INTA | CIRN Instituto de Recursos Biológicos

Módulo I

CONCEPTOS BÁSICOS DE ESTADÍSTICA

Dra. Andrea P. Goijman

Curso de Posgrado: "Métodos cuantitativos de detección imperfecta para el análisis de poblaciones y comunidades de fauna silvestre"

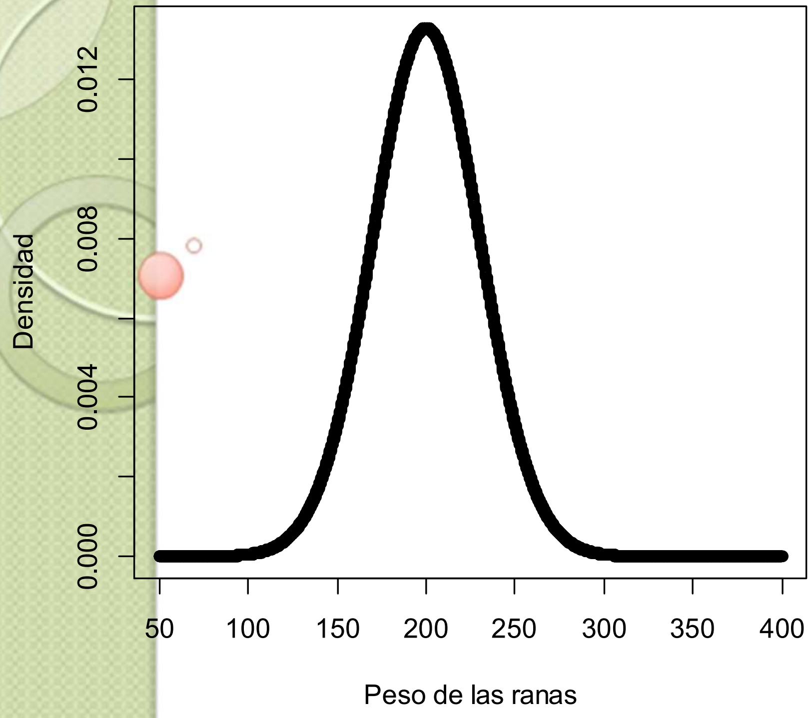
Dpto. de Ciencias Naturales, UNRC

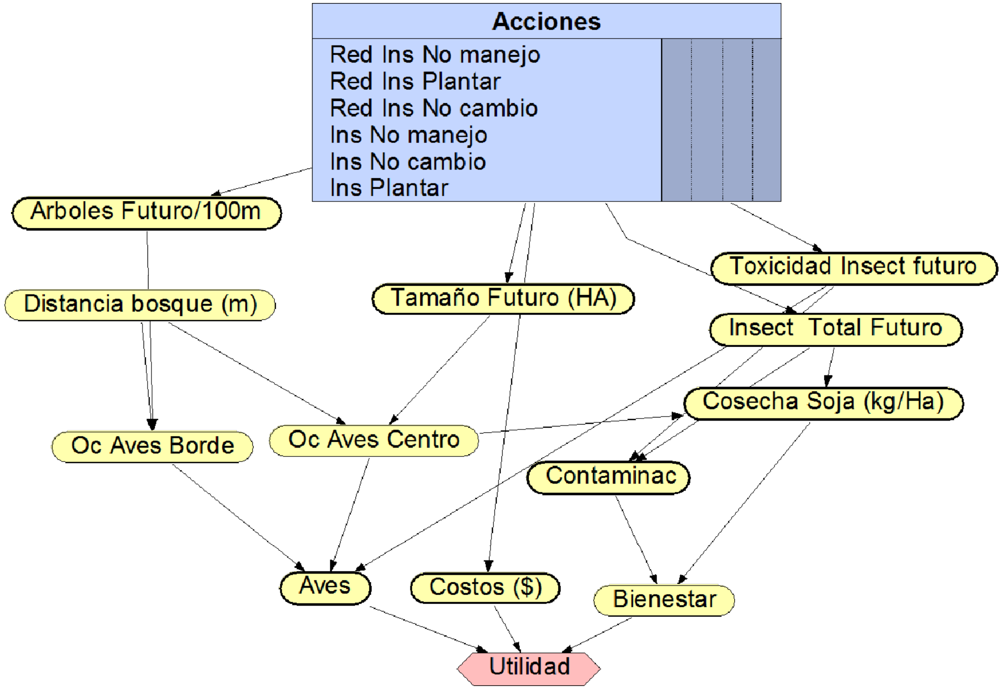
27 de Junio – 1 de Julio 2016

INTA Ministerio de Agroindustria Presidencia de la Nación

¿QUÉ ES UN MODELO?

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





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

Ministerio de Agroindustria
Presidencia de la Nación

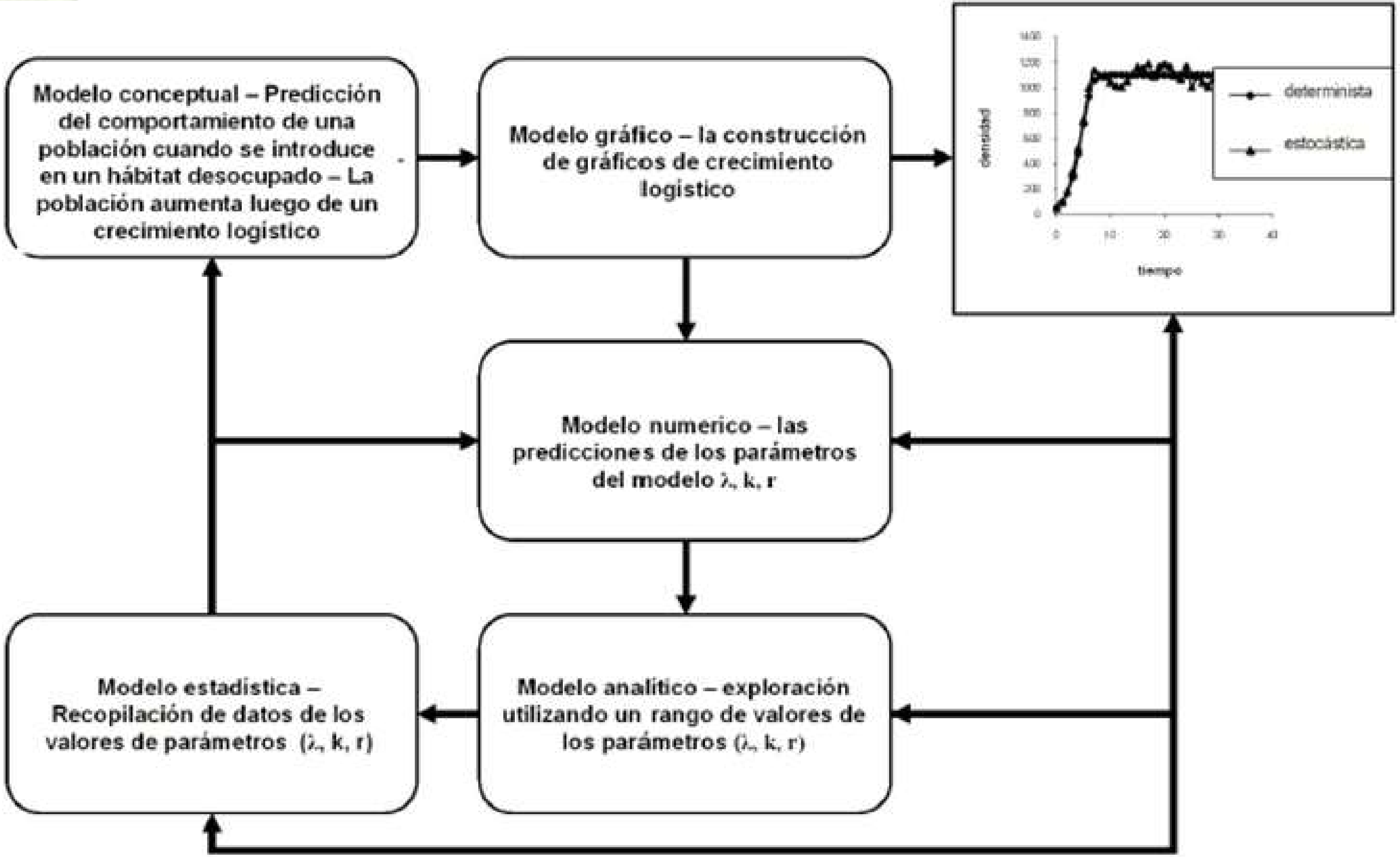


Figura 2.1. Diagrama de flujos de las realimentaciones de diversos tipos de modelos que pueden utilizarse para comprender mejor un problema en la biología de la conservación.

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

Ministerio de Agroindustria
Presidencia de la Nación

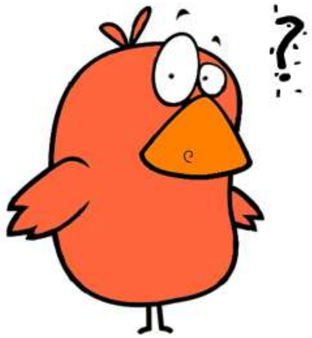
MODELOS

- 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

INTA

Ministerio de Agroindustria
Presidencia de la Nación

CLAVES PARA ELABORAR MODELOS

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

¿Detectabilidad?

¿Densidad?

¿Riqueza?

¿Supervivencia?

¿Extinción?

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)

DISTRIBUCIÓN DE PROBABILIDAD

- Naturaleza estocástica del mundo natural explicada por medio de **variables aleatorias**
- La combinación los factores que no podemos medir puede ser representado 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=1* de 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

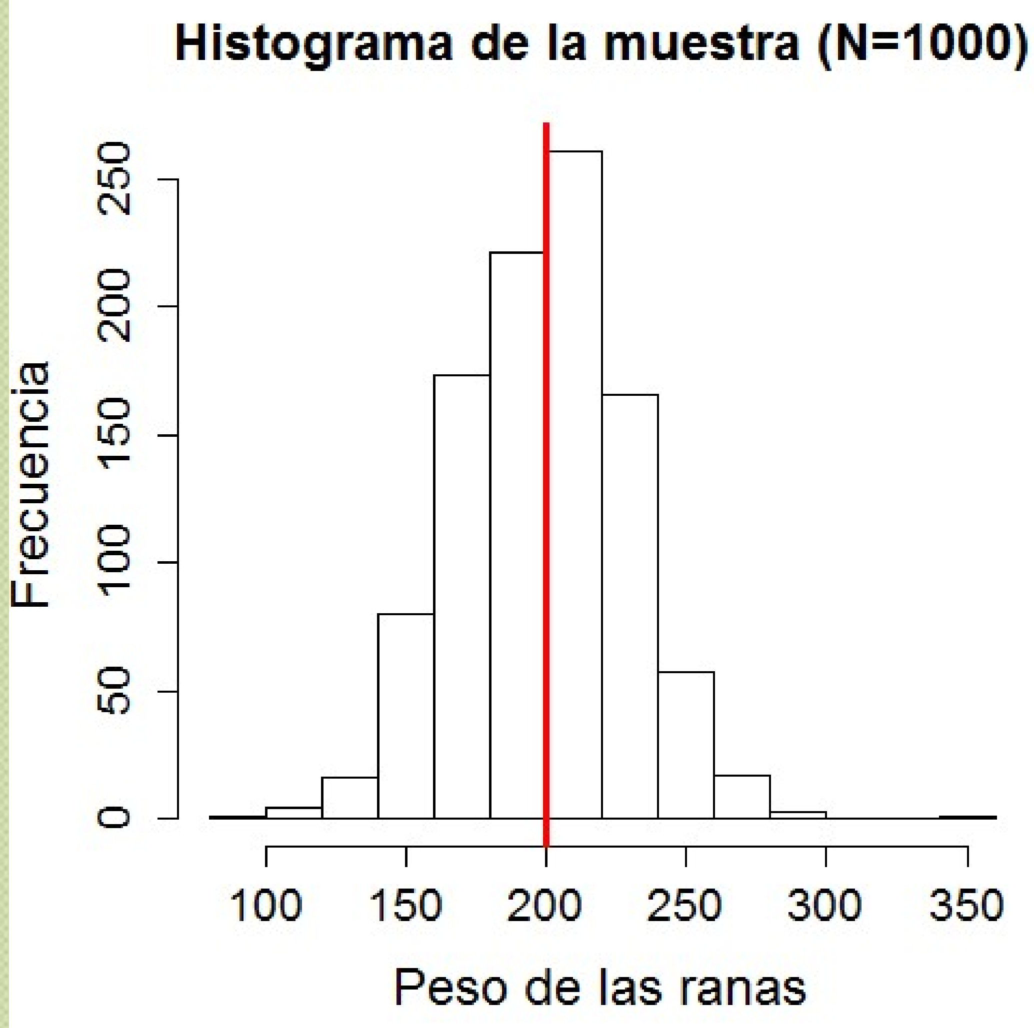
- Es un modelo que describe la relación entre los valores de una variable aleatoria y la probabilidad de asumir esos valores
- Describe todas las posibles posibilidades de ocurrencia, para que la suma de todas las probabilidades sea 1.

DISTRIBUCIÓN DE PROBABILIDAD

- 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.

Ministerio de Agroindustria
Presidencia de la Nación

DISTRIBUCIÓN DE FRECUENCIAS



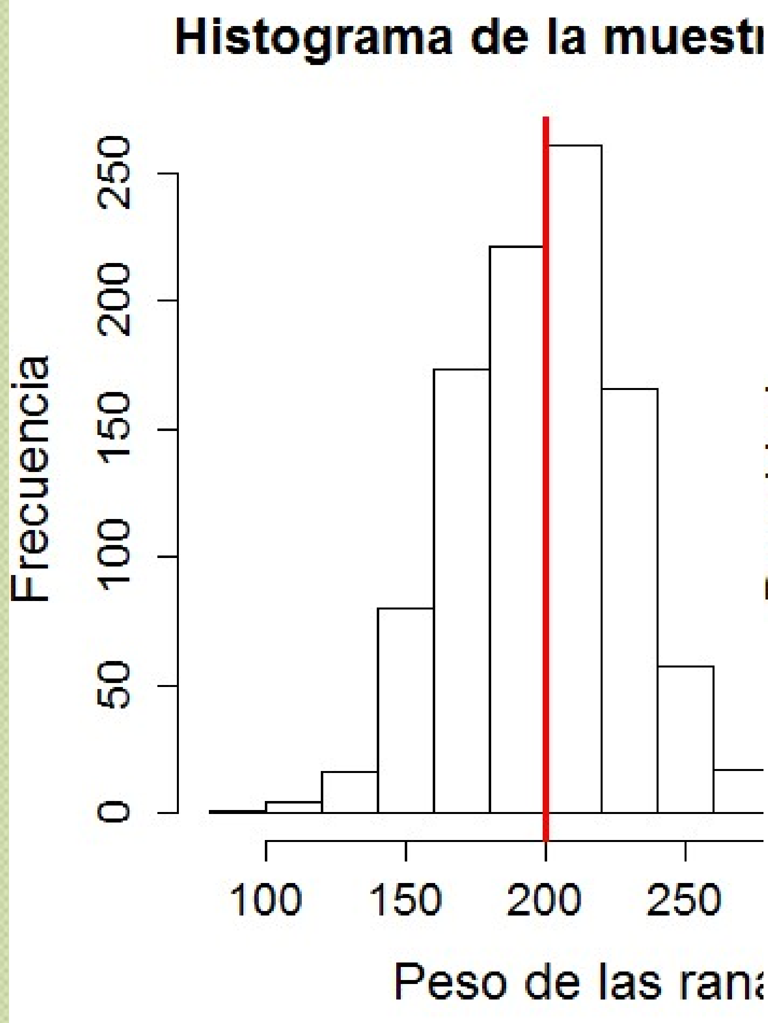
# numero de muestras
n <- 1000

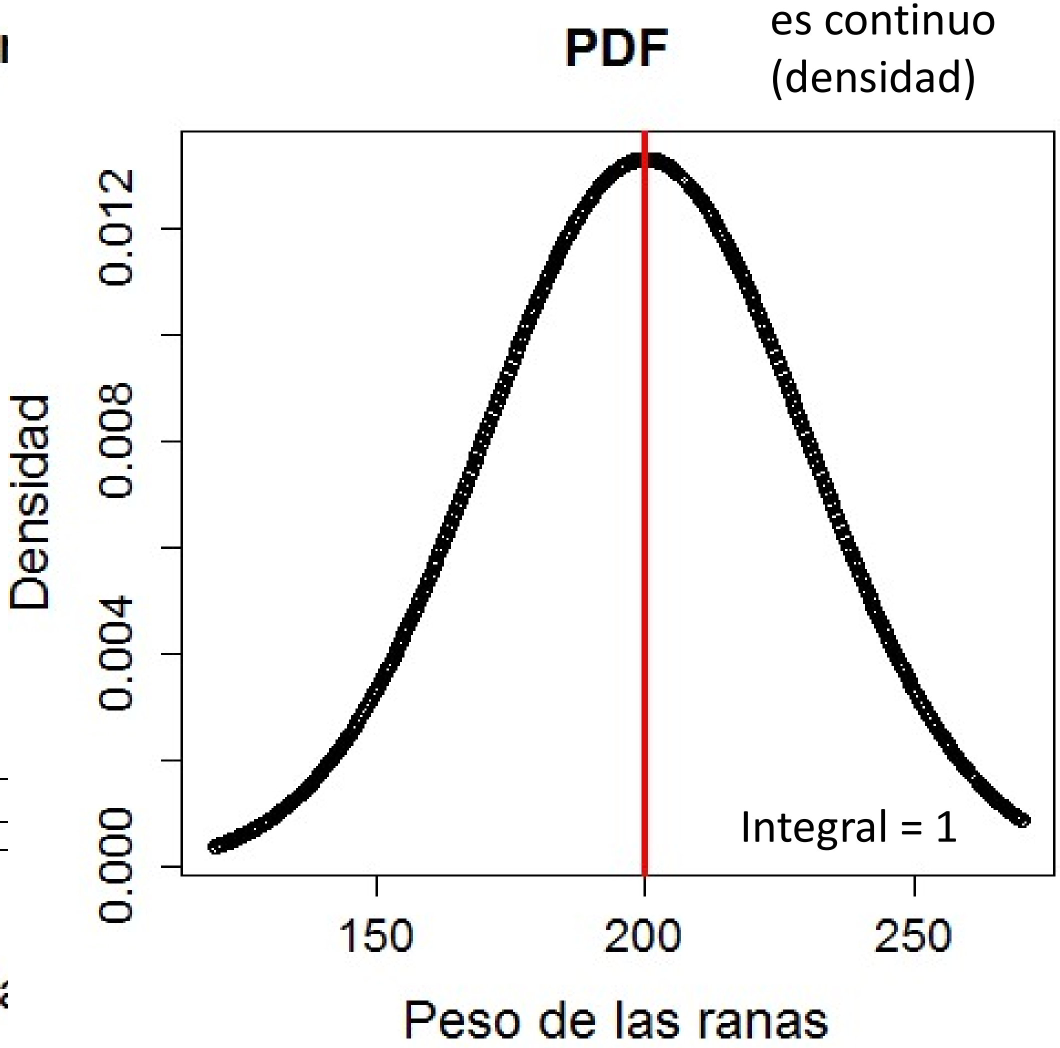
# media del peso de las ranas
mean <- 200

# SD del peso de las ranas
sd <- 30

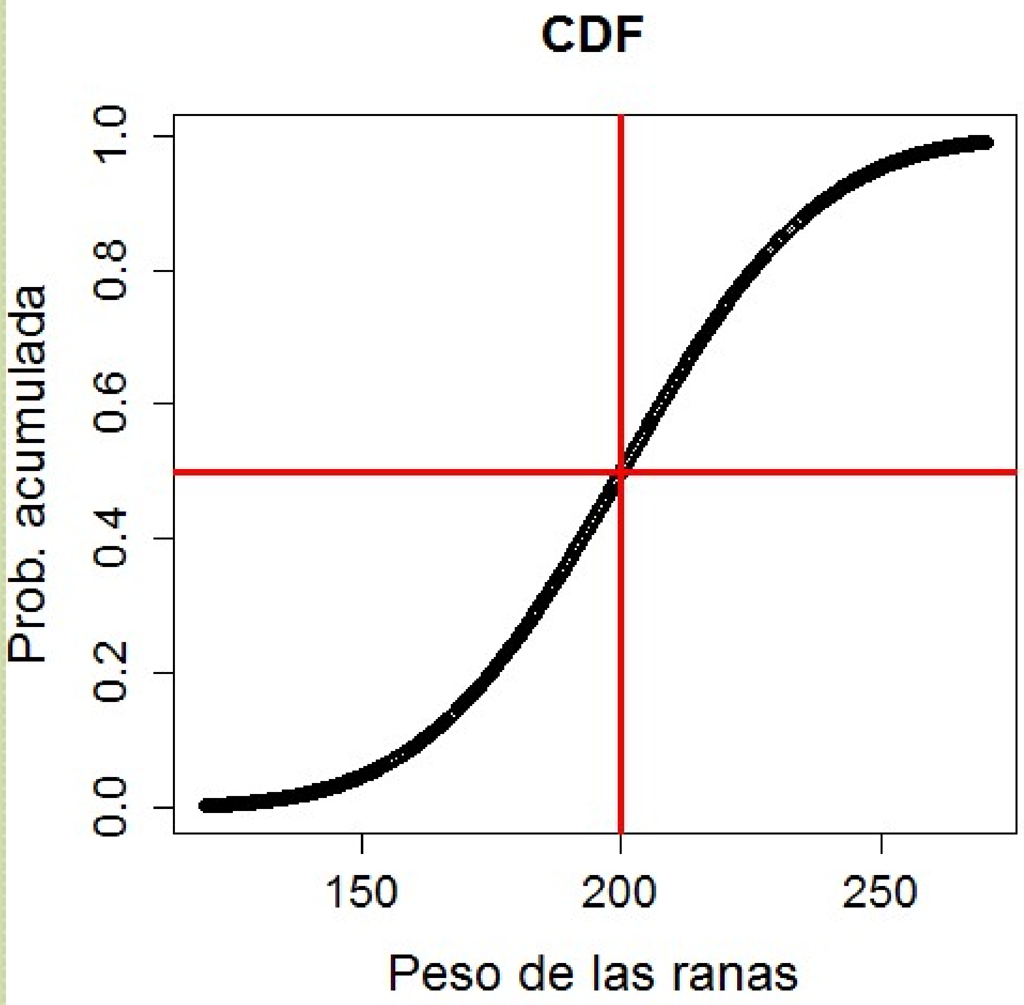
FUNCION DE DENSIDAD DE PROBABILIDAD (PDF)

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





FUNCION PROBABILIDAD DE DISTRIBUCION (CDF)



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

DISTRIBUCIÓNES DE PROBABILIDAD

- **Discretas**

Bernoulli: Dos valores posibles, 1 evento

Binomial: Dos valores posibles,>1 evento

Multinomial >1 valor posibles,>1 evento

Poisson: Valores discretos, no negativos

- **Continuas**

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

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

Beta: $0 < x < 1$

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

ESTIMACION DE PARÁMETROS

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

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

Densidad

Detectabilidad

Riqueza

Captura

Supervivencia

INTA

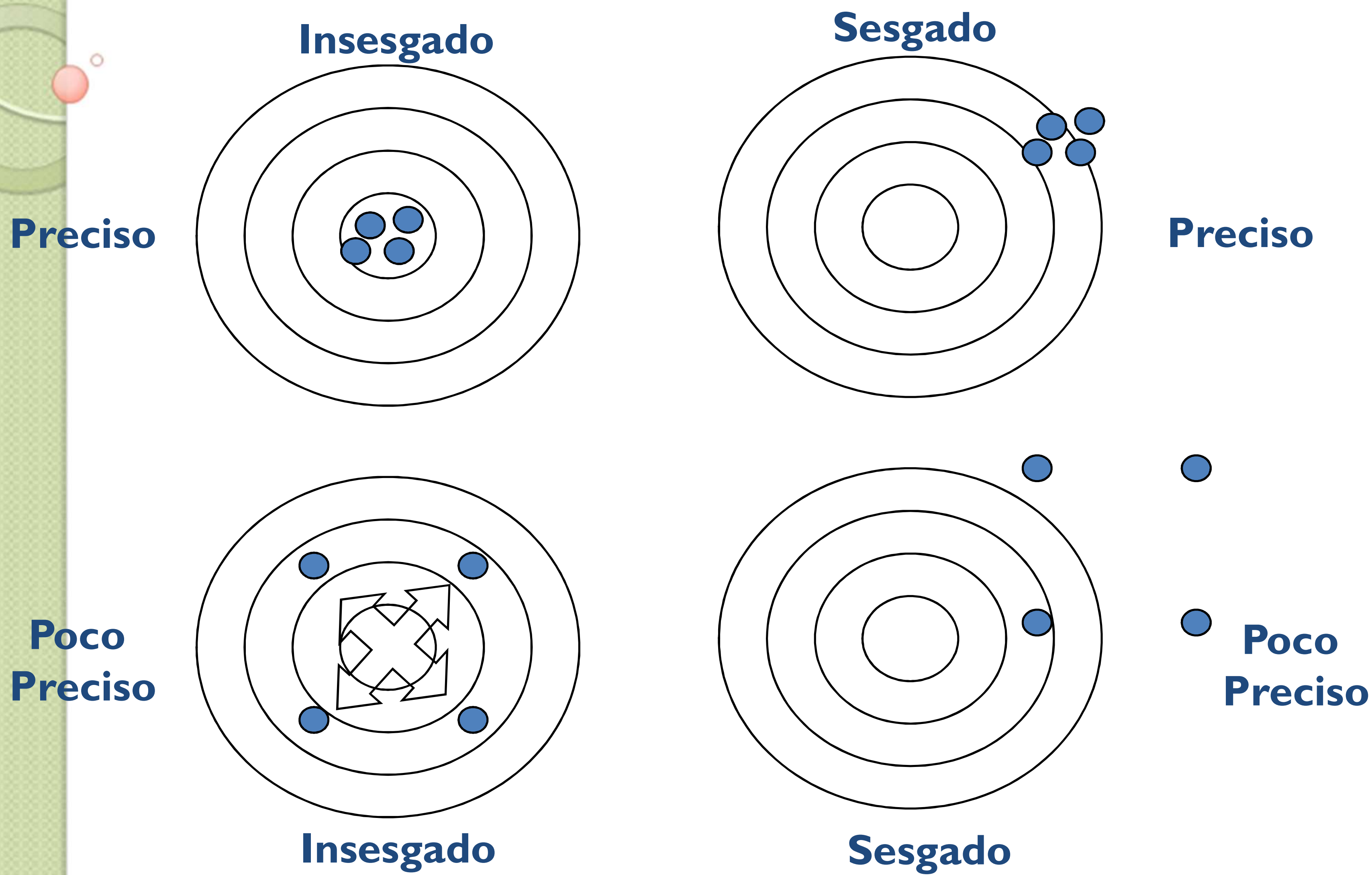
Ministerio de Agroindustria
Presidencia de la Nación

ESTIMACION DE PARÁMETROS

Medidas de comportamiento estadistico 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



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

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?

$$L(\theta|x)$$

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

Ministerio de Agroindustria
Presidencia de la Nación

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

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

Ministerio de Agroindustria
Presidencia de la Nación

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

Ejemplo Binomial

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

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

n = 10 trampas de ratones

x = 0 capturado, x = 0 no capturado

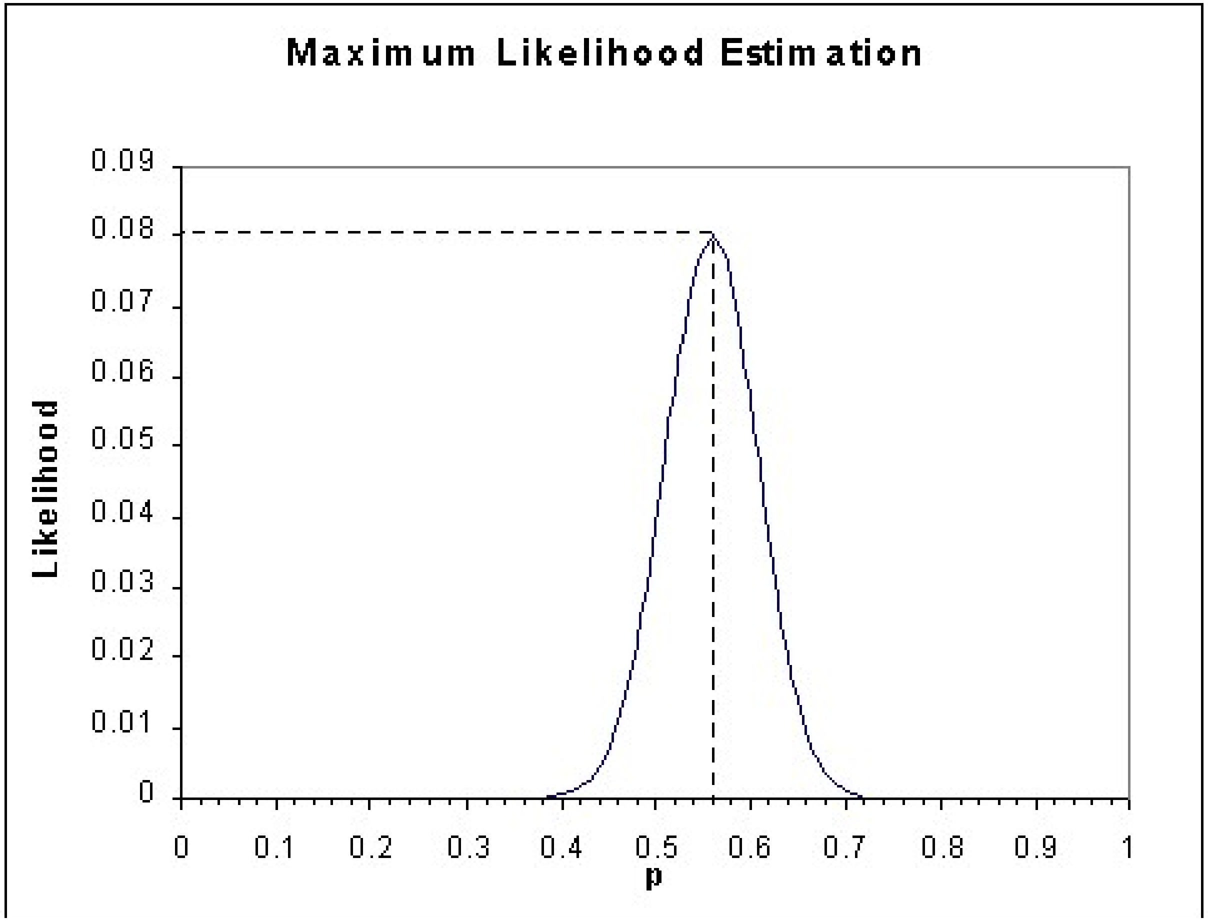
x= {0,1,1,1,0,1,1,0,0,1}

¿Cuál es la probabilidad p de captura?

MÉTODO DE MÁXIMA VEROSIMILITUD

¿Cuál es la probabilidad p de captura?

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



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)

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

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

Ministerio de Agroindustria
Presidencia de la Nación

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.
- Williams, B., J. Nichols, and M. Conroy. 2002. Analysis and Management of Animal Populations: Modeling, Estimation, and Decision Making. Academic Press, San Diego, CA.

INTA Ministerio de Agroindustria Presidencia de la Nación