

Curso de Capacitación

“ Doce (12) puntos de verificación para la vigilancia de enfermedades en organismos acuáticos: una nueva aproximación para asistir equipos multidisciplinarios en países en desarrollo ”



Organización de las Naciones Unidas para la Alimentación y la Agricultura

ICA
Instituto Colombiano Agropecuario



SESIÓN 6. Lista de chequeo 9 Análisis de Datos

Fernando Mardones
femardones@uc.cl

Lista de chequeo 9.

Análisis de datos

- Análisis exploratorio de datos (AED) es el primer paso en el proceso del análisis.
- Objetivos principales
 - Resumir el dato
 - Identificar y evaluar anomalías
 - Preparar la información para análisis más elaborados
- Regla general: se destina más tiempo en el análisis exploratorio que los análisis finales
- AED es un proceso lento, iterativo e involucra análisis que acota el problema
- Después del AED, la satisfacción es la comprensión del dato y hacer el procesp final más suave y fácil.

Análisis de datos

Objetivos

- Cuantificar prevalencia (aparente) y distribución de enfermedad
- Identificar observaciones extremas y como manejarlos
- Identificar que variables son respuestas (outcomes o variable dependiente) y cuales son factores explicatorios (variables independientes).
- Robustez en las asociaciones y conclusiones de los análisis
- Estimación de parámetros e incertidumbres
- Opciones para análisis estadísticos finales



Proceso

Análisis exploratorio de datos

- Procesar el dato en un formato adecuado para análisis
 - Planilla Excel y uso de tablas dinámicas
- Evaluar la calidad del dato incluyendo la evaluación de observaciones omitidas, valores poco probables, errores y valores extremos.
- Modificación del dato si es necesario
 - Transformaciones
 - Recodificación
 - Corrección de errores
 - Agrupamiento, etc.
- Llevar a cabo el AED (uno o más)
 - Estadísticas descriptivas
 - Gráficos de una variable
 - Gráficos de dos variables
 - Análisis de asociación entre dos variables



Estadísticas descriptivas

- **Datos cualitativos (categóricos)**
- Tablas de frecuencia con números y proporciones (o porcentaje)

Table 10.1. Frequency of AGID serology results and body condition score for ovine paratuberculosis in 224 confirmed infected sheep.

Condition score	AGID result		Total
	Positive (%)	Negative	
1	16 (94%)	1	17
2	16 (55%)	13	29
3	17 (14%)	103	120
4	6 (10%)	52	58
Total	55 (25%)	169	224

Estadísticas descriptivas

Datos cuantitativos

- Calcular estadísticas de resumen para cada variables numéricas.
- Estas estadísticas incluyen la media (promedio), mediana, desviación estándar, mínimo, máximo, correlación entre variables y otras medidas.
- Es muy útil mostrar estas estadísticas en tablas de resumen.



Ejemplo Peso Promedio Alevinaje

Observación	Huila	Córdoba	Meta	Cauca
1	18	12	18	8
2	18	11	19	8
3	13	8	12	6
4	18	9	11	2
5	15	10	18	5
6	11	7	13	7
7	19	12	11	2
8	19	9	14	3
9	15	10	16	7
10	11	11	13	8
11	11	9	12	2
Promedio	15	10	14	5
Mediana	15	10	13	6
Mínimo	11	7	11	2
Máximo	19	12	19	8
Desviación Estándar	3.3	1.6	3.0	2.6

Técnicas cuantitativas

Parámetro

Descripción

Estimadores de ubicación

Promedio aritmético (μ)	Suma de los valores divididos por el número de valores (n)
Promedio truncado	El promedio aritmético calculado después de descartar una fracción inferior o superior (típicamente 5% o 0.05)
Promedio Winsorizado	El promedio aritmético es calculado después que los valores truncados sean reemplazados por los valores menor y mayor que quedan para el cómputo
Mediana	El valor de al medio
Mínimo, máximo	Los valores más bajos y altos

Estimadores de dispersión

Varianza (σ^2)	Desviación promedio de las observaciones con respecto al promedio
Desviación estándar (σ)	Raíz cuadrada de la varianza
Desviación mediana absoluta	La diferencia mediana de las observaciones con respecto a la mediana
Rango intercuartil	Diferencia entre el tercer (75%) y el primer cuartil (25%) de una distribución o de las observaciones ordenadas

Precisión y confianza

Error estándar del promedio de la muestra $y(s \downarrow y)$	Precisión del estimador y
--	-----------------------------

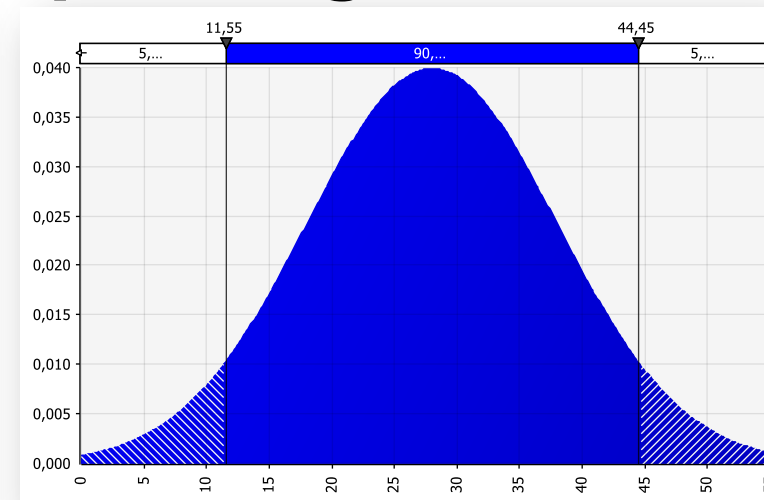
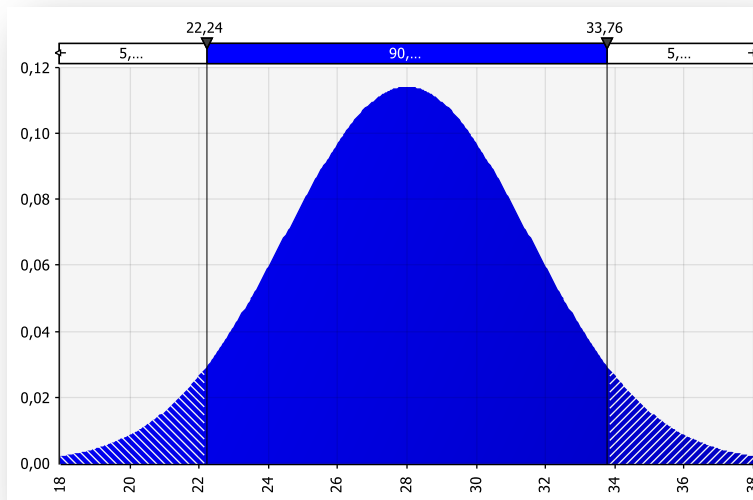
Intervalo de confianza 95% de μ	Intervalo con un 95% de probabilidad de contener el verdadero promedio
-------------------------------------	--

Descripción variables cuantitativas

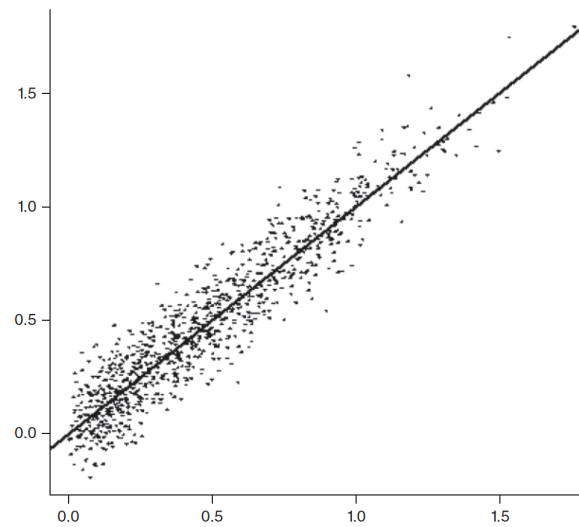
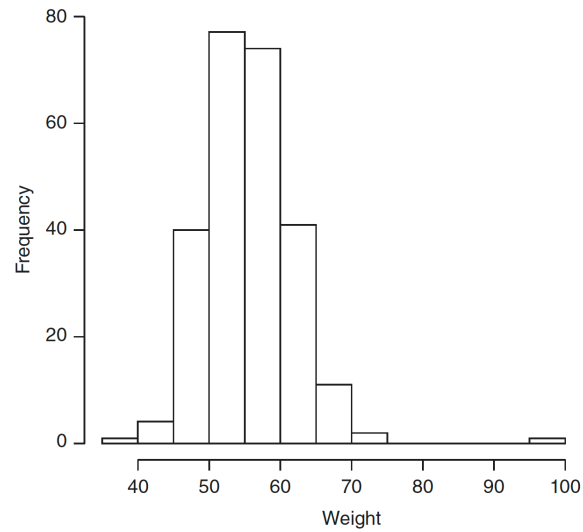
- Problema de usar una sola medida (posición) para “resumir” una distribución (conjunto de datos)

PUEDE SER ALTAMENTE ENGAÑOSO!!

Peso alevines: $\mu = 28$ gramos



Técnicas gráficas



Descripción variables cuantitativas

- Diagrama de caja y bigotes
 - Contiene toda la distribución
 - Me, IQR (Media), valores extremos/externos...

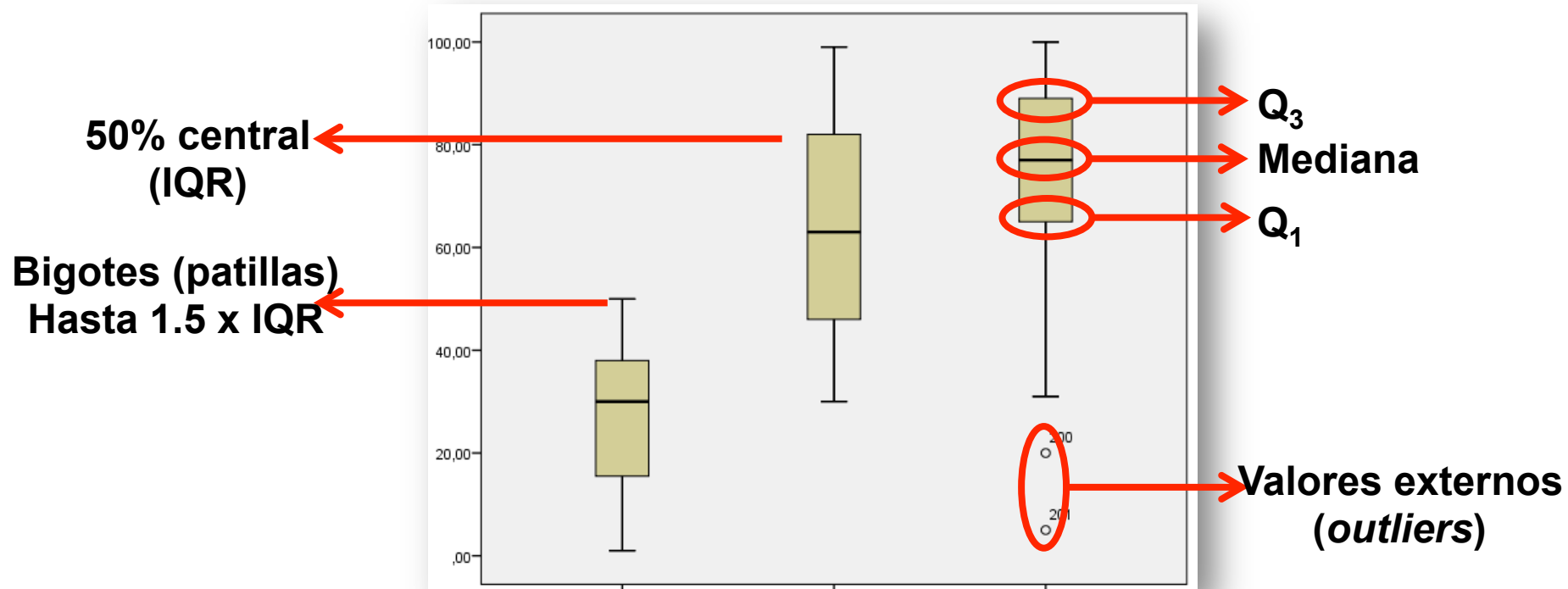
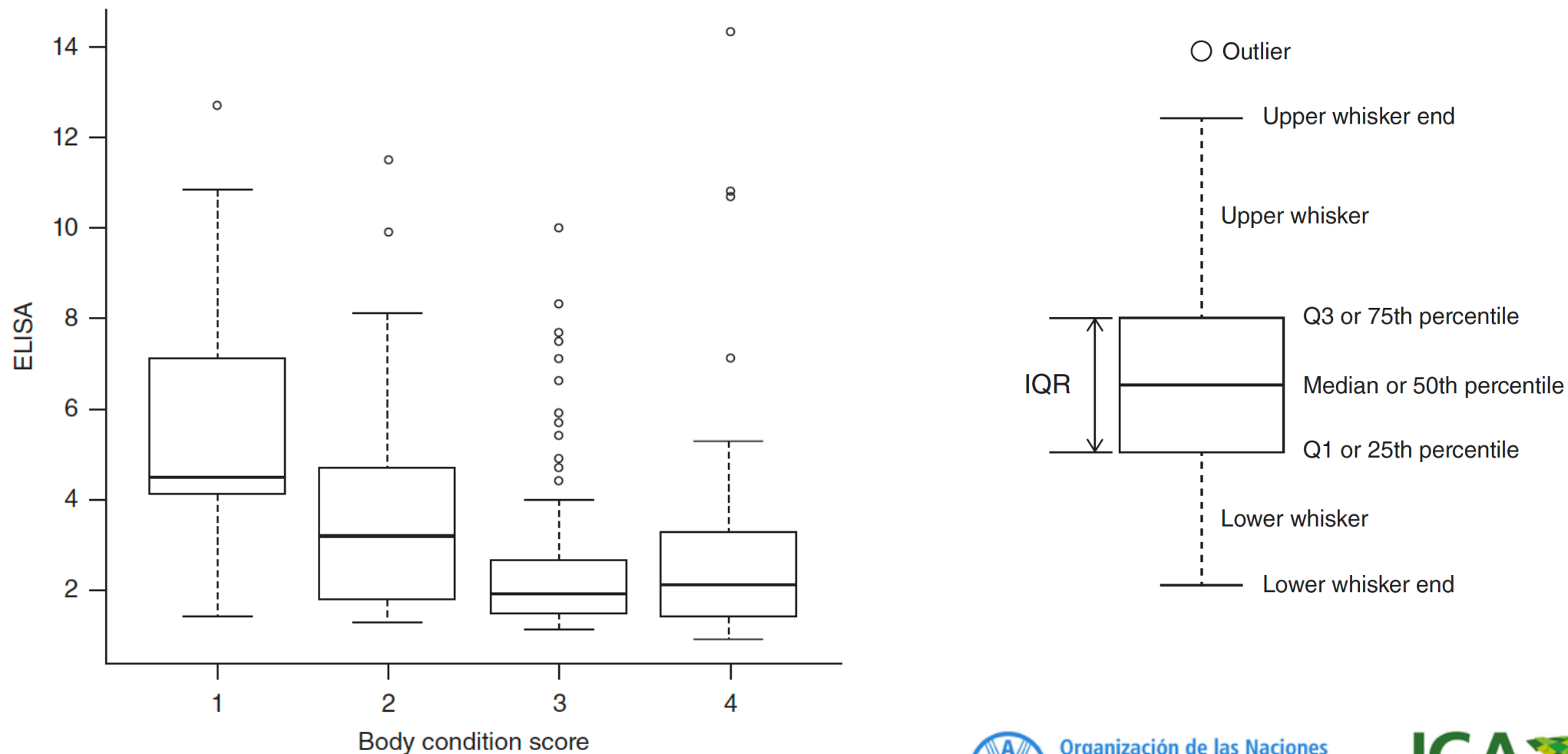


Gráfico de Cajas y Bigotes (Boxplot)



Medición de frecuencia de enfermedades

- Medir la “cantidad” de enfermedad es un paso esencial en la vigilancia y comprender la enfermedad.
- Por ejemplo, conociendo la prevalencia de TiLV podemos determinar que tan extenso es el problema, comparar la prevalencia entre regiones o grupos o monitorear un programa de control de TiLV.

Medición de enfermedad

- Medidas de frecuencia de un evento: tipos de datos
 - Prevalencia: medida transversal
 - Número de casos / total de individuos
 - Incidencia: datos longitudinales (tiempo)
 - Riesgo (incidencia acumulada):
 - Número de nuevos casos/total individuos comienzo
 - Tasa de incidencia
 - Número de casos/individuos-tiempo expuestos



Prevalencia

- Proporción donde el numerador corresponde a la frecuencia de la enfermedad.
- Denominador es el total de la población.
- Generalmente se multiplica por 100 para expresar en porcentajes.
- En nuestro muestreo de vigilancia, es una prevalencia aparente.
- Si tenemos información de censo, entonces aproximamos a la prevalencia verdadera (si tenemos una definición de caso consistente)



Estudios transversales

Estudios de prevalencia donde se obtiene una “foto” en un tiempo particular. Se determina en forma simultanea quien está expuesto (o factor de riesgo) y quien está enfermo (TiLV).

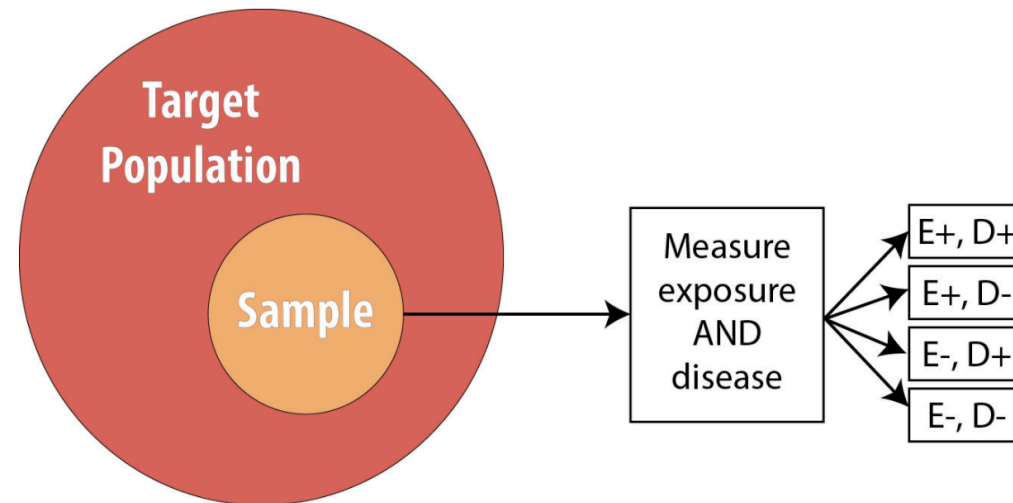


Tabla de 2 x 2

Tipo	Estatus de TiLV		Total
	Enfermo	No Enfermo	
Alevín	a	b	a+b
Engorda	c	d	c+d
Total	a+c	b+d	n =a+b+c+d

Determinamos medidas de asociación

- Por ejemplo, Odds Ratio (o razón de los cruzados).
- Medida que mide la fuerza de asociación
- Muy utilizado en estudios epidemiológicos de varios tipos.

Estimación del Odds Ratio (OR)

Odds Ratio

$$\text{Odds Ratio} = \frac{a/b}{c/d} = \frac{ad}{bc} \text{ or } \frac{a/c}{b/d} = \frac{ad}{bc}$$

	Diseased	Not diseased	Total
Exposed	a	b	$a+b$
Not exposed	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$



Organización de las Naciones
Unidas para la Alimentación
y la Agricultura



Ejemplo

Densidad	Estatus de TiLV		Total
	Enfermo	No Enfermo	
Alta	43	11	54
Óptima	3	18	21
Total	46	27	75

$$OR = ad/bc = (43 \times 18) / (11 \times 3) = 774 / 33 = 23.5$$

Also try <http://www.winepi.net/>

Interpretación del OR

- OR varía desde 0 a infinito
- Valores menores a 1 (y significativos a un valor de $p < 0.05$) sugiere que el factor es protector de TiLV.
- El valor nulo de $OR = 1$, indica que no hay asociación.
- Valores mayores a 1 (y significativos a un valor de $p < 0.05$) sugieres que existe una asociación con un potencial factor de riesgo.
- El OR se puede evaluar estadísticamente con la prueba del Chi-cuadrado y/o con el intervalo de confianza.

Interpretación del OR

En nuestro ejemplo, el riesgo de desarrollar TiLV fue de **23.5 veces mayor** en aquellas granjas que estaban a mayores densidades que aquellas granjas que no estuvieron expuestas a una mayor densidad.

Curva epidémica

- Si es posible, se debe capturar la curva epidémica ante la incursión de una enfermedad exótica.
- Esta curva resume el patrón temporal de la enfermedad, entregando una visualización de la escala o magnitud del evento y la tasa en la cual ocurren nuevos casos.
- Representa el establecimiento de nuevos casos (o brotes) ya sea como histograma o gráfico de barras, líneas, etc.

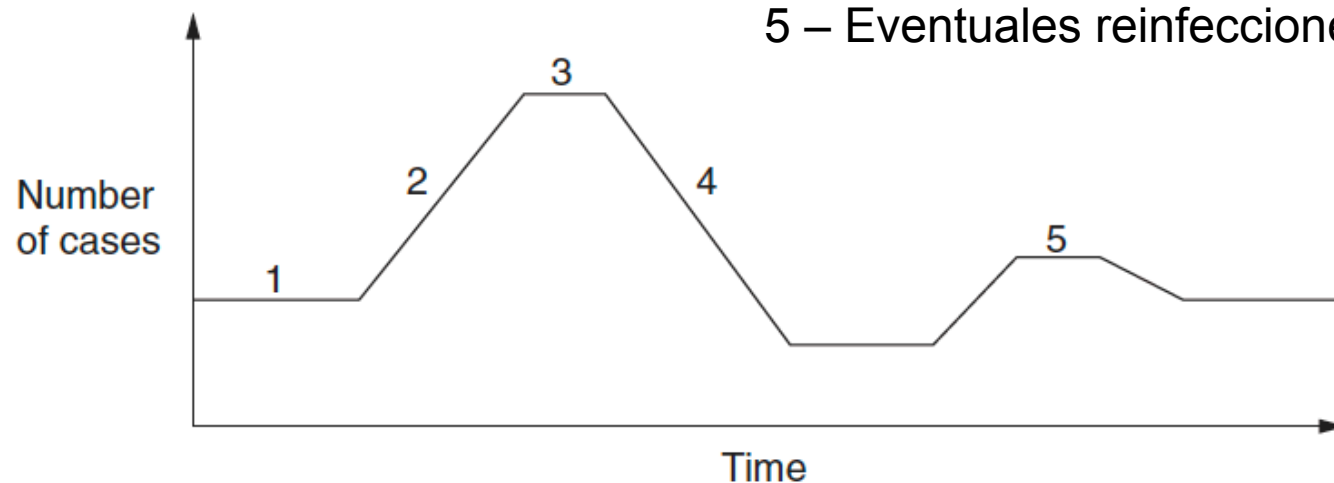
Curva epidémica (5 etapas)

1- Aumento de casos por sobre esperado.

2- Pendiente ascendente revela tipo de exposición o modo de transmisión.

3 – Longitud del plateau y pendiente de descenso (4) se relaciona con la disponibilidad de animales susceptibles. Esto depende de muchos factores, densidades de cultivo, introducción en poblaciones, cambios en mecanismos de transmisión, proporción de animales inmunes.

5 – Eventuales reinfecciones o diseminación a otras zonas.

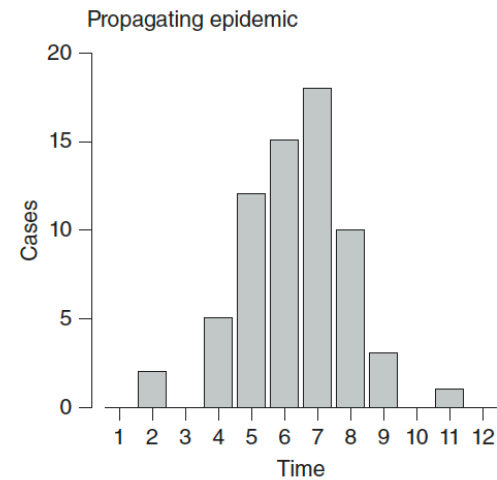
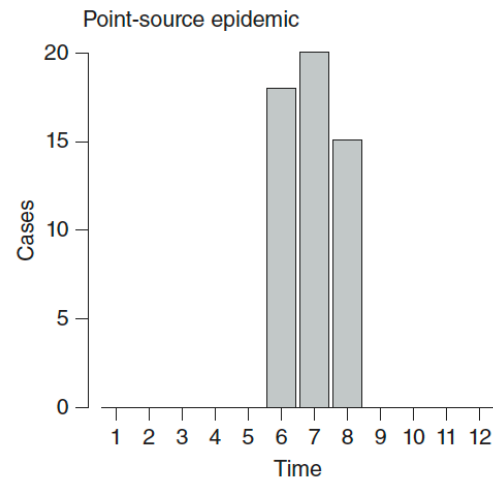
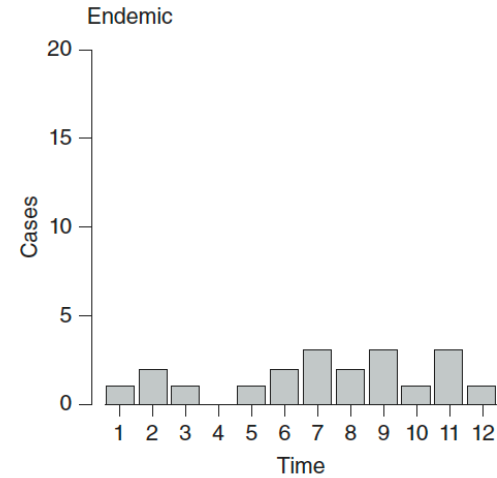
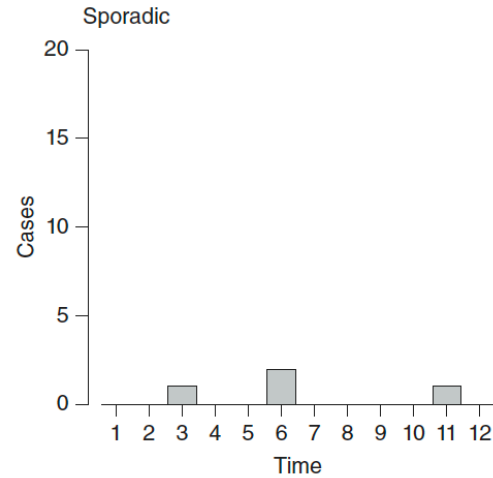


Organización de las Naciones
Unidas para la Alimentación
y la Agricultura



Instituto Colombiano Agropecuario

Tipos de curvas epidémicas



Curso de Capacitación

“ Doce (12) puntos de verificación para la vigilancia de enfermedades en organismos acuáticos: una nueva aproximación para asistir equipos multidisciplinarios en países en desarrollo ”



Organización de las Naciones Unidas para la Alimentación y la Agricultura

ICA
Instituto Colombiano Agropecuario



SESIÓN 6. Lista de chequeo 9 Análisis de Datos

Fernando Mardones
femardones@uc.cl