Módulo 2

CONCEPTOS BÁSICOS DE ESTADÍSTICA

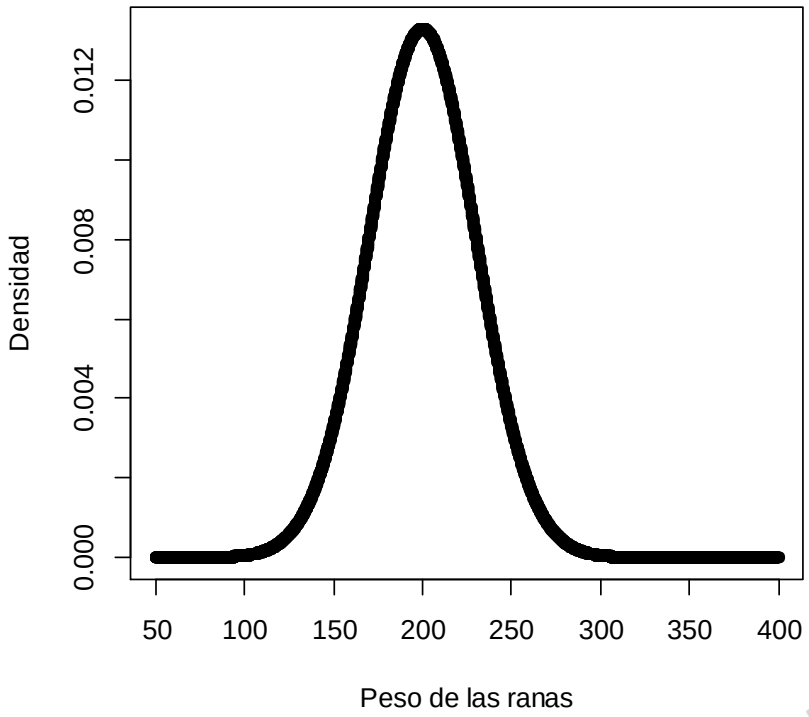
Curso de Posgrado: "Modelado y estimación de ocupación para poblaciones y comunidades de especies bajo enfoque Bayesiano"

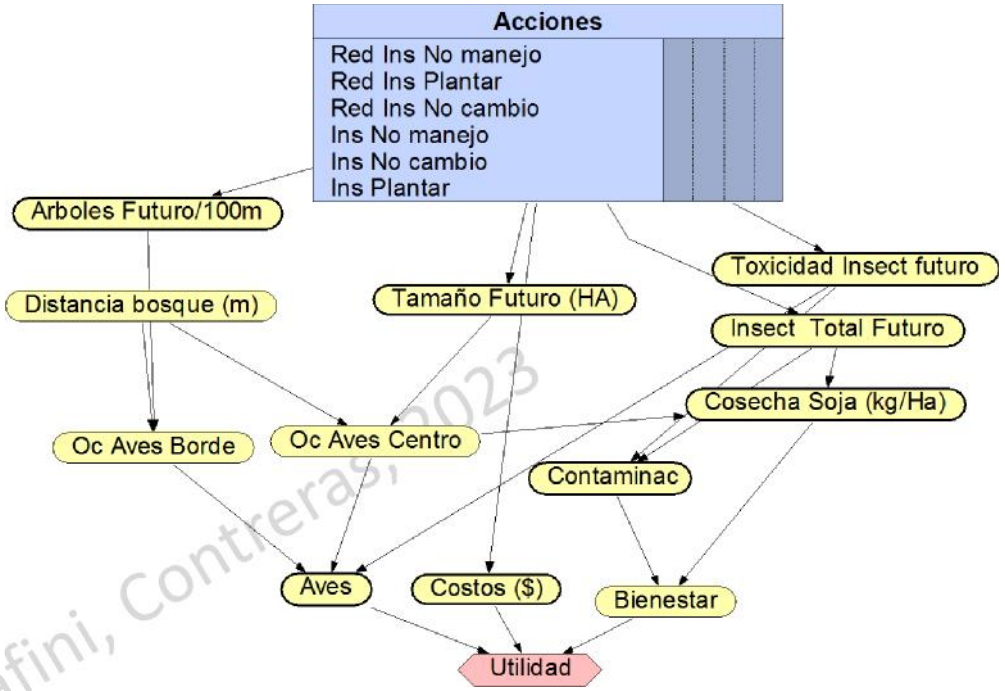
CCT CONICET Mendoza

24 - 28 Abril 2023

¿QUÉ ES UN MODELO?

- Abstracción de la realidad
- Los usamos todos los días
  - Conceptuales
  - Físicos
  - Gráficos
  - Analíticos
  - Numéricos
  - Empíricos o estadísticos





Model*i*	Prior weight $p(m_i)$	Likelihood[1] $p(x\|m_i)$	Posterior weight $p(m_i\|x)$
No effect	0.00001	0.00257	0.00000
Bird group	0.00001	0.02219	0.00000
Tree	0.57830	0.58101	0.16351
Tree + Bird group	0.42160	4.07702	0.83649
Forest	0.00001	0.18890	0.00000

[1]Likelihood values where multiplied by 1E+237 to eliminate excessive zeroes

$$\text{logit}(\psi_i) = \alpha_{psi} + \beta_{x1} * x1_i$$

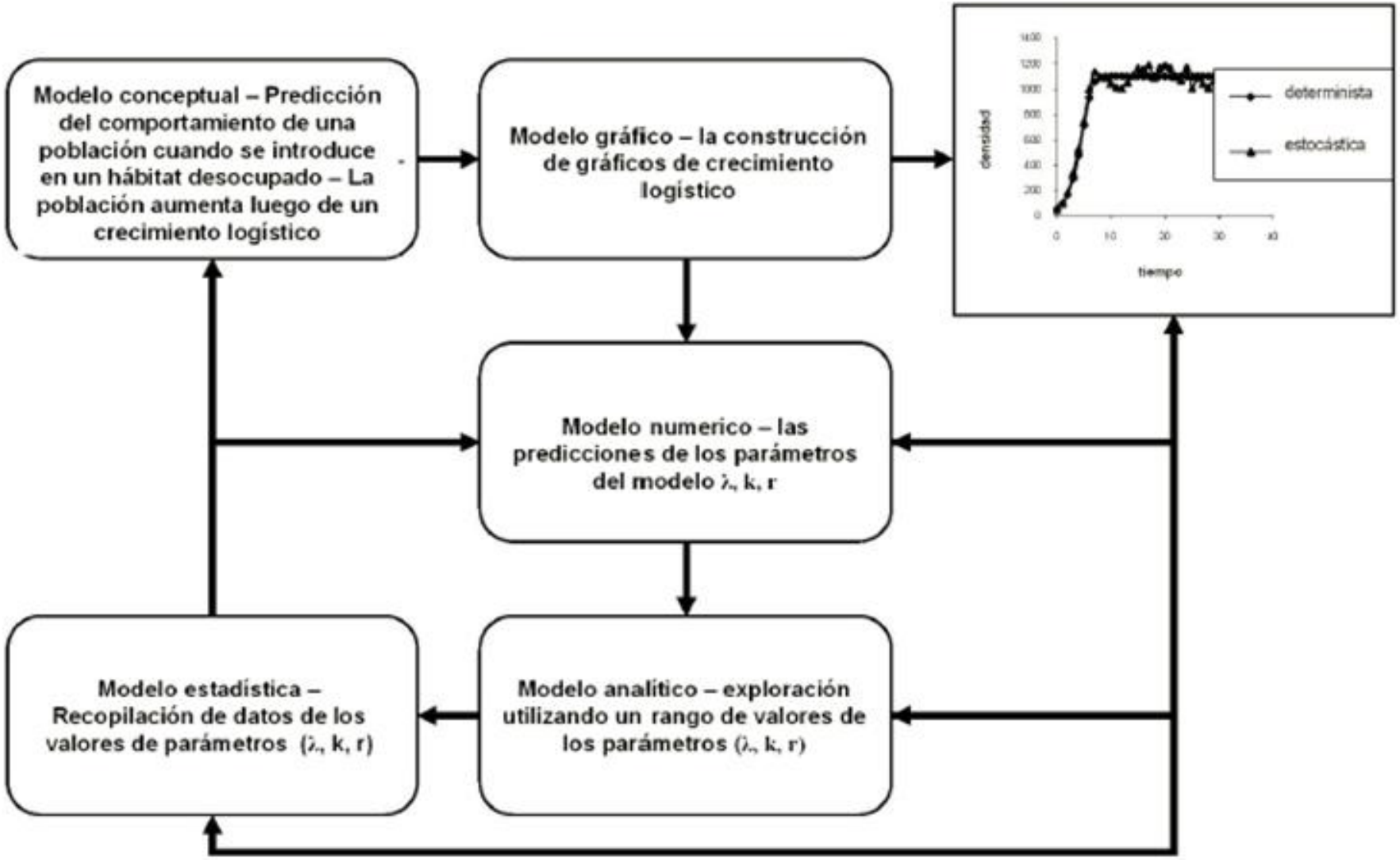


Figura 2.1. Diagrama de flujos de las realimentaciones de diversos tipos de modelos que pueden utilizarse para comprender mejor un problema en la biología de la conservación.

(Conroy & Carroll 2009, Conroy et al. 2015)

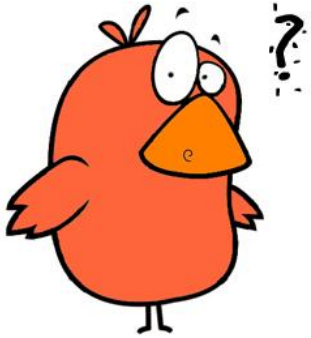
MODELOS

- Cualitativo
  - ej. descripción general de un área
- Cuantitativo
  - Resultado preciso

- Discretos
  - ej. abundancia
- Continuos
  - ej. densidad

MODELOS

- Determinístico
  - No hay incertidumbre
- Estocástico
  - Distribuciones de probabilidad

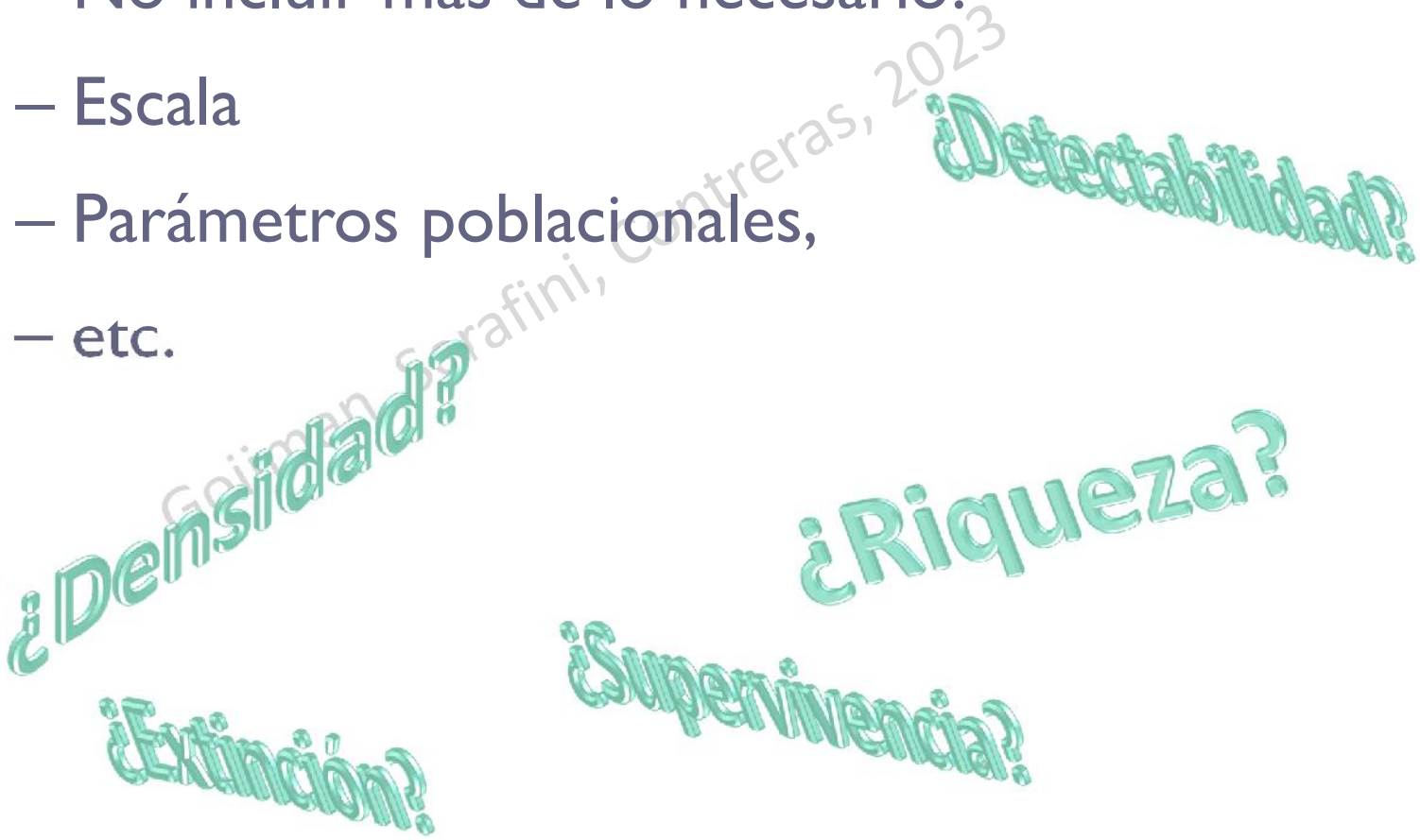


A + B = C ✓

A + B + X = C ± Incertidumbre

CLAVES PARA ELABORAR MODELOS

- Definir claramente el objetivo
  - No incluir mas de lo necesario!
  - Escala
  - Parámetros poblacionales,
  - etc.



CLAVES PARA ELABORAR MODELOS

- Definir componentes
  - Parámetros: lo que tratamos de estimar
    - Constante o variable
    - Fijo o aleatorio
  - Variables
    - Respuesta o dependiente: lo que tratamos de modelar
    - Predictiva o independiente: a la derecha de la ecuación. Explicatoria.

MODELOS ESTADISTICOS

- Producir inferencias **confiables** para explicar el mundo natural – Resultados **replicables** y **defendibles**
  - Datos colectados siguiendo un diseño apropiado.
  - Analizar datos con un modelo apropiado: tener en cuenta el diseño y usar los principios de probabilidad y estadística para hacer inferencias válidas

MODELOS ESTADÍSTICOS

Estos modelos son construidos alrededor de valores aleatorios o **estadísticos** que son observados como datos de una muestra.

Estadísticos: Cualquier función de los datos muestreales (ej. media, varianza, percentiles)

MODELOS VERSUS REALIDAD

- En ciencias biológicas no podemos esperar encontrar la verdad exacta, o alcanzar la realidad con un set finito de datos.
  - Dimensiones infinitas vs. Muestras finitas

- El **error observacional o de medición** se refiere a la diferencia entre un valor medido de una cantidad y su valor verdadero

MODELOS VERSUS REALIDAD

- Inferencia basada en un buen modelo aproximado.
  - inferencia condicional a los datos
- No existe un único modelo que explique la realidad…

¿Cómo encontramos el modelo que "mejor" la explica?

DISTRIBUCIÓN DE PROBABILIDAD

- Naturaleza estocástica del mundo natural explicada por medio de **variables aleatorias**

  Es una característica que exhibe una variabilidad entre unidades o elementos con dicha característica

- Los posibles valores de una variable **aleatoria** tiene valores posibles que pueden ser representados con una abstracción matemática: **distribución de probabilidad**

- La distribución de probabilidad le asigna a cada evento de una variable aleatoria, una probabilidad de ocurrir.

DISTRIBUCIÓN DE PROBABILIDAD

- Una probabilidad puede pensarse como una medida de incertidumbre de un evento aleatorio
  - Si *X* tiene *P*=1de ocurrir , estamos seguros que *X* ocurre
  - Si *P=0* entonces estamos seguros que *X* no ocurre
  - *Si P=0.5* estamos igual de seguros que *X* ocurre y que no

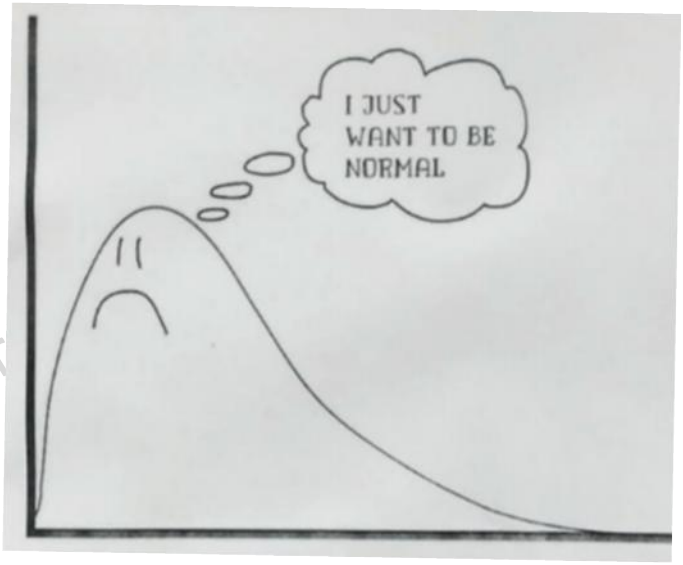
El valor "*X*" es una variable aleatoria

DISTRIBUCIÓN DE PROBABILIDAD

- Es un modelo que describe la relación entre los valores de una variable aleatoria y la probabilidad de asumir esos valores
- Describe todas las posibles posibilidades de ocurrencia, para que la suma de todas las probabilidades sea 1.

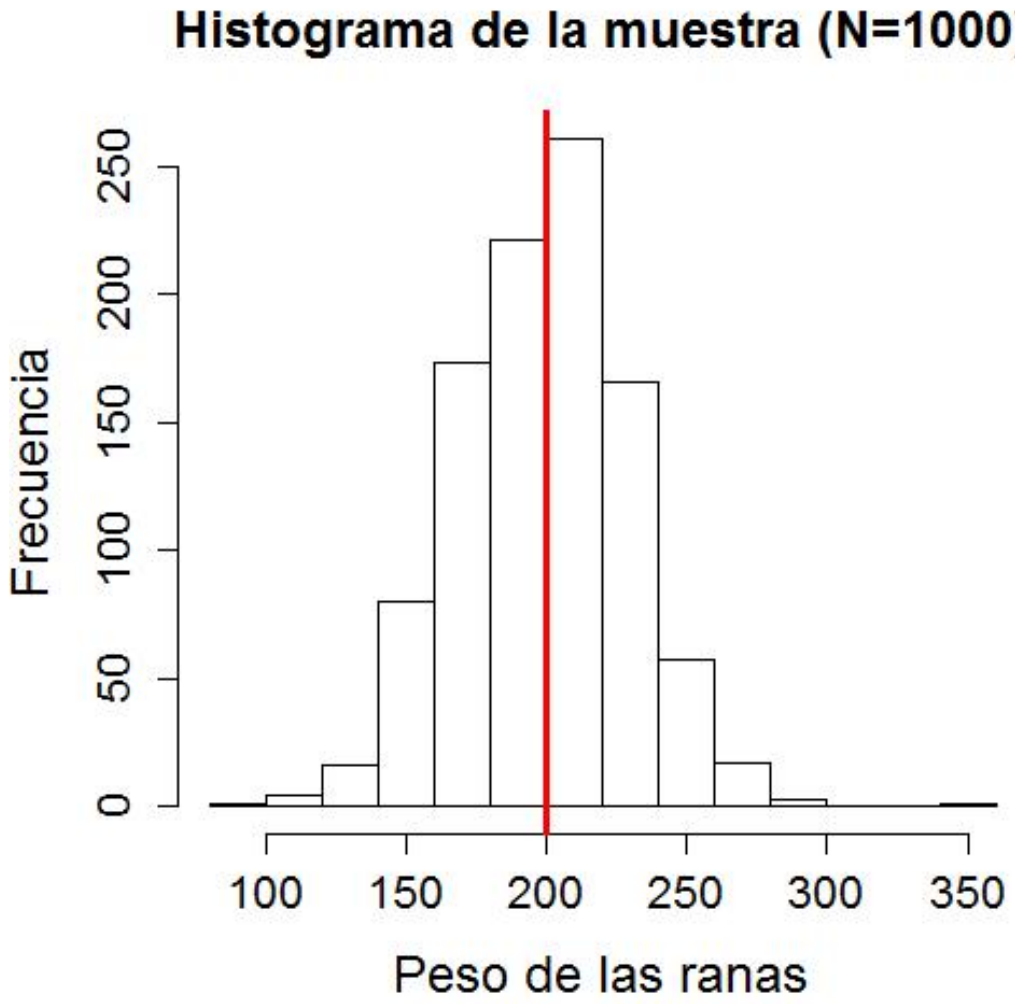
DISTRIBUCIÓN DE PROBABILIDAD

- El error de observación o de medición generalmente es tratado en la estadística como normalmente distribuido con media 0



- En ecología de poblaciones y comunidades las cantidades focales como abundancia o riqueza son típicamente medidas en conteos que son menores a lo que realmente hay (nos "perdemos" individuos)
- Este error necesita otros tipos de distribuciones

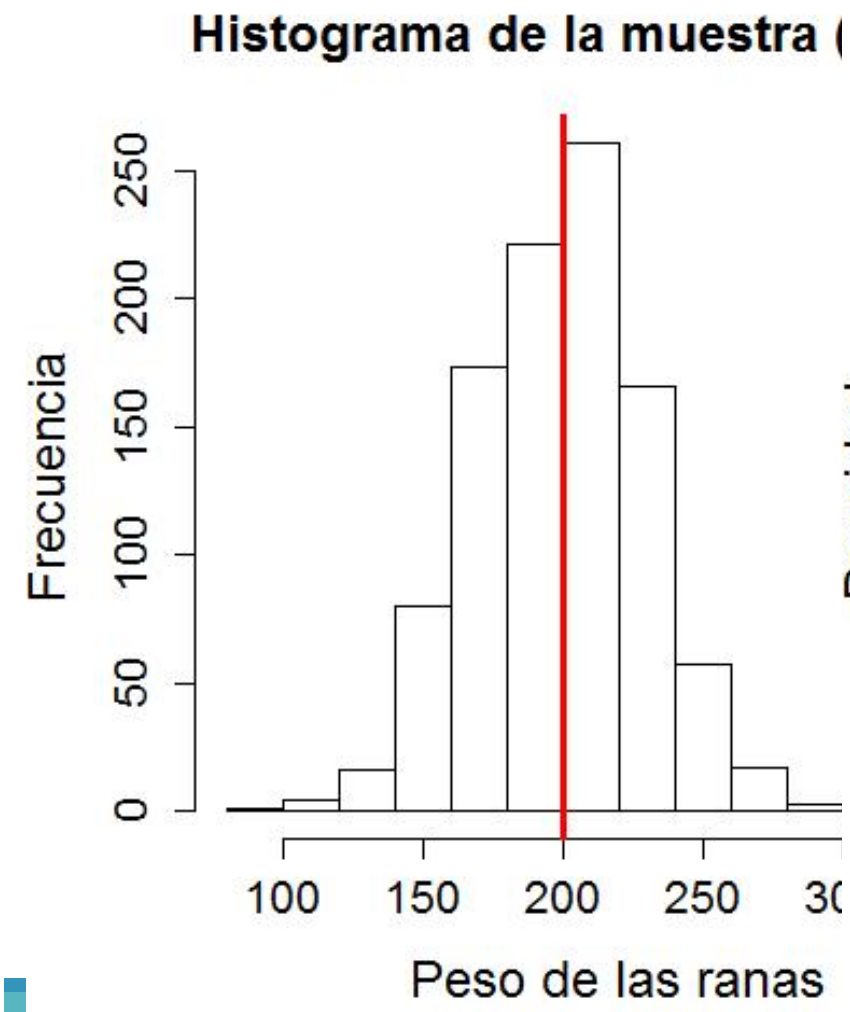
DISTRIBUCIÓN DE FRECUENCIAS

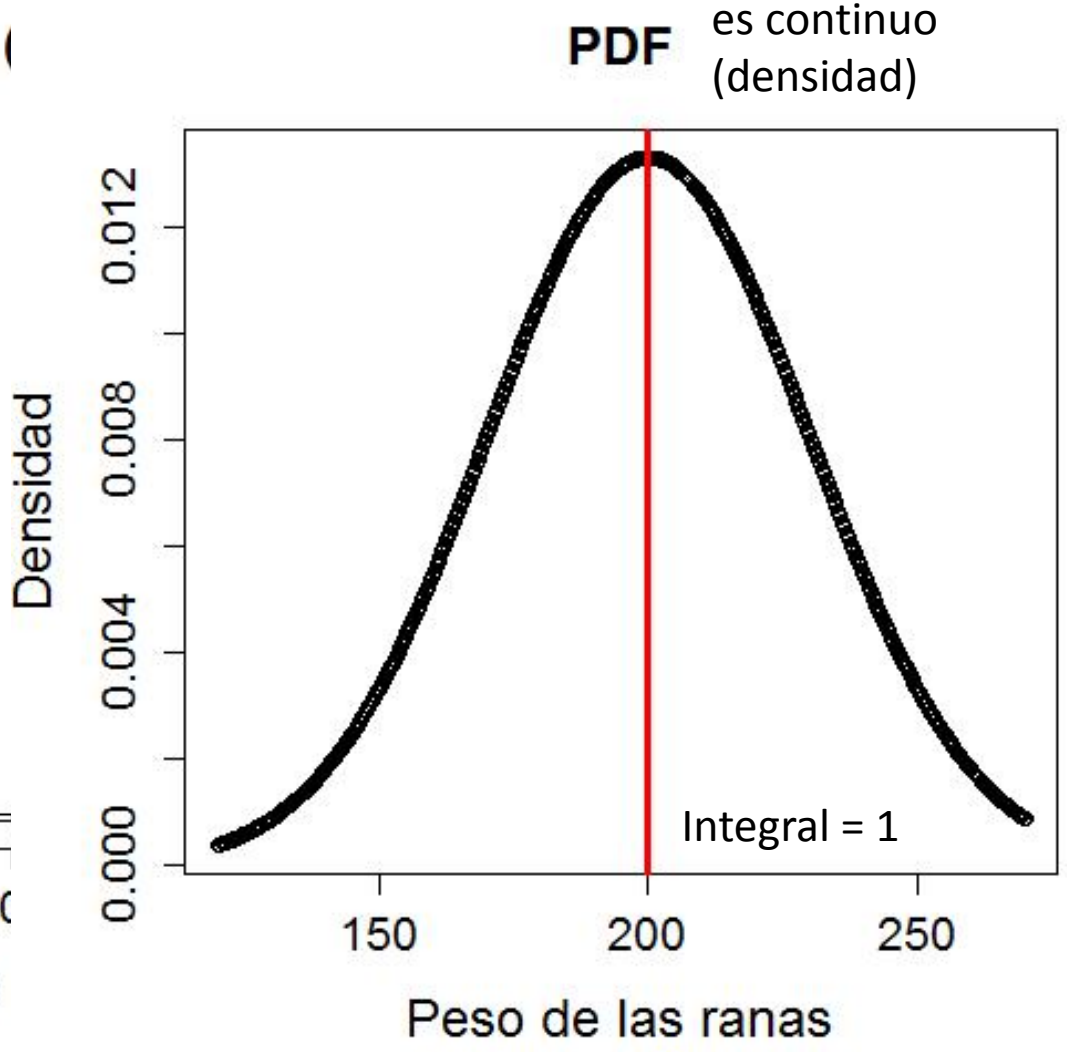



```
# numero de muestras
n <- 1000

# media del peso de las ranas
mean <- 200

# SD del peso de las ranas
sd <- 30
```

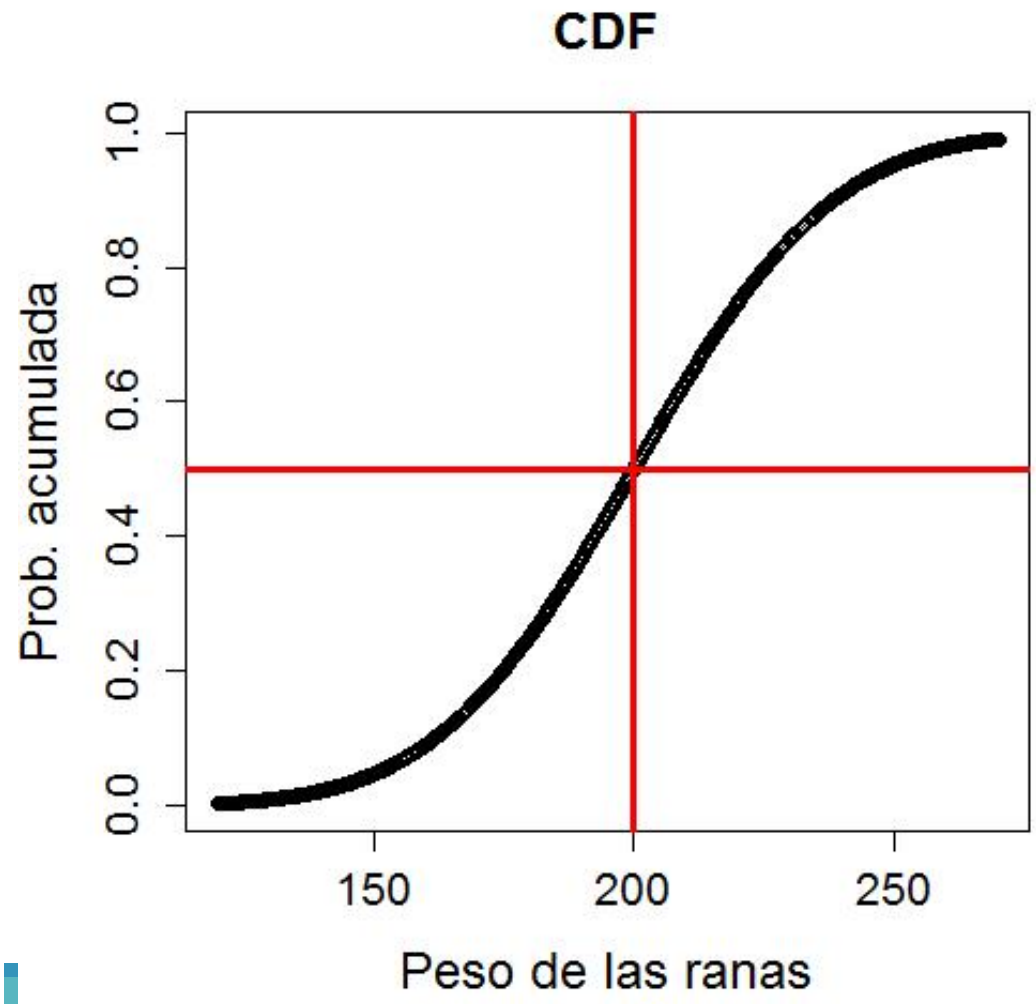
FUNCION DE DENSIDAD DE PROBABILIDAD (PDF)

No hay valor exacto de pb. pq es continuo (densidad)





FUNCION PROBABILIDAD DE DISTRIBUCION (CDF)



Probabilidad que una variable aleatoria *X* (continua) sea menor o igual a un valor particular

FUNCIONES DE PROBABILIDAD

Las funciones de densidad (o masa) de una probabilidad (PDF o PMF)

f(y) dependen de una o mas cantidades, llamadas **parámetros**

DISTRIBUCIÓNES DE PROBABILIDAD

- Variables aleatorias discretas

Bernoulli: Dos valores posibles, 1 evento

Binomial: Dos valores posibles, >1 evento

Multinomial >1 valor posibles, >1 evento

Poisson: Valores discretos, no negativos

puede tomar solo un número contable de valores distintos como 0, 1, 2, 3, 4, 5... 100, 1 millón, etc.

DISTRIBUCIÓNES DE PROBABILIDAD

- Variables aleatorias continuas

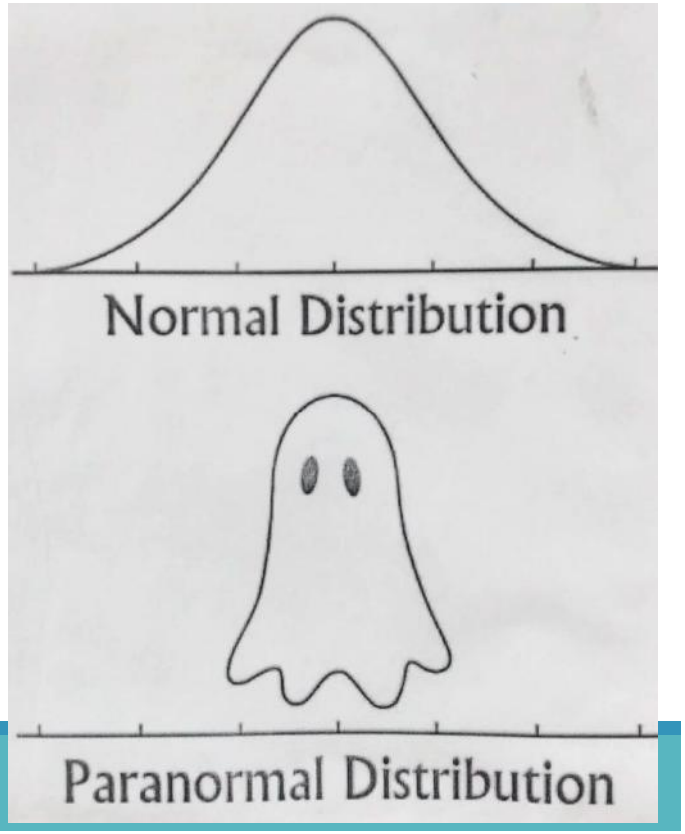
Uniforme: Probabilidad uniforme, $a \leq x \leq b$

Normal: $(-\infty, +\infty)$

Beta: $0 < x < 1$

Gamma: $0 \leq x < +\infty$

puede tomar un número infinito de valores posibles.



ESTIMACION DE PARÁMETROS

Un **estimador** de un parámetro poblacional, se basa en un muestreo aleatorio.

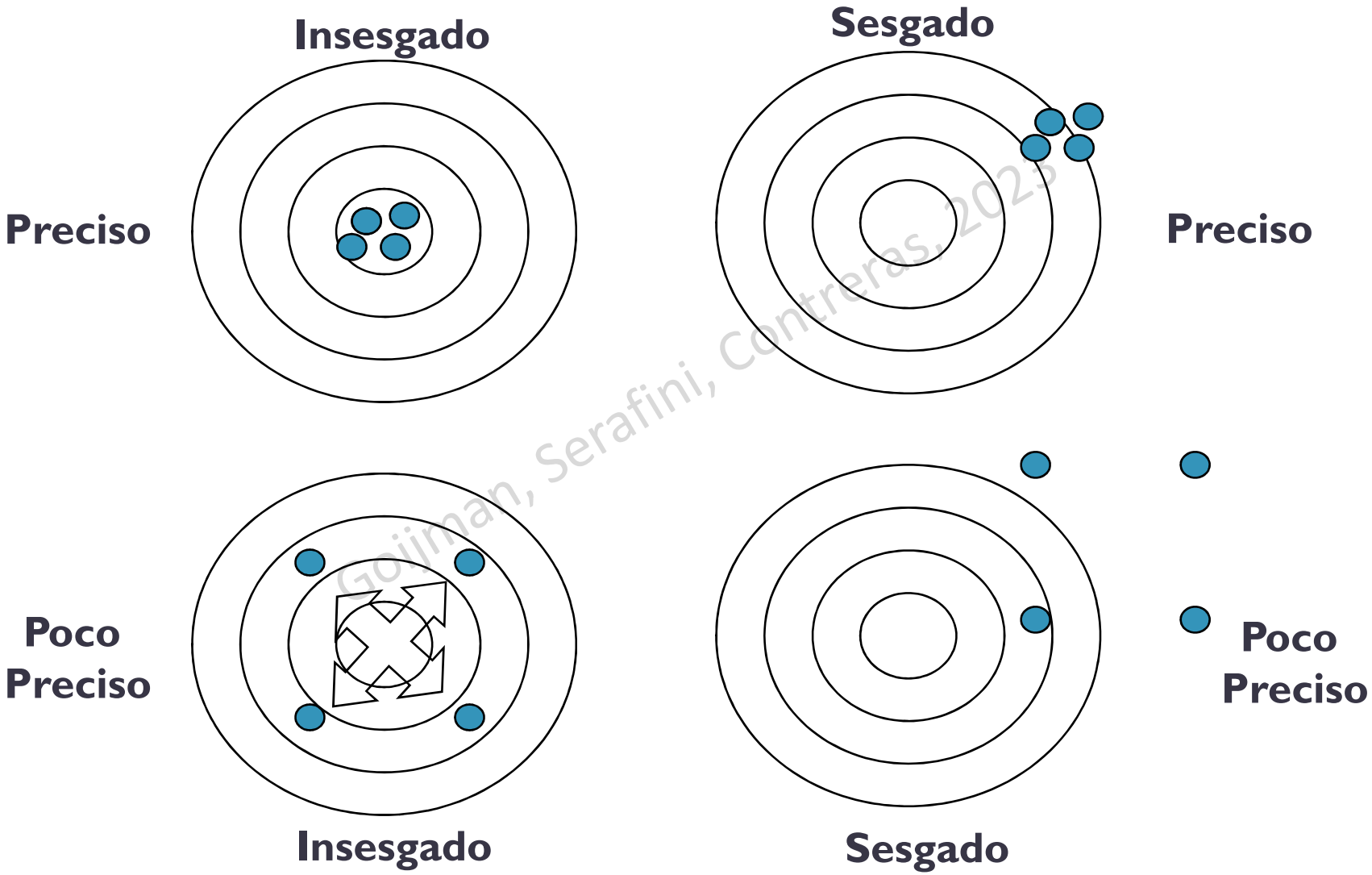
Estimador es una **variable aleatoria** con una **distribución estadística.**

ESTIMACION DE PARÁMETROS

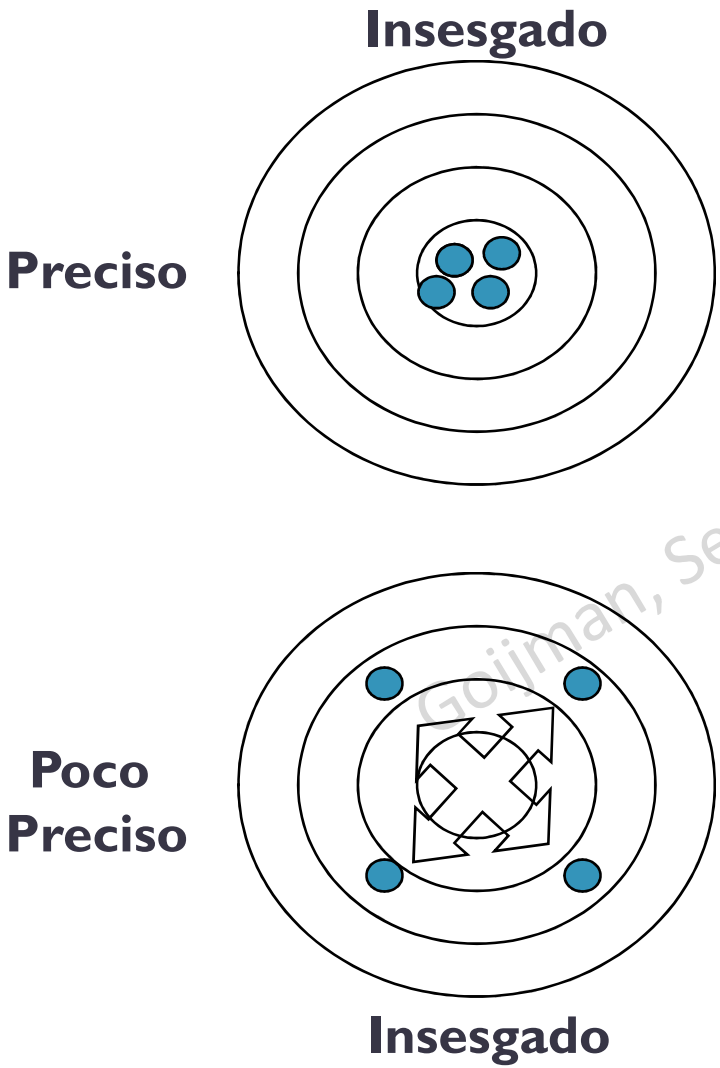
Medidas de comportamiento estadístico de un estimador:

- **Precisión**: Es una medida del error del muestreo
- **Sesgo:** Diferencia entre el valor observado y la realidad
- **Exactitud:** Precisión + Sesgo

ESTIMACION DE PARÁMETROS



ESTIMACION DE PARÁMETROS



Si tratamos a una variable aleatoria continua de distribución normal (ej. Peso, altura), puede o no ser preciso, pero los errores a cada lado de la distribución en mediciones sucesivas son iguales (insesgado)

ESTIMACIÓN DE PARÁMETROS

Para conteos de variables discretas (abundancia o "presencia/ausencia")

- No es lo mismo contar de más o de menos
- Al contar de menos (no detectar un organismo) – **Falsos negativos**
- "Presencia/ausencia" *(y)* es representada por una Bernoulli (2 valores posibles, 1 evento)

$$y \sim Bernoulli\ (p)$$

- No es insesgado (no se cancelan los errores)

BONDAD DE AJUSTE

- Diferencia entre valores esperados bajo un modelo y lo observado
- Ejemplo, test de Chi cuadrado

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

χ^2 (estimado) > χ^2 (crítico/tabla)

No hay ajuste

MÉTODOS PARA ESTIMACIÓN DE PARÁMETROS

- Métodos frecuentistas

Estiman la **verosimilitud** de observar los datos

Se basan en la **frecuencia esperada** de que esos datos sean observados si el mismo procedimiento de recolección de datos y análisis fuese implementado muchas veces

- TEST DE HIPÓTESIS NULA (verosimilitud de observar datos extremos $p \leq 0,05$)
- MÉTODOS DE TEORÍA DE LA INFORMACIÓN

MÉTODO DE TEORÍA DE INFORMACIÓN (*Information theoretic*)

- Definición *a priori* del set de modelos candidatos (hipótesis)
- Los datos se utilizan pare evaluar el soporte relativo de diferentes modelos.
- El mejor modelo es aquel que pierde la menor cantidad de información.
- **Compromisos** entre el **ajuste del modelo** (+parámetros) y la **varianza del estimador** (-parámetros = parsimonia) por medio de una optimización.

PROBABILIDAD y VEROSIMILITUD

- **Función de probabilidad**

Parámetros, modelo, tamaño muestral → CONOCIDO

¿Cuál es la probabilidad de observar un evento X?

$$f(x|\theta)$$

- **Función de verosimilitud *("Likelihood")***

Datos (observados), modelo (asumido) → CONOCIDO

¿Cuál/es son los parámetros que relacionan los datos al modelo?

$$L(\theta|x)$$

$$L(\theta|datos, modelo)$$

ESTIMACIÓN DE PARÁMETROS

METODO DE MAXIMA VEROSIMILITUD

- Con los datos colectados queremos estimar los valores de los parámetros que los explican
- Seleccionar los valores de los parámetros para maximizar la función de verosimilitud

$$L(\boldsymbol{\theta}|\boldsymbol{datos}, \boldsymbol{modelo})$$

MÉTODO DE MÁXIMA VEROSIMILITUD
MLE ("*Maximum likelihood estimation*")

Ejemplo Binomial

$$\text{VEROSIMILI TUD}: \quad L(p \mid n, x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$L(p \mid n, x) = \binom{n}{x} p^x (1-p)^{n-x}$$

n = 10 trampas de ratones

x = 0 capturado, x = 0 no capturado

x= {0,1,1,1,0,1,1,0,0,1}

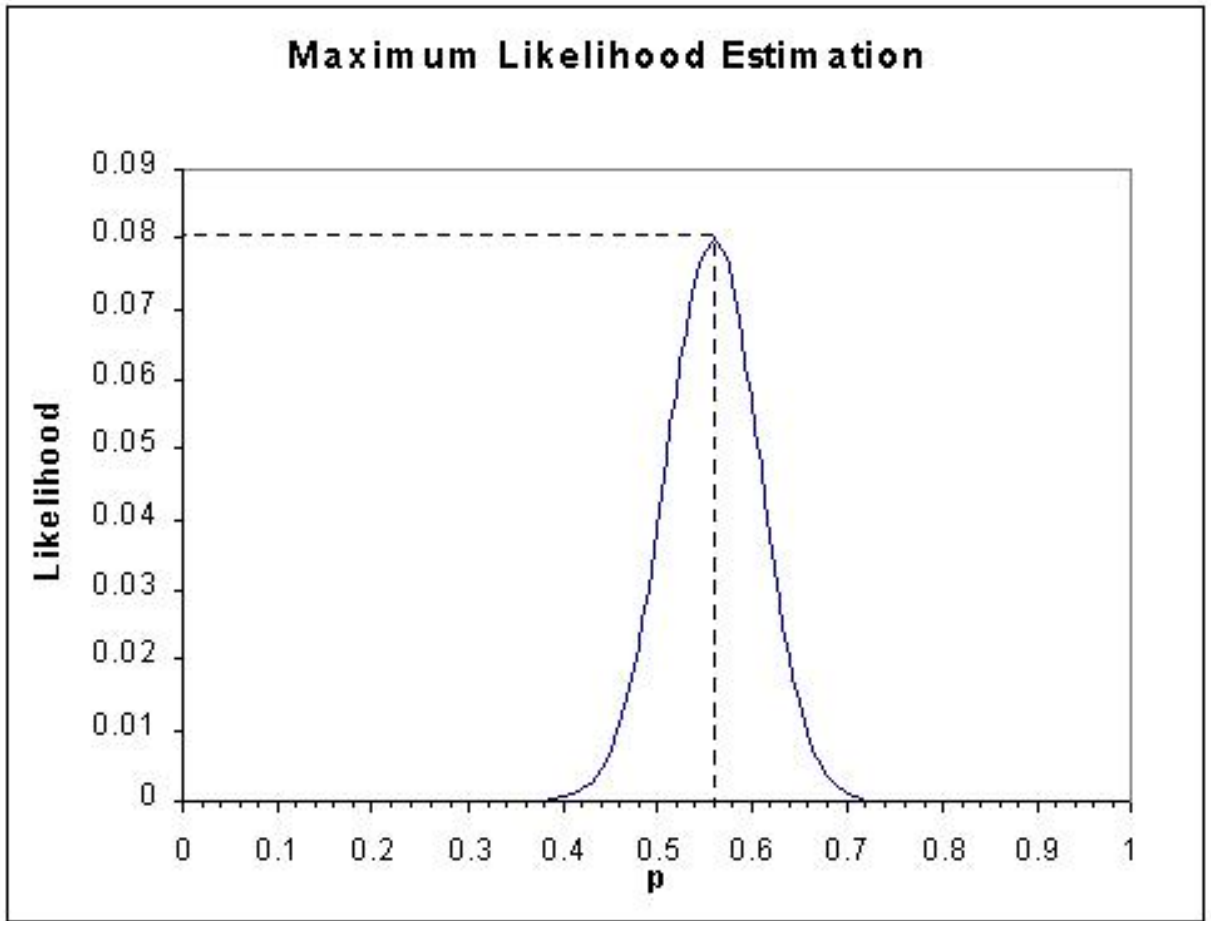
¿Cuál es la probabilidad p de captura?

MÉTODO DE MÁXIMA VEROSIMILITUD

¿Cuál es la probabilidad p de captura?

$$L(p \mid n, x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

2023



MÉTODO DE MÁXIMA VEROSIMILITUD

¿Cuál es la probabilidad p de captura?

$$L(p \mid n, x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

Hoja de cálculo

METODO DE MAXIMA VEROSIMILITUD (MLE)

1) Aplico ln (p sigue siendo igual)

$$L(p \mid n, x) = \binom{10}{6} p^6 (1-p)^4$$

$$\ln L(p \mid n, x) = \ln \binom{10}{6} + 6 \ln p + 4 \ln(1-p)$$

2) Derivo con respecto a p (busco máximo)

$$\frac{\ln L(p)}{\partial p} = \frac{6}{p} - \frac{4}{1-p} = 0$$

$$\boxed{\hat{p} = 6/(6+4) = 0.6}$$

REFERENCIAS

- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference : a practical information-theoretic approach. 2nd edition. Springer, New York.
- Conroy, M. J., and J. P. Carroll. 2009. Quantitative conservation of vertebrates. Wiley-Blackwell, Chichester, West Sussex, UK ; Hoboken, NJ.
- Kéry, M. 2010. Introduction to WinBUGS for Ecologists: A Bayesian Approach to Regression, ANOVA and Related Analyses. Access Online via Elsevier.
- Marc Kery & J. Andy Royle. 2016. Applied hierarchical modeling in ecology. Modeling distribution, abundance and species richness using R and BUGS. Volume 1: Prelude and Static models. Academic Press
- Williams, B., J. Nichols, and M. Conroy. 2002. Analysis and Management of Animal Populations: Modeling, Estimation, and Decision Making. Academic Press, San Diego, CA.