

**MATERIA**

Epidemioloxía, Medicina Preventiva e Policía Sanitaria I

**unidade  
didáctica  
3**

**TITULACIÓN**

Grao en Veterinaria

# Epidemioloxía e Medicina Preventiva. Análise e exposición de datos

**Ceferino M. López Sáñez**

**Sanidade Animal**

**Departamento de Patoloxía Animal**

**Facultade de Veterinaria de Lugo**

**unidadesdidácticas**  
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



**DESCATALOGADO**

Esta obra atópase baixo unha licenza Creative Commons BY-NC-SA 3.0. Calquera forma de reprodución, distribución, comunicación pública ou transformación desta obra non incluída na licenza Creative Commons BY-NC-SA 3.0 só pode ser realizada coa autorización expresa dos titulares, salvo excepción prevista pola lei. Pode acceder Vde. ao texto completo da licenza nesta ligazón:

<http://creativecommons.org/licenses/by-nc-sa/3.0/es/legalcode.gl>

© Universidade de Santiago de Compostela, 2016

**Deseño e maquetación**

J. M. Gairí

**Edita**

Servizo de Publicacións e Intercambio Científico  
da Universidade de Santiago de Compostela  
[usc.es/publicacions](http://usc.es/publicacions)

ISBN

978-84-16533-93-0

DOI

<http://dx.doi.org/10.15304/9788416533930>

**MATERIA: Epidemioloxía, Medicina Preventiva e Policía Sanitaria I**

**TITULACIÓN: Grao en Veterinaria**

PROGRAMA XERAL DO CURSO

Localización da presente unidade didáctica

O programa desta materia abarca 3 bloques temáticos que se correspondería con 3 Unidades didácticas, as prácticas e titorías. Esta Unidade Didáctica cobre a unidade 3 dos contidos, cunha duración de 9 horas.

**UNIDADE 1: Introducción e perspectiva histórica. Conceptos básicos.**

- Tema 1. Desenvolvemento histórico e perspectivas de futuro
- Tema 2. Concepto de poboación. Causalidade
- Tema 3. Bioseguridade. Deseño dos programas sanitarios e produtivos
- Tema 4. Probabilidade. Aleatoriedade. Cálculo da probabilidade

**UNIDADE 2. Mostraxe. Diagnóstico. Medidas epidemiolóxicas. Fontes; tipos de datos e a súa obtención**

- Tema 5. Mostraxe. Determinación do tamaño da mostra
- Tema 6. Tipos de medidas epidemiolóxicas; estandarización das medidas
- Tema 7. Diagnóstico. Validación dunha proba diagnóstica. Sensibilidade. Especificidade
- Tema 8. Deseño de estudos epidemiolóxicos
- Tema 9. Elaboración de cuestionarios

**UNIDADE 3. Análise e exposición de datos**

- Tema 10. Estatística descritiva
- Tema 11. Estatística analítica
- Tema 12. Epidemioloxía estatística. Medidas de asociación e impacto
- Tema 13. Confusión e interacción

## ÍNDICE

---

### PRESENTACIÓN

### OBXECTIVOS

- Obxectivos xerais da materia
- Obxectivos específicos desta Unidade Didáctica,

### OS PRINCIPIOS METODOLÓXICOS

### CLASES TEÓRICAS

### CLASES INTERACTIVAS

### TITORÍAS

### ACTIVIDADES E TRABALLOS

### CONTIDOS BÁSICOS

Tema 10. Estatística descritiva. Conceptos básicos. Distribucións. Presentación dos datos

- 10.1. Estatística descritiva
  - 10.1.1. Medidas que describen unha variable aleatoria
- 10.2. Distribucións estatísticas
  - 10.2.1. Distribución Normal
  - 10.2.2. Outras distribucións estatísticas. Distribucións de probabilidade discretas
  - 10.2.3. Transformación a unha Normalidade aproximada
- 10.3. Presentación dos datos numéricos como gráficos
  - 10.3.1. Gráfico de sectores
  - 10.3.2. Diagrama de caixas
  - 10.3.3. Histograma
  - 10.3.4. Gráfico de columnas ou barras
  - 10.3.5. Gráficos de liñas

Tema 11. Estatística analítica. Introducción. Probas para detectar diferenzas. Probas para detectar correlacións. Probas para detectar asociacións

- 11.1. Estatística analítica: inferencia estatística
  - 11.1.1. Formulación da hipótese (Contraste de hipótese):
  - 11.1.2. Fixar o nivel de significación (establecemento das marxes de erro aceptados)
  - 11.1.3. Seleccionar un método de análise estatística
  - 11.1.4. Tomar unha decisión (acéptase ou rexeita a hipótese nula)

Tema 12. Epidemioloxía estatística. Medidas de asociación e impacto. Estimadores de risco: risco relativo, risco atribuíble, fracción etiolóxica, “odds ratio”

12.1. Medidas de asociación → RISCO

12.2. Medidas de impacto

Tema 13. Concepto de confusión e interacción do efecto. Técnicas de control da confusión

13.1. Confusión (do latín “confundere”, mesturar)

13.1.1. *Control da confusión*

13.2. Interacción

13.2.1. *Avaliación da interacción:*

## **AVALIACIÓN**

## **BIBLIOGRAFÍA**

### **ANEXO 1. Exemplo sobre a Actividade 4**

RESPOSTAS ao anexo 1

### **ANEXO 2. Transformación das variables para achegalas á Normalidade**

### **ANEXO 3. Comandos para realizar as probas estatísticas sinaladas no tema 11 utilizando o paquete estatístico R**

## PRESENTACIÓN

---

Esta materia está incluída no Módulo de Ciencias Clínicas e Sanidade Animal no grao de Veterinaria.

O interese desta materia no grao está asociado á crecente importancia da Epidemioloxía desde un punto de vista profesional. A lexislación Europea é moi restritiva sobre o uso de fármacos en animais, non só para o tratamento etiolóxico das enfermidades infecciosas e parasitarias, senón tamén e sobre todo evitando o uso de promotores de crecemento de orixe antibiótica. A veterinaria clínica ten que evolucionar cara á Medicina Preventiva, o que supón a presenza dun profesional non só Veterinario Clínico, senón tamén Epidemiolóxico.

A porcentaxe de veterinarios especializados en Epidemioloxía ten que aumentar, ben como profesionais libres, contratados por cooperativas ou empresas de maior envergadura ou na Administración Pública. O seu obxectivo será o de controlar enfermidades de importancia económica. En primeiro lugar os brotes que supoñan o peche das fronteiras a exportacións a outros países ou que leven a un sacrificio masivo de animais se chegan a aparecer; ademais, o control das zoonoses. Para o seu control, as persoas máis adecuadas son aquelas que recibisen unha formación epidemiolóxica.

Outro obxectivo moi importante, e cada día máis, é o benestar animal. Neste punto é moi importante un coñecemento profundo da Epidemioloxía e Medicina Preventiva para facer posible unha forma de traballo rendible, pero que cubra tamén a preocupación pola comodidade dentro da produción.

Nesta unidade didáctica descríbense conceptos básicos de matemáticas e estatística que son imprescindibles dentro da Epidemioloxía e Medicina Preventiva, sobre todo centrándonos en que análise se usa e como se interpreta. Así mesmo explícanse as diferentes técnicas de detección da confusión, que pode levar a erro ao epidemiólogo, interpretando asociacións falsas entre variables como verdadeiras, e a maneira de controlalas.

A duración da impartición desta Unidade Didáctica é de 9 horas.

## OBXECTIVOS

---

### Obxectivos xerais da materia

- Obxectivo 1. Entender o concepto poboacional, base dos estudos epidemiolóxicos, e a súa prolongación en espazo e tempo, para poder xerar programas sanitarios destinados á Medicina Preventiva.
- Obxectivo 2. Coñecer os indicadores de saúde e a súa valoración estatística, así como os determinantes de enfermidade e a súa cuantificación como factores de risco nas poboacións.
- Obxectivo 3. Coñecer e entender que mecanismos participan na transmisión e mantemento da enfermidade nunha poboación.

- Obxectivo 4. Coñecer as fontes de datos accesibles ao veterinario para obter información, así como a metodoloxía para realizar estudos observacionais en diferentes situacións ambientais. Ademais, deberá adquirir as habilidades precisas para, non só obter, senón tamén codificar, almacenar e analizar estatisticamente toda a información.
- Obxectivo 5. Coñecer e comprender a achega da epidemioloxía á loita fronte ás enfermidades e as súas posibilidades de modelización e prevención, así como a súa achega en relación á saúde pública.
- Obxectivo 6. Adquirir a terminoloxía adecuada da materia que se imparte.
- Obxectivo 7. Aprender a planificar programas sanitarios e produtivos co fin de rendibilizar as explotacións gandeiras.

#### **Obxectivos específicos desta Unidade Didáctica,**

- Obxectivo 1. Comprender a variabilidade dos seres vivos e a importancia que esta supón na comparación de grupos.
- Obxectivo 2. Coñecer as distribucións estatísticas máis importantes, especialmente a Normal, e a súa importancia na interpretación dos resultados; coñecer as formas que se poden aproximar á distribución Normal outras distribucións.
- Obxectivo 3. Comprender o significado dos erros de inferencia (Tipo I e Tipo II) e interpretar o nivel de significación sobre a hipótese nula.
- Obxectivo 4. Coñecer os métodos estatísticos que permiten separar os efectos causais das variacións aleatorias, presentando as técnicas estatísticas básicas que permiten descubrir a asociación entre enfermidade e os factores estudados.
- Obxectivo 5. Aprender a estimar o risco de padecer unha enfermidade asociado aos factores estudados.
- Obxectivo 6. Coñecer a existencia da confusión entre factores e a importancia que ten controlar esta confusión para coñecer a verdadeira causalidade dunha enfermidade.

O obxectivo específico 1 encádrase no obxectivo xeral 1 da materia Epidemioloxía, Medicina Preventiva e Policía Sanitaria 1; os obxectivo 2 e 3 encádranse no obxectivo xeral 2. O obxectivo 4 relaciónase cos obxectivos xerais 2, 3 e 5. O obxectivo 5 relaciónase cos 5 e 7 xerais e o obxectivo específico 6 relaciónase cos 3 e 7 xerais.

#### **OS PRINCIPIOS METODOLÓXICOS**

---

Este bloque trata conceptos matemáticos e estatísticos básicos na Epidemioloxía; é fundamental comprendelas. Este bloque é moi importante para as clases interactivas 2 e 4 e o Traballo 2. O alumno dispón dos apuntamentos no curso virtual, a partir dos cales terán que realizar de forma autónoma problemas para comprender a forma de aplicar os algoritmos matemáticos que nos ofrecen os paquetes estatísticos, para logo analizar os resultados, que os alumnos terán que explicar. Así mesmo, nas clases interactivas 2 e 4 explícanse as ideas básicas para poder utilizar o paquete estatístico

R e os comandos para as rutinas básicas estatísticas. Na plataforma virtual dispónse de ficheiros cos que pode realizar os exercicios tamén na súa casa ou onde dispoñan dun computador conectado a rede. No caso de que se presente algunha dificultade, están a disposición dos alumnos o foro de dúbidas do curso na plataforma virtual. A maior dificultade atópase en cambiar os costumes que se aplican nas aplicacións gráficas e adaptarse a unha aplicación CLI ("*command line interface*"), que funciona nun terminal baixo comandos. É difícil adaptarse a unha sintaxe ríxida e evitar erros de escritura.

### **CLASES TEÓRICAS**

---

Serán de tipo expositivo, de carácter obrigatorio e impartiranse nas aulas a un grupo completo. A súa duración será aproximadamente de 50 minutos. Nesta Unidade Didáctica son 9 horas.

Darase o material teórico previamente á clase. Na exposición a través de sistemas multimedia introduciranse problemas e preguntas que poñan a atención dos alumnos en casos reais que suceden na vida diaria nas explotacións.

Ao achegarse un profesor á finalización da unidade, realizaranse un control de comprensión dunha duración aproximada a 30 minutos (véxase Anexo 1).

### **CLASES INTERACTIVAS**

---

As clases interactivas 2, 4 e 5 están pensadas sobre a unidade III. Estas prácticas realizaranse nas aulas de informática con grupos de 20 alumnos, cada un cun computador. Nas clases 2 e 4 aprenderán a manexar os datos e realizar gráficos en follas de cálculo; posteriormente introducirán os datos en R, onde aprenderán os comandos básicos para obter gráficos, valores descritivos de las variables e a realizar probas estatísticas do Traballo 2 na clase interactiva 5.

### **TITORÍAS**

---

As titorías daranse na primeira hora das clases interactivas 2 e 3. O seu obxectivo e explicar o traballo que os alumnos terán que realizar para superar os Traballos 1 e 2.

### **ACTIVIDADES E TRABALLOS**

---

Conforme ós principios metodolóxicos expostos, as actividades propostas para o traballos dos contidos desta Unidade Didáctica serían os seguintes:

*Actividade 4:* dirixida a traballar cos temas desta Unidade Didáctica; consistirá nunha proba de comprensión do tipo que se mostra no Anexo 1. Realizarase de xeito individual, nas sesións expositivas e destinarase á súa realización 30 minutos ó rematar a sesión.

*Traballo 2:* farase ó terminar a Unidade 3, na clase interactiva 5, e utilizaranse contidos desta Unidade e da Unidade 1.



## CONTIDOS BÁSICOS

---

### Tema 10. Estatística descritiva. Conceptos básicos. Distribucións. Presentación dos datos

A variabilidade é inherente a todos os datos recollidos nas observacións. As medicións biolóxicas son especialmente susceptibles a variabilidade e na Epidemioloxía ten unha gran importancia en dous aspectos, na representatividade dunha mostra e na comparación entre grupos (factor estudado, factor que confunde, variación natural).

Respecto da *representatividade dunha mostra*, un dos factores que condiciona o tamaño dunha mostra é o grao de dispersión dos datos obtidos na medición das variables estudadas —variabilidade—, como se observa no Tema 5 sobre a mostraxe, dentro do Unidade 2 desta materia.

No caso da *comparación entre grupos*, para estudar os datos biolóxicos necesitanse técnicas estatísticas que separen os efectos dos factores estudados da variabilidade natural e da confusión. Isto leva ao estudo da probabilidade, que xa tratamos no tema 4, dentro do Bloque I.

Para seguir adecuadamente esta Unidade imos describir inicialmente unhas definicións básicas para poder entender os conceptos seguintes:

- **Variable:** suceso observable que pode variar. As variables poden ser cualitativas ou cuantitativas
- **Variable estudo:** calquera variable considerada nunha investigación
- **Variable resposta** (tamén denominada dependente) é aquela que é afectada por outra, independente. O feito de considerar a unha variable dependente e a outra independente depende do sentido da pregunta orixinada no obxectivo da investigación
- **Poboación:** conxunto de todos os individuos que constitúen o obxecto dun determinado estudo e sobre os que se queren obter certas conclusións
- **Mostra:** subconxunto de medidas que constitúe unha parte representativa dunha poboación se non é posible estudar todos os individuos da poboación
- **Parámetro:** cantidade que pode ser distinta en diferentes situacións, pero que é fixa para cada poboación no momento que se está considerando
- **Estatístico:** estimación, máis ou menos precisa, dun parámetro, obtida dunha mostra
- **Serie de datos:** colección de datos
- **Datos brutos:** medidas iniciais que forman a base de análise

#### 10.1. Estatística descritiva

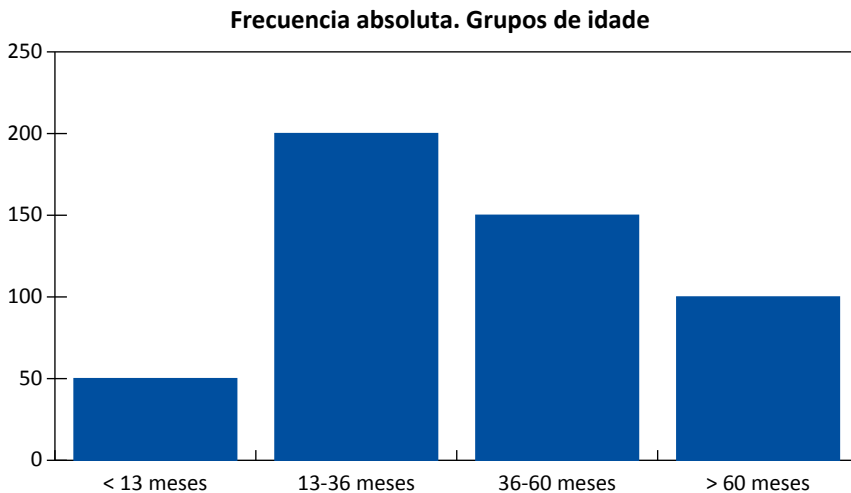
A estatística descritiva pode ser definida como a rama da Estatística que **resume** as observacións da poboación estudada; recolle, ordena e representa un conxunto de datos co fin de describir as características desta poboación. A cuantificación da frecuencia da enfermidade e a descrición da súa distribución no espazo e tempo é a aplicación máis clara da estatística descritiva na Epidemioloxía.

A descrición dos datos adoita comezar mediante **táboas de frecuencias** (Taboa 1) que mostran como se distribúen. A **frecuencia absoluta** (suma de casos/individuos correspondentes a cada categoría) pódese acompañar coa **frecuencia relativa** sobre o total. Pode ser interesante incluír as **frecuencias acumuladas**, que nos permiten comprobar o número ou proporción que representan cada categoría e as inferiores.

**Taboa 1. Táboa de frecuencias. Grupos de idade nun rabaño de cabras**

Idade	Frecuencia absoluta	Frecuencia relativa	Frec. Absoluta acumulada	Frec. relativa acumulada
< 13 meses	50	0,1	50	0,1
13-36 meses	200	0,40	250	0,5
36-60 meses	150	0,30	400	0,80
> 60 meses	100	0,20	500	1
Total	500	1		

Os datos representados na táboa de frecuencias poden resumirse nun **histograma de frecuencias**.



### 10.1.1. Medidas que describen unha variable aleatoria

Unha vez coñecida a distribución de frecuencias das variables en estudo resulta conveniente caracterizar esas variables mediante os estatísticos descritivos, que son as **medidas de posición** (ou tendencia central) e **medidas de dispersión**.

10.1.1.1. Medidas de posición (ou tendencia central)

Indican un valor media ao redor do que se distribúen o resto dos valores. Os estimadores máis utilizados son:

- *Media aritmética*: suma de todos os valores da poboación ou mostra dividida polo número total de individuos. Indícanse como  $\mu$  (media poboacional) ou  $\bar{x}$  (media da mostra).
- *Mediana ( $Q_2$ )*: valor por baixo do cal se atopa a metade das observacións
- *Moda*: valor da ocorrencia máis frecuente

Nunha distribución Normal, media, mediana e moda son iguais.

10.1.1.2. Medidas de dispersión

Miden o grao de variabilidade das observacións (a separación da media).

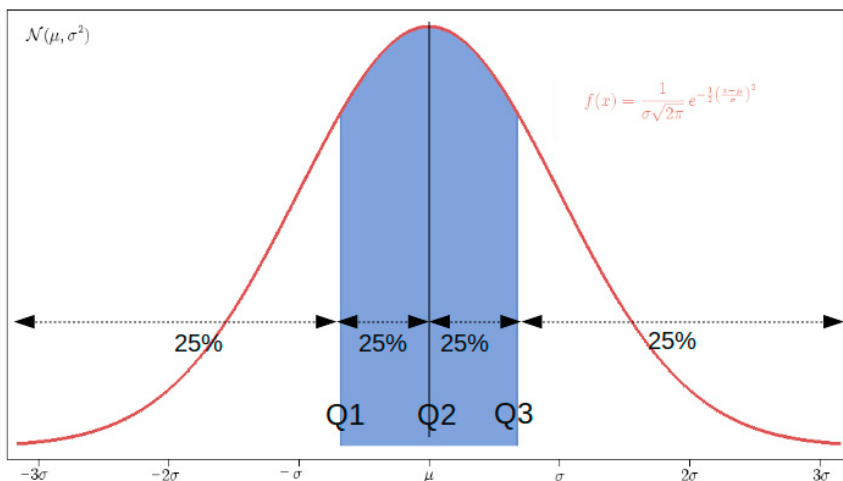
- *Varianza*: media aritmética das diferenzas cuadráticas das observacións con respecto a media. Indícanse de maneira diferente se se trata da varianza poboacional  $-\sigma^2-$  ou varianza da mostra  $-s^2-$ .

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

- *Desviación estándar*: para expresar a medida de dispersión —varianza— na mesma dimensión que as observacións tomamos a súa raíz cadrada. Exprésase como  $\sigma$  —poboacional— ou  $s$  —da mostra—.

$$s = \sqrt{s^2}$$

- *Cuartís*: valores que dividen o conxunto de observacións en catro partes de igual número de casos. Exprésanse como  $Q_1, Q_2$  e  $Q_3$ , sendo o segundo cuartil  $-Q_2-$  a mediana.



[https://es.wikipedia.org/wiki/Cuantil#/media/File:Lqr\\_with\\_quantile.png](https://es.wikipedia.org/wiki/Cuantil#/media/File:Lqr_with_quantile.png)

- *Rango, percorrido ou amplitude*: distancia entre o valor máis alto e o máis baixo das observacións. A súa utilidade é limitada, xa que non distinguen diferentes grupos e vese afectado polos casos extremos

### 10.1.1.3. Estimación de intervalos de confianza. Precisión dunha media

Cada  $\bar{x}$  —media da mostra— é unha estimación puntual da media poboacional  $\mu$ ; cada mostraxe pode dar unha medida diferente. O **erro estándar da media** pódese calcular mediante a división da desviación estándar pola raíz cadrada do número de observacións (tamaño da mostra).

$$EE = \frac{s}{\sqrt{n}}$$

Interésanos coñecer a precisión da estimación realizada sobre a media e obter o rango no que se localiza, para o que calculamos o **intervalo de confianza** utilizando o erro estándar e o valor crítico ( $Z_{\alpha/2}$ ) correspondente ao intervalo desexado, que no caso de 95% é 1,96 e para 99% é 2,576. Os intervalos de confianza dependen da dispersión da poboación, non da media.

- Intervalo de confianza dunha media: rango que contén o verdadeiro valor da media da poboación cunha probabilidade determinada.

$$\text{I.C. 95\%} = \bar{x} \pm \frac{1,96 \times s}{\sqrt{n}}$$

$$\text{I.C. 99\%} = \bar{x} \pm \frac{2,576 \times s}{\sqrt{n}}$$

- Intervalo de confianza dunha proporción: a proporción dunha mostra é tamén unha estimación puntual que varía en cada mostraxe. O rango que contén o verdadeiro valor da proporción calcúlase desde a desviación estándar dunha proporción, que se calcula segundo a raíz cadrada da proporción pola súa complementaria en tantos por un. As fórmulas de cálculo dos intervalos de confianza son:

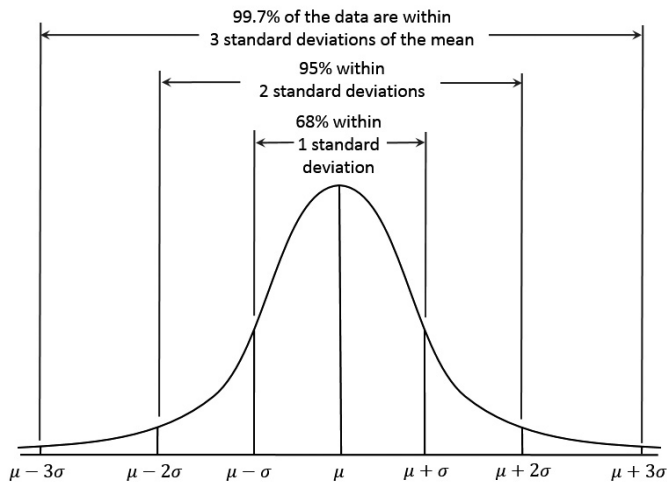
$$\text{I.C. 95\%} = p \pm 1,96 \sqrt{\frac{p \times (1-p)}{n}}$$

$$\text{I.C. 99\%} = p \pm 2,576 \sqrt{\frac{p \times (1-p)}{n}}$$

## 10.2. Distribucións estatísticas

### 10.2.1. Distribución Normal

A distribución Normal, distribución de Gauss ou distribución gaussiana é unha función de densidade de frecuencia acampanada e simétrica que xera unha curva denominada campá de Gauss.



[https://en.wikipedia.org/wiki/Normal\\_distribution#/media/File:Empirical\\_Rule.PNG](https://en.wikipedia.org/wiki/Normal_distribution#/media/File:Empirical_Rule.PNG)

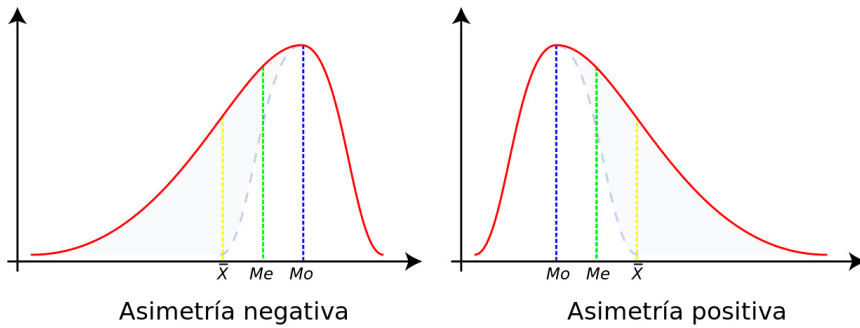
A Normal é a distribución de probabilidade máis importante. Multitude de variables aleatorias continuas, incluídas as variables biolóxicas, seguen unha distribución normal ou se aproximan a ela. Ademais, case calquera distribución de probabilidade, tanto discreta como continua, pódese aproximar por unha Normal baixo certas condicións. Caracterízase por:

- A media aritmética, a mediana e a moda da distribución son iguais e localízanse no centro. Así, a metade da área baixo a curva atópase á dereita deste punto central e a outra metade está á esquerda
- A distribución de probabilidade normal é simétrica ao redor da súa media
- A curva normal descende suavemente en ambas as direccións a partir do valor central. É asíntótica; achégase ao eixo X pero xamais chega a tocalo. É dicir, as “colas” da curva esténdense de maneira indefinida en ambas as direccións
- A distribución Normal queda definida por 2 parámetros, a media e a desviación estándar; é dicir, describe a probabilidade ao redor da media nunha distribución  $N(\mu, \sigma^2)$

**As análises de datos están baseados na normalidade da distribución dos datos!**

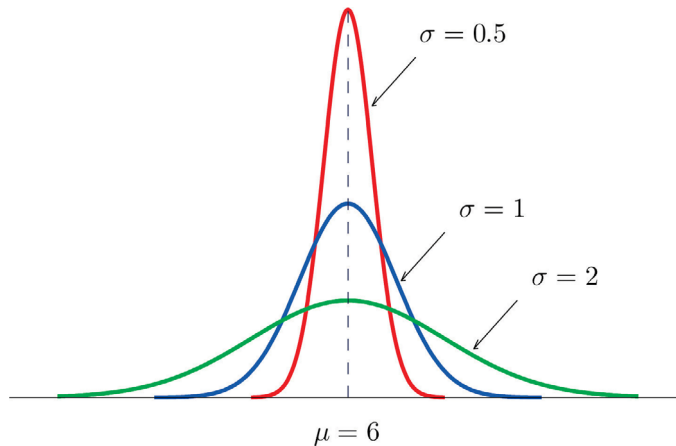
**Asimetría e curtose:** desviacións da normal

A *asimetría* é a medida que indica a simetría da distribución respecto da media aritmética. Os coeficientes de asimetría indican se hai o mesmo número de elementos a esquerda e dereita da media.



[https://commons.wikimedia.org/wiki/File:Posiciones\\_relativas\\_de\\_par%C3%A1metros\\_centrales.svg](https://commons.wikimedia.org/wiki/File:Posiciones_relativas_de_par%C3%A1metros_centrales.svg)

A *curtose*, ou apuntamento, mide a frecuencia de acumulo ao redor da media, sinalando o escarpada ou achatada que está unha distribución. O coeficiente de curtose indica a cantidade de datos que hai preto da media; a maior coeficiente, máis apuntada será a curva —leptocúrtica—. Unha distribución platicúrtica (apuntamento negativo) indica que se recolleron máis valores extremos, xa que nas colas hai máis casos acumulados; unha distribución leptocúrtica —apuntamento positivo— adoita sinalar un “arranxo” de datos, por eliminación de casos extremos.



Imaxe realizada a partir da <http://statwiki.ucdavis.edu/@api/deki/files/72/14fc6536402c36855bba94c6a80f7143.jpg?revision=1>

10.2.2. Outras distribucións estatísticas. Distribucións de probabilidade discretas

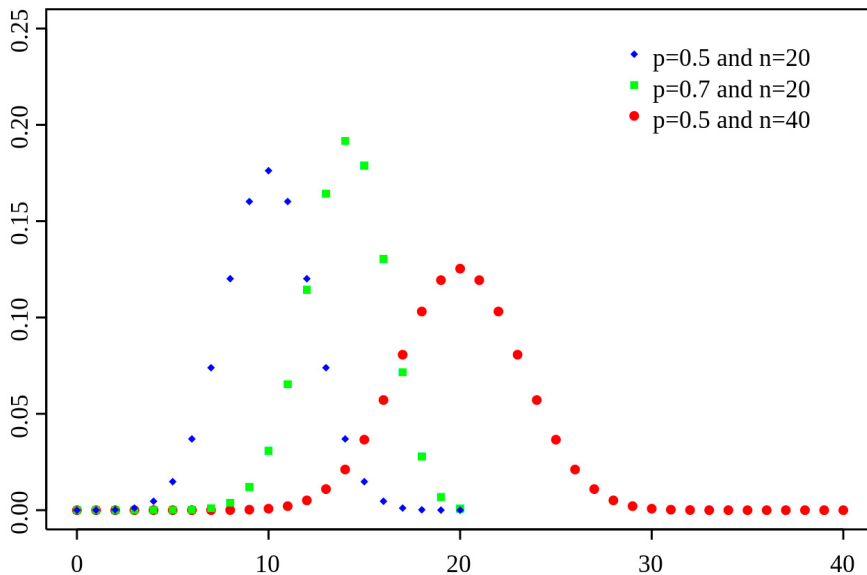
Non todas as variables seguen unha distribución Normal; algunhas das medidas ás que non se lles pode aplicar unha Normalidade son as seguintes:

- Recontos (por exemplo, número de carrachas por animal)
- Datos ordinais que só teñen un pequeno número de intervalos na escala

10.2.2.1. Distribución binomial

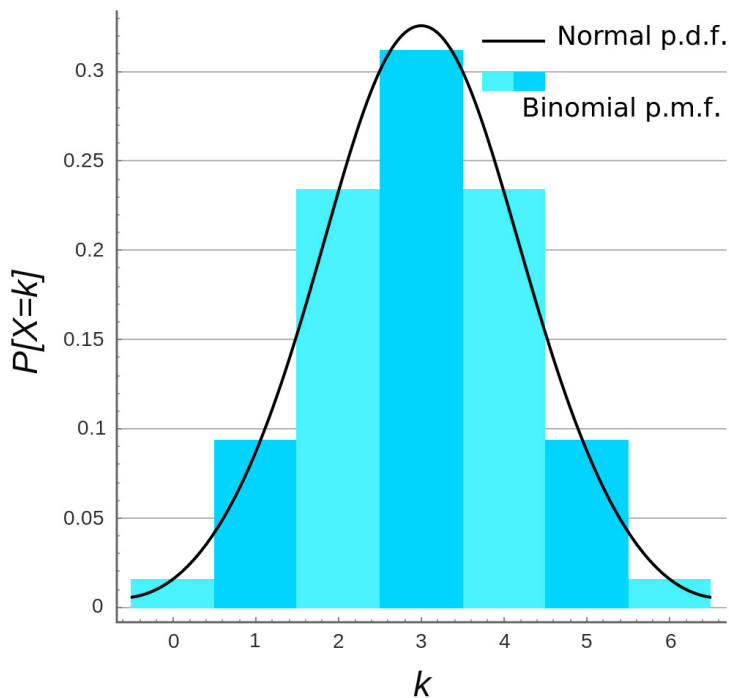
Distribución de probabilidade discreta que conta o número de éxitos nunha secuencia de  $n$  ensaios independentes entre si, cunha probabilidade fixa  $p$  de ocorrencia do éxito entre os ensaios. Representa datos discretos con só dous posibles resultados.

- Clasificación en dous grupos (M/F, vivos/mortos...), mesmo nunha variable medible
- Definida por  $B(n,p)$



[https://upload.wikimedia.org/wikipedia/commons/thumb/7/75/Binomial\\_distribution\\_pmf.svg/1280px-Binomial\\_distribution\\_pmf.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/7/75/Binomial_distribution_pmf.svg/1280px-Binomial_distribution_pmf.svg.png)

As distribucións Normal e binomial son similares pero non idénticas e esixen análises estatísticas diferentes —comparación de frecuencias: test Chi cadrado  $\chi^2$ —. Se  $n$  é suficientemente grande, entón a asimetría da distribución non é demasiado grande. Neste caso unha aproximación razoable a  $B(n,p)$  vén dada pola distribución Normal.



[http://upload.wikimedia.org/wikipedia/commons/thumb/b/b7/Binomial\\_Distribution.svg/1008px-Binomial\\_Distribution.svg.png](http://upload.wikimedia.org/wikipedia/commons/thumb/b/b7/Binomial_Distribution.svg/1008px-Binomial_Distribution.svg.png)

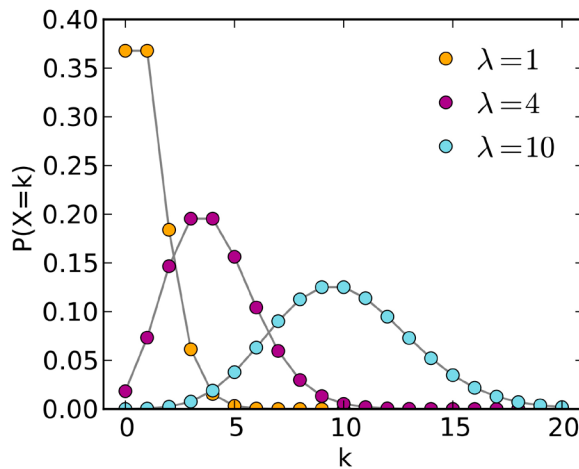
#### 10.2.2.2. Distribución de Poisson

Distribución de probabilidade discreta que expresa, a partir dunha frecuencia de ocorrencia media, a probabilidade de que ocorra un determinado número de eventos durante certo período de tempo ou espazo. Concretamente, é aplicable á ocorrencia de sucesos discretos de que teñen lugar aleatoriamente e con probabilidades baixas na natureza —sucesos “raros”— e a recontos. Por exemplo:

- Células sanguíneas nun hematócrito
- Nº partículas virais que infectan unha célula nun cultivo
- Distribución espacial e temporal dunha enfermidade
- Aparición aleatoria de casos (tempo ou superficie)

A aparición dunha desviación significativa desta distribución indica unha desviación da aleatoriedade. A distribución de Poisson está caracterizada polo parámetro  $\lambda$  (lambda), ou reconto medio por unidade de superficie ou tempo. Para valores suficientemente grandes de  $\lambda$ , unha distribución normal, con media  $\lambda$  e varianza  $\lambda$  (desviación estándar  $\sqrt{\lambda}$ ), é unha excelente aproximación á distribución Normal.





[http://upload.wikimedia.org/wikipedia/commons/thumb/1/16/Poisson\\_pmf.svg/2000px-Poisson\\_pmf.svg.png](http://upload.wikimedia.org/wikipedia/commons/thumb/1/16/Poisson_pmf.svg/2000px-Poisson_pmf.svg.png)

### 10.2.3. Transformación a unha Normalidade aproximada

As escalas naturais de medida poden producir distribucións non Normais e para poder aplicar as probas estatísticas que necesitan a normalidade dos valores analizados existen técnicas de transformación á Normal (véxase Anexo 2).

- Elevando a variable a unha potencia sinxela
- Converténdoa en logaritmos. Existen variables para as cales os logaritmos dos datos ( $\log x + 1$ ) distribúense normalmente. Por exemplo:
  - Curvas epidémicas
  - Períodos de incubación
  - Os casos novos (números discretos) tratados como continuos, xa que son números altos

#### *Aproximación á Normal das distribucións Binomial e de Poisson*

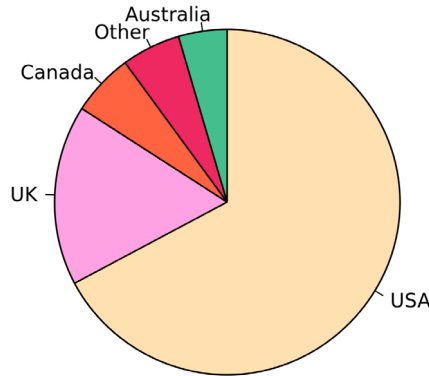
A distribución Binomial pódese aproximar á Normal se a mostra recollida é grande e p non está demasiado próximo a 0 ou 1.

A distribución de Poisson pódese aproximar á normal se as mostran son grandes.

10.3. Presentación dos datos numéricos como gráficos

10.3.1. Gráfico de sectores

Representa porcentaxes ou proporcións

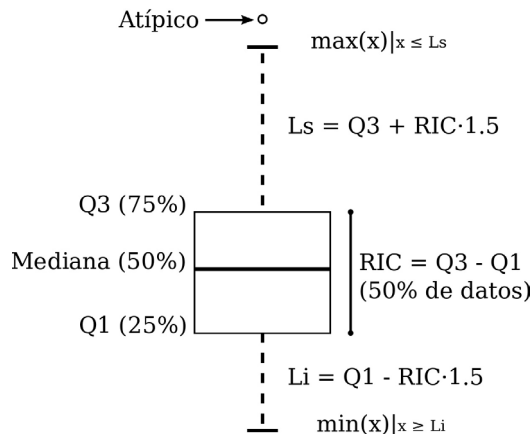


[https://upload.wikimedia.org/wikipedia/commons/thumb/d/db/English\\_dialects1997.svg/1059px-English\\_dialects1997.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/d/db/English_dialects1997.svg/1059px-English_dialects1997.svg.png)

Comando en R: pie()

10.3.2. Diagrama de caixas

Indica a media ou mediana, o rango intercuartil e os valores extremos. É moi útil para comparar variables cuantitativas entre si.

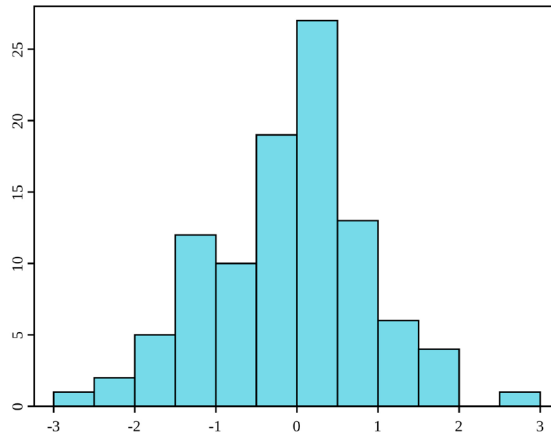


<http://commons.wikimedia.org/wiki/File:Boxplot.svg>

Comando en R: boxplot()

### 10.3.3. Histograma

Gráfico de frecuencias

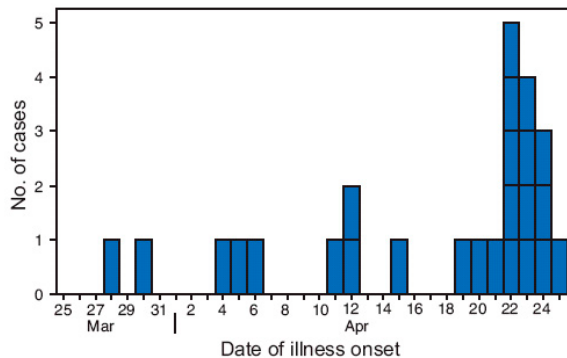


[http://upload.wikimedia.org/wikipedia/commons/thumb/8/8e/Histogram\\_example.svg/2000px-Histogram\\_example.svg.png](http://upload.wikimedia.org/wikipedia/commons/thumb/8/8e/Histogram_example.svg/2000px-Histogram_example.svg.png)

Comando en R: `hist()`, `stem()`

### 10.3.4. Gráfico de columnas ou barras

Forma de representar un conxunto de datos ou valores, conformado por barras rectangulares de lonxitudes proporcionais aos valores representados. Estes gráficos son usados para comparar dous ou máis valores ou comparar un valor en diferentes categorías. Poden orientarse vertical —columnas— ou horizontalmente —barras—.



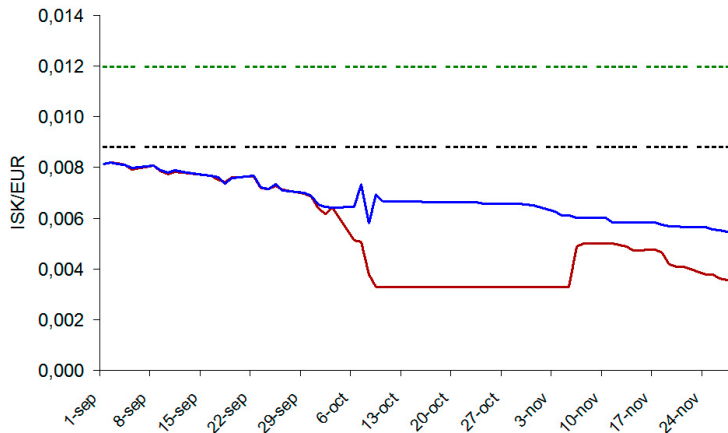
[https://upload.wikimedia.org/wikipedia/commons/a/a2/CDC\\_MMWR\\_m8d0428a2f.png](https://upload.wikimedia.org/wikipedia/commons/a/a2/CDC_MMWR_m8d0428a2f.png)

Comando en R: `barplot()`

### 10.3.5. Gráficos de liñas

Representan variables (eixo y) ao longo do tempo (eixo x). Poden utilizarse para o control de niveis indesexables na produción como indicadores de enfermidade, segundo uns niveis aceptables e un nivel inaceptable ao que non se pode chegar. A limitación deste sistema de control está en dous niveis:

- Que variables hai que controlar
- A variabilidade natural nos seres vivos



[http://upload.wikimedia.org/wikipedia/commons/f/f5/ISK\\_exchange\\_rate\\_Sep-Nov\\_2008.png](http://upload.wikimedia.org/wikipedia/commons/f/f5/ISK_exchange_rate_Sep-Nov_2008.png)

Comando en R: plot()

## Tema 11. Estatística analítica. Introducción. Probas para detectar diferenzas. Probas para detectar correlacións. Probas para detectar asociacións

### 11.1. Estatística analítica: inferencia estatística

O obxectivo da estatística analítica é detectar a diferenza, asociación, relación ou dependencia entre 2 ou máis variables de estudo. Os pasos a seguir dentro dunha análise estatística son:

- Formulación dunha hipótese de traballo
- Fixar o nivel de significación
- Seleccionar un método de análise estatística
- Tomar unha decisión

Nunca debemos esquecer dous feitos moi importantes:

- *Unha asociación estatística NON é o único criterio de causalidade.* Para probar unha inferencia causal —atribuír a presenza dunha enfermidade a un ou máis factores de risco— é necesario que se cumpran de forma consecutiva as tres propiedades causais esenciais descritas nos criterios

de Susser (1977-1995), asociación, secuencia temporal e dirección causal (Unidade Didáctica I, páxina 14).

- Diferenzas estatisticamente significativas *poden non ser biolóxica, clínica ou patolóxicamente significativas.*

Pasos da análise estatística:

11.1.1. Formulación da hipótese (Contraste de hipótese):

Enfrontamos dúas hipótese mutuamente excluíntes, a hipótese nula e a alternativa. A **hipótese nula** ( $H_0$ ) é a hipótese de traballo a rexeitar e na que se asume que non existen diferenzas, asociacións ou relacións entre as variables analizadas. A **hipótese alternativa** ( $H_1$ ) é aquela na que existe algunha diferenza, asociación ou relación.

Segundo as variables analizadas formúlense diferentes hipóteses nulas:

- Se enfrontamos dúas variables cuantitativas, o valor dunha non depende da outra
- Se unha variable é cuantitativa e outra cualitativa, non existen diferenzas nas medias dos grupos definidos pola cualitativa (uni ou bilaterais)
- Se son dúas variables cualitativas, a distribución das frecuencias observadas non difire das esperadas

11.1.2. Fixar o nivel de significación (establecemento das marxes de erro aceptados)

Fixase un *nivel de significación* ( $\alpha$  ou erro tipo I), que é a probabilidade de rexeitar a hipótese nula sendo certa. O nivel de significación especifícase a priori (0,05 ou 0,01).

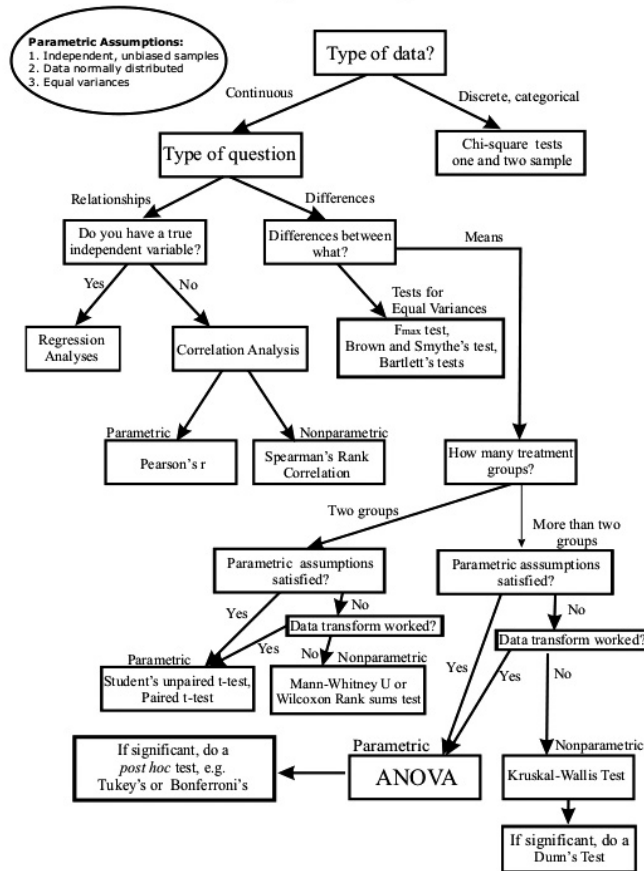
Existe un *segundo erro*, denominado  $\beta$  ou erro tipo II, que é a probabilidade de aceptar a hipótese nula sendo falsa. Xéranse por tanto outros dous conceptos, o *nivel de confianza*, ou probabilidade de aceptar  $H_0$  sendo certa, e que é o complementario do nivel de significación ( $1-\alpha$ ; 95 ou 99%) e a *potencia do contraste*, ou probabilidade de rexeitar  $H_0$  sendo falsa, que é o complementario do erro de tipo II ( $1-\beta$ ). A tamaño de mostra fixo, ao aumentar  $\alpha$ , diminúe  $\beta$

DECISIÓN	ESTADO REAL	
	$H_0$ certa	$H_0$ falsa
Aceptar $H_0$	Nivel de confianza ( $1-\alpha$ )	Erro $\beta$ ou tipo II
Rexeitar $H_0$ (= aceptar $H_1$ )	Erro $\alpha$ ou tipo I	Potencia ( $1-\beta$ )

11.1.3. Seleccionar un método de análise estatística

En función da escala de medición utilizada nas variables analizadas selecciónase a **proba de contraste estatístico** máis adecuada, que calculará o **estatístico de contraste**. Podemos basearnos en táboas orientadoras como esta:

**Flow Chart for Selecting Commonly Used Statistical Tests**

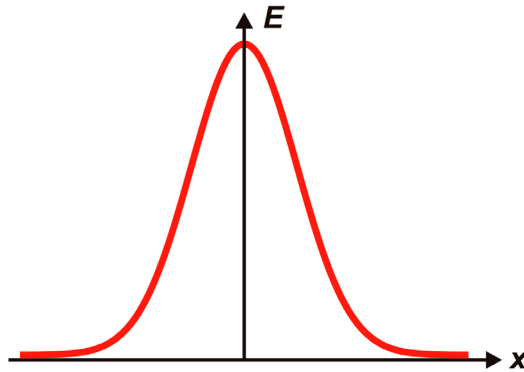


<http://abacus.bates.edu/~ganderso/biology/resources/statistics.html>

As probas estatísticas utilizadas para demostrar a asociación poden ser paramétricas ou non paramétricas. As probas paramétricas reciben ese nome porque estiman parámetros da poboación estudada. Ademais, débese cumprir a homocedasticidade das variables en estudo para que poidan ser analizadas mediante probas paramétricas; isto supón que:

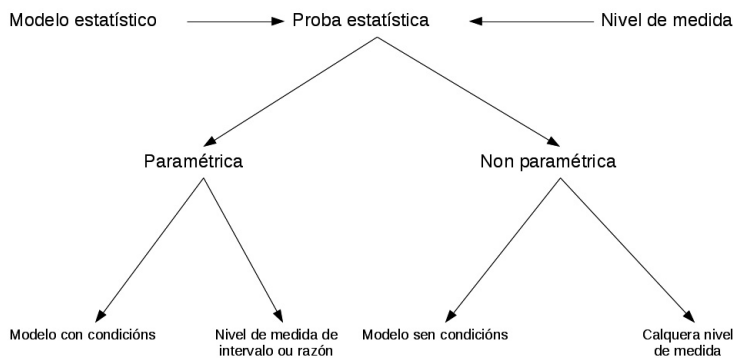
- Polo menos unha das variables ten unha escala cuantitativa (discreta ou continua)

- Os valores obtivéronse dunha poboación normalmente distribuída



[https://commons.wikimedia.org/wiki/File:Gaussian\\_distribution\\_thick\\_lines.svg](https://commons.wikimedia.org/wiki/File:Gaussian_distribution_thick_lines.svg)

- As varianzas dos grupos en análises deben ser homoxéneas  
Se as variables non cumpren estas condicións pódense transformar para achegarse á normalidade (Véxase Anexo 2) ou se utilizan probas non paramétricas, que non presupoñen ningún suposto previo para ser aplicadas.



A mesmo tamaño de mostra as probas paramétricas presentan maior potencia de análise; con todo, as probas non paramétricas pódense aplicar en calquera situación. Non debemos esquecer que en Epidemioloxía a estatística aplícase fundamentalmente para descubrir a CAUSALIDADE.

*As probas básicas que podemos aplicar son as seguintes:*

- *Dúas variables cuantitativas non emparelladas*

Proba paramétrica	Coefficiente de correlación lineal de Pearson; regresión
Proba non paramétrica	Coefficiente de correlación lineal de Spearman

— *Dúas variables cuantitativas emparelladas*

Proba paramétrica t de Student para mostras dependentes ou emparelladas  
Proba non paramétrica rangos emparellados e sinalados de Wilcoxon

— *Unha variable cuantitativa e unha cualitativa dicotómica*

Proba paramétrica t de Student para mostras non emparelladas  
Proba non paramétrica proba U de Mann Whitney; suma de rangos de Wilcoxon

— *Unha variable cuantitativa e unha cualitativa*

Proba paramétrica Análise da varianza (ANOVA)  
Proba non paramétrica proba de Kruskal-Wallis

— *Dúas variables cualitativas*

Proba non paramétrica Chi-cadrado  
Proba exacta de Fisher (frecuencias esperadas baixas)

## OUTRAS PROBAS ESTADÍSTICAS

— *Métodos de dependencia* - causalidade

Análise discriminante, regresión loxística, correlacións canónicas, análises de ecuacións estruturais...

— *Métodos de interdependencia* – redución de dimensión

Compoñentes principais, análise factorial, “cluster”...

— *Evolución temporal*

Análise de supervivencia, series temporais...

Os comandos que se aplicarán na práctica 2 e 4 para realizar estas probas utilizando o paquete estatístico R expónse no Anexo 3.

## 11.1.4. Tomar unha decisión (acéptase ou rexeita a hipótese nula)

O estatístico de contraste e os graos de liberdade permítenos calcular o valor de significación (p valor) que hoxe se obtén directamente nos programas estatísticos e que se compara co nivel de confianza. Con valores iguais ou inferiores ao nivel de significación establecido —xeralmente en ciencias biolóxicas e de saúde 0,05— rexeitamos a hipótese nula, xa que a probabilidade de que sexa certa é moi baixa. Con valores de p superiores, aceptamos a hipótese nula

Unha asociación (significativa) da presenza da enfermidade co/os factores estudados pode supor que eses sexan causais, se se demostra ademais a secuencia temporal e a dirección causal.



**Tema 12. Epidemioloxía estatística. Medidas de asociación e impacto. Estimadores de risco: risco relativo, risco atribuíble, fracción etiolóxica, “odds ratio”**

**12.1. Medidas de asociación → RISCO**

A detección dunha asociación significativa entre a presenza dunha enfermidade e un factor estudado pode non ser suficiente para coñecer o risco que xera, é dicir, é necesaria unha medida da diferenza na frecuencia de presentación dunha enfermidade debida a ese factor.

Por tanto, definimos **risco** como o grao en que un factor causa unha enfermidade —grao de asociación coa enfermidade—. O **risco absoluto** é un sinónimo de incidencia, ou ritmo de aparición da enfermidade, e é a taxa da que se deduce o risco relativo e o atribuíble.

- Risco relativo – razón da taxa de incidencia entre os expostos e a taxa de incidencia dos non expostos ao factor estudado.

	Expostos ao factor	Non expostos ao factor
Presenza da enfermidade	a	c
Ausencia da enfermidade	b	d

$$RR = \frac{\text{Taxa de incidencia nos expostos}}{\text{Taxa de incidencia nos non expostos}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a \cdot (c+d)}{c \cdot (a+b)}$$

Nalgúns casos —estudos caso-control— non se pode extraer o risco relativo, xa que non se coñece a incidencia da poboación afectada; obtense unha estimación mediante o “**ODDS RATIO**” ou **OR**, chamado tamén razón de vantaxes ou cociente de produtos cruzados.

$$OR = \frac{\text{Razón de vantaxe nos expostos}}{\text{Razón de vantaxe nos non expostos}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \cdot d}{c \cdot b}$$

**ODDS RATIO (OR)** - Estimador do risco relativo

- Cálculo do intervalo de confianza sobre o OR mediante aproximación logarítmica.

O Intervalo de confianza obtense do erro estándar, que para un OR se calcula

$$EE(LnOR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Os operadores para o nivel superior e inferior extráese da operación

$$LnOR \pm Z * EE(LnOR)$$

sendo o valor crítico Z para o nivel de confianza 95% 1,96 e para 99% 2,576, do mesmo xeito que se indicou no tema 10.

O nivel superior e inferior obteranse mediante a exponenciación deses valores

$$e^{LnOR \pm Z \cdot EE(LnOR)}$$

**Exemplo:** Infección por protostrongídeos en ovino en función da infección por *Dictyocaulus filaria* (factor de risco; López e col. 2011. *Vet. Parasitol*, **178**: 108-114).

	Expostos a <i>D. filaria</i>	Non expostos a <i>D. filaria</i>	Total
Infectados por protos.	55	187	242
Non infectados por protos.	168	1683	1851
Total	223	1870	2093

$$RR = (55/223)/(187/1870) = 2,47$$

ou a súa estimación

$$OR = (55/168)/(187/1683) = 2,95$$

- Risco relativo > 1 asociación estatística positiva entre o factor e a enfermidade
- Risco relativo < 1 asociación negativa (efecto protector)
- Risco relativo = 1 non existe asociación entre o factor e a enfermidade

O Risco relativo ou o OR nou mide a probabilidade, se non a fortaleza da asociación do factor e o efecto (enfermidade). Para coñecer a probabilidade, é dicir, a significación estatística do resultado do OR, aplícase un test Chi-cadrado ou constrúese un intervalo de confianza para comprobar a súa significación/validez.

### 12.2. Medidas de impacto

— Risco atribuible ou taxa atribuible (denominado coa letra grega delta -  $\delta$ )  
Mide a parte do risco absoluto —ou incidencia— atribuible a un factor e presenta o beneficio potencial esperado se se reduce a exposición ao factor.

$$\delta = \text{taxa de incidencia da expostos} - \text{taxa de incidencia do non expostos}$$

No caso que estamos a dar de exemplo:

$$\delta = 55/223 - 187/1870 = 0,247 - 0,1 = 0,147$$

**Fracción etiolóxica** ou proporción atribuible ( $\lambda$  - lambda)

Proporción atribuible do total da frecuencia da enfermidade —risco absoluto ou incidencia— á exposición a un factor causal determinado

$$\lambda = (OR - 1) / OR$$

No exemplo que estamos a utilizar:

$$\lambda = (2,95 - 1) / 2,95 = 0,661$$

As dúas medidas indican o impacto da eliminación dun factor causal sobre a redución da incidencia.

Hai que ter en conta:

- Os estimadores de risco son enunciados de PROBABILIDADE
- Non todos os individuos expostos ao factor presentan a enfermidade
- A enfermidade aparecerá tamén nalgúns individuos non expostos

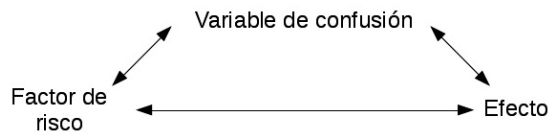
### Tema 13. Concepto de confusión e interacción do efecto. Técnicas de control da confusión

#### 13.1. Confusión (do latín “confundere”, mesturar)

Situación na que a medición do efecto dunha exposición sobre un risco al térase debido á asociación con outros factores, dando lugar a unha asociación aparente entre as variables, sendo imposible separar de forma lóxica a contribución de calquera factor illado sobre un efecto.

A *variable de confusión* ou equívoco, para selo, debe cumprir:

- Está relacionada coa enfermidade,
- Mostra asociación co factor ou variable estudado,
- Pero non é un paso intermedio entre o factor e a enfermidade; non é un efecto causal.



##### 13.1.1. Control da confusión

Existen diferentes formas de controlar a confusión:

#### 1. Exclusión

Consiste nunha mostraxe restritiva, escollendo unha soa categoría ou nivel da variable de confusión, co que se elimina o seu efecto. Esta forma de controlar a confusión pode facer que a mostraxe non sexa representativa da poboación en estudo.

#### 2. Apareamento, tamén denominado “matching”

Escóllense primeiro o grupo sometido ao efecto e logo o grupo control restrinxindo para que presenten os mesmos niveis de variables de confusión que os expostos.

#### 3. Análise dos datos observacionais

Utilízase a técnica de Mantel-Haenszel. Calcúlase o OR cru de todos os casos, e logo calcúlanse os OR axustados a cada categoría da variable de confusión. Se a diferenza entre o 1º e o 2º é grande, podemos afirmar que existe confusión.

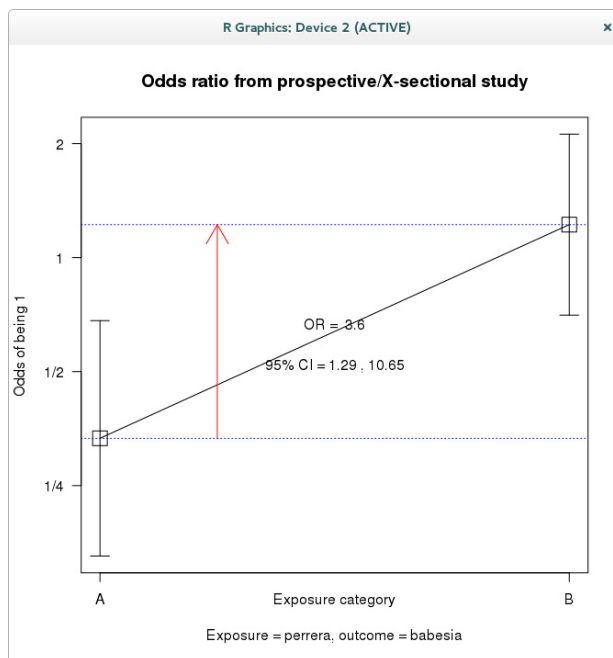
Exemplo con datos teóricos, non baseados nun caso real:  
Analizamos dúas canceiras —Canceira A e B— localizadas na mesma cidade, buscando animais infectados por *Babesia* e atopamos os seguintes resultados.

	Canceira A	Canceira B	Total
Negativos a <i>Babesia</i>	30	18	48
Positivos a <i>Babesia</i>	10	22	32
Total	40	40	80

O OR cru sería

$$\text{OR cru} = 30 \times 22 / 18 \times 10 = 3,66 \text{ (IC 95\% 1,29-10,65)}$$

A súa significación demóstrase polo intervalo de confianza, que **NON CONTÉN 1** e polo resultado da proba Chi-cadrado ( $\chi^2 = 7,5$ , g.l. = 1,  $P = 0,006$ ).



Poderíase pensar que a canceira utilizada pode ser un factor de risco, sendo maior a posibilidade de que enfermen os animais se se atopan na canceira B. Se as condicións son similares, podería deberse a unha confusión. Por exemplo, as canceiras poden estar en zonas diferentes, e ter ao seu cargo animais de diferente

orixe. Supoñamos que analizamos o tipo de animais que son e para que os utilizan os seus donos; por exemplo, clasificando a eses mesmos animais en función de se son de caza ou de compañía:

Aptitude	Compañía	Caza	Total
Negativos a <i>Babesia</i>	36	12	48
Positivos a <i>Babesia</i>	4	28	32
Total	40	40	80

Neste caso o OR cru sería

$$\text{OR cru} = 36 \times 28 / 4 \times 12 = 21 \text{ (IC 95\% 5,51-94,77)}$$

Serán ambos os factores, Canceira e Aptitude, factores de risco ou nos atopamos nun caso de **CONFUSIÓN?**.

Aplicaremos unha análise Mantel-Haenszel, que supón a segmentación do efecto entre os diferentes factores. Se o segmentamos a táboa resultante sería:

Canceira/Aptitude	Caza sans	Caza enfermos	Compañía sans	Compañía enfermos
<i>Canceira A</i>	3	7	27	3
<i>Canceira B</i>	9	21	9	1
Total	OR = 1		OR = 1	

Unha vez segmentados non existe risco asociado á Canceira. Neste caso estamos nun caso de confusión, sendo a verdadeira causa a aptitude, e a variable de confusión a canceira; mostra unha asociación aparente —OR = 3,66— mais non é un risco para a enfermidade.

$$\text{OR cru} = 3,66 \quad \text{OR}_{\text{MH}} \text{ ou axustados} = 1$$

### 13.2. Interacción

Nalgúns casos prodúcese **interacción** entre diferentes factores estudados, é dicir, prodúcese unha modificación do efecto de exposición nos distintos niveis de estratificación. A interacción Indica un efecto causal das variables estudadas e débese describir ou explicar.

#### 13.2.1. Avaliación da interacción:

Calcular o OR de cada estrato

Se existe diferenza → interacción (describir)

## AVALIACIÓN

---

Realizarase unha avaliación continua sobre os alumnos que se realizará da seguinte maneira:

### Actividades

Ao longo das exposicións dos temas efectuaranse un pequeno control aleatorio da unidade 3 (Actividade 4). Aplicaranse no momento en que un profesor termine de dar o bloque conceptual. Este control, como os outros efectuados ao longo do semestre, ten a función de controlar a asistencia dos alumnos e valorar a súa aprendizaxe progresiva. Será do tipo que se mostran no Anexo 1 e buscarán posibles solucións ante situacións asociadas aos temas dados.

Valor individual até 0,5 puntos. Total da unidade 3, 0,5 puntos.

A avaliación xeral inclúe tamén outras 3 actividades que se realizan no Bloque 1 (Actividade 1 e 2) e no Bloque 2 (Actividade 3), todas elas co mesmo valor individual de 0,5 puntos, para un total de 2 puntos polas 4 Actividades.

### Traballos na aula

Tras as horas de tutoría, nas clases interactivas 2 e 3, presentaranse ao alumnado a través da plataforma virtual un boletín de problemas similar a os que terán que realizar (Traballo 1 e 2) nas prácticas 3 e 5.

Para realizar estes boletíns forneceráselles aos alumnos un conxunto de datos relacionados con algunha enfermidade ou alteración dunha poboación animal da que os alumnos terán que descubrir a causalidade, detectar que factores están a influír, demostralo estatisticamente e presentar un informe coas recomendacións adecuadas para solucionar o problema.

O valor será de 2 puntos para o Traballo 1 e 6 para o Traballo 2. Total dos traballos, 8 puntos, nos que esta Unidade Didáctica non terá unha cualificación individualizada.

Os alumnos que mediante os controis aleatorios e boletíns superasen os 5 puntos, superarían a materia sen necesidade de asistir ao exame teórico.

### Exame final

Os alumnos que non puidesen superar a materia a través dos controis e boletíns poderán realizar un exame similar a suma dos Traballos 1 e 2 ao final do curso. O seu valor, até 10 puntos, servirá para facer media co valor dos controis e boletíns. Nel exporanse cuestións que teñen que ver co contido traballado nesta Unidade.

## BIBLIOGRAFÍA

---

Na bibliografía específica desta unidade podemos destacar da bibliografía xeral os seguintes títulos:

BARNETT, V.; FERIDUN TURKMAN, K. (1995). Statistics for the environment. John Wiley & Sons, Chichester, UK.

- BERTHOLD, M.R.; BORGELT, C.; HÖPPNER, F.; KLAWONN, F. (2010). *Guide to Intelligent Data Analysis*. Springer-Verlag London Limited.
- BORCARD, D.; GILLET, F. Y LEGENDRE, P. (2011). *Numerical Ecology with R*. Springer Science and Business Media, LLC, New York, USA.
- CAO ABAD, R. y col. (2001). *Introducción a la estadística y sus aplicaciones*. Ediciones Pirámide, Madrid.
- CONTRERAS, A; SÁNCHEZ, A.; CORRALES, J.C. (2001). *Epidemiología veterinaria*. ICE. Universidad de Murcia.
- COOK, D. Y SWAYNE, D.F. (2007). *Interactive and Dynamic Graphics for Data Analysis with R and Ggobi*. Springer Science and Business Media, LLC, New York, USA.
- CRESSIE, N.A.C. (1993). *Statistics for spatial data*. John Wiley & Sons, Inc. New York.
- DE BLAS, I.; RUIZ-ZARZUELA, I.; BAYOT, B.; FERREIRA, C. (2007). *Manual de Epidemiología Veterinaria*. Universidad de Zaragoza.
- DOHOO, I.; MARTIN; W., STRYHN, H. (2009). *Veterinary Epidemiologic Research*. Segunda edición. VER Inc., Charlottetown, Canada.
- DURR, P.; GATRELL, A. (2004). *GIS and spatial analysis in veterinary science*. CAB International, Wallingford, Oxfordshire, UK.
- GARDENER, M. (2012). *Statistics for Ecologists Using R and Excel. Data Collection, Exploration Analysis and Presentation*. Pelagic Publishing, Exeter.
- (2014). *Community Ecology. Analytical Methods Using E and Excel*. Pelagic Publishing, Exeter.
- GORDIS, L. (2005). *Epidemiología*. Tercera edición. Elsevier España, Madrid.
- GRAFEN, A.; HAILS, R. (2002). *Modern Statistics for the life sciences*. Oxford University Press, Oxford.
- KAHN, H. y SEMPOS, C. (1989). *Statistical methods in epidemiology*. University Press, Oxford.
- LUQUE MARTÍNEZ, T. Ed. (2000). *Técnicas de análisis de datos en investigación de mercados*. Ediciones Pirámide (Grupo Anaya, S.A.), Madrid.
- MORTON, E.R.; HEBEL, J.R.; McCARTER, R.J. (1993). *Bioestadística y epidemiología*. 2ª edición. McGraw-Hill, México, D.F.
- PEÑA, D. (1991). *Estadística. Modelos y métodos*. 1. Fundamentos. Alianza Editorial, Madrid.
- PETRIE, A.; WATSON, P. (1999). *Statistics for veterinary and animal science*. Blackwell Science, Ltd., Oxford.
- PFEIFFER, D.U. (2008). *Spatial Analysis in Epidemiology*. Oxford University Press
- ROTHMAN, K.J. (1987). *Epidemiología moderna*. Díaz de Santos, Madrid.
- (2012). *Epidemiology. An introduction*. Second edition. Oxford University Press, New York.
- , GREENLAND, S.; LASH, T.L. (2008). *Modern Epidemiology*. Third edition. Lippincott Williams & Wilkins, Philadelphia, USA.
- SNEDECOR, G.W.; COCHRAN, W.G. (1980). *Statistical Methods*. The Iowa State University Press, Ames, U.S.A.
- STEVENS, M.Henry H. (2009). *A Primer of Ecology with R*. Springer Science and Business Media, LLC, New York, USA.

- SOKAL, R.R.; ROHLF, F.J. (1984). Introducción a la bioestadística. Editorial Reverte, S.A., Barcelona.
- THRUSFIELD, M. (1990). Epidemiología Veterinaria. ACRIBIA, Zaragoza.
- (2005). Veterinary Epidemiology. Third Edition. Blackwell Scientific Publications Ltd., Oxford.
- VIEDMA, J.A. (1976). Bioestadística. Métodos estadísticos en medicina y biología. Gráfica Internacional, Madrid.

**Páxinas Web**

<http://aeema.vet-alfort.fr/>  
[http://es.wikipedia.org/wiki/Factor\\_de\\_confusi%C3%B3n](http://es.wikipedia.org/wiki/Factor_de_confusi%C3%B3n)  
<http://es.wikipedia.org/wiki/Sesgo>  
<http://www.oie.int>  
<http://davidmlane.com/hyperstat/index.html>  
<http://ordination.okstate.edu/>  
<http://www.cdc.gov/epo/>  
<http://www.math.yorku.ca/SCS/StatResource.html>  
<http://www.provalisresearch.com/>  
<http://www.sas.com/technologies/analytics/statistics/index.html>  
<http://www.sportsci.org/resource/stats/index.html>  
<http://www-01.ibm.com/software/analytics/spss/>  
<http://www.statgraphics.com/>  
<http://www.statpages.org/>  
<http://www.statsci.org/free.html>  
<http://www.statsoft.com/textbook/stathome.html>  
<http://www.systat.com/>  
<http://www.stata.com/>  
<http://www.statsoft.com/Products/STATISTICA/Product-Index>

**Páxinas Web (R-cran)**

<http://cran.r-project.org/>  
<http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>  
[http://cran.r-project.org/doc/contrib/Paradis-rdebuts\\_en.pdf](http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf)  
<http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>  
<http://cran.r-project.org/faqs.html>  
<http://cran.r-project.org/doc/manuals/R-intro.pdf>  
<http://cran.r-project.org/doc/manuals/R-data.pdf>  
<http://cran.r-project.org/doc/manuals/R-admin.pdf>



**ANEXO 1. EXEMPLO SOBRE A ACTIVIDADE 4**

Nome: ..... DNI: .....

Nunha granxa de gando ovino de carne as ovellas empezan a presentar signos clínicos como tose, dificultades respiratorias, mucosidade líquida no fociño... Tras a chamada por parte do gandeiro o veterinario examina aos animais e toma mostras de sangue, soro e feces de 20 animais da granxa, 15 afectados e 5 aparentemente sans sen sintomatoloxía e envíaos ao laboratorio INVESAGA. O laboratorio envía os resultados de sangue (hemograma, análise bioquímica...) e feces. Este último indica a presenza das seguintes formas parasitarias nas feces observadas; ovos de nematodos gastrointestinais en 18 animais (media de eliminación 150 ovos por gramo de feces), larvas de *Dictyocaulus filaria* en 17 dos animais (media de eliminación, 10 l.p.g.) e larvas de nematodos pulmonares protostronxílicos en 16 dos animais (media de eliminación 230 l.p.g.). Ao mesmo tempo, o veterinario observa que os animais de máis idade presentan unha sintomatoloxía máis marcada e, segundo os datos do laboratorio, están máis parasitados.

Nos últimos estudos realizados sobre a causalidade das pneumonías por protostronxílicos viuse que os factores de risco que inflúen máis para a aparición da infección nesta rexión e nesas condicións de explotación son, por orde de importancia: 1. A infección previa por *Dictyocaulus*; 2. A idade dos animais; 3. A introdución de femias doutras explotacións; 4. Manexo conxunto con cabras; 5. Tratamento exclusivo con benzimidazois, fronte á maior eficacia das lactonas macrocíclicas

Un exemplo dos datos utilizados para este estudo é esta táboa, en relación ao efecto da idade na infección por nematodos protostronxílicos:

	Idade en meses		TOTAL
	<49 meses	>48 meses	
Protostronxílicos			
Negativo	1015	836	1851
Positivo	82	160	242
TOTAL	1097	996	2093

Mediante unha proba Chi cadrado obtivo

$$\chi^2 = 36,829, \text{ g.l.} = 1, P = 1,289 \times 10^{-9}$$

Pregunta 11. Explique por favor este resultado desde un punto de vista estatístico.

P12. Tendo en conta o que sabe ata este momento, podemos considerar a idade dos animais como un factor de risco?

P13. Que tipo de proba é un Chi cadrado? Considéraa adecuada para os obxectivos de esta proba? Por que?

P.14. Calcule a razón de vantaxes e explique o seu resultado. Podemos considerala significativa?

P.15. Calcule as medidas de impacto e explique o seu resultado.

**RESPOSTAS ao anexo 1**

*Pregunta 11. Explique por favor este resultado desde un punto de vista estatístico.*

A probabilidade observada é moi inferior a 0,05, o nivel de significación admitido nas probas biolóxicas e médicas, así que consideramos que a idade se asocia de maneira significativa coa prevalencia da enfermidade, sendo esta maior na idade máis avanzada (vantaxe  $160/836=0,191$ ) que nos animais máis novos (vantaxe  $82/1015=0,081$ ).

*P12. Tendo en conta o que sabe ata este momento, podemos considerar a idade dos animais como un factor de risco?*

Co que obtivemos na pregunta 11 só demostramos a existencia dunha asociación idade-prevalencia. Temos que demostrase os demais criterios para considerar a idade causa (temporalidade, mecanismos biolóxicos e fisiolóxicos, factores patolóxicos...).

*P13. Que tipo de proba é un Chi cadrado? Considéraa adecuada para os obxectivos de esta proba? Por que?*

Chi-cadrado é unha proba estatística non paramétrica. É a proba adecuada para demostrar asociación entre variables categóricas; ningunha das analizadas neste caso é unha variable cuantitativa (a idade foi segmentada en 2 grupos), así que esta proba é adecuada.

*P.14. Calcule a razón de vantaxes e explique o seu resultado. Podemos considerala significativa?*

OR = 2,37.

Inicialmente, ao partir de a, que son os 1015 negativos no grupo de menor idade, dinos que a probabilidade de estar san é 2,37 veces maior nos animais novos que na de máis idade. Pero iso tamén quere dicir, á inversa, que a probabilidade de infectarse é 2,37 veces maior nos animais de máis idade respecto dos máis novos.

Si podemos considerala significativa, xa que a súa significación tamén vén dada pola significación do Chi-cadrado que vimos previamente.

*P.15. Calcule as medidas de impacto e explique o seu resultado.*

Risco atribuible =  $0,191-0,081=0,11$

É dicir, parte da incidencia (0,11) do grupo de máis idade de debe precisamente a iso, á súa idade; ademais tamén significa canto menos infectado está o grupo máis novo.

Fración etiolóxica =  $(2,37 - 1) / 2,37 = 0,58$

Teoricamente, e sen máis información, poderíamos atribuír o 58% da prevalencia ao factor idade.

## ANEXO 2. TRANSFORMACIÓN DAS VARIABLES PARA ACHEGALAS Á NORMALIDADE

Gran parte das análises estatísticas usadas na ciencia dependen da normalidade da distribución das variables estudadas. Con todo, moitas variables asociadas aos seres vivos non cumpren ese requisito. En moitas ocasións, para axustar os nosos datos a unha distribución normal, transformámoslos cunha operación matemática. As operacións máis frecuentes son:

- Logaritmo
- Raíz cadrada
- Arcoseno (angular)
- Recíproco
- Transformación logarítmica

Útil cando dispomos de datos cuxo rango é maior que un grao de magnitude (tamaños moi diferentes). Para evitar o problema dos valores 0, xeralmente a transformación é

$$x_t = \log_{10}(x + 1)$$

Segundo a natureza dos datos, poden ser máis adecuadas outras bases no logaritmo (natural ...)

En Calc (Folla de cálculo de [LibreOffice](#)) transfórmase de forma sinxela coas fórmulas

$$x_t = \log(\text{celda})$$

- Transformación mediante raíz cadrada

Ten utilidade nos recontos. Para mellorar os resultados adóitase sumar 0,5 a cada valor previamente á raíz para mellorar o achegamento á normal mediante a transformación.

En calc a fórmula é

$$x_t = \text{raiz}(\text{celda})$$

- Transformación angular (arcoseno)

Aplícase cando enfrontámonos a variables que presentan límites naturais, nas que os límites da distribución non nos permiten achegarnos á normalidade. Tamén as porcentaxes (0-100) e as proporcións (01). Transfórmase a raíz cadrada da proporción. Se son porcentaxes, divídense entre 100 e realízase a raíz da proporción. En calc, como funciona con radians (do mesmo xeito que R) transformámoslas así:

$$x_t = \text{aseno}(\text{raiz}(\text{celda}) * 180 / \pi())$$

$$x_t = \text{aseno}(\text{raiz}(\text{celda}/100) * 180 / \pi())$$

Recuperaremos o valor orixinal mediante

$$x = \text{seno}(x_t * \pi() / 180)^2$$

$$x = (\text{seno}(x_t * \pi() / 180)^2) * 100$$

- Transformación recíproca

Útil cando unha das variables é unha serie de dilucións.

$$x_t = 1/x$$

**ANEXO 3. COMANDOS PARA REALIZAR AS PROBAS ESTADÍSTICAS SINALADAS NO TEMA 11 UTILIZANDO O PAQUETE ESTADÍSTICO R**

---

```
# PROBAS DE HOMOCEDASTICIDADE
shapiro.test() # PROBA DE NORMALIDADE
var.test() # COMPARA VARIANZAS DE DÚAS MOSTRAS NORMAIS
bartlett.test() # COMPARA VARIANZAS ENTRE GRUPOS

# PROBAS DE DIFERENZAS. 2 CATEGORÍAS
t.test() # DIFERENZAS DE MEDIAS ENTRE DOUS GRUPOS
t.test(x,y,paired=TRUE) # DIFERENZAS ENTRE DÚAS MOSTRAXES RELACIONADAS
wilcox.test() # Mann-Whitney U test - EQUIVALENTE NON PARAMÉTRICO

# PROBAS DE ASOCIACIÓN
chisq.test() # FRECUENCIAS
fisher.test() # PROBA EXACTA DE FRECUENCIAS

# PROBAS DE CORRELACIÓN
cor()
cor.test() # method="pearson"      "spearman"      "kendal"
lm() # REGRESIÓN LINEAL

# PRUEBAS DE DIFERENZA. MÁIS DE 2 CATEGORÍAS - ANOVA
oneway.test()
anova(lm())
aov()
model.tables(OBXETOANOVA, type="means") # COÑECER A TÁBOA DE MEDIAS

# PROBAS POSTHOC
TukeyHSD()

# EQUIVALENTES NON PARAMÉTRICOS
kruskal.test() # proba equivalente non paramétrica
friedman.test() # proba equivalente non paramétrica para mostraxes relacionadas
```



Unha colección orientada a editar materiais docentes de calidade e pensada para apoiar o traballo do profesorado e do alumnado de todas as materias e titulacións da universidade

unidadesdidácticas  
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA