

Curso de epidemiología veterinaria

Andres Perez

Center for Animal Disease Modeling and Surveillance,
School of Veterinary Medicine,
University of California, Davis



Universidad Complutense de Madrid
Madrid, Mayo de 2007

Presentación del curso

Objetivo:

Repensar en un sentido epidemiológico el ámbito de desempeño laboral.

Adquirir habilidad en la identificación de problemas epidemiológicos y herramientas para su tratamiento.

Presentación del curso

Organización

Curso de posgrado con evaluación final

4 módulos presenciales (Mie, Jue, Vie, Lun)

1 seminario (Mar)

Agradecimientos

Parte del material presentado en este curso fue diseñado originalmente o fue preparado con la colaboración de

1. Mark Stevenson (Massey U, NZ)
2. Mark Thurmond (UC Davis, USA)
3. Tim Carpenter (UC Davis, USA)
4. Bimal Chhetri (Nepal)
5. Rebecca Garabed (UC Davis, USA)
6. Fernando Mardones (UC Davis, USA)
7. Alfredo Martinez (Laboratorios Azul)

Universidad Complutense de Madrid
Mayo 2008
Curso Epidemiología Veterinaria

Introducción. Conceptos básicos

Andres Perez

Repaso de conceptos

Evolución de la epidemiología

Primeros estudios formales (1840s):

John Snow
Ignas Semmelweis

Repaso de conceptos

Epidemiología

Estudio de enfermedad en poblaciones

Inferencias poblacionales referidas a enfermedades

Que es Epidemiología?

Por que algunos animales se enfermaron y otros no?
Por que algunos rodeos sufrieron la enfermedad y otros no?
Cual es la probabilidad de que un animal se enferme?
Cual es la probabilidad de que la enfermedad se encuentre en un rodeo?
Cual es la probabilidad de que la enfermedad se encuentre en un animal/rodeo por dia?
Cual es la velocidad de transmision de la enfermedad entre animales y entre rodeos?
Cual es la probabilidad de detectar la enfermedad en un rodeo o animal, usando determinados tests diagnosticos?
Que perdidas causo la enfermedad?
Cual es la estrategia mas efectiva y barata para controlarla?

Ejes centrales

1. Definición de la hipótesis y objetivo
2. Definición de 'enfermos' y 'población'
3. Diseño y recolección de datos

Relación entre estadística y epidemiología

1. Población
2. Muestra
3. Inferencia
4. Variable
5. Estadístico y parámetro
6. Probabilidad



LAPLACE

"En el fondo, la teoría de probabilidades es sólo sentido común expresado con números".

"Es notable que una ciencia que comenzó con las consideraciones de juegos de azar había de llegar a ser el objeto más importante del conocimiento humano. Las cuestiones más importantes de la vida constituyen en su mayor parte, en realidad, solamente problemas de probabilidad".

Epidemiología descriptiva

Medidas de frecuencia

- Tasas, índices y proporciones
- Incidencia y prevalencia
- Riesgo

Cocientes, Tasas, Proporciones

- Son tres clases de indicadores o estimadores matemáticos
- Se utilizan frecuentemente para relacionar el número de casos de una enfermedad con el tamaño de la población analizada

Cociente

- Se obtiene al dividir una cantidad por otra. Estas cantidades pueden estar relacionadas o ser totalmente independientes.
- Usualmente se expresa como: $\frac{x}{y} \times 10^n$

Ejemplo: Número de abortos por mil nacimientos vivos

$$\frac{\# \text{ abortos}}{\# \text{ nac vivos}} \times 1000$$

- Es un término general que incluye a los índices y las tasas

Proporción

- Un cociente en el que el numerador está incluido en el denominador
 - Se expresa como: $\frac{x}{y} \times 10^n$
- donde, 10^n es habitualmente 100

Ejemplo: Número de nacidos débiles sobre el total de nacidos vivos

$$\frac{\# \text{ nac debiles}}{\text{nac debiles} + \text{nac normales}} \times 100$$

- La respuesta se interpreta habitualmente como un porcentaje

Tasa

- Una medida de la rapidez en que sucede un evento
- Se expresa como: $\frac{x}{y} \times 10^n$

Ejemplo: Numero de casos de encefalitis que ocurren por cada 1000 animales en Santa Fe por año (TASA ANUAL DE MORBILIDAD)

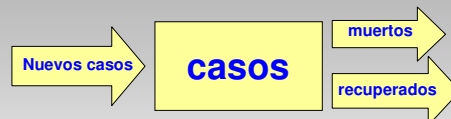
$$\frac{\# \text{ casos de encefalitis en SF en 1 a}}{\# \text{ animales en SF en 1 a}} \times 1000$$

- Para cada tasa deben especificarse el tiempo, el lugar y la población examinada

Medidas de frecuencia de la enfermedad

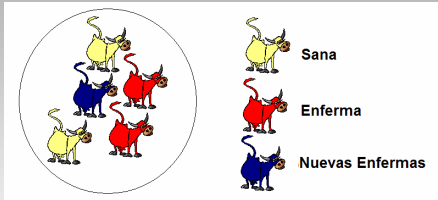
- **Incidencia (I):** Medida de los **nuevos** casos de una enfermedad que se desarrollan en un periodo de tiempo
- **Prevalencia (P):** Medida de casos **existentes** de una enfermedad en un momento determinado o a través de un periodo de tiempo

Prevalencia vs. Incidencia



- La prevalencia es una medida del grupo de casos dentro de una población
- La incidencia describe el influjo de nuevos miembros en ese grupo de casos
- Las tasas de fatalidad y recuperación describen la salida de miembros de ese grupo

Prevalencia vs. Incidencia



Hay un nuevo caso por día y permanecen 2 días enfermos. Si observamos un rodeo de 5 animales durante 1 día, cuál es la prevalencia y cuál es la incidencia diaria?

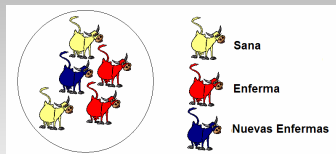
Prevalencia=60%; Incidencia diaria=20%

Medidas de frecuencia de la enfermedad

- La **Prevalencia** es una **Proporción** y la **Incidencia** es una **Tasa**
- Ambas estiman la probabilidad de ocurrencia de un suceso

Por que la prevalencia es una proporción y estima probabilidad?

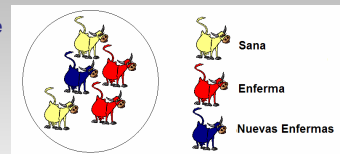
- Porque el numerador está incluido en el denominador



•Prevalencia = Proporción de enfermos =
Probabilidad de encontrar un enfermo = 60%

Por que la incidencia es una tasa y estima probabilidad?

- Porque describe que tan rápido ocurren los eventos en una población de interés



• Incidencia = Tasa de nuevos casos en un periodo de tiempo = Probabilidad de encontrar un nuevo enfermo en un día = 20%

Medidas de frecuencia de enfermedad

- Prevalencia
- Incidencia / Riesgo

Medidas de frecuencia de enfermedad

- Prevalencia
- Incidencia / Riesgo

Prevalencia

- Mide la existencia de casos en una población
- Es el indicador primario de un estudio de corte
- Se la suele dividir en prevalencia puntual y en prevalencia en un periodo de tiempo, pero el tiempo puntual no existe

Prevalencia

- Estimación Tradicional
- Estimación Bayesiana

Prevalencia

- Estimación Tradicional
- Estimación Bayesiana

Estimacion de prevalencia

$$P = \frac{C}{N}$$

C = # casos de enfermedad
N = Total poblacional

Prevalencia

Ejemplo: Estimar prevalencia de trichomoniasis bovina en la provincia de Buenos Aires (Datos del Lab. Azul, 1993-2003)

Método Tradicional:

Total = Casos + Controles

Prevalencia: Casos/Total

Intervalo de confianza:

$\sqrt{(z^2 \times \text{Prevalencia} \times (1 - \text{Prevalencia})) / \text{Total}}$

Numero de Partidos: 29

Prevalencia

Nombre	Casos	N	Prev	2.50%	97.50%
25 de Mayo	8	12	0.666667	0.399944	0.933389
Azul	476	599	0.794658	0.762308	0.827008
Bolivar	12	17	0.705882	0.489282	0.922482
Castelli	7	8	0.875	0.645823	1
Coronel Suarez	4	12	0.333333	0.066611	0.600056
Daireaux	67	77	0.87013	0.795044	0.945216
Dolores	5	8	0.625	0.28952	0.96048
General Alvear	24	33	0.727273	0.575319	0.879227
General Belgrano	8	14	0.571429	0.312199	0.830658
General Guido	11	15	0.733333	0.509541	0.957126
General Juan Madariaga	10	11	0.909091	0.739201	1
General Lamadrid	135	208	0.649038	0.584177	0.7139
General Paz	1	1	1	0	1

Prevalencia

- Estimación Tradicional
- Estimación Bayesiana

Prevalencia

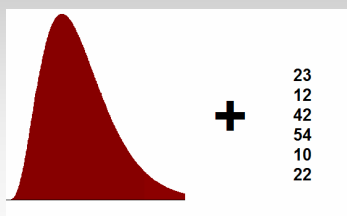
Método Bayesiano:

Distribución a posteriori de la prevalencia es consecuencia de una distribución a priori más el efecto de los datos observados

Prevalencia

Estadística bayesiana:

Distr. a priori + Observado = Distr. a posteriori



Trichomoniasis bovina

Estimar prevalencia:

Método Bayesiano:

A priori = Datos de prevalencia por partido estimados en el plan toro con definición similar de casos y controles. La distribución se asume Beta.

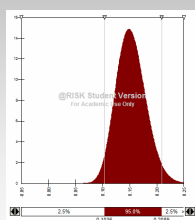
Ej: Veinticinco de Mayo

Establecimientos muestreados en Plan Toro: 175

Positivos: 26

Trichomoniasis bovina

Distribución a priori



+ Datos Lab Azul $\xrightarrow{(+ Se y Sp)}$ Distribución a posteriori

Resultado

Nombre	2.5%	Mediana	97.5%	PrevAz
25 de Mayo	0.1338	0.1843	0.2427	0.666667
Azul	0.6703	0.7043	0.7369	0.794658
Balcarce	0.06153	0.1001	0.1504	0.4
Bolivar	0.5505	0.737	0.8797	0.705882
Castelli	0.1535	0.2433	0.3536	0.875
Coronel Suarez	0.07726	0.187	0.3468	0.333333
Daireaux	0.5521	0.6333	0.7113	0.87013
Dolores	0.2219	0.296	0.3785	0.625
General Alvear	0.4756	0.6077	0.7295	0.727273
General Belgrano	0.4135	0.5097	0.6049	0.571429
General Guido	0.2193	0.2909	0.3708	0.733333
General Juan Madariaga	0.3985	0.5465	0.6885	0.909091
General Lamadrid	0.555	0.6213	0.6838	0.649038
General Paz	0.1184	0.1731	0.2396	1

Estadística bayesiana

Introducción de la subjetividad o información previa
Menor importancia de los intervalos de confianza
Posibilidad de trabajar con muestras más chicas
Posibles aplicaciones:
• Interpretación de tests diagnósticos
• Estrategias de muestreo y vigilancia
• Análisis de distribución espacial

Medidas de frecuencia de enfermedad

- Prevalencia
- Incidencia / Riesgo

Incidencia

- Mide la existencia de nuevos casos en una población en un determinado periodo de tiempo
- Es la base para la estimación del riesgo

Riesgo

Probabilidad de que un individuo con una cierta característica como:

Edad

Raza

Sexo

experimente un cambio en su estado sanitario durante un periodo de tiempo

Se asume que el individuo:

- No está enfermo al comienzo del periodo
- No muere de otras causas durante el periodo de tiempo.

Riesgo

$$0 \leq \text{Riesgo} \leq 1$$

$$0\% \leq \text{Porcentaje} \leq 100\%$$

Es necesario especificar el periodo de tiempo

Ejemplo:

El riesgo anual de que una vaca lechera desarrolle mastitis es del 5% ($R=0.05$)

Incidencia acumulada

$$IC = \frac{I}{N}$$

I = # de nuevos casos en un periodo de tiempo

N = # total de individuos al comienzo del periodo de tiempo estudiado

Mide la frecuencia de adición de nuevos casos y siempre se estima para un periodo de tiempo determinado

Ejemplo

Los registros de Dinamarca indican que de 10.000 vacas lecheras (N) observadas durante 3 años, 600 desarrollan mastitis (I)

$$IC = \frac{I}{N}$$

IC = 600 / 10.000 = 6% en 3 años

Riesgo de que una vaca desarrolle mastitis en Dinamarca el transcurso de un año: 2%

Incidencia acumulada

- Es la forma mas común de estimar riesgo
- Es siempre una proporción
- Asume una cohorte constante
- Cuando se la estima en periodos de tiempo breves, como una epidemia, se la suele llamar "tasa de ataque"
- La fórmula no refleja cambios poblacionales en cohortes dinámicas.
- Suele ser necesario ajustarla para deserción o seguimiento discontinuo de individuos

Riesgo ajustado

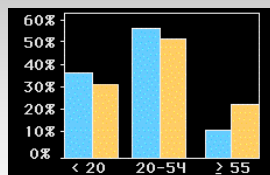
- En 1996 murieron 427 personas por cada 100.000 habitantes de Alaska
- En el mismo año murieron 824 personas por cada 100.000 habitantes de Arizona
- El riesgo de morir en Alaska es del 0.4 %, mientras que el riesgo de morir en Arizona es del 0.8% (CASI EL DOBLE!!!!)
- Es más peligroso vivir en Arizona?

Riesgo ajustado

- Estas dos tasas se denominan crudas o brutas porque representan la proporción global de muertes en cada población
- Las tasas crudas NO controlan por las diferencias entre las poblaciones en factores como edad, raza o sexo, que pueden influir en la mortalidad.
- Sin considerar estos factores podemos arribar a conclusiones erróneas

Riesgo ajustado

- El gráfico muestra la distribución poblacional de Alaska (celeste) y Arizona (naranja)



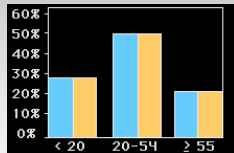
- ¿Cuál de las dos poblaciones es más vieja? Y eso qué implica?

Riesgo ajustado

- El clima seco y árido de Arizona atrae a pobladores de mayor edad que Alaska, en cuyo clima frío y húmedo suelen establecerse parejas jóvenes
- Las personas de mayor edad se encuentran a un mayor riesgo de morir
- La diferente distribución etaria distorsiona las comparaciones. Por lo tanto es necesario ajustar por edad para poder comparar ambas poblaciones.

Riesgo ajustado

- El ajuste (o estandarización) se realiza igualando las distribuciones de ambas poblaciones



- Considerando los porcentajes observados en los grupos etarios, cual sería el número total de casos en Alaska si la estructura poblacional fuera la misma que en Arizona?

Riesgo ajustado

- La tasa de mortalidad estandarizada en ambos estados es idéntica (0.8%).
- Como se ajustan (o estandarizan) las tasas?

Riesgo ajustado

Riesgo mensual de enf x en dos poblaciones:

	Población A	Población B
Población (Enf/Total = Riesgo)	40/200 = 20%	30/200 = 15%
Machos (Enf/Total m = Riesgo)	40/100 = 40%	30/60 = 50%
Hembras (Enf/Total h = Riesgo)	0/100 = 0%	0/140 = 0%

- El riesgo parece ser mayor en la población A
- Sin embargo el riesgo es mayor en machos que en hembras
- Hay una proporción mayor de machos en A que en B
- Es posible que el riesgo aparentemente mayor en A se deba a que hay más machos?

Riesgo ajustado

Riesgo mensual de enf x en dos poblaciones:

	Población A	Población B
Poblacion (Enf/Total = Riesgo)	40/200 = 20%	30/200 = 15%
Machos (Enf/Total m = Riesgo M)	40%*(100+60)	50%*(100+60)
Hembras (Enf/Total h = Riesgo H)	0%*(100+140)	0%*(100+140)
Riesgo Estandarizado	(64+0)/400 = 16%	(80+0)/400 = 20%

Efectivamente el riesgo estandarizado es mayor en la Población B que en la A

Riesgo ajustado

- Se nos ocurren ejemplos prácticos en veterinaria?

Algunas definiciones

- Factor de riesgo
- Confounding (confusión)
- Bias (sesgo)

Algunas definiciones

Factor de riesgo

- Es una variable que se asocia con una mayor probabilidad de ocurrencia de una enfermedad
- Por ejemplo, ser macho en el ejemplo anterior

Algunas definiciones

Confounding (confusión)

- Es una variable que se asocia con un factor de riesgo y la enfermedad, por lo que confunde las conclusiones
- Por ejemplo, vivir en la población B en el ejemplo anterior

Algunas definiciones

Bias (sesgo)

- Es un error en las estimaciones debido a no controlar algunos factores
- Por ejemplo, si elegimos las poblaciones A y B basados en nuestra conveniencia, y no en un muestreo aleatorio, puede ser que incurramos en un "sesgo de selección"

Algunas definiciones

El estudio de los factores de riesgo relacionados con enfermedad, controlando los sesgos y las confusiones, es uno de los ejes de la epidemiología y será el tema común a discutir en los módulos del presente curso

Universidad Complutense de Madrid
Mayo 2008
Curso Epidemiología Veterinaria

Asociación entre enfermedad y factores epidemiológicos

Andres Perez

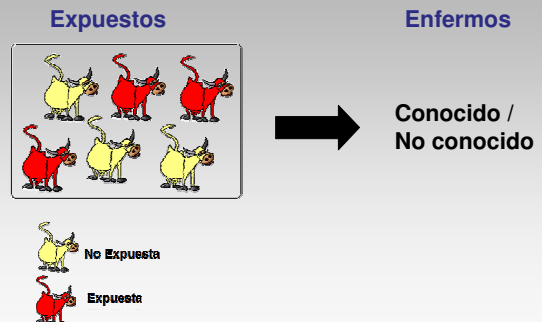
Diseño de estudios epidemiológicos

- Diseños experimentales vs. Observacionales
- Direccionalidad de los estudios
- Estudios clínicos
- Repaso de los estudios observacionales

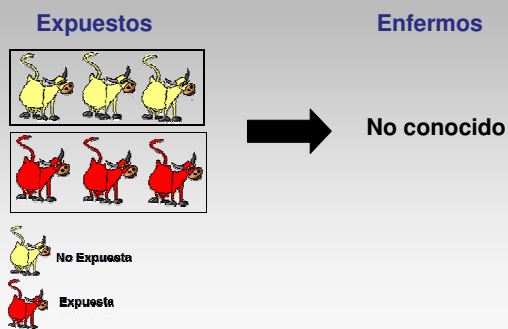
Tipos de estudios epidemiológicos

- Experimentales
- Observacionales

Estudios observacionales



Estudios experimentales



Tipos de estudios experimentales

- Estudios clínicos: Estudiar el efecto (eficacia) de un tratamiento
- Intervenciones: Estudiar el efecto de un programa de (control / prevención).

Tipos de estudios observacionales

- Descriptivos: Sugerir hipótesis
- Analíticos: Evaluar hipótesis

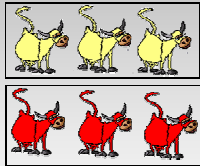
Direccionalidad de los estudios

Exposición → Enfermedad

Sin embargo, la direccionalidad del estudio (el lugar temporal en el que se ubica el observador) puede ser distinta

Direccionalidad de los estudios

Exposición → Enfermedad



?

ESTUDIOS LONGITUDINALES PROSPECTIVOS

- Estudios de cohorte
- Estudios clínicos

Direccionalidad de los estudios

Exposición → Enfermedad

?



ESTUDIOS LONGITUDINALES RETROSPECTIVOS

- Estudios de casos-controles
- Estudios de cohorte retrospectivos

Direccionalidad de los estudios

Exposición • • Enfermedad

? = ?

ESTUDIOS TRANSVERSALES O DE CORTE

Direccionalidad de los estudios

Si el estado sanitario es conocido cuando comienza el estudio y se analizan las unidades expuestas/no expuestas en el pasado, el estudio es **RETROSPECTIVO**

Direccionalidad de los estudios

Si las unidades expuestas / no expuestas son conocidas cuando comienza el estudio, pero aun no se sabe cuales se enfermarán y cuáles no, el estudio es **PROSPECTIVO**

Direccionalidad de los estudios

Si no existe direccionalidad en el análisis, el estudio es **DE CORTE** o **TRANSVERSAL**

Direccionalidad de los estudios

Los estudios **retrospectivos** y **transversales** son más **baratos**, pero muy susceptibles de ser influenciados por **sesgos de selección** y no es posible estimar **causalidad**

Por qué?

Direccionalidad de los estudios

Además, la información en los estudios **transversales** y **retrospectivos** suelen obtenerse de datos de hospitales, registros productivos, encuestas, etc., por lo que la información suele ser de menor calidad que en los estudios **prospectivos**.

Son más susceptibles de verse afectados por **sesgos de información**

Tipos de estudios epidemiológicos

Experimentales

Observacionales

Tipos de estudios epidemiológicos

Experimentales

Observacionales

Estudios clínicos

Son estudios longitudinales, prospectivos, experimentales.

Que significa cada una de estas características?

El objetivo de los estudios clínicos es evaluar la eficacia de una intervención terapéutica (curar, controlar) o preventiva, comparando los beneficios de uno o más tratamientos

Estudios clínicos

Los estudios clínicos se caracterizan por ser:

Aleatorizados

A ciego

Susceptibles de conflictos éticos

Intencionales en términos analíticos

Estudios clínicos

Los estudios clínicos se caracterizan por ser:

Aleatorizados:

Los individuos son idénticos en todos sus aspectos (excepto por la exposición al tratamiento) y ubicados al azar en cada grupo

Estudios clínicos

Los estudios clínicos se caracterizan por ser:

A ciego

El paciente o el investigador (ciego) o el paciente y el investigador (doble ciego) desconocen que individuo pertenece a cada grupo de tratamiento.

Estudios clínicos

Los estudios clínicos se caracterizan por ser:

Susceptibles de conflictos éticos

- Qué individuo recibe tratamiento y que individuo no?
- Efectos negativos de tratamientos experimentales
- Considerar reglas para detener el estudio
- Aprobación de comités éticos

Estudios clínicos

Los estudios clínicos se caracterizan por ser:

Intencionales en términos analíticos

El investigador “analiza lo que aleatoriza”

Ejemplo de estudios clínicos

Comparar la eficacia de dos tratamientos antiparasitarios en terneros (un tratamiento nuevo y un tratamiento estándar)

Ho) Tratamiento Nuevo $\leq$ Nuevo Standard
H1) Tratamiento Nuevo $>$ Nuevo Standard

Por que un tratamiento estándar y no un grupo no tratado?

Ejemplo de estudios clínicos

Aleatorización: Individuos de idénticas características (edad, raza, etc) se distribuyen al azar en dos grupos iguales utilizando un generador de números aleatorios. A campo:

Animal ID	G.N.A.	Grupo
1	19	2
2	27	2
...	...	...
n	32	1

Ejemplo de estudios clínicos

A ciego: En el laboratorio de parasitología se preparan $n/2$ tratamientos estándar y $n/2$ nuevos tratamientos y se los distribuye a los clínicos para su aplicación. El clínico obtiene las muestras y las envía a los parasitólogos para su evaluación. El parasitólogo examina las muestras, desconociendo cual de ellas proviene de animales con el tratamiento estándar y cuales de animales con el nuevo tratamiento.

Ejemplo de estudios clínicos

Conflictos éticos: Los animales tratados con el nuevo tratamiento son monitoreados diariamente para detectar posibles efectos nocivos para la salud. Puede elegirse un comité ético que evalúe el trabajo.

Ejemplo de estudios clínicos

Intencionalidad analítica: Se compara la carga parasitaria media de ambos grupos x días luego de aplicado el tratamiento utilizando un test no paramétrico. Se compara el peso promedio de ambos grupos utilizando un test de t de Student. No puede compararse el efecto del tratamiento en distintos sexos, ya que ese efecto se fijó al elegir los grupos.

Ejemplo de estudios clínicos

Resultado:
Si la carga parasitaria media es mayor en el grupo tratado con el nuevo tratamiento que en el grupo tratado con el tratamiento estándar ($P < 0.05$) se rechaza la hipótesis nula y se concluye que el nuevo tratamiento supera al tratamiento estándar.

Intervenciones

- Estudiar el efecto de un programa de (control / prevención) en una comunidad.
- Esencialmente el mismo diseño que un estudio clínico, pero interviniendo en una comunidad.
- Ejemplo: construir letrinas en algunas casas y evaluar la carga parasitaria en niños en hogares con y sin letrinas (Rinne, Ecuador 2005)
- Planes experimentales en vida silvestre

Tipos de estudios epidemiológicos

Experimentales

Observacionales

Estudios observacionales

El resultado de estudios observacionales simples típicamente se organiza utilizando tablas de 2 x 2

La estructura de las tablas 2 x 2 es siempre similar, independientemente del tipo de estudio observacional

Estudios observacionales

	Expuestos	No Expuestos	Total
Enfermos	A	B	A + B
Sanos	C	D	C + D
Total	A + C	B + D	A + B + C + D

Estudios observacionales

Transversales

Longitudinales de casos y controles

Longitudinales de cohorte

Estudios observacionales

Transversales

Longitudinales de casos y controles

Longitudinales de cohorte

Estudios transversales

No existe dimensión temporal

Los animales se agrupan en expuestos y no expuestos; enfermos y sanos en un periodo concreto de tiempo

Típicamente se desconoce el n de las categorías (expuestos, no expuestos, sanos, enfermos) al comienzo del estudio.

Estudios transversales

Ejemplo: Un veterinario recibe el llamado de un establecimiento con 500 bovinos. Al llegar encuentra 50 animales enfermos. 40 de ellos se encuentran en un monte, junto con otros 160 animales, mientras que los restantes 10 se encuentran con los otros 290 animales en potreros de mejor calidad.

Recordemos

Factor de riesgo

- Es una variable que se asocia con un mayor riesgo de ocurrencia de una enfermedad

Estudios transversales

Cual es el factor de riesgo?

Encontrarse en el monte

Por qué es un estudio transversal?

Porque no existe una delimitación clara del tiempo. El veterinario agrupa animales enfermos y sanos; expuestos y no expuestos, en un periodo de tiempo puntual (cuando visita el establecimiento)

Estudios transversales

	Expuestos	No Expuestos	Total
Enfermos	40	10	50
Sanos	160	290	450
Total	200	300	500

Estudios transversales

El indicador recomendado para estudios transversales se denomina **razón de las Prevalencias (RP)**

La **razón de las proporciones (OR)**, usado en algunos estudios y que veremos mas adelante, suele sobreestimar las asociaciones si la prevalencia es alta

Estudios transversales

	Expuestos	No Expuestos	Total
Enfermos	A	B	A+B
Sanos	C	D	C+D
Total	A+C	B+D	A+B+C+D

Razón de las Prevalencias = $[A/(A+C)] / [B/(B+D)]$

RP = 1: no hay asociación entre el factor de exposición y la enfermedad

RP > 1: el factor de exposición se asocia con la ocurrencia de la enfermedad (factor de riesgo)

RP < 1: el factor de exposición se asocia con no ocurrencia de la enfermedad (factor protector)

Estudios transversales

	Expuestos	No Expuestos	Total
Enfermos	40	10	50
Sanos	160	290	450
Total	200	300	500

Razón de las Prevalencias = $(40/200) / (10/300) = 6$

Cómo se interpreta este resultado?

Estudios transversales

	Expuestos	No Expuestos	Total
Enfermos	40	10	50
Sanos	160	290	450
Total	200	300	500

Razón de las Prevalencias = $(40/200) / (10/300) = 6$

El riesgo de que un bovino que pastorea en el monte se enferme es 6 veces el riesgo de que se enferme un bovino que no pastorea en el monte

Estudios transversales

RP > 1, entonces podemos concluir que el factor de riesgo se asocia con la enfermedad?

Es esta asociación estadísticamente significativa?

Debemos estimar el intervalo de confianza de RP y comprobar si el límite inferior del intervalo es mayor que 1

Estudios transversales

Entonces:

1. Si el límite inferior del intervalo de confianza de RP es mayor que 1, entonces el factor representa un riesgo significativo para la enfermedad.
2. Si el límite superior del intervalo de confianza es inferior a 1, entonces el factor representa una protección significativa para la enfermedad.
3. Si el intervalo de confianza incluye a 1, entonces no podemos concluir que el factor represente un riesgo o una protección para la enfermedad

Estudios transversales

Cómo se estima el intervalo de confianza?

Existen varias formulas, por ejemplo una aproximación logarítmica, donde z representa el valor de t para el nivel de confianza deseado (z=1.96 para un intervalo de confianza del 95%)

$$e^{\ln RP \pm Z \sqrt{\{C / [A(A+C)] + D / [B(B+D)]\}}}$$

En nuestro ejemplo: IC 95% de RP = 3.1 – 11.7

Ahora podemos concluir que la asociación es estadísticamente significativa!!

Estudios transversales

Cómo se interpreta que IC 95% de RP = 3.1 – 11.7?

Podemos afirmar con un 95% de confianza que los animales en el monte se encuentran a un riesgo entre 3 y 12 veces mayor de sufrir la enfermedad que los animales que no se encuentran en el monte.

En un trabajo científico, esto suele reportarse como “el riesgo de que los animales situados en el monte se enfermen fue 6 veces mayor (IC=3.1-11.7; p<0,05) que el de los animales que no se encontraban en el monte”

Estudios transversales

Una yapa...

La razón de las prevalencias (RP) suele utilizarse para estimar la fracción atribuible y la fracción etiológica.

Estudios transversales

La **fracción etiológica** es la proporción de enfermos expuestos en los cuales la enfermedad puede atribuirse a la exposición al factor: $FE = (RP - 1) / RP = 0.83$

En el 83% de los enfermos en el monte puede atribuirse la enfermedad al hecho de que estén en el monte.

Estudios transversales

La **fracción atribuible** es la proporción de enfermos en una población que se deben a la exposición al factor: $FE \times A / (A+B) = 0.83 \times 4/5 = 0.67$

En el 67% de los bovinos enfermos en el establecimiento puede atribuirse la enfermedad al hecho de estar en el monte.

Estudios observacionales

Transversales

Longitudinales de casos y controles

Longitudinales de cohorte

Estudios de casos controles

Existe dimensión temporal retrospectiva

Los animales se agrupan en enfermos y sanos y luego se investiga si estuvieron expuestos o no a un factor considerado protector o de riesgo

Típicamente se conoce el n de las categorías sano y enfermo al comienzo del estudio, pero se desconoce el de expuestos y no expuestos

Estudios de casos controles

Ejemplo: El veterinario, preocupado por su hallazgo, decide investigar sus registros en otros campos. Agrupa los animales como enfermos o sanos de acuerdo a la sintomatología que acaba de identificar y luego investiga con los propietarios para saber si en el ultimo mes pastorearon en montes o en pasturas de mejor calidad. Luego resume sus hallazgos en una tabla de 2 x 2.

Estudios de casos controles

Por qué es un estudio longitudinal de casos controles?

Porque el veterinario define sus casos (enfermos) y controles (sanos) en un periodo de tiempo y luego investiga sus registros retrospectivamente para saber si los animales estuvieron expuestos o no a un factor

Estudios de casos controles

	Expuestos	No Expuestos	Total
Enfermos	89	23	112
Sanos	254	798	1052
Total	343	821	1164

Estudios de casos controles

El indicador recomendado para estudios de casos-controles se denomina **razón de las proporciones**, más conocido por sus siglas en ingles como **OR (odds ratio)**

Estudios de casos controles

	Expuestos	No Expuestos	Total
Enfermos	A	B	A+B
Sanos	C	D	C+D
Total	A+C	B+D	A+B+C+D

Odds ratio = (A/C) / (B/D)

OR = 1: no hay asociación entre el factor de exposición y la enfermedad

OR > 1: el factor de exposición se asocia con la ocurrencia de la enfermedad (factor de riesgo)

OR < 1: el factor de exposición se asocia con no ocurrencia de la enfermedad (factor protector)

Estudios casos controles

	Expuestos	No Expuestos	Total
Enfermos	89	23	112
Sanos	254	798	1052
Total	343	821	1164

Odds ratio = (89/254) / (23/798) = 12.2

El riesgo de que un bovino que pastorea en el monte se enferme es 12 veces el riesgo de que se enferme un bovino que no pastorea en el monte

Estudios de casos controles

En forma similar al RP de los estudios transversales:

1. Si el limite inferior del intervalo de confianza de OR es mayor que 1, entonces el factor representa un riesgo significativo para la enfermedad.
2. Si el limite superior del intervalo de confianza es inferior a 1, entonces el factor representa una protección significativa para la enfermedad.
3. Si el intervalo de confianza incluye a 1, entonces no podemos concluir que el factor represente un riesgo o una protección para la enfermedad

Estudios de casos-controles

Cómo se estima el intervalo de confianza?

También existen varias formulas, para mantener la consistencia, mostramos la aproximación logarítmica, donde z representa el valor de t para el nivel de confianza deseado (z=1.96 para un intervalo de confianza del 95%)

$$e^{\ln OR \pm Z \cdot \sqrt{(1/A + 1/B + 1/C + 1/D)}}$$

En nuestro ejemplo: IC 95% de OR = 7.5 – 19.6

Ahora podemos concluir que la asociación es estadísticamente significativa!!

Estudios de casos-controles

Cómo se interpreta que IC 95% de OR = 7.5 – 19.6?

Es similar al intervalo de confianza del RP de los estudios transversales.

Cómo se interpreta?

Estudios de casos controles

Para tener en cuenta:

En los estudios de casos controles, la fracción etiológica y la fracción atribuible se estima y se interpreta en forma análoga a los estudios transversales

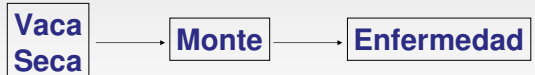
Estudios de casos controles

Ejemplo: El veterinario decide comentar el hallazgo con un colega. El colega, escéptico, le hace un comentario interesante. Las vacas secas suelen pastorear en los campos de peor calidad, mientras que las vacas que están en lactancia suelen hacerlo en campos de mejor calidad. No será que la relación aparente entre el monte y la enfermedad, se debe a una confusión entre el efecto monte y el estadio reproductivo de la vaca?

Recordemos

Confounding (confusión)

- Es una variable que se asocia con un factor de riesgo y la enfermedad, por lo que confunde las conclusiones



Estudios de casos controles

Estratificación:

La estratificación de casos y controles permite controlar por factores potenciales de confusión

Estudios casos controles

Datos originales (sin estratificar)

	Expuestos	No Expuestos	Total
Enfermos	89	23	112
Sanos	254	798	1052
Total	343	821	1164

Estudios casos controles

Datos estratificados

	Casos		Controles	
	Expuestos	No Expuestos	Expuestos	No Expuestos
Estrato 1	56	3	200	101
Estrato 2	23	20	54	697
Total	89	23	254	798

Casos = Enfermos; Controles = Sanos; Estrato 1 = vacas secas; Estrato 2 = vacas lactando; Expuestos = en el monte; No expuestos = en otras pasturas

Estudios de casos controles

Estratificación:

Los ORs para cada factor, ajustado por estrato, pueden estimarse pesando cada $\ln(\text{OR})$ por la inversa de su variancia o usando el test de Maentel-Haenszel. Como resultado obtenemos estimaciones de OR del factor de riesgo ajustado para cada estrato y para ambos estratos juntos ("pooled" OR)

Estudios de casos controles

En nuestro ejemplo:

OR(MH): 11.8 (5.9 – 23.8)

OR(vacas secas): 9.4 (2.9-30.9)

OR(vacas lactando): 14.8 (7.7-27.7)

De hecho, las vacas en lactación que se encontraban en el monte se encuentran a mayor riesgo que las vacas secas que se encontraban en el!

Estudios de casos controles

Nota:

Los estudios transversales también pueden estratificarse.

En este caso, el test de Maentel-Haenszel se usa para controlar factores de confusión en la estimación de las RPs.

Estudios observacionales

Transversales

Longitudinales de casos y controles

Longitudinales de cohorte

Estudios de cohorte

Existe dimensión temporal prospectiva (o en algunos casos retrospectiva)

Los animales se agrupan en expuestos y no expuestos a un factor considerado protector o de riesgo y luego se examina si se enferman o no

Típicamente se conoce el n de las categorías expuesto y no expuesto al comienzo del estudio, pero se desconoce el de enfermos y no enfermos

Estudios de cohorte

Ejemplo: Nuestro veterinario ya pudo probar que el monte se asocia con la enfermedad, pero aun no puede comprobar que es el monte quien causa la enfermedad. No sabe, por ejemplo, si los animales que estaban enfermos en el monte se enfermaron allí o ya estaban enfermos al momento de trasladarlos

Estudios de cohorte

IMPORTANTE:

Los estudios de cohorte pueden no ser suficientes para demostrar causalidad, pero debido a que son los únicos estudios observacionales en los que estamos seguros que el factor de riesgo antecede a la enfermedad, suelen ser un requisito indispensable en el camino a demostrarla

Estudios de cohorte

Ejemplo: El veterinario decide ahondar en su investigación y planea una visita al campo antes de que se roten los animales nuevamente. Comprueba que todos los animales están sanos al comienzo de la introducción, tantos los que se destinan al monte como los que van a otras pasturas, y vuelve a examinarlos un mes después donde determina cuáles se enfermaron y cuáles no

Estudios de cohorte

	Expuestos	No Expuestos	Total
Enfermos	80	5	85
Sanos	160	300	460
Total	240	305	540

Estudios de cohorte

El indicador recomendado para estudios de cohorte se denomina **riesgo relativo (RR)**

El **RR** es el mejor estimador del riesgo que un factor representa para una enfermedad, es aplicable solo cuando se conoce que el factor antecede a la enfermedad y tiende a ser menor que el valor de OR.

Estudios de cohorte

	Expuestos	No Expuestos	Total
Enfermos	A	B	A+B
Sanos	C	D	C+D
Total	A+C	B+D	A+B+C+D

$$\text{Riesgo Relativo} = [A/(A+C)] / [B/(B+D)]$$

RR = 1: no hay asociación entre el factor de exposición y la enfermedad

RR > 1: el factor de exposición se asocia con la ocurrencia de la enfermedad (factor de riesgo)

RR < 1: el factor de exposición se asocia con no ocurrencia de la enfermedad (factor protector)

Estudios de cohorte

NOTA:

El **RR** se estima en forma similar a la **RP**, sin embargo, se suele hacer una distinción en el nombre para indicar las características de cada uno de los estudios.

Cual es la importancia practica de esta distinción?

Estudios de cohorte

	Expuestos	No Expuestos	Total
Enfermos	80	5	85
Sanos	160	300	460
Total	240	305	545

$$\text{Riesgo Relativo} = (80/240) / (5/305) = 20.3$$

El riesgo de que un bovino que pastorea en el monte se enferme es 20 veces el riesgo de que se enferme un bovino que no pastorea en el monte

Estudios de cohorte

NOTA:

En forma análoga a los estudios transversales, también es posible estimar la fracción etiológica, la fracción atribuible, el intervalo de confianza del riesgo relativo y la estratificación de las variables.

Estudios de cohorte

NOTA:

Una variación muy útil de los estudios de cohorte con medición de incidencia acumulada, son los que miden la tasa de incidencia en días. En estos casos, la unidad de sanos/enfermos son los días en cada categoría. En nuestro ejemplo, supongamos que todos los animales enfermos permanecieron en este estado durante un día y que el estudio duró un mes. Entonces:

Estudios de cohorte

Incidencia acumulada

	Expuestos	No Expuestos
Enfermos	80	5
Sanos	160	300

Tasa de incidencia

	Expuestos	No Expuestos
Enfermos	80*1	5*1
Sanos	160*30 + 80*29	300*1 + 5*29

Estudios de cohorte

QUE SIGUE LUEGO DE UN ESTUDIO DE COHORTE?

El paso posterior suelen ser los estudios experimentales, de acuerdo a lo que ya discutimos. Mediante los estudios experimentales, se puede investigar su grupos homogeneizados de expuestos al factor desarrollan la enfermedad. La estimación del RR suele ser el objetivo de este tipo de estudios.

Software: WinEpiscope

1. Como parte del plan de control de paratuberculosis se intenta determinar si existe una asociación con alimentar a los terneros con leche de vacas adultas de la explotación. De 40 explotaciones lecheras donde se diagnosticó la enfermedad, en 35 los terneros se alimentan con leche de vaca. Por otra parte, de 200 establecimientos que son libres de la enfermedad, 100 utilizan leche de vaca. Existe una asociación entre alimentar a las vacas con leche de vaca y ocurrencia de paratuberculosis?
2. Como parte del programa nacional de vigilancia se deciden revisar los registros del laboratorio de microbiología del Instituto Nacional de Epidemiología. En este laboratorio se registraron en los últimos 10 años 50 diagnósticos positivos de tuberculosis de un total de 500 muestras en las que se intentó el aislamiento. Al cruzar los resultados con las fichas de ingreso se descubre que 300 muestras (40 positivas) provenían de trabajadores rurales, mientras que las 200 restantes (10 positivas) provenían de pacientes con otras profesiones. Estime el riesgo relativo que tiene un trabajador rural de desarrollar tuberculosis comparado con trabajadores de otras profesiones.

El machete

	Expuestos	No Expuestos	Total
Enfermos	A	B	A+B
Sanos	C	D	C+D
Total	A+C	B+D	A+B+C+D

Fracción etiológica: Proporción de enfermos expuestos en los cuales la enfermedad puede atribuirse a la exposición al factor: $FE = (I - 1) / I$

Fracción atribuible: Proporción de enfermos en una población que se deben a la exposición al factor: $FE \times A/(A+B)$

	Expuestos	I
Transversal	$Exp/NE = Enf/S$	$RP = [A/(A+C)]/[B/(B+D)]$
Caso-control	$E/S \rightarrow Exp/NE$	$OR = (A/C)/(B/D)$
Cohorte	$Exp/NE \rightarrow E/S$	$RR = [A/(A+C)]/[B/(B+D)]$

$I < 1$: **Protectivo**

$I > 1$: **Riesgo**

$I = 1$: **No asociado**

Estimación del status sanitario y elección de la muestra

Estimación del status sanitario

Estatus sanitario

Hasta ahora solo mencionamos “enfermos” y “sanos”, pero es necesario hacer algunas consideraciones con respecto a como se dividen las categorías

Estatus sanitario

Dependiendo del nivel de análisis:

Individual: Cuál es la probabilidad de que un individuo catalogado como positivo (negativo) esté verdaderamente enfermo (sano)?

Conceptos de Sensibilidad y Especificidad

La probabilidad depende de la prevalencia!

Estatus sanitario

Dependiendo del nivel de análisis:

Grupal: Cuál es la probabilidad de que un grupo (rodeo, rebaño, establecimiento) catalogado como positivo (negativo) esté verdaderamente enfermo (sano)?

Depende de la Sensibilidad, Especificidad, Prevalencia esperada y tamaño poblacional

Estatus sanitario

Cameron AR, Baldock FC. Two-stage sampling in surveys to substantiate freedom from disease. *Prev Vet Med.* 1998, 34(1):19-30.

Cameron AR, Baldock FC. A new probability formula for surveys to substantiate freedom from disease. *Prev Vet Med.* 1998, 34(1):1-17.

Estimación del tamaño muestral

Algunas definiciones

- **Población:** es el "universo" o grupo objeto de estudio
- **Censo:** muestreo del total de la población
- **Muestra:** elementos de la población objeto de estudio que se seleccionan para representarla
- **Unidad de análisis:** un miembro de la población (animal, rebaño, ciudad, etc).
- **Marco de muestreo:** Lista que identifica las potenciales unidades de análisis en una población

Algunas definiciones

- **Representatividad de la muestra:** Grado en el que la muestra se asemeja a la población en estudio con respecto a sus elementos fundamentales.
- **Poder de generalización de la muestra:** Grado en el que se puede inferir características de una población basándose en las características de la muestra.

Algunas definiciones

- **Error de muestreo al azar (random sampling error):** Discrepancias entre la muestra y la población como resultado de usar un muestreo en lugar de un censo.
- **Errores sistemáticos:** Discrepancias (sesgo) entre la muestra y la población originado por un error de diseño o aplicación del método. Ej: incorrecta definición de las poblaciones, del marco de muestreo, o de los procesos de selección.

Pasos del muestreo

Es necesario seleccionar:



Pasos del muestreo

Definir la población objetivo:

- Guarda relación íntima con los objetivos del trabajo
- Las conclusiones del trabajo son solo aplicables a la población en estudio
- Definir la población temporal y espacialmente, además de otras características de interés
- Ejemplo: Felinos machos de la ciudad de Casilda durante el bienio 2004-2005.

Pasos del muestreo

Diseñar el marco de muestreo:

- Realizar un listado comprensivo y extensivo de todas las unidades dentro de la población objetivo (censos, directorios, registros, RENSPA).
- En algunos estudios suelen ser útiles las encuestas para evaluar si el individuo califica o no para el estudio.
- Errores en el marco de muestreo: imposibilidad de identificar algunos individuos de EXISTENCIA CONOCIDA. Ej: la dirección de un establecimiento esta incorrecta, o se perdió un registro de un animal

Pasos del muestreo

Diseñar el marco de muestreo:

- PARA TENER EN CUENTA:

Si el marco de muestreo es desconocido (no puede realizarse), entonces no es posible realizar un muestreo probabilístico (al azar). Esto no invalida la calidad del trabajo, pero debe ser tenido en cuenta por el investigador y discutido como un potencial factor de sesgo en los resultados (sesgo de selección).

Pasos del muestreo

Determinar el método de muestreo:

- Muestreo probabilístico: se obtienen unidades muestrales al azar del marco de muestreo.
- Muestreo no probabilístico: las unidades muestrales no se obtienen al azar, por ejemplo, porque no es posible diseñar un marco de muestreo.

Pasos del muestreo

Determinar el método de muestreo:

- EJEMPLO:

Se desea analizar la seroprevalencia de una enfermedad en la población canina de la ciudad de Rosario. Se obtienen muestras de todos los animales que concurren al hospital, se analizan y se estima el porcentaje de positivos.

Preguntas:

El muestreo fue realizado al azar o no?

Podemos decir que la estimación resultante es una medida de la prevalencia?

Pasos del muestreo

El muestreo fue realizado al azar o no?

No, ya que los individuos son susceptibles de verse afectados por sesgos de selección (por ejemplo, que los propietarios geográficamente mas cercanos al hospital concurren SISTEMATICAMENTE con mas frecuencia que los mas alejados). *Qué pasa si seleccionáramos al azar ALGUNOS de los individuos que concurren al hospital?*

Podemos decir que la estimación resultante es una medida de la prevalencia?

Debido a que el muestreo no fue realizado al azar, en sentido estricto no es una prevalencia (probabilidad) sino una medición de las chances (odds) o tasa de positividad

Pasos del muestreo

Determinar el método de muestreo:

• Métodos no probabilísticos:

Por conveniencia: Es el mas usado en países en desarrollo. Se "muestrea lo que se puede".

Por experto: Un experto juzga qué se debe muestrear y qué no (ej. Opinión de expertos)

Bola de nieve: Cascada de expertos.

Una modificación aplicable a todos ellos: establecer una cuota por sesión de muestreo (ej, 5 perros por día). Esto permite una cierta aleatorización temporal (o espacial) incluso en estudios no probabilísticos.

Pasos del muestreo

Determinar el método de muestreo:

• Métodos probabilísticos:

Cada unidad experimental tiene una probabilidad conocida de ser elegida:

Simple: La probabilidad de cada unidad es igual (Eg: generador de números aleatorios)

Pasos del muestreo

Determinar el método muestral:

• Métodos probabilísticos:

Cada unidad experimental tiene una probabilidad conocida de ser elegida:

Simple: La probabilidad de cada unidad es igual (Eg: generador de números aleatorios)

Sistemático: se selecciona un intervalo igual a n/N y un numero de orden al azar para comenzar a muestrear

Estratificado: se clasifica la población en subpoblaciones y luego se aplica cualquiera de los métodos descriptos

En grupos o dos estadios: cuando existen jerarquías organizacionales

Pasos del muestreo

Determinar el tamaño de muestreo:

• Métodos no estadísticos:

Arbitrario: Porcentaje de la población

Conveniencia: Económica, logística

Convencional: Tradición, estudios pasados

Pasos del muestreo

Determinar el tamaño de muestreo:

• Métodos estadísticos:

Depende del objetivo del estudio:

1. Estimar o comparar una proporción (eg. Prevalencia)
2. Estimar o comparar un promedio
3. Estudios de casos y controles
4. Estudios de cohorte y experimentales
5. Estimar presencia/ausencia de enfermedad

Pasos del muestreo

Muestreo probabilístico para estimar prevalencia:

Depende de la prevalencia esperada, error aceptado, nivel de confianza deseado y tamaño de la población.

Prevalencia esperada: Aproximación, basada en estudios previos. Es rara? Es común?

Nivel de confianza: Probabilidad de que el rango de valores sea el correcto

Error aceptado: Porcentaje de error admitido en la estimación

Pasos del muestreo

Muestreo probabilístico para estimar prevalencia:

$$n = \{t * \sqrt{[P*(1-P)] / L} \}^2 / \{1 + \{t * \sqrt{[P*(1-P)] / L} \}^2 / N\}$$

Prevalencia esperada (P), error aceptado (L), nivel de confianza deseado (t) y tamaño de la población (N).

Pasos del muestreo

Muestreo probabilístico para estimar prevalencia:

Ejemplo: Se desea estimar la prevalencia de tuberculosis en un establecimiento lechero de 500 animales, con un error máximo del 5% y un intervalo de confianza del 95%. Basados en estudios previos, se espera una prevalencia del 10 %

$$\begin{aligned} n &= \{t * \sqrt{[P*(1-P)] / L} \}^2 / \{1 + \{t * \sqrt{[P*(1-P)] / L} \}^2 / N\} \\ &= \{1.96 * \sqrt{[0.1*(1-0.1)] / 0.05} \}^2 / \{1 + \{1.96 * \sqrt{[0.1*(1-0.1)] / 0.05} \}^2 / 500\} \\ &= 108.3 \sim 109 \end{aligned}$$

Pasos del muestreo

Muestreo probabilístico para estimar prevalencia:

Que pasa si, por ejemplo, queremos comparar dos poblaciones para saber si presentan la misma prevalencia?

En este caso, el tamaño muestral depende de la prevalencia esperada en cada población (p1 y p2), el nivel de confianza (error de tipo I) y la potencia de la prueba (error de tipo II) [se usa el valor de t de St para el intervalo de confianza (Z1) y la potencia (Z2)]

$$n = [(Z1 + Z2) / (p1 - p2)]^2 * [p1 * (1-p1) + p2 * (1-p2)]$$

Qué es el error de tipo I y el error de tipo II?

Pasos del muestreo

Muestreo probabilístico para estimar prevalencia:

Ejemplo: Se desea estimar si el numero de establecimientos positivos a trichomoniasis en la cuenca del Salado es superior que la prevalencia en la zona sur de Santa Fe. Se espera un 30% de establecimientos positivos en el Salado y un 5% en Santa Fe y se pretende un nivel de confianza y potencia del 95% y 85% respectivamente.

El resultado puede estimarse para tests de una cola o dos colas. En nuestro ejemplo, n = 31 (una cola) y n = 39 (dos colas).

Qué diferencia hay entre un test a una y dos colas?

Pasos del muestreo

Muestreo probabilístico para estudios de casos-controles y cohorte

Es más complejo. En general, suelen determinarse por conveniencia los estudios de casos-controles y por convención los estudios de cohorte.

Existen sin embargo algunas aproximaciones que permiten estimar el n necesario, en caso de observarse algunas situaciones esperadas.

Pasos del muestreo

Muestreo probabilístico para estudios de casos-controles

$$n = \frac{\left(Z(a) \sqrt{\left(1 + \frac{1}{c} \right) pm (1 - pm) + Z(b) \sqrt{p1 (1 - p1) + \frac{p0 (1 - p0)}{c}}} \right)^2}{(p1 - p0)^2}$$

Siendo:

$$p1 = \frac{p0 \cdot RR}{1 + p0 (RR - 1)} \quad pm = \frac{p1 + c \cdot p0}{1 + c}$$

donde:

- Z(a) = valor de la t, de Student para el nivel de confianza especificado
- Z(b) = valor de la t, de Student para la potencia especificada
- p0 = proporción esperada de exposición entre los sanos (controles)
- p1 = proporción esperada de exposición entre los enfermos (casos)
- pm = proporción esperada de exposición en la población (enfermos y sanos)
- 1-pm = proporción esperada de no expuestos en la población
- c = relación esperada entre enfermos y sanos
- OR = Odds Ratio estimado de suficiente importancia

Requiere proveer estimaciones del OR considerado biológicamente significativo, la proporción de controles expuestos al factor, el numero de casos por controles deseado, el nivel de confianza y la potencia

Pasos del muestreo

Muestreo probabilístico para estudios de cohorte

$$n = \frac{(p1 \cdot q1 + p2 \cdot q2) \cdot K}{(p1 - p2)^2} \quad p2 = \frac{p1}{RR}$$

donde:

- n = número de animales necesarios para cada grupo
- K = (Z(a) + Z(b))²
- Z(a) = valor de la t, de Student para el nivel de confianza especificado
- Z(b) = valor de la t, de Student para la potencia especificado
- p1 = proporción esperada en animales no expuestos
- q1 = 1 - p1
- p2 = la incidencia mínima a detectar en los animales expuestos (basada en el RR que queremos detectar)
- q2 = 1 - p2
- RR = Riesgo Relativo estimado de suficiente importancia

Requiere proveer estimaciones del RR considerado biológicamente significativo, la proporción esperada de enfermos no expuestos al factor, el nivel de confianza y la potencia

Pasos del muestreo

Trabajo a campo:

- Considerar número de veces que se contactará en caso de no respuesta
- Mantener registro de negativas y perdidas
- Organización de la agenda de trabajo
- Entrenamiento y práctica (pre-test)
- Asegurarse tener la logística y materiales necesarios
- Guardar buenos registros, incluyendo información de contacto! (teléfono, email, etc)

Estudio de caso

Los estudios de casos-controles son los más comunes en medicina veterinaria. Suelen ser mas baratos y convenientes. Como vimos, uno de los problemas mas importantes de este tipo de estudios es controlar los factores de confusión. Si bien vimos como controlar por un factor usando el test de Maentel-Haenszel, que sucede cuando hay mas de un factor de confusión? Aun mas, muchas veces existe mas de un factor de riesgo influyendo en la ocurrencia de la enfermedad. Como estimar el impacto de cada uno de ellos en la ocurrencia de la enfermedad?

Factores de riesgo asociados con casos clínicos de moquillo canino en Casilda, Santa Fe

El moquillo canino es una enfermedad causada por el virus del Distemper Canino

Aunque se considera una de las enfermedades infecciosas caninas mas importantes de la Argentina, no existe información actualizada acerca de los factores de riesgo de la enfermedad.

El objetivo del trabajo fue identificar factores de riesgo asociados con manifestación clínica de la enfermedad en el área de Casilda

Factores de riesgo asociados con casos clínicos de moquillo canino en Casilda, Santa Fe

Fuente de datos:

Entre Junio del 2001 y Mayo del 2002, se recogieron los datos referidos a presencia o no de signos clínicos de moquillo en animales tratados en el hospital de clínica de la facultad, edad, raza, sexo, propietario (si/no), status vacunal, frecuencia y tipo de pases y estación.

Factores de riesgo asociados con casos clínicos de moquillo canino en Casilda, Santa Fe

Análisis:

Primero se utilizó un análisis logístico bivariado para detectar aquellos factores potencialmente relacionados con la enfermedad ($P < 0.2$).

Este tipo de análisis es similar a los estudios de regresión, pero la variable dependiente toma solo dos valores posibles (sano o enfermo en este caso).

Variables independientes: factores de riesgo
Variable dependiente: status

Factores de riesgo asociados con casos clínicos de moquillo canino en Casilda, Santa Fe

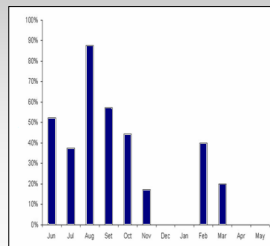
Análisis:

Aquellos factores de riesgo con un $P < 0.2$ fueron incluidos en un modelo de regresión logística multivariada, usando un algoritmo de pasos hacia adelante.

La regresión multivariada identifica el factor de riesgo que mejor explica la existencia de enfermedad. Luego se intenta agregar al modelo el factor que continúa en importancia y se evalúa si el ajuste del modelo mejora al agregar este factor. Esta técnica es muy útil para controlar *confounding*.

Factores de riesgo asociados con casos clínicos de moquillo canino en Casilda, Santa Fe

Resultados: Tasa de positividad del 35% (n=121)



Result

Variable	Categoría	Casos	Controles	Riesgo (OR)	p
Sexo	Macho	25	41	1,25	0,566
	Hembra	18	37	1*	
Edad (años)	1	28	29	8,37	0,006
	2	8	9	8,87	
	3	1	10	0,87	
	4	2	4	4,33	
	≥ 5	3	26	1*	
Hábito	Callejero	9	9	2,03	0,170
	Doméstico	34	69	1*	
Frecuencia de ejercicio	Nunca	6	24	0,31	0,128
	1 vez por semana	6	5	1,45	
	1-6 veces por semana	4	5	0,99	
	1 vez por día	6	16	0,46	
	> 1 vez por día	21	26	1*	
Forma de ejercicio	Libre	28	42	0,95	0,929
	Con correa	1	10	1*	
Vacunación previa	No	33	82	1,65	0,248
	Si	10	24	1*	
Vacunación en el último año	No	38	63	1,81	0,286
	Si	5	15	1*	
Estación	Otoño	9	8	4,22	0,159
	Invierno	17	23	2,77	
	Primavera	13	32	1,52	
	Verano	4	15	1*	
Raza	Indefinida/Cruza	30	45	1,83	0,161
	Pura	12	33	1*	

Resultados: Multivariado

Variable	Categoría	Casos	Controles	Riesgo (OR)	IC 95%	p
Edad (años)	1	27	28	8,37	2,27; 30,80	0,004
	2	8	8	11,53	2,82; 47,17	
	3	1	10	0,90	0,08; 10,19	
	4	1	4	4,90	0,52; 46,44	
	≥ 5	3	26	1*	...	
Hábito	Callejero	9	9	3,84	1,08; 13,71	0,04
	Doméstico	32	66	1*	...	

* categoría de referencia

Ajuste: Hosmer-Lemeshow ($P > 0.05$)
Interacción: NS ($P > 0.05$)

Factores de riesgo asociados con casos clínicos de moquillo canino en Casilda, Santa Fe

Discusión:

• Aunque la asociación con estación no fue significativa, el riesgo de ser diagnosticado en Otoño-Invierno fue 3-4 veces mayor.

• La asociación fue significativa con edad (menos de 1 año) y hábito (callejeros).

Estudio de caso

En relación al trabajo presentado, responda las siguientes preguntas:

- 1.Cuál es la definición del caso y el instrumento diagnóstico utilizado?
2. Que tipo de diseño de estudio epidemiológico fue utilizado?
- 3.Cuál fue el método de muestreo utilizado? Es al azar?
4. Por qué algunas causas que se hubiera esperado se asocien significativamente con la enfermedad no lo hicieron?
5. Cuales son las falencias del trabajo y qué opciones recomendaría para mejorarlo?

Ejercicio

En función de los temas discutidos:

1. Imagine una hipótesis de asociación de enfermedad que le gustaría investigar en su área de trabajo
2. Sugiera un diseño para la investigación. (*Nota: Debe ser factible!*)
3. Elija un método de muestreo. Defina un tamaño muestral por conveniencia y fundaméntelo.
4. Pretenda que realizó el estudio y obtuvo los resultados. Estime el indicador de asociación que corresponde. Que conclusión sugiere?
5. Describa potenciales sesgos y factores de confusión.

Universidad Complutense de Madrid
Mayo 2008
Curso Epidemiología Veterinaria

Introducción al análisis espacial y los sistemas de información geográficos

Andres Perez

Introducción

Conceptos básicos

- Mapas
- GIS o SIG
- Formatos de datos espaciales
- Georreferencias
- Temas de epidemiología espacial

Introducción

Conceptos básicos

- Mapas
- GIS o SIG
- Formatos de datos espaciales
- Georreferencias
- Temas de epidemiología espacial

Mapas

• Para qué sirven los mapas?

1. Los mapas identifican lo que hay en un lugar
2. Los mapas te dicen donde estás
3. Los mapas identifican relaciones
 - Por ej. los geógrafos estudian densidades poblacionales en áreas urbanas, para ayudar el desarrollo público
 - Por ej. Los epidemiólogos relacionan el patrón de eventos de enfermedades con factores ambientales o poblacionales

Mapas

- **Para qué sirven los mapas? (cont.)**
 4. Permiten combinar y superponer datos espaciales
 - por ej.: Dónde colocar un lugar destinado a residuos
 5. Determinan la mejor vía/ruta entre un lugar y otro
 - por ej.: el transporte público
 6. Modelan (predicen) eventos futuros
 - por ej.: derrame de químicos tóxicos

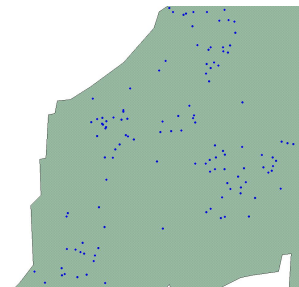
Mapas

- **Qué son los mapas?**
 - abstracciones de la realidad
 - interfase entre datos geográficos y percepción
 - dentro de un GIS, los mapas son una presentación dinámica de datos geográficos

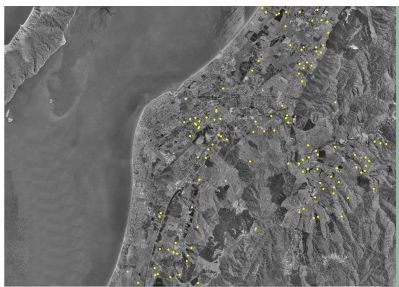
Mapas

- **Cómo pueden presentar la información geográfica los mapas?**
 - como datos discretos
 - como imágenes
 - como superficies

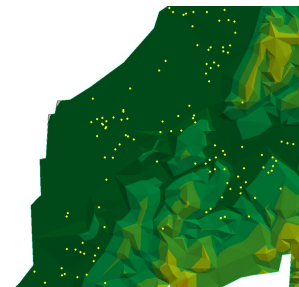
Mapas: El área en estudio es un dato discreto (polígono) y los establecimientos son datos discretos (puntos)



Mapas: El área en estudio es una imagen (satelital) y los establecimientos son datos discretos (puntos)



Mapas: El área en estudio es una superficie y los establecimientos son datos discretos (puntos)



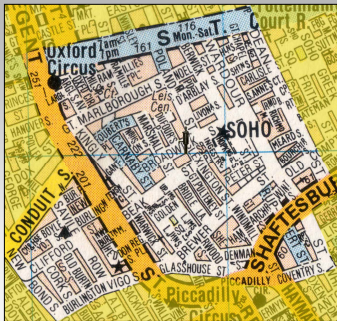
Mapas

- Objetivos
 - muestran una ubicación
 - indican atributos del lugar
- Como pueden mostrar los atributos?
 - un texto descriptivo (etiqueta)
 - un valor codificado
 - un valor numérico discreto
 - un valor numérico continuo

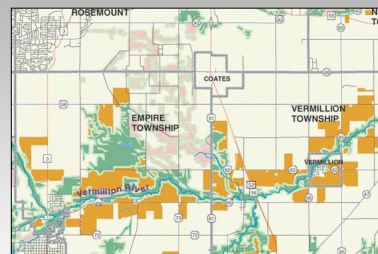
Mapas: Polígonos descriptivos + texto descriptivo (etiquetas) (Provincias, Argentina)



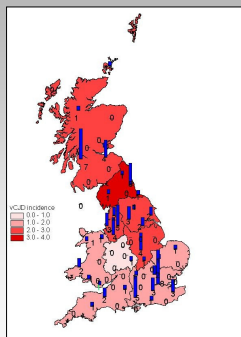
Mapas: Polígonos descriptivos con atributos + texto descriptivo (etiquetas) (Direcciones en Soho)



Mapas: Polígonos clasificados cualitativamente (color) + texto descriptivo (etiquetas) (tipos de suelo, Illinois)



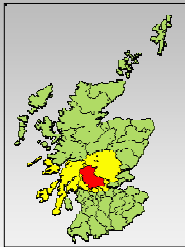
Mapas: Polígonos clasificados cuantitativamente + texto descriptivo (etiquetas) (Casos/millón de vCJD por condado, UK)



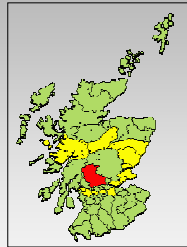
Mapas

- Cuáles son las formas de demostrar relaciones espaciales en mapas?
 - Adyacencia
 - Inclusión
 - Intersección
 - Proximidad

Adyacencia: Los distritos de Escocia marcados en amarillo son adyacentes al distrito marcado en rojo.

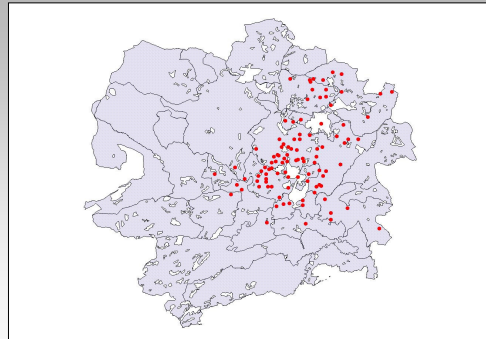


Primer orden de adyacencia

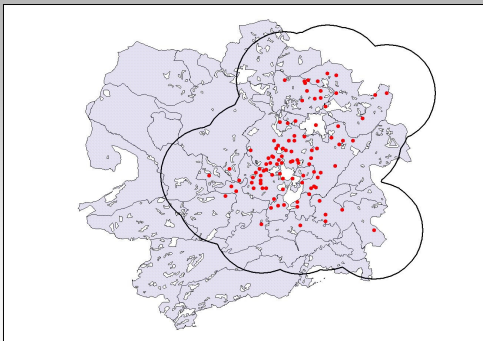


Segundo orden de adyacencia

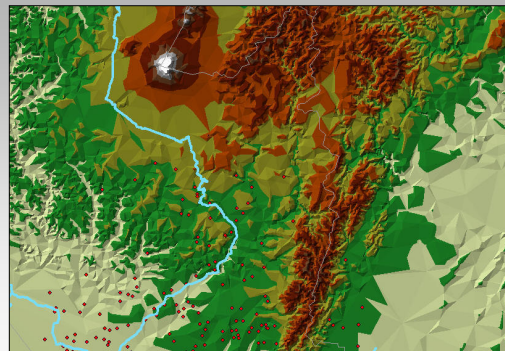
Inclusión: Establecimientos del condado de Cheshire con diagnóstico de FA durante el brote de 1967 – 1968 en UK



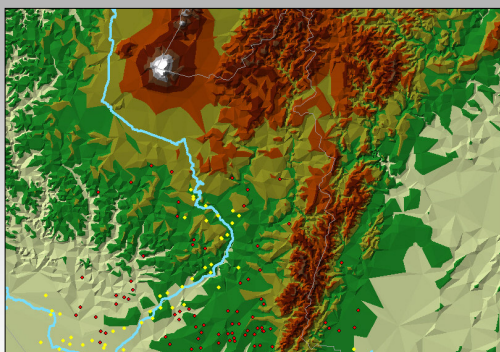
Intersección: Buffer de 5km alrededor de los establecimientos del condado de Cheshire con diagnóstico de FA (UK, 1967–1968)



Proximidad: Establecimientos ovinos en Rangitikea, NZ.



Proximidad: Establecimientos ovinos en Rangitikea, NZ. Se han seleccionado los establecimientos a <5 Km de la ruta principal



Introducción

Conceptos básicos

- Mapas
- GIS o SIG
- Formatos de datos espaciales
- Georreferencias
- Temáticas de epidemiología espacial

Sistemas de información geográfica

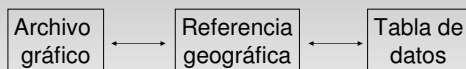
- **Qué es?**
 - hardware y software requeridos para guardar, mostrar y analizar datos espaciales
 - mapeo del siglo 19 en formato del siglo 21

Sistemas de información geográfica

- **Qué es?**
 - hardware y software requeridos para guardar, mostrar y analizar datos espaciales
 - mapeo del siglo 19 en formato del siglo 21

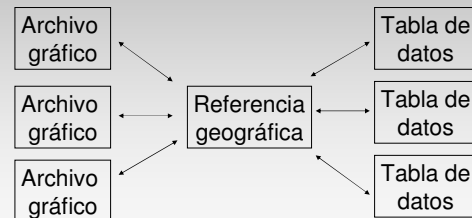
Sistemas de información geográfica

- **Qué es?**



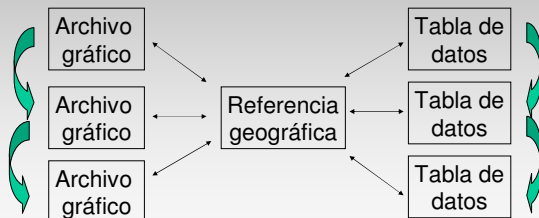
Sistemas de información geográfica

- **Qué es?**



Sistemas de información geográfica

- **Qué es?**



Sistemas de información geográfica

- **Tabla de Datos 1:**
Localización de establecimientos con y sin brotes de fiebre aftosa en Cumbria, UK

FID	Shape	ID	XCOORD	YCOORD	STATUS	DAYS	POLY
0	Point	1	342400	541400	0	241	17
1	Point	2	354000	536300	0	241	34
2	Point	3	370010	544000	0	241	63
3	Point	4	377800	544000	0	241	63
4	Point	5	361200	532400	0	58	34
5	Point	6	363600	531300	0	69	43
6	Point	7	361210	532400	0	241	34
7	Point	8	361170	530410	0	76	35
8	Point	9	343000	530330	0	241	18
9	Point	10	340000	530300	0	241	25
10	Point	11	344000	543200	0	241	17
11	Point	12	356400	536200	0	241	34
12	Point	13	343000	528000	0	241	15
13	Point	14	361400	537400	0	241	35
14	Point	15	356900	539000	1	42	34
15	Point	16	351100	522600	0	63	24
16	Point	17	340000	524000	0	70	6
17	Point	18	341600	543200	0	72	17
18	Point	19	344000	535170	0	79	18
19	Point	20	373000	521200	0	241	50

Sistemas de información geográfica

- Tabla de Datos 1:

Localización de establecimientos con y sin brotes de fiebre aftosa en Cumbria, UK

FID	Shape	ID	XCOORD	YCOORD	STATUS	DAYS	POLY
0	Point	1	342400	541400	0	241	17
1	Point	2	354900	536300	0	241	34
2	Point	3	378610	544500	0	241	63
3	Point	4	377880	544500	0	241	63
4	Point	5	361200	532400	0	58	34
5	Point	6	363800	531300	0	69	43
6	Point	7	361210	532400	0	241	34
7	Point	8	361170	539410	0	76	35
8	Point	9	342600	530330	0	241	16
9	Point	10	346900	536300	0	241	25
10	Point	11	344000	543200	0	241	17
11	Point	12	356400	536200	0	241	34
12	Point	13	343900	528600	0	241	15
13	Point	14	361400	537400	0	241	35
14	Point	15	369900	533600	1	42	34
15	Point	16	351100	523600	0	63	24
16	Point	17	340800	524600	0	70	6
17	Point	18	341920	543320	0	72	17
18	Point	19	344550	535170	0	79	16
19	Point	20	373000	521200	0	241	50

Sistemas de información geográfica

- Tabla de Datos 1:

Identificación (ID) de la granja

FID	Shape	ID	XCOORD	YCOORD	STATUS	DAYS	POLY
0	Point	1	342400	541400	0	241	17
1	Point	2	354900	536300	0	241	34
2	Point	3	378610	544500	0	241	63
3	Point	4	377880	544500	0	241	63
4	Point	5	361200	532400	0	58	34
5	Point	6	363800	531300	0	69	43
6	Point	7	361210	532400	0	241	34
7	Point	8	361170	539410	0	76	35
8	Point	9	342600	530330	0	241	16
9	Point	10	346900	536300	0	241	25
10	Point	11	344000	543200	0	241	17
11	Point	12	356400	536200	0	241	34
12	Point	13	343900	528600	0	241	15
13	Point	14	361400	537400	0	241	35
14	Point	15	369900	533600	1	42	34
15	Point	16	351100	523600	0	63	24
16	Point	17	340800	524600	0	70	6
17	Point	18	341920	543320	0	72	17
18	Point	19	344550	535170	0	79	16
19	Point	20	373000	521200	0	241	50

Sistemas de información geográfica

- Tabla de Datos 1:

Identificación geográfica de la granja

FID	Shape	ID	XCOORD	YCOORD	STATUS	DAYS	POLY
0	Point	1	342400	541400	0	241	17
1	Point	2	354900	536300	0	241	34
2	Point	3	378610	544500	0	241	63
3	Point	4	377880	544500	0	241	63
4	Point	5	361200	532400	0	58	34
5	Point	6	363800	531300	0	69	43
6	Point	7	361210	532400	0	241	34
7	Point	8	361170	539410	0	76	35
8	Point	9	342600	530330	0	241	16
9	Point	10	346900	536300	0	241	25
10	Point	11	344000	543200	0	241	17
11	Point	12	356400	536200	0	241	34
12	Point	13	343900	528600	0	241	15
13	Point	14	361400	537400	0	241	35
14	Point	15	369900	533600	1	42	34
15	Point	16	351100	523600	0	63	24
16	Point	17	340800	524600	0	70	6
17	Point	18	341920	543320	0	72	17
18	Point	19	344550	535170	0	79	16
19	Point	20	373000	521200	0	241	50

Sistemas de información geográfica

- Tabla de Datos 1:

Status de la granja (0: caso; 1: control)

FID	Shape	ID	XCOORD	YCOORD	STATUS	DAYS	POLY
0	Point	1	342400	541400	0	241	17
1	Point	2	354900	536300	0	241	34
2	Point	3	378610	544500	0	241	63
3	Point	4	377880	544500	0	241	63
4	Point	5	361200	532400	0	58	34
5	Point	6	363800	531300	0	69	43
6	Point	7	361210	532400	0	241	34
7	Point	8	361170	539410	0	76	35
8	Point	9	342600	530330	0	241	16
9	Point	10	346900	536300	0	241	25
10	Point	11	344000	543200	0	241	17
11	Point	12	356400	536200	0	241	34
12	Point	13	343900	528600	0	241	15
13	Point	14	361400	537400	0	241	35
14	Point	15	369900	533600	1	42	34
15	Point	16	351100	523600	0	63	24
16	Point	17	340800	524600	0	70	6
17	Point	18	341920	543320	0	72	17
18	Point	19	344550	535170	0	79	16
19	Point	20	373000	521200	0	241	50

Sistemas de información geográfica

- Tabla de Datos 2:

Número de granjas

CU_parish	VALUE	COUNT	CODE	x	y	FARMS
551091	31330	10	31330	384010	551091	10
551093	31399	1	31399	395551	551093	1
554564	31372	1	31372	390088	554564	1
551761	31633	7	31633	367443	551761	7
549444	31411	9	31411	376962	549444	9
492322	48389	7	48389	395292	492322	7
488482	48409	7	48409	383279	488482	7
494248	48410	18	48410	384557	494248	18
492620	48414	6	48414	391413	492620	6
500323	48387	4	48387	395497	500323	4
500803	48369	31	48369	387611	500803	31
511944	12692	4	12692	394684	511944	4
517765	12694	4	12694	395623	517765	4
532368	12005	30	12005	383584	532368	30
525871	12697	6	12697	390447	525871	6
519051	12699	6	12699	393360	519051	6
547308	12210	5	12210	393898	547308	5
518193	12601	1	12601	397482	518193	1
523560	12602	17	12602	385868	523560	17

Sistemas de información geográfica

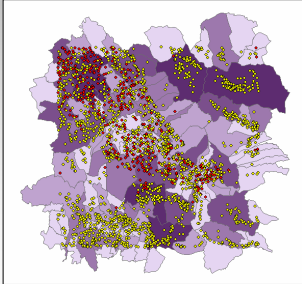
- Tabla de Datos 2:

Localización geográfica del condado

CU_parish	VALUE	COUNT	CODE	x	y	FARMS
551091	31330	10	31330	384010	551091	10
551093	31399	1	31399	395551	551093	1
554564	31372	1	31372	390088	554564	1
551761	31633	7	31633	367443	551761	7
549444	31411	9	31411	376962	549444	9
492322	48389	7	48389	395292	492322	7
488482	48409	7	48409	383279	488482	7
494248	48410	18	48410	384557	494248	18
492620	48414	6	48414	391413	492620	6
500323	48387	4	48387	395497	500323	4
500803	48369	31	48369	387611	500803	31
511944	12692	4	12692	394684	511944	4
517765	12694	4	12694	395623	517765	4
532368	12005	30	12005	383584	532368	30
525871	12697	6	12697	390447	525871	6
519051	12699	6	12699	393360	519051	6
547308	12210	5	12210	393898	547308	5
518193	12601	1	12601	397482	518193	1
523560	12602	17	12602	385868	523560	17

Sistemas de información geográfica

- Visualización de los datos en ambas tablas:



Sistemas de información geográfica

- Los SIG permiten el manejo de datos complejos en forma sencilla
- Tienen una gran capacidad de visualización de datos con un componente geográfico y muchas variantes y opciones
- Además posee una cierta capacidad analítica propia y en combinación con otros software, una capacidad analítica muy completa

Introducción

Conceptos básicos

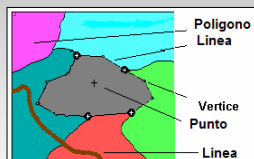
- Mapas
- GIS o SIG
- Formatos de datos espaciales
- Georreferencias
- Temáticas de epidemiología espacial

Formato de datos espaciales

- Vectores
 - representación por puntos, polígonos y líneas rectas
 - polígonos = series de vértices conectados por líneas rectas
- Raster
 - espacio geográfico dividido en un conjunto de celdas
 - atributos asignados a cada celda

Formato de datos espaciales

- Vectores



Formato de datos espaciales

- Vectores:

En la siguiente figura mostrando las rutas principales de Irán y los aislamientos de virus de fiebre aftosa de los últimos años, que vectores tienen formato de punto, polígonos y líneas?

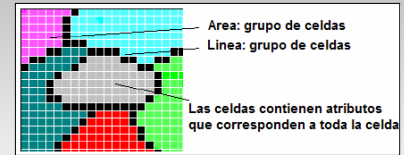
Formato de datos espaciales

- Vectores



Formato de datos espaciales

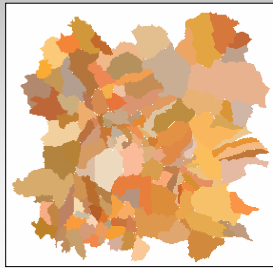
- Rasters



Formato de datos espaciales

- Rasters

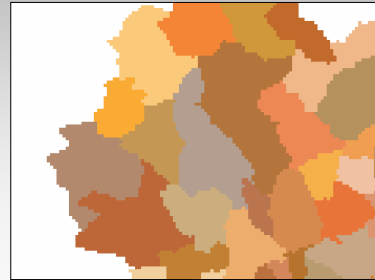
Condados de Cumbria



Formato de datos espaciales

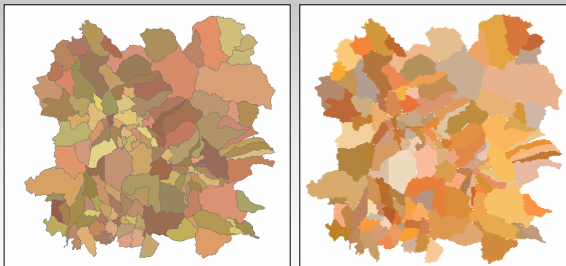
- Rasters

Condados de Cumbria



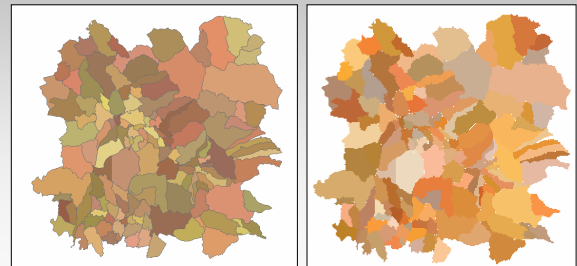
Formato de datos espaciales

- **Nota:** Los mismos datos pueden visualizarse en formato raster o vector!



Formato de datos espaciales

- **Pregunta:** Cual le parece mejor? Por qué?



Formato de datos espaciales

- **Vectores:** Mejor estética
- **Rasters:** Más correctos (mejores para cálculos)

Por qué los rasters son más correctos?

Introducción

Conceptos básicos

- Mapas
- GIS o SIG
- Formatos de datos espaciales
- Georreferencias
- Temáticas de epidemiología espacial

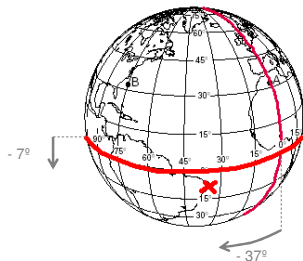
Georreferenciar

- **Qué es?**
 - proceso de asignar el lugar geográfico a la información
 - muchos sistemas diferentes

Georreferenciar

- **Latitud y longitud**
 - horizontal = líneas de latitud o paralelos
 - vertical = líneas de longitud o meridianos
 - expresados en grados decimales o DMS

Latitud y longitud



Georreferenciar

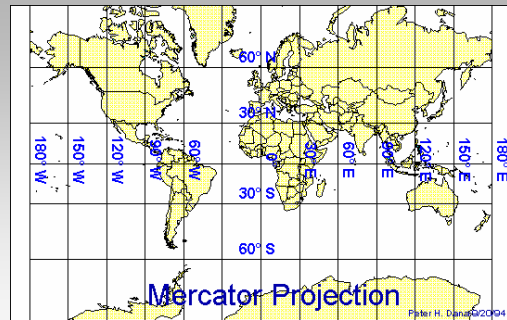
- **Elipsoides**
 - porque la Tierra no es perfectamente esférica, se definen los *elipsoides*
 - los mapas publicados se basan en un elipsoide específicamente definido
 - el Sistema Geodésico Mundial de 1984 (WG S84) es ahora aceptado como 'estándar'

Georreferenciar

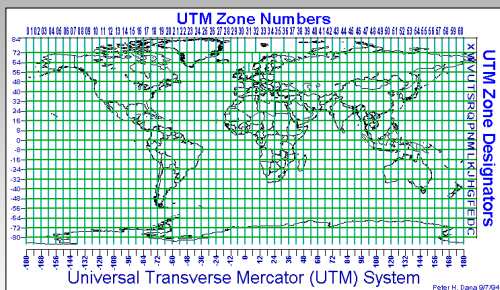
Proyecciones:

- Transpolar las localizaciones en el elipsoide a un plano
- Existen distintos sistemas de proyección

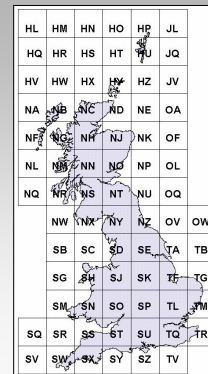
Mapa del mundo: proyección Mercator



Mapa del mundo: UTM



La British National Grid (similar Panafotosa y SENASA)



Georreferenciar

Proyecciones:

- Las distintas proyecciones tienen ventajas y desventajas (preservar áreas, preservar distancia, estética, etc).
- a veces necesitamos convertir una proyección a otra, por eso es útil conocer los diferentes sistemas

Introducción

Conceptos básicos

- Mapas
- GIS o SIG
- Formatos de datos espaciales
- Georreferencias
- Temas de epidemiología espacial

Temas de epidemiología espacial

- Autocorrelación
- Falacia ecológica
- El problema de la unidad de superficie modificable ('MAUP')

Temas de epidemiología espacial

- Autocorrelación
- Falacia ecológica
- El problema de la unidad de superficie modificable ('MAUP')

Autocorrelación

- La primera ley de la geografía de Tobler
 - 'todo se relaciona a todo, pero las cosas cercanas están mas estrechamente relacionadas que las cosas distantes'
- Lo que significa
 - que los datos pueden ser generalizados
 - que los datos no son independientes

Autocorrelación

Definición

- Tendencia de una variable a ser similar a través del espacio
- Luego veremos en el modulo siguiente que esto es también un atributo importante de los datos temporales

Autocorrelación espacial moderada



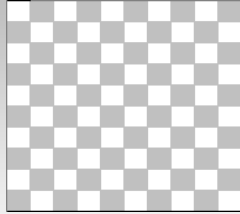
Raster mostrando celdas casos (grises) y controles (blanco)

Autocorrelación espacial extrema



Raster mostrando celdas casos (grises) y controles (blanco)

Sin autocorrelación espacial



Raster mostrando celdas
casos (grises) y controles (blanco)

Autocorrelación

Medida

- Veremos mas adelante que la autocorrelación puede ser una medida muy útil para evaluar las enfermedades

Temas de epidemiología espacial

- Autocorrelación
- Falacia ecológica
- El problema de la unidad de superficie modificable ('MAUP')

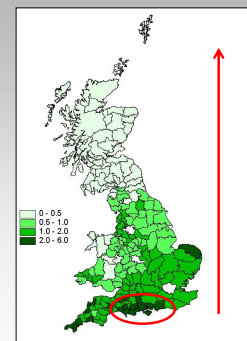
Falacia ecológica

- Qué es un estudio ecológico?

Falacia ecológica

- En geografía médica los estudios muchos estudios son realizados a nivel de *areas*
 1. por razones administrativas
 2. por confidencialidad
- Estos estudios permiten apreciar tendencias espaciales pero no permite hacer inferencias a nivel de 'sujetos individuales' (en caso de hacerlo, se denomina "falacia ecológica")

Tasa de Mortalidad Estandarizada de EEB en ganado Británico nacido antes del 30/06/1988



Tendencia de
la TME de EEB

Área de mayor riesgo, sin
embargo, no podemos
asegurar que TODAS las
granjas estén a mayor riesgo

Temas de epidemiología espacial

- Autocorrelación
- Falacia ecológica
- El problema de la unidad de superficie modificable ('MAUP')

Unidad de superficie modificable

• Definición

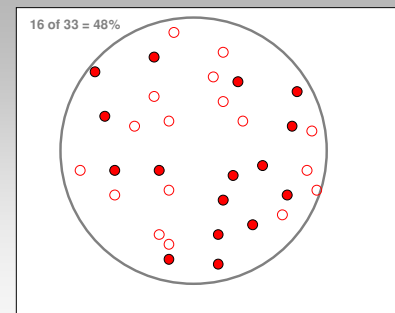
- problema que surge de la imposición de límites artificiales en fenómenos geográficos continuos
- efecto escala
- efecto 'zona'

Unidad de superficie modificable

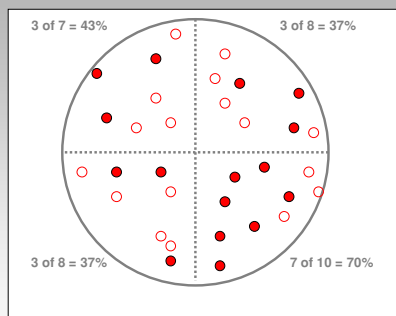
• Efectos escala

- variación en resultados numéricos, que ocurren debido al número de zonas empleadas en un análisis
- a medida que el número de zonas $\uparrow$, la variación en los resultados numéricos $\uparrow$

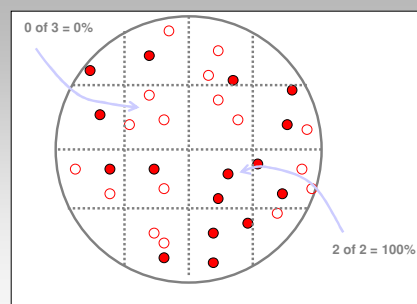
El problema de la unidad de superficie modificable: efecto escala



El problema de la unidad de superficie modificable: efecto escala



El problema de la unidad de superficie modificable: efecto escala

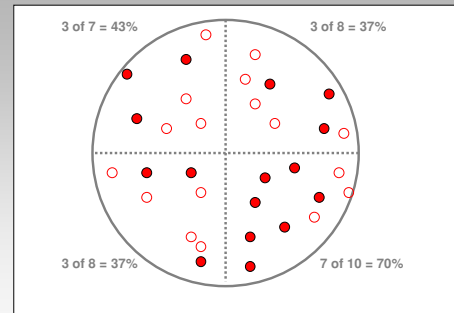


Unidad de superficie modificable

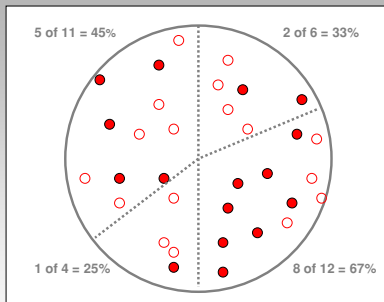
- Efecto 'zona'

- variación en resultados numéricos que ocurre como consecuencia de definiciones de límites artificialmente establecidos

El problema de la unidad de superficie modificable: efecto zona



El problema de la unidad de superficie modificable: efecto zona



Unidad de superficie modificable

- Soluciones

- no existen recetas sencillas!
- atención con el efecto que el MAUP puede tener en los análisis
- usar variedad de definiciones de límites y evaluar la sensibilidad a los resultados

El primer objetivo del análisis espacial es la visualización de datos (etapas pre-hipotéticas de un trabajo)

Visualización

- Objetivos

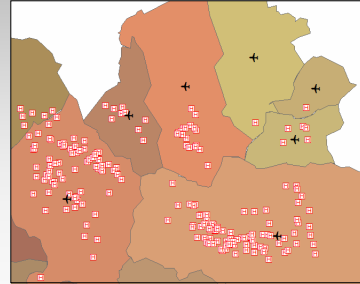
- Los métodos de visualización varían entre puntos (vectores puntos) o áreas (rasters o vectores polígonos)
- Algunos símbolos son convencionales

Visualización

- Símbolos convencionales
- Puntos
- Áreas

Visualización

- **Símbolos convencionales**
Por ejemplo, rutas, ciudades, etc



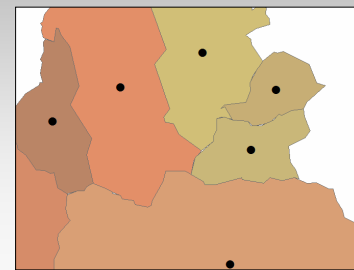
Visualización

Puntos

- Valor único
- Categóricos
- Numéricos:
 - 1 variable:
 - Intensidad de color
 - Tamaño graduado
 - Tamaño proporcional
 - Más de una variable

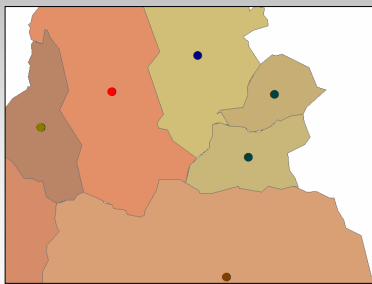
Visualización

Puntos: Valor único



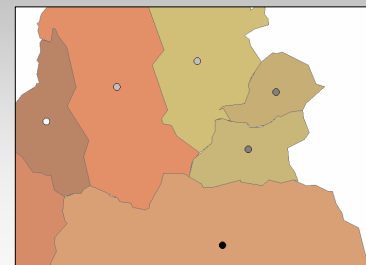
Visualización

Puntos: Categóricos



Visualización

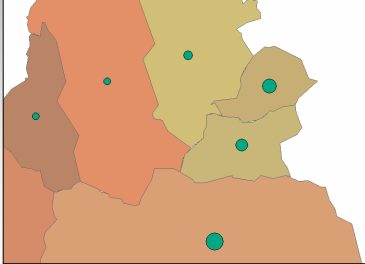
Puntos: Numérico; intensidad de color



Visualización

Puntos:

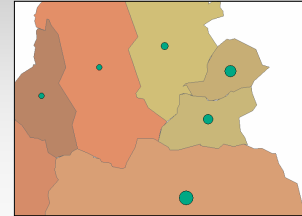
Numérico; tamaño graduado y proporcional



Visualización

Puntos:

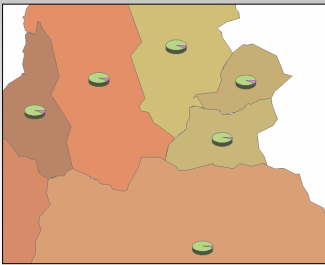
Cuál es la diferencia entre una visualización por puntos graduados y una por puntos proporcionales?



Visualización

Puntos:

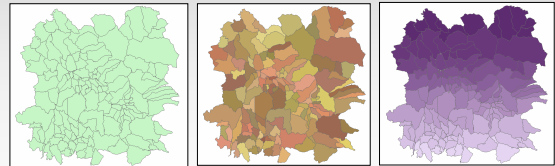
Numérico, >1 variable: gráficos (tortas, barras,etc)



Visualización

Áreas

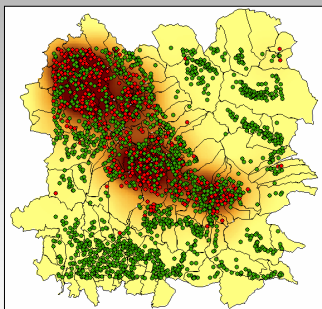
- Valor único: Mismo color
- Categóricos: Distintos colores
- Numéricos: Intensidad de color



Visualización

Preguntas:

- Cuántas hojas ve superpuestas en este gráfico?
- De qué tipo parecen ser? (vectores o rasters)
- Cuál es el sistema de visualización?



Explorando ArcGIS.....

Análisis geoestadístico

Andres Perez

Tests estadísticos de análisis espacial

1. Introducción
2. Tipos de tests
3. Ejemplo

Tests estadísticos de análisis espacial

1. Introducción
2. Tipos de tests
3. Ejemplo

Introducción

- Origen de la epidemiología: investigación de grupos de casos
- Investigar si hay un exceso de casos
- Medios gráficos: curvas y mapas

Introducción

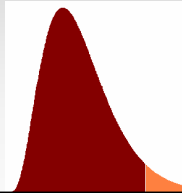
- Generación vs. corroboración de hipótesis
- Mas precisión al estudio, facilitan comparación

Introducción

- Comparar distribuciones observadas con la esperada en caso de distribución al azar
- Hipótesis nula: azar / uniformidad de casos

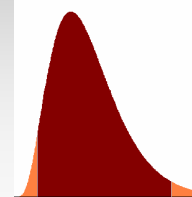
Introducción

Hipótesis alternativa:
Test de una cola: los casos observados son (significativamente) mas (o menos) que los esperados



Introducción

Hipótesis alternativa: test de dos colas:
los casos observados son (significativamente) mas/menos que los esperados



Tests estadísticos de análisis espacial

1. Introducción
2. Tipos de tests
3. Ejemplo

Tipos de Tests

Historia reciente

- 1950 autocorrelación
- 1954 nearest neighbour index
- 1990 test de Cuzick y Edward
- 1991 análisis de segundo orden
- 1997 scan statistic

Tipos de Tests

De acuerdo a la definición espacial

1. Datos individuales
2. Datos agrupados

De acuerdo al propósito

1. Globales
2. Locales
3. Focalizados

Tipos de Tests

De acuerdo a la definición espacial

1. Datos individuales
2. Datos agrupados

De acuerdo al propósito

1. Globales
2. Locales
3. Focalizados

Tipos de tests

Datos individuales

La unidad de estudio son individuos (rodeos, animales, etc) con una localización específica

Tipos de tests

Datos agrupados

En una determinada unidad geográfica se localiza más de un individuo

Tipos de tests

Ejemplo:

ID	Lat	Long	Casos	Población	Status
1	38.1	63.5	57	200	+
2	40.2	64.1	0	250	-
...	...	...	...	...	...
n	38.1	65.2	23	112	+

Tipos de tests

Datos individuales:

ID	Lat	Long	Casos	Población	Status
1	38.1	63.5	57	200	+
2	40.2	64.1	0	250	-
...	...	...	...	...	...
n	38.1	65.2	23	112	+

Tipos de tests

Datos agrupados:

ID	Lat	Long	Casos	Población	Status
1	38.1	63.5	57	200	+
2	40.2	64.1	0	250	-
...	...	...	...	...	...
n	38.1	65.2	23	112	+

Tipos de Tests

De acuerdo a la definición espacial

1. Datos individuales
2. Datos agrupados

De acuerdo al propósito

1. Globales
2. Locales
3. Focalizados

Tipos de tests

Tests globales

Detectar agrupamiento (significativo) de casos en el estudio, sin precisión de la localización del agrupamiento

Ej: Cuzick y Edwards', K-function, Autocorrelación, Grimson's.

Tipos de tests

Tests locales

Detectar agrupamiento (significativo) de casos en localizaciones específicas dentro del área en estudio

Ej: Besag y Newell's, G-test, Spatial scan statistic.

Tipos de tests

Tests focalizados

Detectar agrupamiento (significativo) de casos en el área de influencia de una localización específica (exposición a una fuente)

Ej: Score test, Spatial scan statistic.

Tests estadísticos de análisis espacial

1. Introducción
2. Tipos de tests
3. Ejemplo

Ejemplos

1. Tuberculosis bovina
2. Tricomoniiasis

Ejemplos

1. Tuberculosis bovina
2. Tricomoniiasis

Ejemplo



Tuberculosis bovina

→ Fuente de datos:

Inspección de los 126 frigoríficos federales del país (Marzo 1995 – Febrero 1997)

→ Datos:

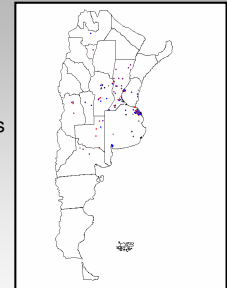
Animales con lesiones compatibles con TB,
Departamento/partido de origen,
Número de animales en la tropa

Tuberculosis bovina

Dep	°S	°O	Examinados	Casos
Azul	59.675	36.775	20150	1163
Bahía Blanca	62.175	38.550	14015	990
Balcarce	58.275	37.725	9533	478
:	:	:	:	:
Vera	60.425	29.050	992	12542

Tuberculosis bovina

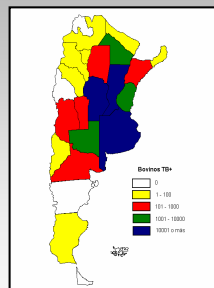
- > 70% de los Frigoríficos
- > 9.72.396 animales inspeccionados (~ 47% del ganado faenado)
- > 128.038 con lesiones compatibles
- > Tasa +: 1,35% (1,30-1,40%)



Tuberculosis bovina

Distribución del Total por Provincia

Buenos Aires	34%
Santa Fe	26%
Córdoba	21%
La Pampa	10%
Entre Ríos	5%
Total	96%



Tuberculosis bovina

1. Autocorrelación I (Moran, 1950):

$$I = \frac{N \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{J \sum_i \sum_j w_{ij} (x_i - \bar{x})^2}$$

N = número de puntos, J = no. de pares de puntos

x_i, x_j = valores de dos puntos vecinos

w_{ij} matriz de ponderación (ej. distancia entre vecinos)

$I > 0$ implica agrupamiento

Tuberculosis bovina

2. Test de Cuzick y Edwards (1990):

$$T_k = \sum_{i=1}^n \delta_i d_i^k$$

$\delta_i = 1$ si i es un caso, en su defecto $= 0$

$d_i^k = 1$ si k ésimo vecino de i es un caso, en su defecto $= 0$

Valores elevados de T_k implican agrupamiento

Tuberculosis bovina

3. Spatial Scan Statistic (Kulldorf, 1997):

- Superposición de ventanas circulares en el mapa
- El centro de las ventanas es el de los distintos puntos
- Cada ventana contiene diferentes grupos de casos
- La relación casos/controles se examina dentro y fuera de cada ventana

Tuberculosis bovina

3. Spatial Scan Statistic (Kulldorf, 1997):

- Modelo de Poisson o Bernoulli
- Poblaciones no homogéneas
- Control de confounding
- Test de una cola para probar significancia

Tuberculosis bovina

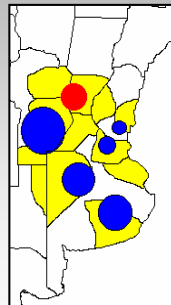
Tests de agrupamiento

- Autocorrelación TB: I de Moran $= 0,009$; $P = 0,089$
... sin evidencia de agrupamiento
- Autocorrelación Población: I de Moran $= 0,137$; $P < 0,001$
... agrupamiento significativo
- Test de Cuzick y Edwards:
... Globalmente agrupamiento no significativo ($P = 0,234$)
... Agrupamiento cuando $k = 1$ ($P = 0,036$); $k = 2$ ($P = 0,038$)

Tuberculosis bovina

Spatial Scan Statistic

- Agrupamiento primario
- Centro: 61,2°S; 31,20°
- Area: 103 km
- Observado: 5793 casos
- Esperado: 4382 casos
- Agrupamientos secundarios
- Áreas Lecheras



Tuberculosis bovina

Discussion

- Falla de la autocorrelación para detectar agrupamientos:
Población heterogénea
- Cuzick and Edwards test:
Definición de casos
Identificación de agrupamientos a nivel distrito

Tuberculosis bovina

Discusion

- El scan statistic permite detectar agrupamientos específicos y evaluar significancia
- Permite controlar por distribución heterogénea de poblaciones
- Modelos flexibles: datos individuales o agregados
- Permite controlar confounders

Tuberculosis bovina

- scan statistic:
identificación y descripción de agrupamientos
- Grupos de tuberculosis:
Cuenca lecheras
- Posibilidad de avanzar en nuevas hipótesis: datos referidos a edad, especie, etc.

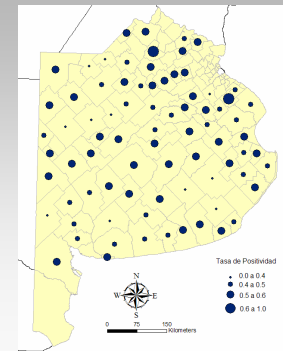


Ejemplos

1. Tuberculosis bovina
2. Tricomoniasis

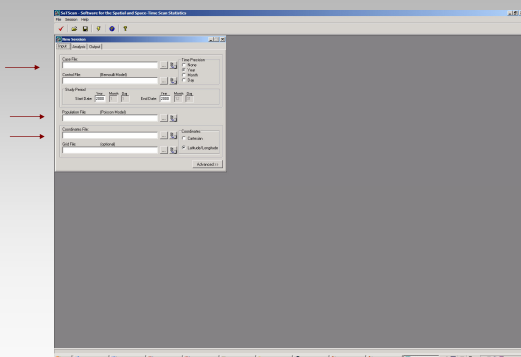
Tricomoniasis

Número de tests realizados y diagnósticos positivos a tricomoniasis

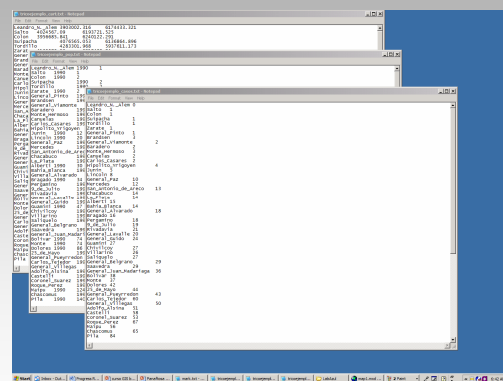


Tricomoniasis

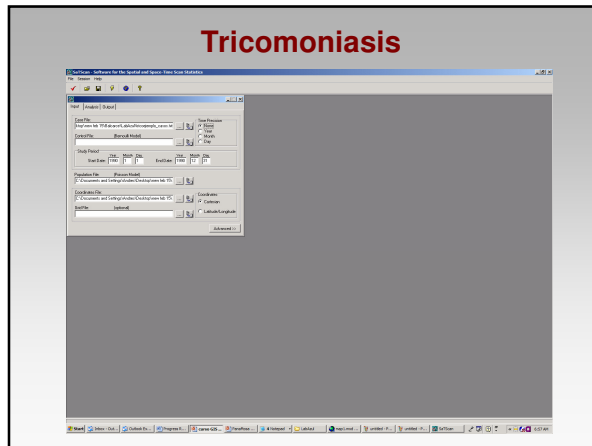
www.satscan.org



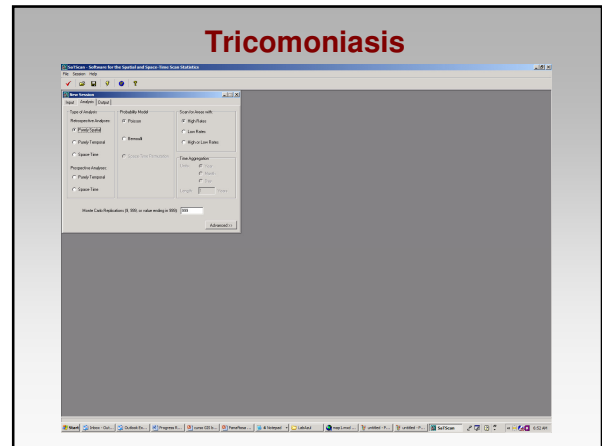
Tricomoniasis



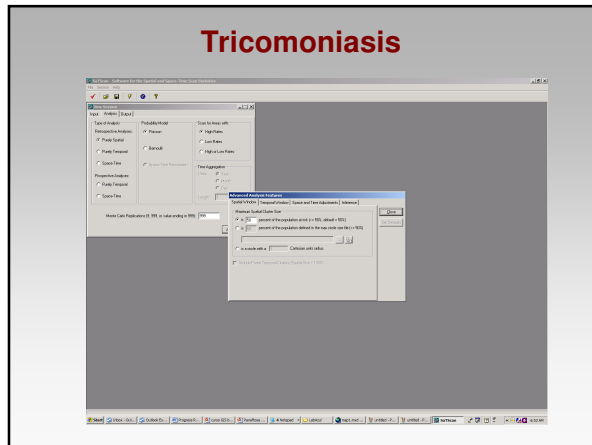
Tricomoniasis



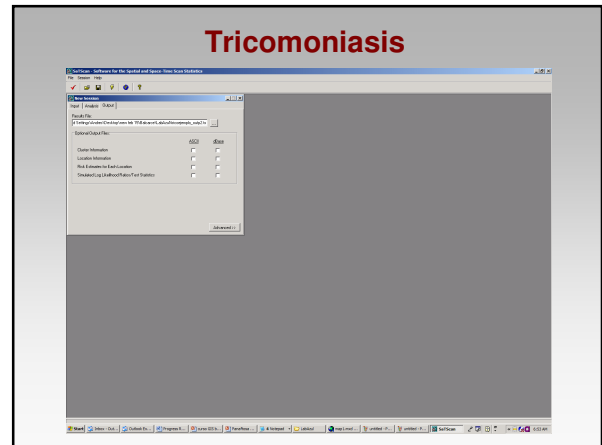
Tricomoniasis



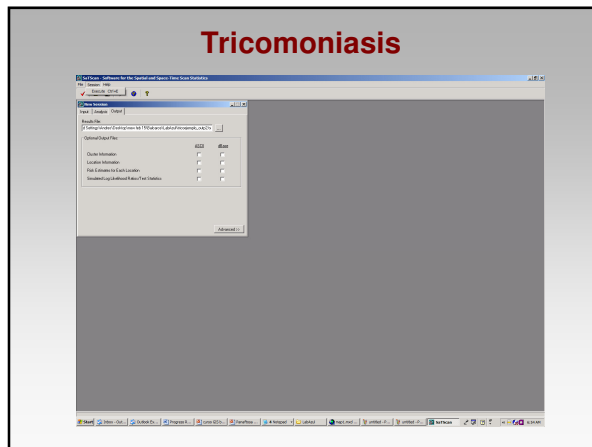
Tricomoniasis



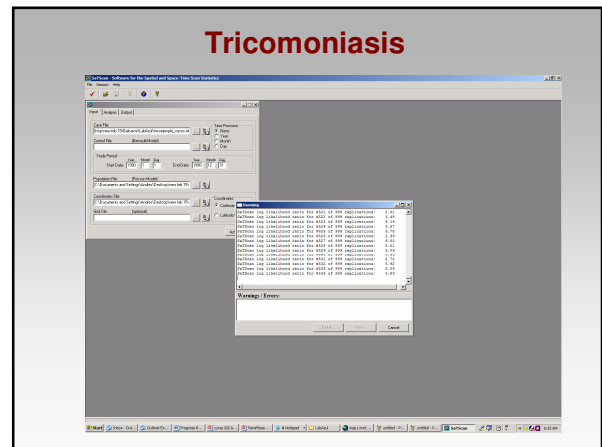
Tricomoniasis



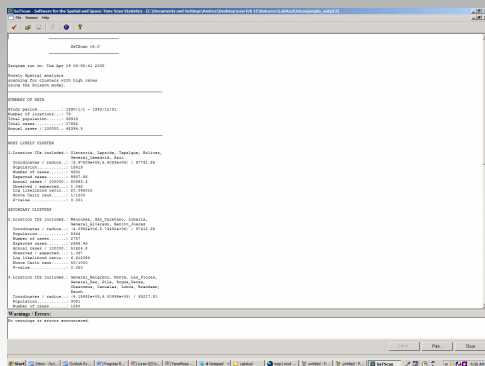
Tricomoniasis



Tricomoniasis

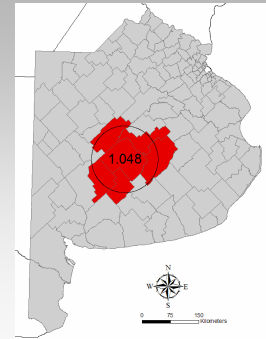


Tricomoniasis



Tricomoniasis

Grupo de partidos con riesgo relativo de diagnosticos positivos de tricomoniasis superior al promedio ($P < 0.01$)



Ejemplos

Ward MP, Perez A. Association between soil type and paratuberculosis in cattle herds. American Journal of Veterinary Research 65:10–14, 2004.

Chhetri BK, Perez AM, Thurmond MC. Factors associated with spatial distribution of FMD in Nepal. En preparacion.

Perez AM, Konig G, Späth E, Thurmond M. Variation in the VP1 gene of serotype A foot-and-mouth disease virus associated with epidemiological characteristics of outbreaks in the 2001 epidemic in Argentina. Journal of Veterinary Diagnostic Investigation. Aceptado

Inclusión del componente espacial en la variable respuesta

Ward MP, Perez AM.

Association between soil type and paratuberculosis in cattle herds.

American Journal of Veterinary Research 65:10–14, 2004.

Introduccion

M. avium subsp. *paratuberculosis* → Enf. de Johne

Costo para USA en 1999: **\$200-250M**
(Ott et al. *Prev Vet Med* 1999;40:179–192)

USDA considera que el control de PTB en USA es de **alta prioridad**

Programa de control voluntario
(United States Animal Health Association)

Persistencia en medio ambiente

8-11 meses en charcos, heces, etc.

6 meses en cursos de agua

Resistente al secado y el congelamiento



Johne's Information Center, University of Wisconsin, www.johnes.org

Asociacion con suelos

estudios transversales
No espaciales



Wisconsin, 115 rodeos bovinos:
asociación con suelos ácidos
(Kopecky. *J Am Vet Med Assoc* 1977;170:320-324)

Michigan, 121 rodeos lecheros:
suelos de bajo pH y alto contenido de Fe
(Johnson and Kaneene. *Am J Vet Res* 1999;60:589-596)

España, 61 rodeos ovinos y caprinos:
suelos de bajo pH y carencia de nutrientes
(Reviriego et al. *Prev Vet Med* 2000;43:43-51)

Objetivo

“ Estimar la asociación entre
características del **suelo** y
distribución de
paratuberculosis en rodeos
bovinos de **Indiana** ”



Poblacion en estudio

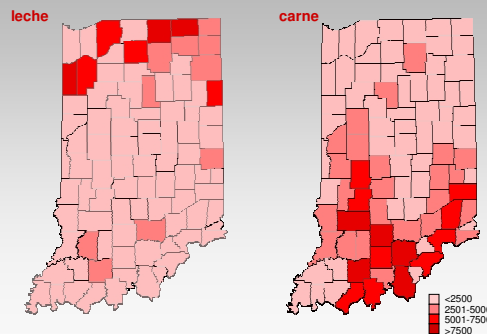
Muestra:

92 rodeos con ≥ 20 bovinos analizados por
ELISA (Ene98 – Dic02) y dirección georreferenciada

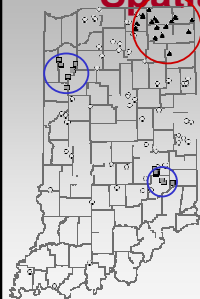
Prevalencia mediana de rodeo: **8.8%** (0-81%)
Número mediano de animales analizados: **50** (20-2936)

Sensibilidad mediana a nivel rodeo: **99%**
(supuestos: ELISA Se=45%, Es=99%, Prevalencia=20%)

Poblacion en estudio



Spatial scan statistic



Cluster	Rodeos	ELISA+	
		Esperado	Observado
1*	17	103	217
2	6	297	157
3	9	65	30

* observado ÷ esperado = 2.11; P=0.01

Analisis

Modelo de regresion logistica

- **casos:** rodeos dentro del cluster (n=17)
- **controles:** rodeos fuera del cluster (n=75)
- **variables:** relacionadas con suelo

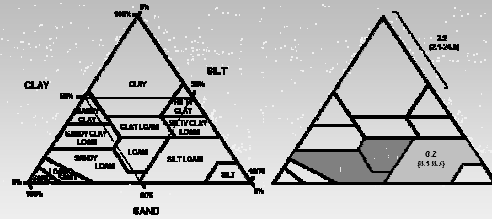
- ▶ pendiente
- ▶ drenaje
- ▶ congelamiento
- ▶ profundidad
- ▶ susceptibilidad a la corrosion
- ▶ textura
- ▶ encharcamiento
- ▶ retencion hidrica
- ▶ limitaciones al uso

Resultado

Asociacion con textura

Variable	OR	95% CI
Franco	3.6	1.0 – 13.2
Franco-arenoso	6.2	1.4 – 27.4
Franco-limoso	0.2	0.1 – 0.7

Textura del suelo



Resultados

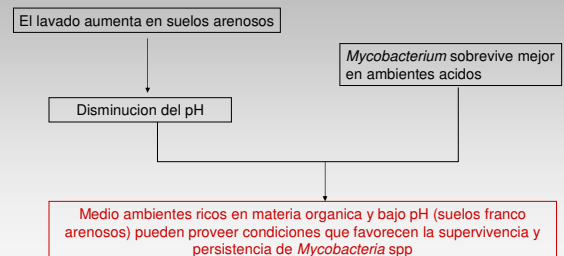
Relacion (univariado)

Tipo	Categoría	Referencia	OR	P
Cont. limo	≤50%	>50%	7.2	0.002
Cont. arena	≤50%	>50%	0.32	0.059
Drenaje	< Pobre	>Moderado	0.30	0.051

Relacion (multivariado)

Bajo contenido de limo → OR = **7.2** (95% CI: 2.1-24.5)

Discusion



Modelos asociativos

1. Inclusión del componente espacial como variable explicativa
2. Inclusión del componente espacial en la variable respuesta
3. Efectos aleatorios: aproximación Bayesiana

Efectos aleatorios: aproximación Bayesiana

Chhetri B.K., Perez A.M., Thurmond M.C.

Spatial analysis and risk factors associated with foot-and-mouth disease outbreaks in Nepal

MPVM, UC Davis (2007). Manuscrito en preparación.

Algunas slides de esta presentación son adaptaciones de slides originalmente diseñadas por Mark Stevenson, Massey University, Nueva Zelanda

Modelo espacial de efectos mixtos

Introducción: Estimación de la TEM

Distrito	VDC+	VDC
A	3	25
B	4	23
C	2	43
...	...	...
Total	89	924

$$= 0.096$$

Modelo espacial de efectos mixtos

Introducción: Estimación de la TEM

Distrito	VDC+	VDC	Esperado	TME
A	3	25	$0.096 \times 25 = 2.4$	
B	4	23	$0.096 \times 23 = 2.2$	
C	2	43	$0.096 \times 43 = 4.1$	
...	...	...	...	
Total	89	924		

$$89 \div 924 = 0.096$$

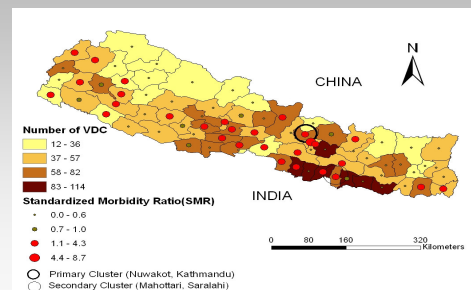
Modelo espacial de efectos mixtos

Introducción: Estimación de la TEM

Distrito	VDC+	VDC	Esperado	TME
A	3	25	$0.096 \times 25 = 2.4$	$3 \div 2.4 = 1.25$
B	4	23	$0.096 \times 23 = 2.2$	$4 \div 2.2 = 1.82$
C	2	43	$0.096 \times 43 = 4.1$	$2 \div 4.1 = 0.49$
...	...	...	...	
Total	89	924		

$$89 \div 924 = 0.096$$

Tasa de Morbilidad Estandarizada (SMR) Fiebre Aftosa, Nepal



Modelo espacial de efectos mixtos

Entonces:

$$\frac{O_i}{E_i} = TME_i$$

$$O_i = E_i \times TME_i$$

- El número de casos observados es igual al esperado nacional multiplicado por un factor de valor variable en cada área que es igual a TME
- Como el esperado nacional es conocido, el problema es estimar TME

Modelo espacial de efectos mixtos

Formalmente:

$$O_i = E_i \times TME_i$$

$$\mu_i = E_i \times RR_i$$

$$\log \mu_i = \log E_i + \log RR_i$$

$$\log \mu_i = \log E_i + (\beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi}) + \varepsilon$$

- el término $\log(E_i)$ es un ajuste para poblaciones de distinto tamaño

Modelo espacial de efectos mixtos

- **Formalmente:**

$$O_i = E_i \times TME_i$$

$$\mu_i = E_i \times RR_i$$

$$\log \mu_i = \log E_i + \log RR_i$$

$$\log \mu_i = \log E_i + (\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) + \varepsilon$$

Los coeficientes de regresion indican la fortaleza de la relacion entre covariable (factor de riesgo!) y enfermedad

ε_i es el riesgo relativo residual en el area i luego de controlar por las variables incluidas en el modelo

Modelo espacial de efectos mixtos

- **Objetivo**

- Además de estimar y controlar por los efectos fijos, estimar los efectos de segundo orden (agrupamiento) no explicados
- Modelo mixto: Efectos fijos + efectos al azar

Modelo espacial de efectos mixtos

- **Método bayesiano:**

- Supongamos que queremos estimar un valor desconocido θ
- Recolectamos datos X
 - Especificamos una distribución a priori de θ como $p(\theta)$
 - Especificamos $p(X | \theta)$ la distribución de X dado θ
 - Aplicamos el teorema de Bayes para determinar $p(\theta | X)$: la distribución de θ , dado X

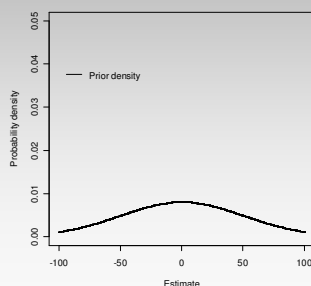
Modelo espacial de efectos mixtos

- **Método bayesiano:**

- La distribución a priori $p(\theta)$ representa nuestra incerteza acerca del valor de θ ANTES de ver los datos
- La distribución a posteriori $p(\theta | X)$ representa nuestra incerteza acerca del valor de θ DESPUES de ver los datos

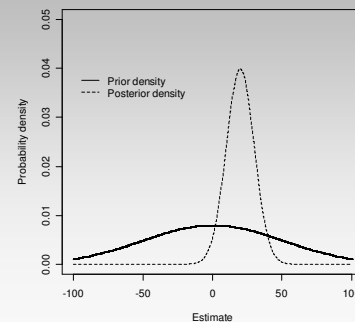
Modelo espacial de efectos mixtos

- **Método bayesiano:**



Modelo espacial de efectos mixtos

- **Método bayesiano:**



Modelo espacial de efectos mixtos

Nuestro modelo de efectos fijos:

$$O_i = E_i \times TME_i$$

$$\mu_i = E_i \times RR_i$$

$$\log \mu_i = \log E_i + \log RR_i$$

$$\log \mu_i = \log E_i + (\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) + \varepsilon$$

Modelo espacial de efectos mixtos

Nuestro modelo de efectos mixtos:

$$O_i = E_i \times SMR_i$$

$$\mu_i = E_i \times RR_i$$

$$\log \mu_i = \log E_i + \log RR_i$$

$$\log \mu_i = \log E_i + (\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) + \varepsilon$$

$$\log \mu_i = \log E_i + (\beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im}) + U_i + S_i + \xi_i$$

Efecto al azar estructurado espacialmente. Depende de especificaciones espaciales

Besag, York y Mollié (1991)

Efecto al azar no estructurado. Estima efectos no espaciales

Recolección de datos

- Cuestionarios enviados a los responsables veterinarios de todos los distritos de Nepal (n = 75)
- Respuestas basadas en información recolectada en 2004
- La unidad de análisis es la VDC.
- Variable respuesta: Caso: al menos un caso clínico de FA observado; Control: otros
- Covariables = 20

Modelo

- Estimación de la R de Spearman's
- Descarte de variables con $P > 0.1$
- Formulación de un modelo bayesiano de efectos mixtos asumiendo una distribución de Poisson para el número de VDC+ y valores a priori no informativos
- Selección del modelo final (DIC)

Valor de la R de Spearman's para la relación entre TEM y variables explicativas

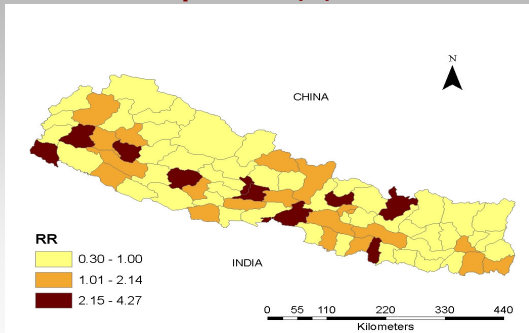
Covariate	Spearman's rho	p-value
Total Population of Humans (totpop)	0.4558	< 0.0001
Population density (popdens)	0.4333	0.0001
Buffalo Population (popbuf)	0.4289	0.0001
Number of technicians (numtech)	0.4257	0.0001
Number of animals slaughtered daily (dailyslaughter)	0.4158	0.0002
Percent of Urban population (urbpop)	0.3688	0.0011
Number of veterinarians (numvet)	0.2915	0.0112
Total population of animals (poptotal)	0.2904	0.0115
Number of village animal health workers (numVAHW)	0.2725	0.0180
Border_China	-0.2632	0.0225
Percent of population having access to veterinary care (perpop)	0.2625	0.0229
Total population of cattle (cattle)	0.2417	0.0367
Road density (road density)	0.2373	0.0403
Total population of goats (goatpop)	0.1863	0.1096
Total population of pigs (pigpop)	-0.1538	0.1877
Number of live animal market (No_liveanimal market)	-0.1216	0.2985
Number of animal imports (No_Anim_YrIn)	-0.1066	0.3627
Number of animal exports (No_Anim_YrOut)	-0.0741	0.5272
Total population of sheep (Sheepop)	-0.0594	0.6126
Border_India	0.0508	0.6651

Modelo final (DIC mas bajo)

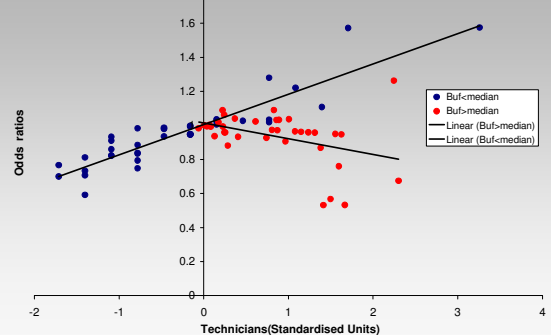
Covariables	OR (IC 95 %)
Población humana	1.1 (0.8-1.7)
Número de búfalos	1.8 (1.2-2.6)
Número de técnicos veterinarios	1.3 (1.0-1.8)
Interacción (búfalos*técnicos veterinarios)	0.6 (0.4-0.8)
DS de efectos no estructurados (U)	0.88 (0.6-1.2)
DS efectos estructurados espaciales (S)	0.04 (0-0.4)

DS S < DS U: Luego de controlar por los efectos fijos, desapareció la evidencia de agrupamiento espacial

Distribucion del riesgo no explicado (U)



The interaction effects of number of technicians and Number of buffaloes for 75 districts.



RESUMEN FINAL

- Tests de agrupamiento espacial
- Modelos de asociación
 - Factor espacial como variable explicativa
 - Factor espacial como determinante de la variable respuesta
 - Factor espacial como efecto aleatorio

Factores de riesgo asociados con evolución viral en la epidemia de fiebre aftosa en Argentina en 2001

A. Perez, G. König, B. Cosentino, E. Späth, M. Thurmond

INTRODUCCION

Antecedentes

1. Eigen
2. Domingo
3. Haydon
4. Yang

INTRODUCCION

Hipótesis

La infección por virus de fiebre aftosa genotípicamente similares resulta en características fenotípicas similares, de tal manera que la estimación de la relación cuantitativa entre la composición genotípica y características fenotípicas del brote permite que al ocurrir un nuevo brote es posible estimar las características genotípicas del virus causal del brote basados en esa relación y las características genotípicas del nuevo brote, aun desconociendo la composición genética del virus causal.

METODOS

Composición genotípica

Distancia genética (número de pares de base diferentes) entre pares de aislamientos

Características fenotípicas

Tasa de ataque del brote (casos clínicos / susceptibles)

Duración del brote (días)

Tiempo entre aislamientos (días)

Distancia entre aislamientos (km)

Dirección sobre el eje norte (km)

Dirección sobre el eje este (km)

METODOS

ID	Características genotípicas							Fen.
	X	Y	D	AR	N	E	T	LG
1	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓	✓	✓
...	...	...	...	...	...	...	...	...
n	✓	✓	✓	✓	✓	✓	✓	✓
n+1	✓	✓	✓	✓	✓	✓	✓	?

METODOS

Modelo mixto de regresión binomial Bayesiana

$$Y_{i,j} = a + b_1 \cdot x_{i,j} + U$$

Y = Distancia genética (VP1) entre pares de aislamientos i y j

x = Variables independientes (características fenotípicas)

1. Distancia (km) entre los aislamientos i y j
2. Tiempo (días) entre los aislamientos i y j
3. Distancia en el eje Norte (km) entre los aislamientos i y j
4. Distancia en el eje Este (km) entre los aislamientos i y j
5. Diferencia en la duración (días) entre los aislamientos i y j
6. Diferencia en la tasa de ataque entre los aislamientos i y j

U = efecto al azar

METODOS

Y	Di	Du	AR	N	E	T
AJUSTE DEL MODELO						
DG _{1,2}	Di _{1,2}	Du ₁ - Du ₂	AR ₁ - AR ₂	DiN _{1,2}	DiE _{1,2}	T _{1,2}
DG _{1,3}	Di _{1,3}	Du ₁ - Du ₃	AR ₁ - AR ₃	DiN _{1,3}	DiE _{1,3}	T _{1,3}
DG _{1,4}	Di _{1,4}	Du ₁ - Du ₄	AR ₁ - AR ₄	DiN _{1,4}	DiE _{1,4}	T _{1,4}
DG _{1,n}	Di _{1,n}	Du ₁ - Du _n	AR ₁ - AR _n	DiN _{1,n}	DiE _{1,n}	T _{1,n}
...	...	...	...	...	...	...
DG _{n-1,n}	Di _{n-1,n}	Du _{n-1} - Du _n	AR _{n-1} - AR _n	DiN _{n-1,n}	DiE _{n-1,n}	T _{n-1,n}
PREDICCIÓN						
DG _{n-1,n}	Di _{n+1,1}	Du _{n+1} - Du ₁	AR _{n+1} - AR ₁	DiN _{n+1,1}	DiE _{n+1,1}	T _{n+1,1}
...	...	...	...	...	...	...
DG _{n-1,n}	Di _{n+1,n}	Du _{n+1} - Du _n	AR _{n+1} - AR _n	DiN _{n+1,n}	DiE _{n+1,n}	T _{n+1,n}

METODOS

Fuente de datos (König et al, 2006)

Ajuste del modelo:

22 aislamientos de 2001

Dos líneas genéticas; número mediano de diferencias nucleotídicas: 6 (rango 0-24)

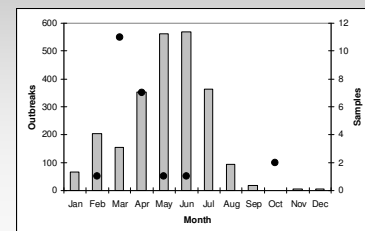
Predicción:

Linaje más probable y distancia genética a los otros 22 aislamientos de 1 aislamiento de 2002.

RESULTADOS

Distribución temporal del muestreo

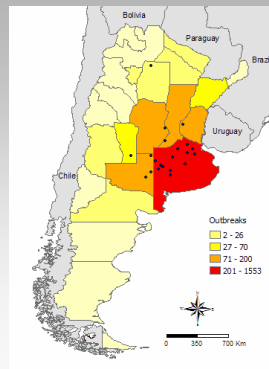
Frecuencia de muestras antes de la mitad de la epidemia (muestras / brotes) vs Frecuencia de muestras después de la mitad de la epidemia (muestras / brotes): $P=0.29$



RESULTADOS

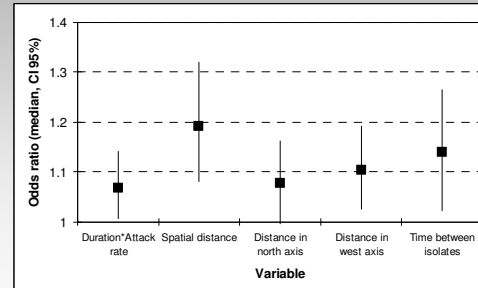
Distribución espacial del muestreo

Frecuencia de muestras en Buenos Aires (muestras / brotes) vs Frecuencia de muestras fuera de Buenos Aires (muestras / brotes): $P=0.85$



Resultados

Modelo que mejor ajusta incluye todas las variables y una interacción (Du*AR)



RESULTADOS

Variable	1 US	ORs	I ^a
N (km)	228.70	1.08 (1.00-1.16)	8%
W (km)	398.47	1.10 (1.03-1.20)	10%
Di (km)	608.89	1.19 (1.08-1.32)	19%
T (días)	123.70	1.14 (1.02-1.26)	14%
AR (casos / susceptibles)	0.34	1.07(1.00-1.14) ^b	7% ^b
Du (días)	16.02		

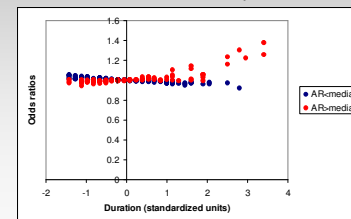
a: Incremento en las chances de 1 sustitución nucleotídica cuando la variable cambia en una unidad estandarizada
b: Valor de la interacción

RESULTADOS

Interacción Du*AR:

Para bajas AR, al aumentar la Du del brote, las chances de sustitución nucleotídica disminuyen levemente

Para altas AR, al aumentar la Du del brote, las chances de sustitución nucleotídica aumentan exponencialmente



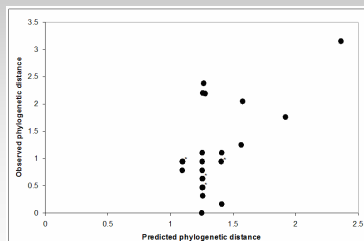
RESULTADOS

Predicción:

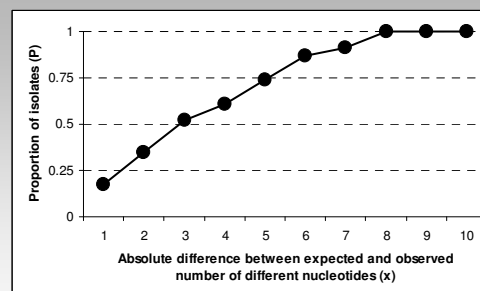
Distancia genética observada vs. predicha ($R=0.6$)

No outliers en pares predicción/observado (Grubb's, $P<0.05$)

Sin diferencias entre predicha y observada (t-test, $P=0.11$)



RESULTADOS



MAPE: 3.8

RESULTADOS

Predicción:

El desvío entre distancia genética observada y predicha fue mayor (Mann-Whitney test, $P < 0.01$) en virus A2001 del subgrupo A (Di media = 1.78) que en los del subgrupo B (Di media = 0.66)

La correlación entre distancia genética observada y predicha para el subgrupo A y B fue $R = -0.1$ y $R = 0.8$, respectivamente.

DISCUSION

El aislamiento de 2002 probablemente perteneció al subgrupo B del virus A2001.

Asociación entre variables fenotípicas y genotípicas

Tiempo

Distancia

Dirección (transmisión local S, E?)

Tasa de ataque y duración (interacción)

Evolución lineal? (log T, sqrt T)

Modelos lineales?