

MATERIA
Bioestatística

TITULACIÓN
Grao en Medicina

unidade
didáctica
2

Fundamentos da teoría da probabilidade

Mercedes Conde Amboage
María Isabel Borrajo
Carmen Cadarso Suárez

Área de Estatística e Investigación Operativa
Departamento de Estatística, 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, 2022

Deseño e maquetación

J. M. Gairí

Edita

Edicións USC

usc.gal/publicacions

ISBN

978-84-19679-03-1

DOI

<https://dx.doi.org/10.15304/9788419679031>

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

Probas diagnósticas: sensibilidade, especificidade, 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 X-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 linear simple**

Introdución aos modelos de regresión
Conceptos previos: vector de medias, coeficiente de correlación e diagrama de dispersión
Modelo de regresión linear 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. Experimento aleatorio, sucesos e espazo mostral
2. Definición de probabilidade
3. Probabilidade condicionada. Independencia de sucesos
 - 3.1. Probabilidade condicionada
 - 3.2. Independencia de sucesos
4. Resultados notables: regra do produto, teorema de probabilidades totais e teorema de Bayes
 - 4.1. Regra do produto
 - 4.2. Teorema de probabilidades totais
 - 4.3. Teorema de Bayes
5. Prevalencia e incidencia dunha enfermidade
6. Probas diagnósticas: sensibilidade, especificidade e valores predictivos
7. Medidas de efecto: risco relativo e odds-ratio

AVALIACIÓN DA UNIDADE DIDÁCTICA

BIBLIOGRAFÍA

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. Na Táboa 1 detállase a listaxe completa de titulacións e materias da Universidade de Santiago de Compostela (USC) 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
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
Dobre Grao en Farmacia e Óptica e Optometría	Matemáticas e estatística I

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 previamente 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 fundamentos da teoría da probabilidade. 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 primeira das tres unidades que conforman o Bloque II. Teoría da probabilidade. A través do seu desenvolvemento preténdese transmitir ao alumnado conceptos básicos como experimento aleatorio, suceso e espazo mostral, así como as ferramentas e condicións do seu uso para poder manexar conceptos e cálculos probabilísticos.

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 que se deben acadar polo alumnado durante esta etapa de formación.

As competencias xerais (denotadas por CG) e específicas (denotadas por CE) vencelladas coa materia de Bioestatística que se recollen na memoria de verificación do título veñen detalladas a continuación. Por coherencia coa memoria de verificación do título, decidíronse manter os seus códigos de nomenclatura. Débese salientar 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 dela. 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 da/o paciente para unha posterior análise, preservando a confidencialidade dos datos;
- CG34:** ter, na actividade profesional, un punto de vista crítico e creativo, con excepcionalismo 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.

Seguindo as competencias arriba indicadas, na guía docente da materia establécese que a finalidade desta é familiarizar o alumnado cos conceptos e técnicas básicas da estatística descritiva, da teoría da probabilidade e da inferencia estatística. Amais, derivanse 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 del derivanse os seguintes obxectivos específicos:

- OE1:** coñecer os conceptos básicos da teoría probabilística;
- OE2:** saber identificar os sucesos nun experimento probabilístico;
- OE3:** coñecer e saber manexar probabilidades condicionadas;
- OE4:** comprender as principais leis que rexen a teoría da probabilidade;
- OE5:** manexar conceptos probabilísticos aplicados ao eido da medicina;
- OE6:** valorar a aprendizaxe autónoma;
- OE7:** amosar iniciativa na resolución de problemas;

OE8: 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	CE32
OE2	CG31, CE32
OE3	CG31, CE32
OE4	CG31, CE32
OE5	CG28, CG35, CE32, CE34
OE6	CG34, CG37
OE7	CG37, CE31
OE8	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 e OE5 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 OE6 a OE8 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 especifícase a liña metodolóxica xeral e a dinámica de traballo que desexamos instaurar neste Bloque II. Teoría da probabilidade. 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 desenvolverán ao longo deste Bloque II 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 recompilou 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.

Como actividades non presenciais asociadas a esta unidade considéranse a resolución de exercicios, as buscas bibliográficas ou a realización das lecturas recomendadas.

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

Actividades presenciais	Horas	Actividades presenciais	Horas
Sesións expositivas	3	Estudo individual	6
Seminarios	1	Resolución de exercicios	1.5
Total horas presenciais	4	Total horas non presenciais	7.5

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. Experimento aleatorio, sucesos e espazo mostral

No mundo real pódense atopar dous tipos de fenómenos: deterministas e aleatorios; os primeiros están rexidos por leis coñecidas e o seu resultado é sempre coñecido de antemán, mentres que os segundos teñen por característica fundamental a incerteza do seu resultado. A teoría da probabilidade ocúpase de entender e estudar esta incerteza que caracteriza os fenómenos aleatorios, asignando a cada un dos posibles resultados do fenómeno un número que indicará o grao no que dito resultado pode ter lugar.

Para poder aplicar correctamente a teoría da probabilidade, precisamos coñecer algúns conceptos fundamentais.

- **Experimento aleatorio:** é aquela proba que, repetida sucesivas veces en idénticas condicións, pode dar lugar a resultados diferentes e imprevisibles.
- **Suceso elemental:** é cada un dos posibles resultados dun experimento aleatorio.
- **Espazo mostral:** é o conxunto de todos os sucesos elementais, xeralmente denótase por Ω .
- **Suceso:** é calquera subconxunto do espazo mostral; en particular Ω denomínase suceso seguro, e o conxunto baleiro, \emptyset , denomínase suceso imposible. Ás veces na literatura úsase o termo “suceso composto” para designar un suceso calquera que non sexa un suceso elemental.

Exemplo 1. *Quérese estudar un xogo de azar consistente no lanzamento de 3 moedas. Nese caso, o experimento aleatorio sería “Lanzar 3 moedas” e o espazo mostral asociado sería*

$$\Omega = \{(ccc), (cc+), (c+c), (+cc), (c++), (+c+), (++c), (+++)\},$$

onde c denota que sae cara e $+$ denota que sae cruz, e téñense eses 8 sucesos elementais.

Pódense definir múltiples sucesos de interese como, por exemplo,

$$\begin{aligned} S_1 &= \text{“Saír polo menos unha cara”} \\ &= \{(ccc), (cc+), (c+c), (+cc), (c++), (+c+), (++c)\} \end{aligned}$$

ou

$$S_2 = \text{“Saír máis dunha cruz”} = \{(c++), (+c+), (++c), (+++)\}.$$

Ambos son sucesos compostos, formados por varios sucesos elementais.

Os sucesos son a unidade fundamental para o desenvolvemento da teoría da probabilidade, e para traballar con eles precísase definir unha serie de operacións.

- **Complementario:** dado un suceso A en Ω , defínese o suceso complementario como aquel que ocorre cando non ocorre A , e denótase por A^c ou \bar{A} .
- **Unión:** dados dous sucesos A e B en Ω , defínese o suceso unión de A e B como aquel que ocorre cando ocorre A ou ocorre B , e denótase por $A \cup B$.
- **Intersección:** dados dous sucesos A e B en Ω , defínese o suceso intersección de A e B como aquel que ocorre cando ocorre A e ocorre B , e denótase por $A \cap B$.
- **Diferenza:** dados dous sucesos A e B en Ω , defínese o suceso diferenza entre A e B como aquel que ocorre cando ocorre A e non ocorre B , e denótase por $A \setminus B$.
- **Sucesos incompatibles:** dados dous sucesos A e B en Ω , dise que A e B son incompatibles se a súa intersección é o conxunto baleiro, é dicir, se $A \cap B = \emptyset$.

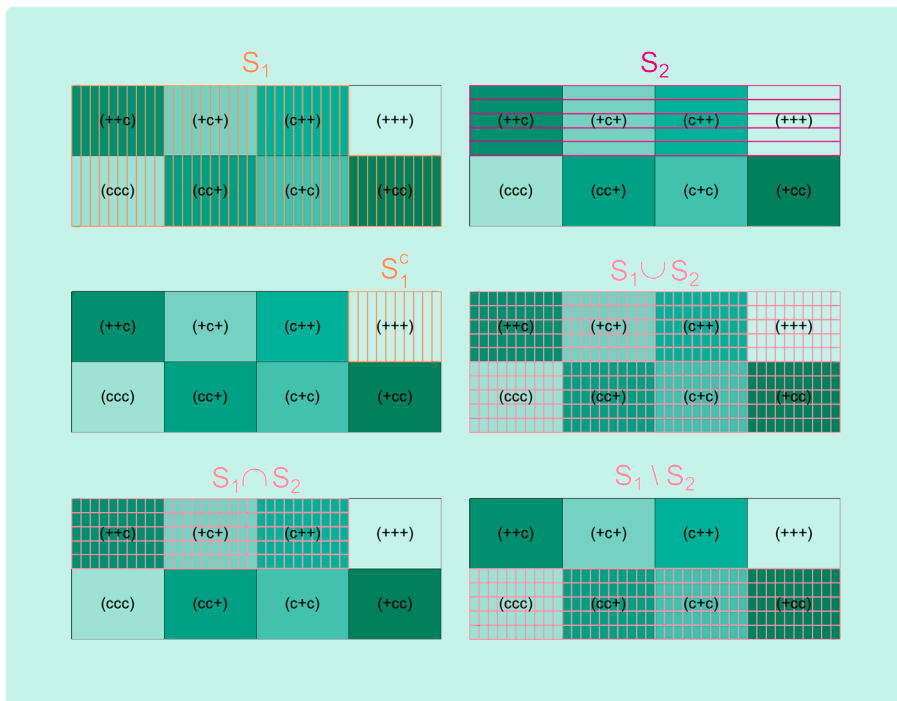
Estas operacións entre sucesos verifican unha serie de propiedades que se deben de coñecer para traballar con elas.

- **Asociativa:** $A \cup (B \cup C) = (A \cup B) \cup C$ e $A \cap (B \cap C) = (A \cap B) \cap C$.
- **Conmutativa:** $A \cup B = B \cup A$ e $A \cap B = B \cap A$.
- **Distributiva:** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ e $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- **Neutro:** \emptyset para a unión, é dicir, $A \cup \emptyset = A$; e Ω para a intersección, é dicir, $A \cap \Omega = A$.
- **Complementario:** $A \cup A^c = \Omega$ e $A \cap A^c = \emptyset$.
- **Leis de De Morgan:** $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$.

Exemplo 2. Retomando o Exemplo 1 sobre o lanzamento de 3 moedas e empregando os sucesos A e B xa definidos, imos realizar algunhas operacións:

- **Complementario:** $S_1^c = \text{“Non saír ningunha cara”} = \{(+++)\}$.
- **Unión:** $S_1 \cup S_2 = \text{“Saír polo menos unha cara” ou “Saír máis dunha cruz”} = \{(ccc), (cc+), (c+c), (+cc), (c++), (+c+), (++c), (+++)\} = \Omega$.
- **Intersección:** $S_1 \cap S_2 = \text{“Saír polo menos unha cara” e “Saír máis dunha cruz”} = \{(c++), (+c+), (++c)\}$.
- **Diferenza:** $S_1 \setminus S_2 = \text{“Saír polo menos unha cara” e non “Saír máis dunha cruz”} = \{(ccc), (cc+), (c+c), (+cc)\}$.
- Ademais, S_1 e S_2 son sucesos compatibles xa que $S_1 \cap S_2 \neq \emptyset$.

A continuación pode verse unha ilustración gráfica destas operacións:



2. Definición de probabilidade

No estudo de experimentos aleatorios non só é importante coñecer os posibles resultados senón saber tamén con que probabilidade se produce cada un deles. A continuación detállanse as distintas posibilidades á hora de definir unha probabilidade segundo a perspectiva que se adopte.

Definición de Laplace

Supóñase que Ω ten un número finito de sucesos elementais e que todos eles ocorren coa mesma probabilidade. Entón, a probabilidade dun suceso A en Ω , $\mathbb{P}(A)$, calcúlase como o cociente entre casos favorables e casos posibles.

Definición frecuentista

Se repetimos n veces un experimento aleatorio e n_A denota o número de veces que ocorre o suceso A , a frecuencia relativa deste suceso defínese como $fr(A) = \frac{n_A}{n}$, onde $0 \leq fr(A) \leq 1$. Ao facer n grande, esta frecuencia estabilízase nun valor e defínese a probabilidade do suceso A como o límite das frecuencias cando $n \rightarrow \infty$.

Definición axiomática de Kolmogorov

Considérese un espazo mostral Ω e unha familia de subconxuntos do mesmo denotada por \mathcal{A} , dise que \mathbb{P} é unha probabilidade en (Ω, \mathcal{A}) se verifica as seguintes condicións:

- a probabilidade do espazo mostral é a unidade, é dicir, $\mathbb{P}(\Omega) = 1$;
- dada unha familia de sucesos $\{A_i\}_{i=1}^{\infty}$, disxuntos dous a dous, $A_i \cap A_j = \emptyset$ se $i \neq j$, entón verifícase que

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i);$$

- dado calquera suceso $A \in \mathcal{A}$, entón, $0 \leq \mathbb{P}(A) \leq 1$.

Exemplo 3. *Retomando o Exemplo 1 consistente no lanzamento de 3 moedas, quérense calcular as probabilidades dos sucesos $S_1 =$ “Saír a lo menos unha cara” e $S_2 =$ “Saír máis dunha cruz”, previamente definidos. Neste escenario, sábese que os sucesos elementais son equiprobables con probabilidade $1/8$, polo que se ten:*

$$\mathbb{P}(S_1) = \frac{7}{8} \quad e \quad \mathbb{P}(S_2) = \frac{4}{8}.$$

Pódense calcular tamén as probabilidades dos sucesos resultantes das operacións realizadas no Exemplo 2.

- *Complementario:* $\mathbb{P}(S_1^c) = 1 - \mathbb{P}(S_1) = 1 - \frac{7}{8} = \frac{1}{8}$.
- *Unión:* $\mathbb{P}(S_1 \cup S_2) = \mathbb{P}(\Omega) = 1$.
- *Intersección:* $\mathbb{P}(S_1 \cap S_2) = \frac{3}{8}$.
- *Diferenza:* $\mathbb{P}(S_1 \setminus S_2) = \mathbb{P}(S_1) - \mathbb{P}(S_1 \cap S_2) = \frac{7}{8} - \frac{3}{8} = \frac{4}{8}$.

3. Probabilidade condicionada. Independencia de sucesos**3.1. Probabilidade condicionada**

Como xa se indicou, a teoría da probabilidade está ligada á incerteza, é dicir, á nosa ignorancia sobre os resultados dun experimento. O feito de que ocorra ou non un determinado suceso (ter ese coñecemento) pode cambiar as probabilidades dos demais. A probabilidade condicionada pon de manifesto o feito de que as probabilidades cambian cando a información dispoñible varía.

Dados dous sucesos A e B en Ω , con $\mathbb{P}(B) > 0$, defínese formalmente a

probabilidade de A condicionada ao suceso B como:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

É importante notar que esta definición é consistente no sentido de que cumpre os axiomas da definición dunha probabilidade. Ademais, no caso particular de que $B = \Omega$, tense que

$$\mathbb{P}(A|\Omega) = \frac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(\Omega)} = \mathbb{P}(A).$$

3.2. Independencia de sucesos

Dados dous sucesos A e B en Ω dise que son independentes se se verifica que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Esta condición resulta equivalente a dicir que $\mathbb{P}(A|B) = \mathbb{P}(A)$ se $\mathbb{P}(B) \neq 0$.

4. Resultados notables: regra do produto, teorema de probabilidades totais e teorema de Bayes

Unha vez definido formalmente o concepto de probabilidade, é preciso dispoñer de directrices sobre como traballar con el. A continuación preséntanse os resultados ou leis esenciais para o manexo de probabilidades, pero antes de centrarnos neles, hai que introducir un novo concepto que será de grande utilidade.

Sistema completo de sucesos: dada unha familia de sucesos $\{A_1, \dots, A_n\}$ en Ω , dise que forma un sistema completo de sucesos se verifica que: $A_i \cap A_j = \emptyset$ se $i \neq j$ e ademais $\cup_{i=1}^n A_i = \Omega$.

Exemplo 4. Retómase Exemplo 1 para visualizar este novo concepto. Nese caso pódense considerar a familia formada por todos os sucesos elementais, que verifican as condicións previamente indicadas e conforman polo

tanto un sistema completo de sucesos:

(++c)	(+c+)	(c++)	(+++)
(ccc)	(cc+)	(c+c)	(+cc)

4.1. Regra do produto

Tendo en conta a definición de probabilidade condicionada, dados dous sucesos A e B en Ω tales que $\mathbb{P}(B) > 0$, pódese calcular a probabilidade do suceso $A \cap B$ como:

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Do mesmo xeito, se consideramos $\{A_1, \dots, A_n\}$ un conxunto de sucesos de Ω , tales que $\mathbb{P}(\cap_{i=1}^n A_i) > 0$, entón a intersección de todos os sucesos A_i pode calcularse como segue:

$$\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|\cap_{i=1}^{n-1} A_i).$$

4.2. Teorema de probabilidades totais

Sexa $\{A_1, \dots, A_n\}$ un sistema completo de sucesos en Ω tal que $\mathbb{P}(A_i) > 0$ para todo $i = 1, \dots, n$. Se consideramos un suceso B en Ω , tense que:

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i).$$

4.3. Teorema de Bayes

Sexa $\{A_1, \dots, A_n\}$ un sistema completo de sucesos en Ω tal que $\mathbb{P}(A_i) > 0$ para todo $i = 1, \dots, n$. Se consideramos un suceso B en Ω tal que $\mathbb{P}(B) > 0$, verificase que:

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}{\sum_{i=1}^n \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)}.$$

5. Prevalencia e incidencia dunha enfermidade

As leis de probabilidade descritas son fundamentais no eido das ciencias da saúde. Dous conceptos especialmente relevantes en epidemioloxía son o de prevalencia e incidencia dunha enfermidade que definimos a continuación.

- **Prevalencia:** proporción de individuos da poboación que presentan unha certa enfermidade;
- **Incidencia:** número de casos novos dunha enfermidade nun período de tempo dado, é dicir, é unha taxa que cuantifica as persoas que contraen unha certa enfermidade nun período de tempo.

6. Probas diagnósticas: sensibilidade, especificidade e valores predictivos

Outro aspecto das ciencias da saúde no que a probabilidade ten unha grande importancia é a avaliación de probas diagnósticas. Unha proba diagnóstica é calquera procedemento que pretende determinar nunha/nun paciente a presenza dunha certa patoloxía non susceptible de ser observada directamente.

Sería desexable poder determinar con absoluta certeza a presenza dunha patoloxía, é dicir, que as probas diagnósticas non presentasen erro ningún. Evidentemente isto non é posible, polo que aparecen os denominados falsos positivos e falsos negativos.

- **Falso positivo:** ocorre cando unha proba diagnóstica indica a presenza da patoloxía, cando en realidade a/o paciente está sá/san.
- **Falso negativo:** ocorre cando unha proba diagnóstica non indica a presenza da patoloxía, cando en realidade a/o paciente está enferma/o.

É interesante poder determinar en que medida unha proba diagnóstica clasifica correctamente aos individuos, é dicir, detecta a presenza da patoloxía en pacientes enfermas/os ou non a detecta en pacientes sas/sans. Para isto precisamos definir dous novos conceptos probabilísticos totalmente vencellados ás probas diagnósticas.

- **Sensibilidade:** é a probabilidade de que unha proba diagnóstica determinada ofrezca un resultado positivo cando a/o paciente está enferma/o, e denótase habitualmente por $\mathbb{P}(+|E)$.
- **Especificidade:** é a probabilidade de que unha proba diagnóstica determinada ofrezca un resultado negativo cando a/o paciente está sá/san e denótase habitualmente por $\mathbb{P}(-|S)$.

Nótase que unha proba diagnóstica será tanto mellor canto máis preto estean a súa sensibilidade e a súa especificidade de valer 1.

Outra cuestión de relevancia é determinar a probabilidade de acerto unha vez que a proba ofrece un resultado. Para isto defínense os valores predictivos.

- **Valor predictivo positivo:** é a probabilidade de que unha/un paciente presente a patoloxía, cando xa se coñece que a proba ofreceu un resultado positivo. Habitualmente denótase por $\mathbb{P}(E|+)$.

- **Valor predictivo negativo:** é a probabilidade de que unha/un paciente non presente a patoloxía, cando xa se coñece que a proba ofreceu un resultado negativo. Habitualmente denótase por $\mathbb{P}(S|-)$.

Nota.- Hai que ter en conta que os valores predictivos dependen da prevalencia da enfermidade, que é algo que non ocorre coa sensibilidade e a especificidade que son valores intrínsecos da proba diagnóstica que se está a avaliar.

Exemplo 5. *Unha proba diagnóstica deseñada para detectar cancro de colo de útero presenta unha probabilidade de falso negativo de 0.05 e unha probabilidade de falso positivo de 0.01. Sábese que a poboación de mulleres obxecto de estudo presenta unha prevalencia de cancro de colo de útero do 4%. Entón:*

- Cal é a probabilidade de que unha muller tomada ao azar resulte positiva na proba diagnóstica?;*
- sabendo que unha muller obtivo un resultado positivo na proba, cal é a probabilidade de que presente cancro de útero?;*
- calcula a sensibilidade e especificidade desta proba diagnóstica.*

O primeiro paso é definir apropiadamente tanto o experimento aleatorio como os sucesos de interese. Tamén é importante facer un esquema dos datos proporcionados no enunciado:

- *Experimento aleatorio: coller unha muller ao azar na poboación de interese, determinar o resultado da proba diagnóstica e determinar se presenta ou non cancro de colo de útero.*
- *Sucesos: + = "A proba diagnóstica resulta positiva", - = "A proba diagnóstica resulta negativa", E = "A muller padece cancro de colo de útero" e S = "A muller non padece cancro de colo de útero".*
- *Datos dispoñibles: $\mathbb{P}(+|S) = 0.01$, $\mathbb{P}(-|E) = 0.05$ e $\mathbb{P}(E) = 0.04$.*

Agora procederase á resolución das cuestións presentadas. En primeiro lugar, para responder ao apartado a) teríamos que:

$$\begin{aligned}\mathbb{P}(+) &= \mathbb{P}(+|E) \cdot \mathbb{P}(E) + \mathbb{P}(+|S) \cdot \mathbb{P}(S) \\ &= (1 - \mathbb{P}(-|E)) \cdot \mathbb{P}(E) + \mathbb{P}(+|S) \cdot (1 - \mathbb{P}(E)) \\ &= (1 - 0.05) \cdot 0.04 + 0.01 \cdot (1 - 0.04) = 0.05.\end{aligned}$$

En segundo lugar, para calcular a probabilidade que se pide no apartado b)

sabemos que:

$$\begin{aligned}\mathbb{P}(E|+) &= \frac{\mathbb{P}(+|E) \cdot \mathbb{P}(E)}{\mathbb{P}(+)} = \frac{(1 - \mathbb{P}(-|E)) \cdot \mathbb{P}(E)}{\mathbb{P}(+)} \\ &= \frac{(1 - 0.05) \cdot 0.04}{0.05} = 0.76.\end{aligned}$$

Finalmente, para responder ao apartado c) debemos calcular:

$$\text{Sensibilidade} = \mathbb{P}(+|E) = 1 - \mathbb{P}(-|E) = 1 - 0.05 = 0.95$$

$$\text{Especificidade} = \mathbb{P}(-|S) = 1 - \mathbb{P}(+|S) = 1 - 0.01 = 0.99.$$

7. Medidas de efecto: risco relativo e odds-ratio

En epidemioloxía empréganse diversas medidas na análise e estudo das patoloxías de interese. En particular, o risco é a probabilidade de ocorrencia dun certo evento (enfermar, morrer, curar, empeorar...). Esta medida absoluta non permite facer comparacións entre distintos escenarios, polo que xorde o concepto de risco relativo.

Risco relativo: defínese como o cociente entre a probabilidade (risco) de ocorrer o evento de interese (enfermar, morrer, curar, empeora ...) nun grupo exposto a un certo factor ou condición e a probabilidade (risco) de ocorrer ese mesmo evento nun grupo sen exposición ao factor ou condición:

$$RR = \frac{\mathbb{P}(\text{evento}|\text{Factor})}{\mathbb{P}(\text{evento}|\text{Non factor})}.$$

A interpretación do RR é sinxela: se $RR > 1$ indica efectos nocivos do factor, é dicir, trátase dun factor de risco; se $RR < 1$ indica efectos beneficiosos do factor, é dicir, trátase dun factor protector. No caso límite de $RR = 1$, pódese afirmar que non hai asociación entre a presenza do factor e o evento de interese.

Na situación particular na que o grupo de suxeitos que sofren o evento é pequeno comparado co número de suxeitos que non o sofren, o risco relativo aproxímase pola denominada odds-ratio (OR). Ilustremos isto cunha táboa de continxencia xenérica:

		Evento	
		Si	Non
Factor	Presente	a	b
	Ausente	c	d

Así teríamos o seguinte:

$$RR = \frac{a/(a+b)}{c/(c+d)} \quad \text{e} \quad OR = \frac{a/b}{c/d} = \frac{ad}{bc}.$$

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 %.

Esta unidade didáctica avaliarase conxuntamente coas outras unidades didácticas que forman parte do Bloque II. Teoría da probabilidade supoñendo un peso do 6 % na nota final da materia. Esta avaliación engloba os oito obxectivos específicos desta unidade e consistirá na entrega dun exercicio realizado de xeito individual por cada alumna/o como traballo non presencial.

BIBLIOGRAFÍA

Bibliografía básica

- BORRAJO, M. I., CONDE-AMBOAGE, M. E CRUJEIRAS-CASAI, R. M. (2021). *Fundamentos da teoría da probabilidade*. Esenciais USC. Universidade de Santiago de Compostela.
- CRUJEIRAS-CASAI, 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) *Estadística 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). *Estadística Aplicada a las Ciencias de la Salud*. Editorial Díaz de Santos.
- 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.
- 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 esta sección pretende proporcionar recursos de calidade para diferentes fins dentro do proceso de aprendizaxe.

— **Academic Skills Kit (ASK)**

https://mas-coursebuild.ncl.ac.uk/lti/content/ASK/default/business/introduction_to_prob/index.html

A Universidade de Newcastle pon a disposición de calquera usuario da Rede cursos de diversas temáticas. En particular, a páxina de introdución á probabilidade ten un curso de iniciación bastante completo, no que se definen todos os conceptos básicos de probabilidade, se propoñen actividades extra, que poden ser vídeos, exemplos ilustrativos, exercicios, outros recursos electrónicos ... Ademais, todos os exercicios propostos son acompañado dunha solución detallada e mesmo hai dispoñibles test de autoavaliación por bloques.

— **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 dedicada 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 da Unidade de Bioestatística 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 Bioestatística do Hospital Ramón y Cajal. Inclúe contido metodolóxico, pero tamén problemas e solucións. Dispón ademais de numerosas referencias bibliográficas de interese.

— **SOCR**

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

SOCR é un recurso interactivo gratuito 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. En relación a esta unidade

dispón de numerosas simulacións de distintos experimentos probabilísticos.

— **Vídeos TED**

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

A oferta de charlas TED é case interminable, do mesmo xeito que os temas que tratan. En relación con esta unidade didáctica, destacamos na súa web unha sección, TED-ED (*TED Education*). No apartado *Data analysis and probability* <https://ed.ted.com/lessons?category=data-analysis-probability>, na que se poden consultar vídeos explicativos curtos sobre distintos conceptos de probabilidade.



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