de Oxford, UK)

**Previamente:** Investigador postdoctoral en la Real Academia de Ciencias de los Países Bajos. Científico de datos postdoctoral en CASUS (Helmholtz-Zentrum Dresden-Rossendorf, Alemania).

**Consultor postdoctoral:** Universitat Autònoma de Barcelona, Agencia Italiana de Cooperación al Desarrollo, Fondo de Población de las Naciones Unidas (UNFPA)

## Publicaciones seleccionadas:

- **Deep Learning Algorithms for the Early Detection of Breast Cancer: A Comparative Study with Traditional Machine Learning.** Informatics in Medicine unlocked.
- How good is good? Bayesian machine-learning estimation of probabilistic benchmarks in noisy datasets and an application to nanofinance+. Systems and Soft Computing 4: 200036.
- Bifurcations in business profitability: An agent-based simulation of homophily in self-financing groups. Journal of Business Research 129: 495-514.
- Which social program supports sustainable grass-root finance? Machine-learning evidence. International Journal of Sustainable Development & World Ecology 27 (5): 389-395.
- Inflation shocks and income inequality: An analysis with genetic algorithms and Bayesian quantile regressions. African Journal of Economic and Management Studies 10 (2): 226-240.
- **Balancing input-output tables with Bayesian slave-raiding ants.** Statistical Journal of the IAOS 33 (4): 943-949.

## Información de contacto

WhatsApp: +31 6 31171933

Email:

[r.m.gonzales.martinez@rug.nl](mailto:r.m.gonzales.martinez@rug.nl)

[gonzalesmartinez@gmail.com](mailto:gonzalesmartinez@gmail.com)

GitHub:

<https://github.com/rogon666/UMSA>

# Contenido del curso

## **(1) Introducción a los Modelos Lineales**

- Definición de modelos lineales.
- Regresión lineal simple y múltiple.
- Métodos de ajuste de modelos lineales
- Laboratorio: Ajuste de modelos lineales en R/Python o el programa de preferencia de los estudiantes.

## **(2) Diagnóstico y Evaluación de Modelos Lineales**

- Diagnóstico de residuos.
- Métricas de evaluación de ajuste.
- Laboratorio: Evaluación de modelos lineales

# Contenido del curso

## **(3) Modelos Lineales Generalizados (GLM)**

- Concepto de GLM.
- Distribuciones familiares en GLM: normal, binomial, Poisson.
- Funciones de enlace.
- Laboratorio: Implementación de GLM en problemas de regresión y clasificación.

## **(4) Estimación Bayesiana**

- Fundamentos de la inferencia Bayesiana.
- Teorema de Bayes.
- Métodos de MCMC (Markov Chain Monte Carlo) para estimación Bayesiana.
- Laboratorio: Estimación Bayesiana de modelos lineales y modelos lineales generalizados

# Contenido del curso

## **(5) Aplicaciones de Modelos Lineales en Machine Learning**

- Integración de modelos lineales en machine learning.
- Regularización en modelos lineales: Ridge, LASSO, Elastic Net.
- Laboratorio: Implementación de algoritmos de regularización en modelos lineales con datos de alta dimensión.

## **(6) Modelos Lineales Generalizados en Machine Learning**

- Uso de GLM en problemas de clasificación y regresión.
- Comparación con otros algoritmos de machine learning.
- Laboratorio: Aplicación de GLM en machine learning.

## Evaluación de la clase

- Prácticas en clase: 4 x 15 puntos = 60 puntos
- Examen de preguntas cerradas (respuesta multiple): 20 puntos
- Trabajo práctico individual guiado: 20 puntos



## Horarios de clase

- Lunes, miercoles y viernes: 7 PM a 10 PM
- Sábados: 8 AM a 2 PM  
4 PM a 8 PM

Lunes y miercoles: presencial, 7PM a 10 PM

Viernes: virtual, 7PM a 10 PM

Sábado: virtual, 8 AM a 12:30 PM - 3 PM a 7 PM

# Software y Hardware

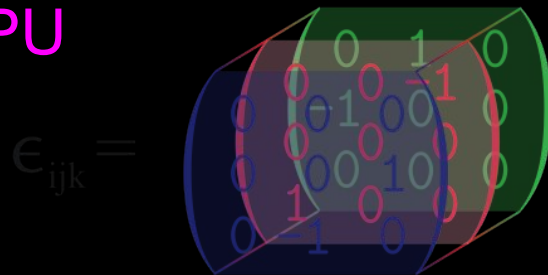
## Software

### Computación en la nube y local

- R
  - R Studio IDE\*
  - R analytic flow IDE
  - Posit cloud (R Studio)
  - Stata
- Python
  - Spyder IDE (Anaconda)
  - Anaconda cloud (Jupyter)
  - Google Colab (Jupyter)
- Matlab

## Hardware

- CPU
- GPU
- TPU



(\*) IDE: integrated development environment

Experience previa con software estadístico



**Software(s) que usan más  
frecuentemente**

# Objetivos de aprendizaje

Desarrollar habilidades para analizar datos cuantitativos con ML y MLG en la práctica:

- Comprender **qué** herramientas cuantitativas están disponibles
- Entender **cuándo** usar esas herramientas
- Saber **cómo** usar esas herramientas

¿Cuáles son sus objetivos?

¿Qué les gustaría discutir y aplicar en la clase?

# Visión de enseñanza y aprendizaje en la clase

Discutir y cubrir el material relacionado con técnicas estadísticas básicas, conocidas y comúnmente aplicadas, pero también métodos estadísticos en las fronteras de la ciencia—en el marco del UDL, la pirámide de Miller y la taxonomía de Bloom—mediante el **alineamiento constructivo y el aprendizaje activo**:

*“Learning [...] has to enable us to work at the boundary of what we know or [...] to go beyond those boundaries, or even reconstruct the very framework of our knowledge.”*

*Fear of a Black Universe* (p. 6, ch. 1, Escape From the Jungle of No Imagination), Stephon Alexander, 2021

# UDL: Diseño universal para el aprendizaje inclusivo

No hay un solo método de enseñanza que sea eficaz para tod@s, por lo que se deben ofrecer múltiples formas de representación, expresión e involucramiento (engagement)

AFFECTIVE NETWORKS:  
THE **WHY** OF LEARNING



## Engagement

For purposeful, motivated learners, stimulate interest and motivation for learning.

RECOGNITION NETWORKS:  
THE **WHAT** OF LEARNING



## Representation

For resourceful, knowledgeable learners, present information and content in different ways.

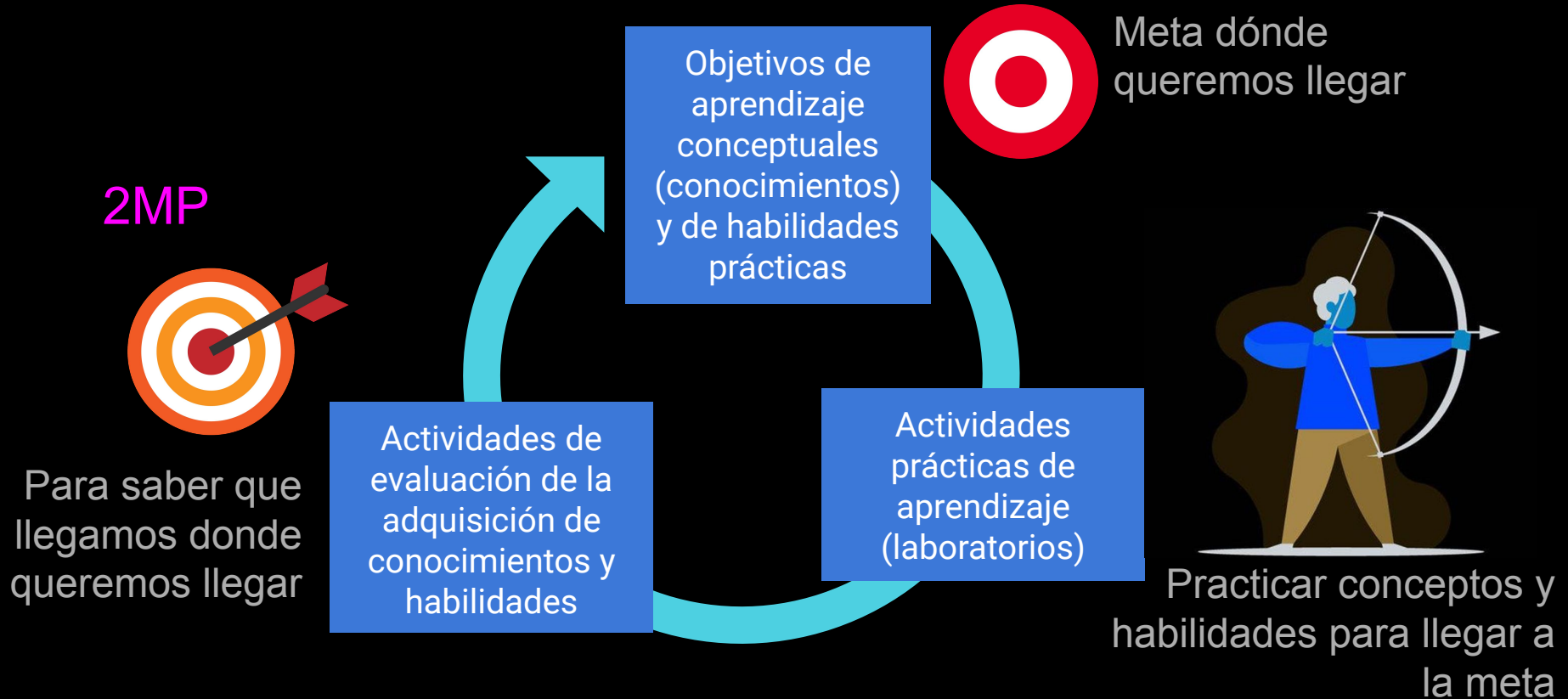
STRATEGIC NETWORKS:  
THE **HOW** OF LEARNING



## Action & Expression

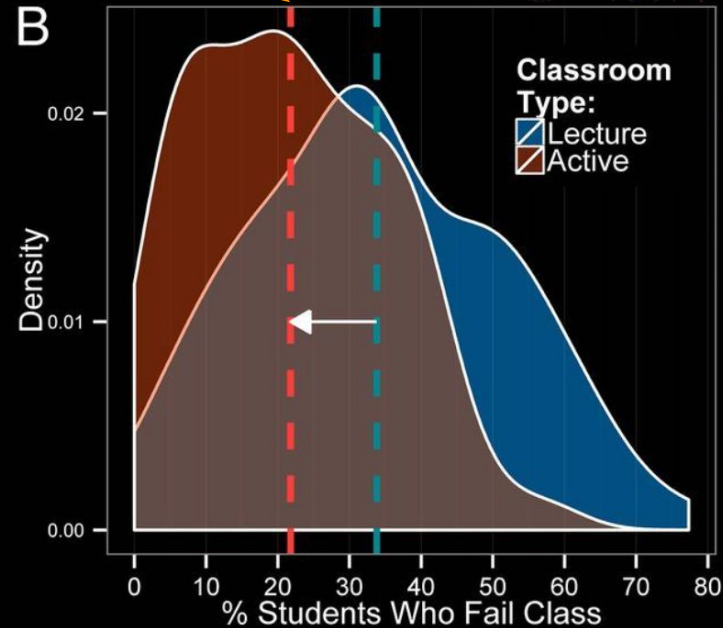
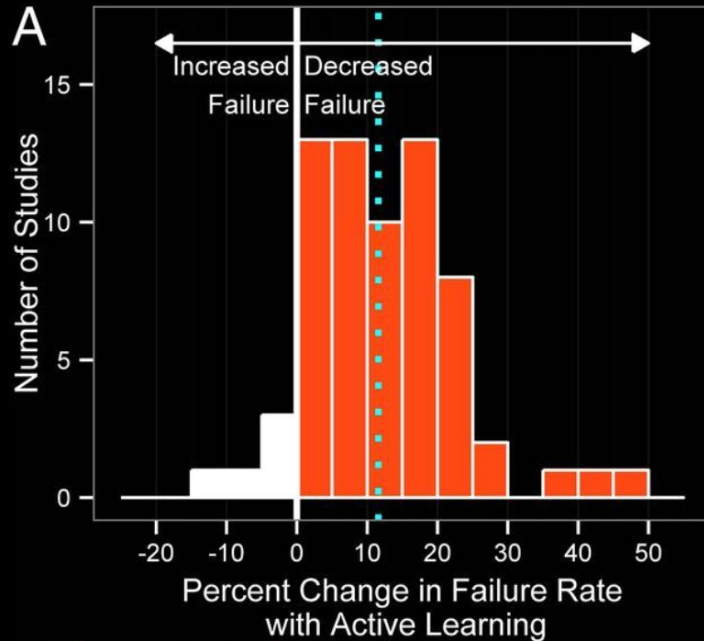
For strategic, goal-directed learners, differentiate the ways that students can express what they know.

# Alineamiento constructivo (objetivos SMART)



# Aprendizaje activo (active learning)

## Why Active Learning?





# Pirámide de Miller para para medir el progreso en la adquisición de habilidades y conocimientos



C: Conocimiento conceptual

H: Habilidad

A: Actitudes (valores, comportamientos y enfoques éticos)

# La Taxonomía de Bloom para esta clase de maestría



Versión revisada para incluir el conocimiento\*:

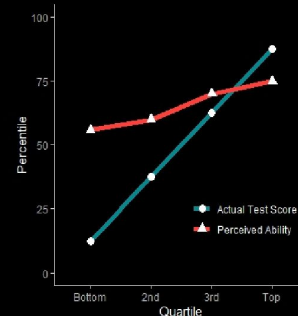
- Factual
- Conceptual
- Procedural
- Metacognitivo

(\*) Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.

# Diagnóstico de conocimientos factuales y metacognitivos

## Objetivo:

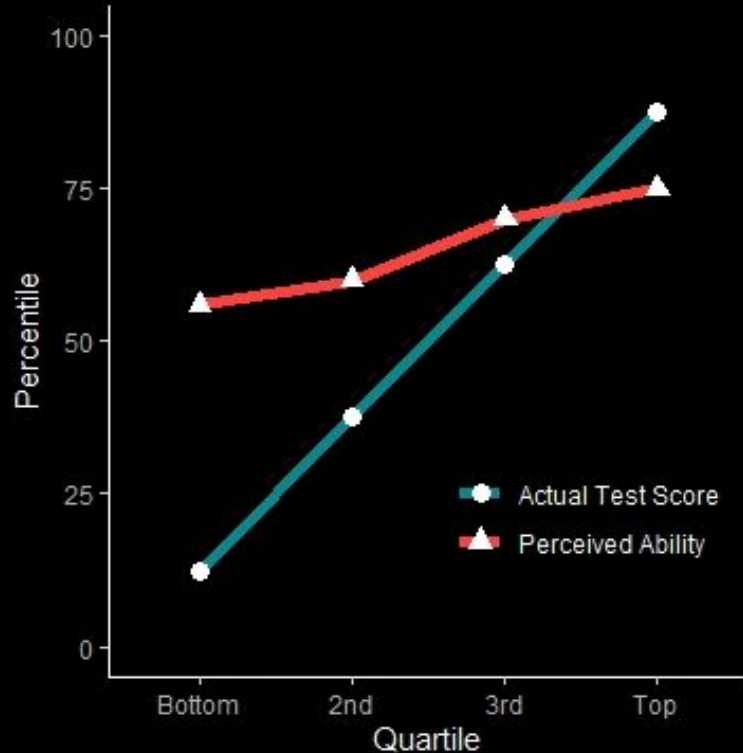
- Evaluar la necesidad de profundizar la información factual para fortalecer los conocimientos conceptuales (sobre las relaciones entre conceptos factuales)
- Diagnosticar la presencia del efecto Kruger-Dunning-Sanchez (aka efecto Dunning-Kruger)



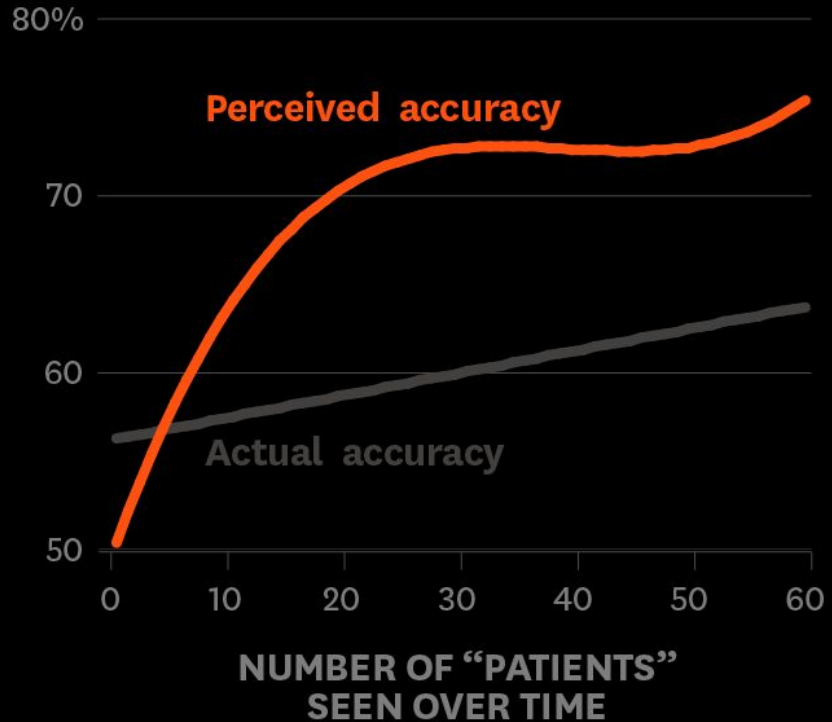
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), 1121-4(1), 10.

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing?. *Journal of personality and Social Psychology*, 11

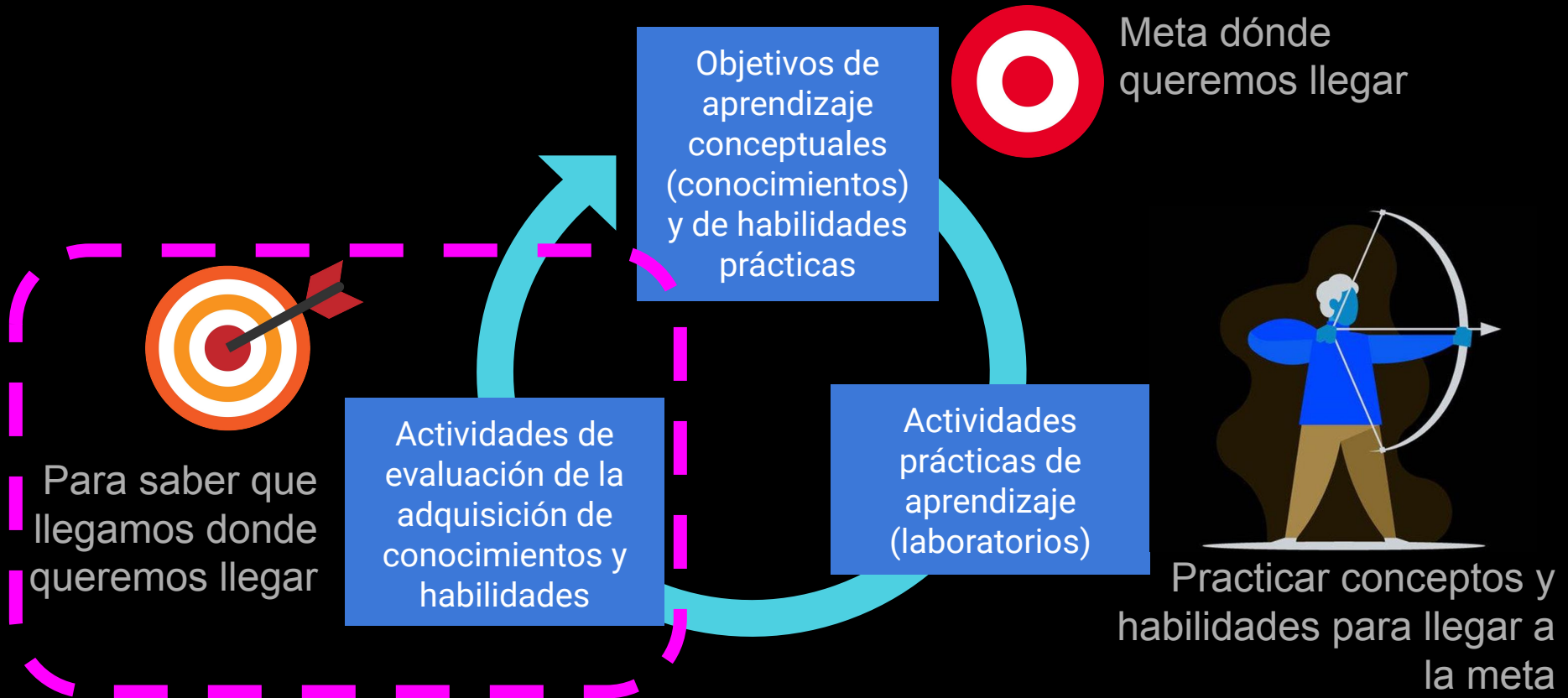
# Kruger-Dunning y Sanchez-Dunning



## CORRECT DIAGNOSES (PERCENTAGE)



# Alineamiento constructivo



# Práctica 1: resolución

```
8 # =====
9 # ~~~~~ Practica 1 resuelta ~~~~~
10 # =====
11
12 # Definiendo semilla y numero de observaciones:
13 set.seed(123)
14 n <- 100 # numero de observaciones
15
16 # Ejercicio 1:
17 beta0 = 2.5
18 beta1 = 3.5
19 e = rnorm(n, mean = 0 , sd = 1)
20 x = rnorm(n, mean = 0.5, sd = 1.5)
21 y = beta0 + beta1*x + e
22 datos1 <- data.frame(x = x, y = y)
23 modelo_lineal_1 <- lm(y ~ x, data = datos1)
24 summary(modelo_lineal_1)
```

# Práctica 1: resolución

```
26 # Ejercicio 2:
27 beta0 = 0.5
28 beta1 = 0.01
29 e = rnorm(n, mean = 0, sd = 15)
30 x = rnorm(n, mean = 0.5, sd = 1.5)
31 y = beta0 + beta1*x + e
32 datos2 <- data.frame(x = x, y = y)
33 modelo_lineal_2 <- lm(y ~ x, data = datos2)
34 summary(modelo_lineal_2)
35
36 # Ejercicio 3:
37 beta0 = 2.5
38 beta1 = 0.1
39 beta2 = -0.5
40 e = rnorm(n, mean = 0, sd = 10)
41 x1 = rnorm(n, mean = 0.5, sd = 1.5)
42 x2 = rnorm(n, mean = -2.5, sd = 1)
43 y = beta0 + beta1*x1 + beta2*x2 + e
44 datos3 <- data.frame(x = x, y = y)
45 modelo_lineal_3 <- lm(y ~ x1 + x2, data = datos3)
46 summary(modelo_lineal_3)
```

# Práctica 1: resolución

```
48 # Ejercicio 4:
49 url <- "https://raw.githubusercontent.com/rogon666/UMSA/main/MLMLG/datos/ejercicio01.csv"
50 download.file(url, destfile = "ejercicio01.csv")
51 datos4 <- read.csv("ejercicio01.csv")
52 datos4$x2_dummy <- ifelse(datos4$x2 == "si", 1, 0)
53 datos4$x2 <- NULL
54 modelo_lineal_4 <- lm(y ~ x1 + x2_dummy, data = datos4)
55 summary(modelo_lineal_4)
```