

MATERIA
Bioestadística

TITULACIÓN
Grao en Medicina

unidade
didáctica
1

Estadística descriptiva

Mercedes Conde Amboage
María Isabel Borrajo
Carmen Cadarso Suárez

Área de Estadística e Investigación Operativa
Departamento de Estadística, Análise Matemática e Optimización
Facultade de Matemáticas

unidadesdidácticas
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



Esta obra atópase baixo unha licenza internacional Creative Commons BY-NC-ND 4.0. Calquera forma de reprodución, distribución, comunicación pública ou transformación desta obra non incluída na licenza Creative Commons BY-NC-ND 4.0 só pode ser realizada coa autorización expresa dos titulares, salvo excepción prevista pola lei. Pode acceder Vde. ao texto completo da licenza nesta ligazón: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.gl>

© Universidade de Santiago de Compostela, 2021

Deseño e maquetación

J. M. Gairí

Edita

Edicións USC

usc.gal/publicacions

DOI

<https://dx.doi.org/10.15304/9788419155047>

MATERIA: Bioestatística
TITULACIÓN: Grao en Medicina
PROGRAMA XERAL DO CURSO

BLOQUE I: ESTATÍSTICA DESCRIPTIVA

Unidade 1. Estatística descritiva

A utilidade da estatística nas ciencias da saúde
Conceptos básicos: poboación, variable aleatoria e mostra
Tipos de variables aleatorias
Táboas de frecuencias
Medidas características
Representacións gráficas

BLOQUE II: TEORÍA DA PROBABILIDADE

Unidade 2. Fundamentos da teoría da probabilidade

Experimento aleatorio, sucesos e espazo mostral
Definición de probabilidade
Probabilidade condicionada. Independencia de sucesos
Resultados notables: regra do produto, probabilidades totais e teorema de Bayes
Prevalencia e incidencia
Probos diagnósticas: sensibilidade, especificidade e valores predictivos
Medidas de efecto: risco relativo e odds-ratio

Unidade 3. Variables aleatorias discretas

Definición de variable aleatoria discreta
Función de masa de probabilidade e función de distribución
Medidas características
Modelos de distribucións discretas

Unidade 4. Variables aleatorias continuas

Definición de variable aleatoria continua
Función de densidade e función de distribución
Medidas características
Modelos de distribucións continuas. A distribución Normal
Aproximación entre distribucións

BLOQUE III: INFERENCIA ESTATÍSTICA

Unidade 5. Introducción á inferencia estatística

Utilidade da inferencia estatística
Conceptos básicos: parámetro, estatístico e estimador
Distribución na mostraxe
Distribucións asociadas á Normal: T de Student e χ^2 -cadrado

Unidade 6. Estimación de parámetros

Definición e propiedades dos estimadores puntuais
Estimación puntual da media, varianza e proporción
Estimación por intervalos de confianza

Unidade 7. Contrastes de hipóteses

Utilidade dos contrastes de hipóteses
Procedemento de contraste
Contrastes sobre unha poboación
Contrastes para comparación de poboacións

Unidade 8. Contrastes para datos categóricos

Introdución
Táboas de continxencia
Contrastes χ^2 -cadrado

BLOQUE IV: MODELOS DE REGRESIÓN**Unidade 9. Modelo de regresión lineal simple**

Introdución aos modelos de regresión
Conceptos previos: vector de medias, coeficiente de correlación e diagrama de dispersión
Modelo de regresión lineal simple
Método de mínimos cadrados
Coeficiente de determinación

Unidade 10. Análise da varianza

Modelo e hipóteses
Táboa de descomposición da variabilidade
Contraste de igualdade de medias
Comparacións múltiples

ÍNDICE

PRESENTACIÓN

COMPETENCIAS E OBXECTIVOS

PRINCIPIOS METODOLÓXICOS

CONTIDOS

1. A utilidade da estatística nas ciencias da saúde
2. Conceptos básicos: poboación, variable aleatoria e mostra
3. Tipos de variables aleatorias
4. Táboas de frecuencias
5. Medidas características
6. Representacións gráficas

AVALIACIÓN DA UNIDADE DIDÁCTICA

BIBLIOGRAFÍA

- Bibliografía básica
- Bibliografía complementaria

RECURSOS NA REDE

PRESENTACIÓN

A bioestatística é segundo o dicionario da RAE (Real Academia Española) a *ciencia que aplica el análisis estadístico a los problemas y objetos de estudio de la biología*. Partindo desta definición, a materia de Bioestatística (con pequenas variacións na súa nomenclatura) é unha materia que se imparte no primeiro curso de todos os graos da rama biosanitaria, tales como Medicina, Odontoloxía, Enfermaría, Bioloxía, Biotecnoloxía, Farmacia, Óptica, Psicoloxía ou Veterinaria.

Esta unidade didáctica caracterízase pola súa transversalidade no senso de que pode ser empregada polo alumnado e profesorado de moi diversas titulacións facendo mínimas adaptacións da mesma. Na Táboa 1 detállase a listaxe completa de titulacións e materias nas que se podería empregar esta unidade.

Táboa 1: Relación de titulacións e materias da rama biosanitaria, impartidas nalgún dos campus da USC, nas que se pode empregar esta unidade didáctica.

Titulación	Materia
Grao en Bioloxía	Bioestatística
Grao en Bioquímica	Bioestatística
Grao en Biotecnoloxía	Bioestatística
Dobre Grao en Farmacia e Óptica e Optometría	Matemáticas e estatística I
Grao en Enfermaría	Estatística e metodoloxía da investigación
Grao en Farmacia	Matemáticas e estatística I
Grao en Nutrición Humana e Dietética	Bioestatística
Grao en Odontoloxía	Bioestatística e introdución á investigación
Grao en Óptica e Optometría	Bioestatística
Grao en Psicoloxía	Análise de datos en psicoloxía
Grao en Veterinaria	Bioestatística

Nestas materias realízase un percorrido polas ferramentas elementais da probabilidade e da estatística focalizándose nas aplicacións biosanitarias en xeral, e no ámbito máis específico de cada titulación en particular. No seu desenvolvemento hai que ter en conta que, por estar asentada no primeiro curso, é necesariamente unha materia de nivelación dos coñecementos do alumnado con procedencias diversas, e á súa vez debe proporcionar unha competencia estatística sólida, pois en moitas das titulacións é o único contacto coa estatística que o alumnado vai ter ao longo deses anos de formación na correspondente facultade ou escola.

A programación docente proposta, da que esta unidade didáctica forma parte, está composta por catro bloques temáticos agrupando diversas unidades segundo se detalla a continuación:

- Bloque I: Estatística descritiva (1 UD);
- Bloque II: Teoría da probabilidade (3 UD);

- Bloque III: Inferencia estatística (4 UD);
- Bloque IV: Modelos de regresión (2 UD).

Esta proposta de programación docente cobre os contidos que se detallan na memoria de verificación de título do Grao en Medicina (pode consultarse na seguinte ligazón: https://pro-assets-usc.azureedge.net/cdn/ff/b4ycnzhOyuydvYl2rdg777pZ-5g_Z4rxapDN7SRNyAw/1632393404/public/plan/2021-09/Medicina_modif_2014.pdf), que son dos máis extensos de todas as titulacións biosanitarias da Universidade de Santiago de Compostela arriba mencionadas. Deste xeito, podemos cubrir os programas de todas as materias, se ben é certo que algunhas unidades non se empregarían en todas as titulacións.

Dentro da programación, con esta unidade didáctica en particular, pretendemos fornecer o alumnado e o profesorado, das materias de primeiro curso das titulacións da rama biosanitaria, cun material tanto teórico como práctico que cubra os contidos sobre estatística descritiva. Un material que sirva de referencia común, e que permita unificar tanto os conceptos básicos requiridos, como a notación e nomenclatura empregadas.

A presente unidade didáctica é a única unidade que conforma o Bloque I. Estatística descritiva. A través do seu desenvolvemento preténdese transmitir ao alumnado conceptos básicos como poboación, variable aleatoria e mostra, así como as ferramentas e condicións de uso das mesmas na análise inicial de calquera posible conxunto de datos co que se poidan atopar.

COMPETENCIAS E OBXECTIVOS

Segundo a memoria de verificación do título de Grao en Medicina pola USC, a comprensión das ciencias básicas, amosar unha actitude ética, o establecemento dunha boa comunicación interpersonal e a adaptación a circunstancias cambiantes son obxectivos fundamentais a acadar polo alumnado durante esta etapa de formación.

As competencias xerais e específicas vencelladas coa materia de Bioestatística que se recollen na memoria de verificación de título veñen detalladas a continuación. Por coherencia coa memoria de verificación de título, decidíronse manter os seus códigos de nomenclatura. Remarcar que nesta listaxe tamén se inclúen as competencias CG32, CE31 e CE37, que a pesar de non vincularse coa materia de Bioestatística na memoria de verificación do título, si se traballan e se desenvolven ao longo da mesma. Polo tanto, as competencias que se traballarán ao longo da materia son:

- CG28:** obter e empregar datos epidemiolóxicos e valorar tendencias e riscos para a toma de decisións sobre saúde;
- CG31:** coñecer, valorar criticamente e saber empregar fontes de información clínica e biomédica para obter, organizar, interpretar e comunicar información científica e sanitaria,
- CG32:** saber empregar as tecnoloxías da información e a comunicación en actividades clínicas, terapéutica, preventiva e investigadora.

- CG33:** manter e utilizar rexistros con información do paciente para unha posterior análise, preservando a confidencialidade dos datos;
- CG34:** ter, na actividade profesional, un punto de vista crítico e creativo, con escepticismo construtivo e orientado á investigación;
- CG35:** comprender a importancia e as limitacións do pensamento científico no estudo, prevención e manexo de enfermidades;
- CG36:** ser capaz de formular hipóteses, recoller e valorar de xeito crítico información para resolver problemas, seguindo o método científico;
- CG37:** adquirir formación básica para a actividade investigadora;
- CE31:** coñecer, valorar criticamente e saber utilizar tecnoloxías e fontes de información clínica e biomédica, para obter, organizar, interpretar e comunicar información clínica, científica e de saúde;
- CE32:** coñecer os conceptos básicos da bioestatística e a súa aplicación ás ciencias médicas;
- CE33:** ser capaz de deseñar e realizar estudos estatísticos sinxelos mediante programas informáticos e interpretar os resultados;
- CE34:** comprender e interpretar datos estatísticos da literatura médica;
- CE37:** manexar con autonomía un ordenador persoal e as aplicacións informáticas máis comúns no campo da biomedicina.

Segundo as competencias arriba indicadas, na guía docente da materia establécese que a finalidade desta é familiarizar ao alumnado cos conceptos e técnicas básicas da estatística descritiva, da teoría da probabilidade e da inferencia estatística. Amais, derívanse os obxectivos xerais da materia:

- OX1:** coñecer a linguaxe estatística básica: proporcionar aos estudantes os coñecementos teóricos básicos que lles permitan comprender os distintos aspectos estatísticos e probabilistas implicados na investigación médico/sanitaria;
- OX2:** coñecer e aplicar algúns métodos estatísticos básicos para representar e analizar conxuntos de datos simples e poder obter conclusións destas análises;
- OX3:** coñecer, expresar e interpretar correctamente os niveis de precisión, confianza e de erro nas conclusións dun estudo estatístico.

Nesta unidade didáctica traballaremos o OX1 e OX2, e deles derívanse os seguintes obxectivos específicos:

- OE1:** coñecer as distintas escalas de medida e a súa utilización na análise estatística;
- OE2:** distinguir entre poboación e mostra;
- OE3:** coñecer e saber aplicar as principais ferramentas de representación gráfica;
- OE4:** deseñar e analizar estudos sinxelos;
- OE5:** saber interpretar axeitadamente os resultados dunha análise descritiva;
- OE6:** manexar un *software* estatístico;
- OE7:** crear bases de datos coherentes e completas, que permitan responder as preguntas de interese;
- OE8:** asumir a necesidade e utilidade da estatística como ferramenta no exercicio profesional da medicina;

- OE9:** desenvolver unha actitude crítica e responsable coa toma de decisións baseada en resultados estatísticos;
- OE10:** ser capaz de traballar en equipo;
- OE11:** valorar a aprendizaxe autónoma;
- OE12:** amosar iniciativa na resolución de problemas;
- OE13:** amosar interese por adquirir novos coñecementos estatísticos.

A relación entre estes obxectivos específicos e as correspondentes competencias vén detallada na Táboa 2.

Táboa 2: Relación entre os obxectivos específicos desta unidade didáctica e as competencias vencelladas á materia segundo a memoria de verificación do Grao en Medicina.

Obxectivos	Competencias
OE1	CE31, CE32
OE2	CE32
OE3	CG32, CE32, CE37
OE4	CG28, CG32, CE33, CE34, CE37
OE5	CE34
OE6	CE33, CE37
OE7	CG28, CG32, CG34, CG36, CE33, CE34, CE37
OE8	CE31, CE34
OE9	CG34, CG35, CE31, CE32
OE10	CG37
OE11	CG37
OE12	CG37, CE31
OE13	CG37, CE31, CE34

Remarcar que os obxectivos OE1 a OE3 están ligados a competencias técnicas (aquelas que implican dominio como experto das tarefas e contidos do seu ámbito de traballo, así como os coñecementos e habilidades necesarios para o seu desempeño); os OE4 a OE7 a competencias metodolóxicas (implican a capacidade de reacción ao aplicar o procedemento adecuado ás tarefas encomendadas e ás irregularidades que xurdan atopando de xeito independente vías de solución, que ademais logo se transfiren a outros eidos da formación ou traballo); e os OE8 a OE13 a competencias actitudinais (vencelladas á colaboración interpersoal de forma comunicativa e construtiva, así como á participación na organización e á capacidade de decidir e asumir responsabilidades).


PRINCIPIOS METODOLÓXICOS

Neste apartado especificase a liña metodolóxica xeral e a dinámica de traballo que desexamos instaurar neste Bloque I: Estatística descritiva da materia de Bioestatística. As actividades propostas clasifícanse en dous grandes grupos atendendo ao carácter presencial ou non das mesmas. Na Táboa 3 pódese ver en detalle a dedicación en horas presenciais e non presenciais que se espera por parte do alumnado para a superación desta unidade didáctica.

As actividades formativas con presenza do profesorado que se desenrolarán ao longo deste Bloque I son:

- **Sesións expositivas.** Para a exposición dos contidos teóricos o profesorado fará uso de presentacións en ordenador que o alumnado terá á súa disposición a través do Campus Virtual da USC. Estas presentacións son esquemas para que o alumnado poida seguir as sesións, pero é recomendable que cada alumna/o tome as súas propias notas e complete os contidos que se expoñen. As sesións expositivas desenvolveranse intercalando problemas prácticos e aplicacións entre as explicacións teóricas. Preténdese enriquecer a participación do alumnado na clase, por medio dun exercicio diario ao finalizar cada sesión que se poida resolver aplicando os coñecementos adquiridos durante a clase correspondente.

Xunto coa presentación de cada sesión, facilitaráselle ao alumnado ao concluír cada tema un material máis desenvolvido. Preténdese que o alumnado compare o material que recopilou durante as sesións con estes documentos, para repasar os conceptos estudados e completar os seus apuntamentos. Ademais, cada tema inclúe o seu correspondente boletín de exercicios nos que o alumnado poderá aplicar as ideas desenvolvidas nas sesións expositivas.

- **Sesións interactivas de seminario.** Estas sesións dedicaranse á resolución por parte do alumnado dos exercicios propostos nos boletíns temáticos. En ocasións nas que sexa necesario clarificar ou incidir en algún dos contidos, os problemas serán resoltos polo profesorado. Preténdese que a/o estudante adquira seguridade na exposición oral de coñecementos e valore o traballo en equipo.
- **Sesións interactivas de prácticas de ordenador.** Estas sesións serán realizadas nas aulas de informática do centro e introducirán o alumnado no manexo do *software* estadístico  de libre distribución. En cada sesión, o alumnado disporá dun guión da práctica que servirá de fío condutor á vez que incluírá exercicios a realizar por parte da/o alumna/o de maneira autónoma.

Como actividades non presenciais asociadas a esta unidade considéranse a elaboración da **memorias de prácticas**, a resolución de exercicios, as buscas bibliográficas ou a realización das lecturas recomendadas. Máis concretamente, nas sesións interactivas de prácticas de ordenador organizaranse parellas de traballo que, con un conxunto de datos reais proporcionado, deberán realizar unha análise estatística descritiva. Nótese que a análise deste conxunto de datos irase incrementando en sesións interac-

tivas de prácticas de ordenador asociadas a outros bloques. Deste xeito, a valoración da memoria de práctica permitirá levar a cabo unha avaliación procesual.

Táboa 3: Dedicación (en horas) do alumnado a esta unidade didáctica segundo o tipo de actividade.

Actividades presenciais	Horas	Actividades presenciais	Horas
Sesiós expositivas	4	Estudo individual	8
Seminarios	2	Resolución de exercicios	3
Prácticas de ordenador	2	Elaboración memoria de prácticas	3
Total horas presenciais	8	Total horas non presenciais	14

CONTIDOS

A definición e estruturación dos contidos da presente unidade didáctica farase en primeiro lugar respectando as indicacións que se especifican na ficha da materia incluída na memoria de verificación do título. Ademais, pretende ser fiel a dous principios chave:

- non hipertrofiar os contidos disciplinarios, configurando un programa acorde co peso curricular da materia, proporcionando ao alumnado as ferramentas necesarias para o correcto desenvolvemento doutras materias da titulación e da súa futura actividade profesional;
- romper o que se coñece como «encefalograma plano», é dicir, dar a todos os contidos o mesmo nivel de importancia. Debemos ser capaces de amosar ao alumnado as chaves da nosa disciplina mediante contido esencial, necesario e de extensión.

Os contidos desta unidade didáctica atópanse distribuídos en seis seccións, que deben de ser tratadas de xeito secuencial no tempo e que detallamos a continuación.

1. A utilidade da estatística nas ciencias da saúde

A variabilidade é un dos elementos esenciais da nosa vida. A conseguinte incerteza que xera esta variabilidade é importante e, en moitos campos, como a medicina, é esencial ter métodos que nos permitan cuantificar esa incerteza e minimizar o seu impacto nas decisións que tomamos.

A estatística ocúpase de ámbitos tan diversos como a recompilación, a presentación, a análise e o uso de datos para a toma de decisións e para a resolución de problemas. Dentro da estatística distínguense fundamentalmente tres ramas:

- **Estatística descritiva:** ocúpase de analizar a información contida nos datos, sen interese en extrapolar eses resultados ou extraer conclusións xerais;
- **Cálculo de probabilidades:** estuda as leis que rexen os fenómenos aleatorios;

- **Inferencia estatística:** a partir da observación dos datos, pretende extraer conclusións aplicables á poboación xeral.

Nesta primeira unidade didáctica estudaranse todas as técnicas relativas á inferencia estatística.

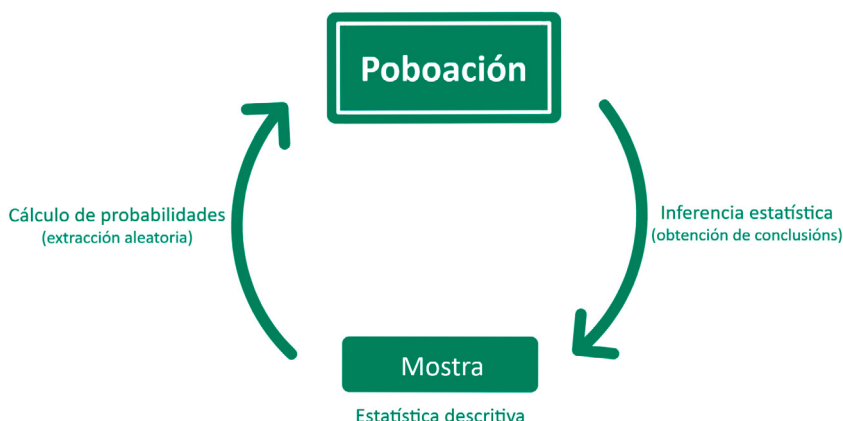
2. Conceptos básicos: poboación, variable aleatoria e mostra

Calquera fenómeno aleatorio sobre o que se poida recoller un conxunto de datos é susceptible de ser analizado con métodos estatísticos. O primeiro paso de calquera análise estatística debe ser unha descrición dos propios datos, procedemento do que se encarga a estatística descritiva.

Para poder aplicar as técnicas correctamente e interpretar os resultados obtidos, débese de manexar unha serie de conceptos básicos que se introducen a continuación.

- **Poboación:** é o conxunto de todos os individuos de interese para o estudo.
- **Variable aleatoria:** é a característica dos elementos da poboación que se pretende analizar.
- **Mostra:** é un subconxunto da poboación sobre o que se coñece os valores da variable aleatoria de interese. Existen diferentes métodos (teoría da mostraxe) para a elección deste subconxunto que deber ser representativo da poboación que se pretende describir. O número de individuos que compoñen este subconxunto denomínase **tamaño mostral**.

Figura 1: Esquema das principais ramas da estatística e a súa función no establecemento da relación poboación-mostra.



Na Figura 1 pódese ver un esquema relacional entre as tres disciplinas da estatística mencionadas na sección anterior, así como os nexos que se establecen entre poboación e mostra.

Exemplo 1 A USC quere analizar a satisfacción do alumnado de grao coa docencia recibida, para o que enquisou a 500 estudantes de diferentes titulacións, pedíndolles que fixesen unha valoración global da docencia recibida cunha puntuación entre 0 e 10. Neste escenario, todo o alumnado de grao matriculado na USC é a **poboación** de interese, as/os 500 estudantes enquisadas/os constitúen a **mostra** e a puntuación valorando a docencia recibida sería a **variable aleatoria**.

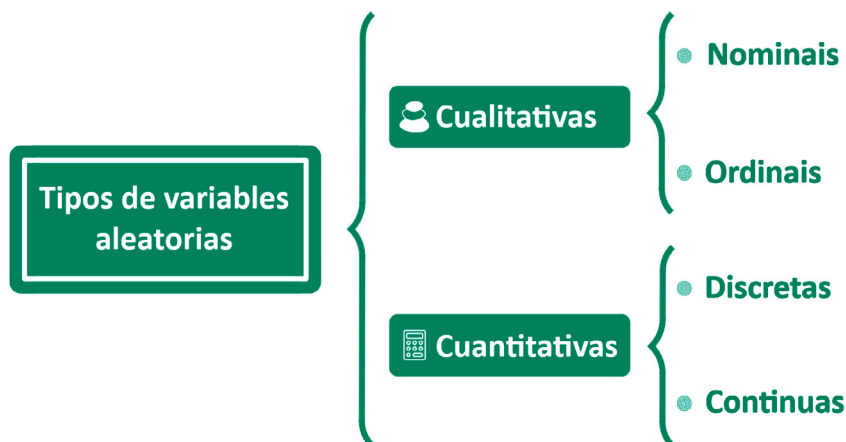
3. Tipos de variables aleatorias

O obxectivo dun estudo de estatística descritiva é a análise de certa característica que se pode medir sobre os individuos da mostra. As posibilidades desta análise varían segundo a natureza desa característica (os distintos tipos de variables aleatorias permiten unhas ou outras operacións), polo que é importante identificar con cal deles se está a tratar.

A clasificación dos distintos tipos de variables aleatorias pode verse resumida na Figura 2, que se basea nas seguintes definicións:

- **Variables cualitativas:** son aquelas variables aleatorias cuxos posibles valores son clases ou categorías; por exemplo a «cor de ollos» (negra, marrón, verde, azul) ou o «nivel de estudos» (primarios, secundarios, universitarios). Dentro das variables cualitativas podemos á súa vez distinguir dous subtipos:
 - **Nominais:** son aquelas nas que os posibles valores non teñen unha orde subxacente, como é o caso da «cor de ollos».
 - **Ordinais:** son aquelas nas que os posibles valores si presentan unha orde predeterminada, como é o caso do «nivel de estudos».
- **Variables cuantitativas:** son aquelas variables aleatorias que toman valores numéricos, é dicir, que se miden en forma de cantidades; por exemplo o «número de cigarros fumados por día» ou a «concentración de hemoglobina en sangue (gr/dl)». De novo, dentro da familia das variables cuantitativas existe unha subclasificación:
 - **Discretas:** son aquelas variables aleatorias que teñen unha cantidade finita ou infinita numerable de posibles valores, como é o caso do «número de cigarros fumados por día».
 - **Continuas:** son aquelas variables aleatorias nas que os posibles valores son unha cantidade infinita (non numerable). Son facilmente identificables porque dados dous posibles valores calquera da variable aleatoria, sempre existirá entre eles outro posible valor da mesma. Un exemplo deste caso sería a «concentración de hemoglobina en sangue (gr/dl)»

Figura 2: Clasificación de variables aleatorias.



4. Táboas de frecuencias

As táboas ou distribucións de frecuencias teñen como fin presentar e describir os datos dunha mostra dun xeito claro e sinxelo, baseándose unicamente no recuento de individuos.

A pesar de que a estrutura das táboas é semellante, existen certos matices en función do tipo de variable aleatoria que se queira resumir. Na Táboa 4 amósase a estrutura xenérica dunha táboa de frecuencias. Para poder completala necesítase coñecer os seguintes conceptos:

- **Frecuencias absolutas (n_i):** é o número de veces que na mostra ocorre cada unha das categorías ou valores.
- **Frecuencias relativas (f_i):** é a proporción de datos, con respecto ao tamaño mostral (n), que hai en cada categoría ou valor. Obtéñense como

$$f_i = \frac{n_i}{n}.$$

- **Frecuencias absolutas acumuladas (N_i):** é o número de veces que ocorre a categoría ou valor i -ésimo e todos os anteriores. Obtéñense como

$$N_i = \sum_{j \leq i} n_j.$$

- **Frecuencias relativas acumuladas (F_i):** é a proporción de datos, con respecto ao tamaño mostral, da categoría ou valor i -ésimo e todos os anteriores. Obtéñense como

$$F_i = \sum_{j \leq i} f_j.$$

Táboa 4: Estrutura xenérica dunha táboa de frecuencias asociada a variables cualitativas ordinais ou cuantitativas discretas.

Variable aleatoria	n_i	f_i	N_i	F_i
Categoría / valor 1	n_1	f_1	N_1	F_1
Categoría / valor 2	n_2	f_2	N_2	F_2
...
Categoría / valor k	n_k	f_k	n	1

É importante remarcar que as frecuencias acumuladas só se poden calcular para variables cuantitativas discretas ou cualitativas ordinais. No caso das cualitativas nominais, ao non existir unha orde subxacente, non saberíamos como facer esa acumulación. Polo tanto, nas correspondentes táboas de frecuencias non existen as dúas derradeiras columnas que se poden ver na Táboa 4.

Tamén é esencial destacar que no caso das variables cuantitativas continuas, hai que facer o que se denomina **categorizar a variable aleatoria**, previamente á elaboración da táboa de frecuencias. Este proceso de categorización da variable aleatoria consiste en definir un número finito de intervalos que cubran todo o rango de valores da mostra, para que cada un dos mesmos constitúa unha categoría, e dese xeito poder contar os datos que hai en cada un deles. Por definición e para que non haxa recontos por duplicado, estes intervalos han de ser dous a dous disxuntos, isto sóese conseguir facendo intervalos pechados pola esquerda e abertos pola dereita, $[a, b)$. Ademais, en tal caso, defínese o que se denomina **marca de clase** (c_i), que se corresponde co punto central de cada un dos intervalos e será como o distintivo ou nome de cada un deles. Esta marca de clase engádese nunha nova columna ná táboa antes das frecuencias absolutas. Podemos ver un exemplo deste tipo na Táboa 5.

Táboa 5: Estrutura xenérica dunha táboa de frecuencias asociada a unha variable cuantitativa continua.

Variable continua	c_i	n_i	f_i	N_i	F_i
Intervalo 1	c_1	n_1	f_1	N_1	F_1
Intervalo 2	c_2	n_2	f_2	N_2	F_2
...
Intervalo k	c_k	n_k	f_k	n	1

5. Medidas características

Para resumir a información que contén unha mostra dunha variable cuantitativa, ademais das táboas de frecuencias, pódense calcular medidas características. Distínguense tres tipos de medidas características:

- **Medidas de posición ou localización.** Estas medidas indican en torno a que valor ou valores se atopan os datos observados. As medidas de posición máis empregadas son:
 - a media aritmética ou a media recortada,
 - a mediana,
 - os cuantís e
 - a moda.
- **Medidas de dispersión.** Estas medidas indican canto se dispersan os datos observados. Dentro das medidas de dispersión destacan:
 - a varianza,
 - a desviación típica,
 - o rango e
 - o rango intercuartílico.
- **Medidas de forma.** Estas medidas permiten estudar se a distribución da mostra considerada se asemella ou non a unha campá de Gauss. Dentro das medidas de forma destacan os coeficientes de:
 - asimetría de Fisher e
 - grao de apuntamento ou curtose.

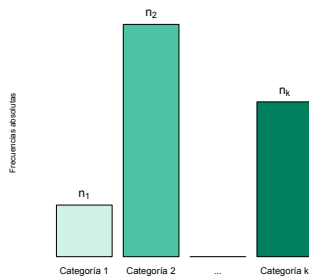
6. Representacións gráficas

Á hora de describir o comportamento dunha mostra, ademais das táboas de frecuencias ou das medidas característica, tamén serán de grande interese as representacións gráficas. De cara a poder empregar estas representacións gráficas, será fundamental ter claro o tipo de variable aleatoria que se está a manexar, posto que en función da clasificación, deberase empregar unha representación ou outra. A continuación, presentaremos as diferentes representacións gráficas que veremos ao longo desta unidade didáctica:

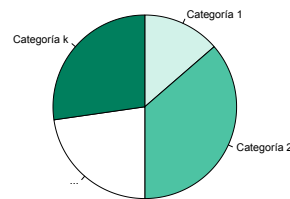
- **Diagrama de barras.** Esta representación só é adecuada para variables cualitativas ou cuantitativas discretas. Para construír un diagrama de barras débense representar no eixe OX as distintas categorías da variable cualitativas ou os valores da variable cuantitativa discreta e levantar sobre eles unhas barras con altura proporcional ás frecuencias absolutas ou relativas da categoría en cuestión. No apartado (a) da Figura 3 pódese ver un exemplo de diagrama de barras.
- **Diagrama de sectores.** Esta representación só é adecuada para variables cualitativas ou cuantitativas discretas. Esta gráfica consiste nun círculo que se divide en tantas partes como categorías ten a variable cualitativa ou valores toma a variable cuantitativa discreta que se está a estudar. Cada unhas destas partes terá un tamaño proporcional á frecuencia de cada unha das categorías ou valores da variable aleatoria de interese. No apartado (b) da Figura 3 pódese ver un exemplo de gráfico de sectores.

- **Histograma.** Esta representación só é adecuada para variables cuantitativas continuas e permite representar as táboas de frecuencias como a que se ilustra na Táboa 5. Para construír un histograma débense representar no eixe OX os distintos intervalos de clase e levantar sobre eles unhas barras con altura proporcional ás frecuencias relativas ou ás densidades do intervalo en cuestión. No apartado (c) da Figura 3 pódese ver un exemplo de histograma.
- **Diagrama de caixa ou *boxplot*.** Esta representación só é adecuada para variables cuantitativas e permite representar varias das medidas características que se presentaron no apartado anterior. No apartado (d) da Figura 3 pódese ver un exemplo de diagrama de caixa onde:
 - Q_1 denota o cuantil 0.25, é dicir, o primeiro cuartil;
 - Q_3 denota o cuantil 0.75, é dicir, o terceiro cuartil;
 - $LI = \max \{ \text{mínimo da mostra}, Q_1 - 1.5 \text{ RIC} \}$ onde RIC denota o rango intercuartílico, e
 - $LS = \min \{ \text{máximo da mostra}, Q_3 + 1.5 \text{ RIC} \}$.

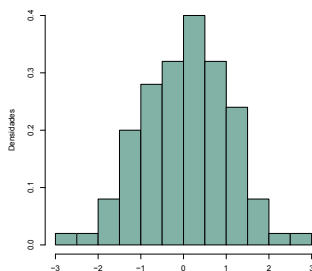
Figura 3: Exemplos das principais representacións gráficas.



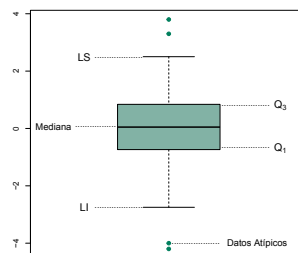
(a) Diagrama de barras



(b) Diagrama de sectores



(c) Histograma



(d) Diagrama de caixa

AVALIACIÓN DA UNIDADE DIDÁCTICA

Na memoria de verificación do Grao en Medicina especificase que en todas as materias do Grao a cualificación de cada alumna/o farase mediante avaliación continua e a realización dun exame final. En concreto, no caso da materia de Bioestatística, a avaliación continua terá un peso do 30 % da nota final, mentres que o exame final terá un peso do 70 %.

A continuación, detállanse as actividades vinculadas a esta unidade didáctica así como o peso de cada unha delas na avaliación final da materia:

- Para a avaliación dos obxectivos OE1-OE3, OE5, OE8, OE9, OE12 e OE13 realizaremos unha proba tipo test a través do Campus Virtual da materia. Esta proba realizarase de xeito individual durante parte dunha sesión de seminario e terá un peso do 2 % na nota final da materia.
- Para a avaliación dos obxectivos OE4-OE7 e OE9-OE11 o alumnado deberá coomezar a elaborar a memoria de prácticas incluíndo unha análise estatística descriptiva do conxunto de datos proporcionado. Esta actividade realizarase en parellas e asóciasele unha dedicación de 2.5 horas de traballo persoal por parte do alumnado e terá un peso do 2.5 % na nota final da materia.

Así mesmo, no exame final da materia poderanse incluír preguntas de tipo teórico e/ou práctico sobre os contidos desta unidade didáctica.

BIBLIOGRAFÍA

Bibliografía básica

- BORRAJO, M. I., CONDE-AMBOAGE, M. E CRUJEIRAS-CASAS, R. M. (2020). *Estatística descriptiva: resumos numéricos e representacións gráficas*. Esenciais USC. Universidade de Santiago de Compostela.
- CRUJEIRAS-CASAS, R. M. E FARALDO-ROCA, P. (2010). *Manual de estadística básica para ciencias de la salud*. Universidade de Santiago de Compostela, Dpto. de Estadística e Investigación Operativa.
- MILTON, J.S. (1994) *Estatística para biología y ciencias de la salud*. (2ª ed). Interamericana, McGraw-Hill.
- ROSNER, B. (2006) *Fundamentals of biostatistics* (6ª edición). Wadsworth Publishing Company. Duxbury Press.

Bibliografía complementaria

- ALTMAN, D. G. (1990). *Practical statistics for medical research*. Chapman & Hall.
- ÁLVAREZ-CÁCERES, R. (2007). *Estatística Aplicada a las Ciencias de la Salud*. Editorial Diaz de Santos.
- SHAHBABA, B. (2012). *Biostatistics with R*. Springer.

- DURRETT, R. (2009). *Elementary probability for applications*. Cambridge University Press.
- COBO, E., MUÑOZ, P., GONZÁLEZ, J. A., BIGORRA, J., CORCHERO, C., MIRAS, F., SELVA, A. E VIDELA, S. (2007). *Bioestadística para no estadísticos*. Elsevier.
- GARCÍA-PÉREZ, A. (2010). *Estadística básica con R*. U.N.E.D.
- MARTIN-ANDRÉS, A. E DE LUNA DEL CASTILLO, J. (1999). *50 ± 10 horas de bioestadística*. Capitel.
- MONTANERO-FERNÁNDEZ, J. E MINUESA-ABRIL, C. (2018). *Estadística básica para ciencias de la salud*. Universidad de Extremadura.

RECURSOS NA REDE

Hoxe en día, tanto alumnado como profesorado teñen acceso a multitude de recursos a través da Rede. Aínda que é certo que na liberdade que reina no mundo dixital hai que saber cales merece a pena consultar. É por iso que se esta sección pretende proporcionar recursos de calidade para diferentes fins dentro do proceso de aprendizaxe.

— **Academic Skills Kit (ASK)**

<https://internal.ncl.ac.uk/ask/numeracy-maths-statistics/statistics/>

A Universidade de Newcastle pon a disposición de calquera usuario da Rede cursos de diversas temáticas. En particular, a páxina de estatística ten un curso de iniciación bastante completo, no que para cada tema comeza coas definicións teóricas básicas, seguidas dalgunha actividade extra, que poden ser vídeos, exemplos ilustrativos, exercicios, outros recursos electrónicos ... Ademais, todos os exercicios propostos son acompañado dunha solución detallada.

— **Fisterra**

<https://www.fisterra.com/formacion/metodologia-investigacion/>

É un sitio web dirixido a profesionais da atención primaria orientada á consulta e formación. Dispón dunha sección adicada á «Metodoloxía da investigación» na que se inclúen conceptos estatísticos de uso común no campo médico. Ademais de explicacións e exemplos, cada tema adoita ir acompañado de ligazóns a recursos electrónicos relacionados e bibliografía vencellada a ese tema.

— **Material didáctico da Unidade de Bioestadística Clínica do Hospital Ramón y Cajal**

http://www.hrc.es/bioest/M_docente.html

Neste sitio web está dispoñible todo o material didáctico producido na unidade de Bioestadística do hospital. Inclúe contido metodolóxico, pero tamén problemas e solucións.

— **SOCR**

<http://www.socr.ucla.edu/>

SOCR é un recurso interactivo gratuíto e accesible a través da Rede para exploración, modelado, análise e interpretación de datos. Foi deseñado pola Universidade de California no 2006. Inclúe ferramentas e recursos baseados en *applets*

interactivos como aplicacións de cálculo e gráficos para a impartición de cursos estatísticos. A súa filosofía é a dunha ferramenta aberta e ampliable para a comunidade educativa. É unha das maiores coleccións de *applets* Java, moi útil para a aprendizaxe autónoma, motivación, modernización e mellora do modelo de ensino estatístico. Promove a aprendizaxe activa e práctica, ca/o estudante como axente centrada/o no proceso de aprendizaxe.

— **Gapminder**

<https://www.gapminder.org/>

Gapminder é unha fundación sen ánimo de lucro con sede en Estocolmo. Como se din eles mesmos, intentan dar unha visión do mundo baseada nos feitos poñendo a disposición do público xeral un método que facilita a comprensión dos datos. Ofrecen, entre outras cousas, unha poderosa ferramenta gráfica para analizar diferentes tipos de datos numéricos de distintos países nunha secuencia temporal. Non se trata dun recurso especialmente dirixido a profesionais sanitarios ou estudantes, máis ben, tenta difundir coñecemento obxectivo. A idea deste recurso é facer entender, ás persoas en xeral e ao alumnado en particular, a necesidade de adquirir coñecementos estatísticos que lles permitan comprender mellor «os números» do mundo que os rodea. Na web pódense atopar bases de datos moi diversas: o número de casos de VIH, a esperanza de vida, a desnutrición infantil, ...

— **Vídeos TED**

<https://www.ted.com/#/>

A oferta de charlas TED é case interminable, do mesmo xeito que os temas que tratan. Relacionado co recurso anterior, é necesario facer unha mención especial ao que probablemente sexa unha das mellores charlas TED relacionadas coa estatística, a do profesor Hans Rosling https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen, co que as/os estudantes, ademais de quedar fascinadas/os, entenderán a importancia non só de facer «boas estatísticas» senón como representalas e expresalas dun xeito adecuado.


— **Numbers needed to mislead, meta-analysis and muddled thinking**

<https://www.youtube.com/watch?v=qrkI3OzYfIY>

Trátase dunha clase maxistral de 40 minutos impartida polo profesor do *Luxembourg Institute Health*, Stephen Senn, na Universidade de Sheffield sobre as conclusións derivadas de análises estatísticas realizadas en ensaios clínicos.

— **Shiny apps**


<http://www.showmeshiny.com/>

Son aplicacións sinxelas feitas con código , pero cunha interface gráfica moi amigable que permite á/ao usuaria/o «xogar» cos números. Os temas que tratan son moi variados, pero hai moitos dedicados a conceptos de estatística básica.

— **StatsTeachR**

<http://www.statsteachr.org/>

É un repositorio en liña de acceso aberto con plans de leccións modulares para

ensinar estatística usando  en niveis de grao e posgrao. Cada módulo céntrase en ensinar un concepto estatístico específico, que vai dende leccións estatísticas introductorias ata temas máis avanzados.



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