

materia
Estatística

unidade didáctica 3

Inferencia estatística

Rosa M^a Crujeiras Casais

Departamento de Estatística e Investigación Operativa
Escola Técnica Superior de Enxeñaría

titulación

Grao en Enxeñaría Informática



VICERREITORÍA DE ESTUDANTES,
CULTURA E FORMACIÓN CONTINUA



unidade didáctica 3

Inferencia estatística

Rosa M^a Crujeiras Casais

Departamento de Estatística e Investigación Operativa
Escola Técnica Superior de Enxeñaría



© Universidade de Santiago de Compostela, 2013



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.g>

Deseño
Unidixital
Servizo de Edición Dixital
da Universidade de Santiago de Compostela

Edita
Vicerreitoría de Estudantes,
Cultura e Formación Continua
da Universidade de Santiago de Compostela
Servizo de Publicacións
da Universidade de Santiago de Compostela

Imprime
Unidixital
Dep. Legal: C 55 - 2013
ISBN 978-84-9887-958-2

ADVERTENCIA LEGAL: reservados todos os dereitos. Queda prohibida a duplicación, total ou parcial desta obra, en calquera forma ou por calquera medio (elec-trónico, mecánico, gravación, fotocopia ou outros) sen consentimento expreso por escrito dos editores.

MATERIA: Estatística
TITULACIÓN: Grao en Enxeñaría Informática
PROGRAMA XERAL DO CURSO
Localización da presente unidade didáctica

Unidade Didáctica I. Estatística Descritiva

ESTADÍSTICA DESCRIPTIVA

Conceptos xerais. Distribucións de frecuencias. Representacións gráficas. Medidas características: posición, dispersión e forma. Estatística descritiva bidimensional. Recta de regresión.

Unidade Didáctica II. Probabilidade

FUNDAMENTOS DE PROBABILIDADE

Experimento aleatorio. Sucesos e espazo mostral. Asignación e definición de probabilidade. Operacións con sucesos. Probabilidade condicionada. Independencia de sucesos. Teoremas clásicos: regra do produto, teorema da probabilidade total, teorema de Bayes.

VARIABLE ALEATORIA

Concepto de variable aleatoria. Variable aleatoria discreta. Función de masa de probabilidade. Función de distribución. Medidas características. Variable aleatoria continua. Función de distribución. Función de densidade. Cálculo de medidas características. Modelos de distribucións discretas. Modelos de distribucións continuas. Variables aleatorias bidimensionais. O modelo de regresión lineal simple.

Unidade Didáctica III. Inferencia Estatística

INTRODUCCIÓN Á INFERENCIA ESTADÍSTICA

Introdución á Inferencia Estatística. Distribucións na mostraxe en poboacións normais. Teorema Central do Límite. Relacións entre distribucións.

ESTIMACIÓN DE PARÁMETROS

Concepto de estimador. Estimación de media, varianza e proporción. Propiedades dos estimadores. Estimación por intervalos de confianza. Intervalos de confianza para unha poboación: intervalos para a media, a varianza e a proporción. Estimación no modelo de regresión lineal simple.

CONTRASTE DE HIPÓTESES

Introdución ao contraste de hipóteses. Procedemento de contraste. Contrastes nunha e en dúas poboacións: media, varianza, proporción, diferenza de medias e cociente de varianzas. Contrastes no modelo de regresión lineal simple. Contrastes de tipo chidadrado. Contrastes de normalidade.

ÍNDICE

| | |
|--|----|
| Presentación | 7 |
| Os obxectivos | 8 |
| Os contidos | 9 |
| 1. Introducción á Inferencia Estatística | 9 |
| 2. Estimación de parámetros | 11 |
| 3. Contraste de hipóteses | 13 |
| Metodoloxía e actividades propostas | 15 |
| Avaliación da UD | 18 |
| Anexo | 19 |
| Bibliografía | 20 |

PRESENTACIÓN

As bases da probabilidade e a estatística, segundo a *Association for Computing Machinery* e a *IEEE Computer Society*, subxacen aos conceptos de fiabilidade, seguridade e dependencia, factores fundamentais no ámbito da Enxeñaría en xeral e da Enxeñaría Informática, en particular. Con esta premisa, as devanditas asociacións elaboraron unhas guías para o deseño dos currículos formativos nos Graos en Enxeñaría Informática, onde se establece como obxectivo fundamental que o alumnado destes títulos adquiera coñecementos dos modelos de probabilidade discretos e continuos, das distribucións na mostraxe, dos métodos de estimación e contraste e do modelado a través da regresión.

No currículo do Grao en Enxeñaría Informática da Universidade de Santiago de Compostela (USC), a Estatística é unha materia do segundo semestre do primeiro curso, pertencente ao módulo de Matemáticas xunto con Álgebra, Fundamentos de Matemáticas, Matemática Discreta (primeiro semestre) e Cálculo e Análise Numérica (segundo semestre). Todas estas son materias básicas e polo tanto de carácter xeral, nas que se debe realizar unha tarefa de nivelación dos coñecementos do alumnado, proporcionando asemade os alicerces metodolóxicos para desenvolvementos posteriores noutras áreas como a representación gráfica ou a xestión empresarial.

Esta Unidade Didáctica (UD) céntrase nos contidos da Inferencia Estatística, cuxas actividades formativas comprenden aproximadamente un terzo das actividades docentes, tanto presenciais como no presenciais da materia. Estatística conta con seis créditos ECTS, cuxos contidos se estruturan en tres UD:

- [UD I] Estatística Descritiva.
- [UD II] Probabilidade.
- [UD III] Inferencia Estatística

Nesta unidade didáctica presentaranse inicialmente os obxectivos da UD III, detallando a continuación os contidos máis relevantes da mesma considerando á súa vez tres bloques temáticos diferenciados (introdución á Inferencia Estatística, estimación de parámetros e contraste de hipóteses). Tamén se describirá a metodoloxía empregada tanto nas sesións expositivas como nas interactivas, así como as actividades de avaliación previstas.

OBXECTIVOS

Dende unha perspectiva global da materia e da titulación, o proxecto formativo fundaméntase nas competencias que debe acadar o alumnado, a nivel técnico, metodolóxico e actitudinal. Estas competencias deben construírse sobre coñecementos sólidos, que non son obxectivos terminais, se ben estes si se poden presentar de maneira específica vinculados a cada UD do programa.

No caso da UD de Inferencia Estatística, os obxectivos específicos que debe acadar o alumnado son os seguintes:

[OB1] Discriminar entre os obxectivos dunha análise estatística, descritiva ou inferencial.

[OB2] Estimar parámetros descoñecidos dunha poboación a partir dunha mostra.

[OB3] Coñecer os principios e aplicación dos contrastes de hipóteses estatísticas.

[OB4] Comparar dúas poboacións a partir de parámetros característicos e descoñecidos das mesmas.

[OB5] Formular problemas reais en termos estatísticos (estimación de parámetros, contraste de hipóteses) e aplicar as técnicas da Inferencia Estatística á súa resolución.

Aínda que as competencias coas que se vinculan estes obxectivos non se presentan de maneira detallada na UD, cabe salientar que os obxectivos específicos considerados están relacionadas con competencias técnicas e metodolóxicas. Con todo, e como se verá no apartado dedicado á metodoloxía docente, algunhas das actividades propostas para o desenvolvemento desta UD contribúen á mellora da capacidade crítica e creativa do alumnado, así como á súa habilidade para o traballo en equipo, sendo estas competencias de tipo actitudinal.

OS CONTIDOS

Para unha selección adecuada dos contidos a incluír na UD, debe terse presente que o obxectivo final da Inferencia Estatística é tirar conclusións sobre una poboación (conxunto homoxéneo de individuos sobre os que se estudan características observables). Estas conclusións sobre o comportamento da poboación poden referirse a aspectos paramétricos, a comportamentos globais da distribución ou a comparativas con outras poboacións ou entre grupos dentro da mesma. Así, as técnicas da Inferencia Estatística poderían clasificarse en dous tipos, atendendo ao seu obxectivo final, se ben ambos os dous están estreitamente relacionados: a estimación de parámetros, de maneira puntual ou mediante intervalos, e o contraste de hipóteses.

Abordar estes problemas dende un enfoque xeral presenta unha gran dificultade matemática, que excedería os coñecementos do alumnado de primeiro curso, polo que os contidos desta UD restrinxiranse, nun primeiro termo, a poboacións normais e ao estudo de proporcións. Malia a imposibilidade dun tratamento amplo do problema da estimación de parámetros, incluíranse algunhas breves mencións ás metodoloxías xerais de estimación, rematando esta UD con algunhas nocións sobre contrastes non paramétricos.

Os contidos desta UD atópanse distribuídos en tres bloques temáticos, consecutivos no tempo, dado que cada un deles senta as bases do seguinte. Cabe destacar que non todos os contidos aquí incluídos deben ser necesariamente tratados en sesións expositivas. No apartado dedicado á metodoloxía docente achegaranse algunhas suxestións para a distribución dos contidos tanto en actividades presenciais como non presenciais. Tamén cabe salientar que a utilización de simulacións durante as sesións expositivas pode facilitar a comprensión dalgúns contidos, como é o caso das distribucións na mostraxe. Indicarase ao longo desta sección onde a ilustración mediante simulación pode resultar de apoio, amosando algunhas saídas gráficas. As gráficas presentadas foron elaboradas co software estatístico R. No Anexo, inclúese unha breve presentación da librería *rpanel*, en concreto, da utilidade *rpcartoons* que permite realizar algunhas ilustracións gráficas relacionadas cos contidos desta UD.

1. Introducción á Inferencia Estatística

Introdución á Inferencia Estatística.
Distribucións na mostraxe en poboacións normais.
Teorema Central do Límite.
Relacións entre distribucións.

O primeiro bloque desta UD comezarase introducindo algúns conceptos xerais (parámetro e estatístico) e explicaranse brevemente algunhas técnicas de mostraxe, en particular, a mostraxe aleatoria simple. Baixo esta condición, introduciranse as distribucións na mostraxe dalgúns estatísticos

de interese, como a proporción mostral e a media e a varianza mostrais en poboacións normais, así como a distribución da diferenza de medias e o cociente de varianzas en poboacións normais. Na Figura 1 amósanse dous exemplos ilustrativos da distribución na mostraxe da media e da varianza mostrais. Ademais da densidade teórica para a media mostral (liña continua), inclúense en trazo discontinuo a estimación non paramétrica tipo núcleo da densidade, que debería ter sido introducida nas sesións prácticas correspondentes á UD I, como ferramenta descritiva.

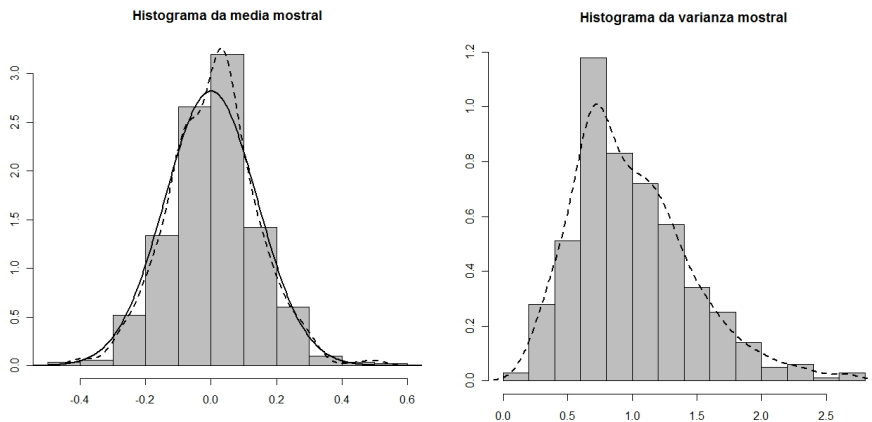


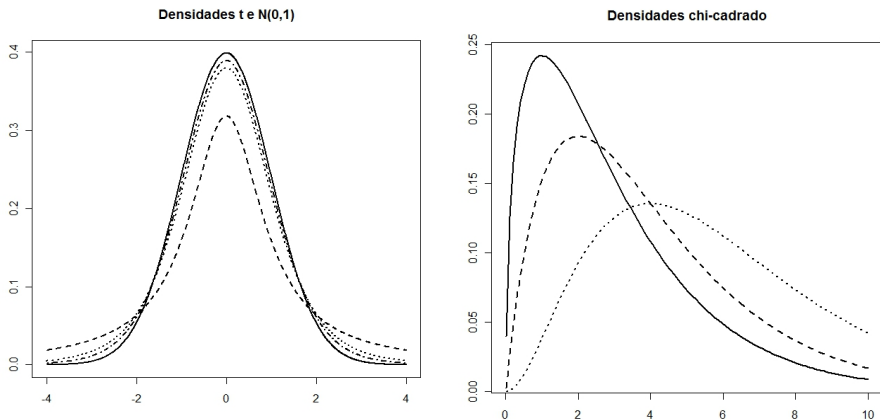
Figura 1. Histograma da media mostral (esquerda) e da varianza mostral (dereita) para mostras dunha poboación normal estándar. Liña sólida: densidade teórica do estimador. Liña de guións: estimación non paramétrica da densidade do estimador.

A distribución da proporción e a media mostrais poden obterse de maneira aproximada a partir do Teorema Central do Límite. No caso de variables normais con varianza coñecida, a distribución da media mostral (en condicións de mostraxe aleatoria simple) resulta das propiedades da distribución normal. A distribución destes estatísticos obtense de maneira sinxela a partir dos conceptos estudados na UD II da materia, correspondente á parte de probabilidade. Con todo, será necesario introducir outros modelos de distribución para determinar a distribución da media mostral cando non se coñece a varianza poboación (distribución t de Student) e a distribución da varianza mostral (distribución chi-cadrado) nunha poboación.

Na Figura 2 (esquerda) represéntanse algunhas densidades tipo t de Student, con distintos graos de liberdade. Cabe salientar a similitude deste tipo de densidades coa normal estándar (de media cero e varianza unidade): trátase de densidades simétricas arredor da orixe, pero con colas máis pesadas que as da normal sendo esta un caso límite ao aumentar os graos de liberdade, que é o parámetro que controla a forma destas

distribucións. No panel dereito da Figura 2 inclúese a representación da densidade chi-cadrado, tamén con distintos graos de liberdade.

A medida que se incrementan, pode observarse que a distribución vaise volvendo simétrica. É importante que se visualice este tipo de aproximacións á normal (no caso da *t* de Student, á normal estándar, e no caso da chi-cadrado, a unha normal de media os graos de liberdade e varianza dúas veces os graos de liberdade).



*Figura 2. Exemplos de densidades *t* de Student (panel esquerdo) e chi-cadrado (panel dereito), con distintos graos de liberdade.*

No contexto de dúas poboacións (normais e independentes), estudarase a distribución da diferenza de medias mostrais, considerando os casos de varianzas coñecidas (distribución Normal) e varianzas descoñecidas (distribución *t* de Student) se os tamaños mostrais son moderados ou grandes, en cuxo caso pode resultar válida a aproximación normal e a distribución *t* de Student coa aproximación de Welch para os graos de liberdade se os tamaños son pequenos. Na mesma situación de independencia, compararanse as varianzas poboacionais a partir do seu cociente. Neste caso, o método para obter información sobre a relación entre as varianzas estará baseado no cociente das cuasivarianzas mostrais (con distribución *F* de Snedecor). A comprensión destes conceptos resulta fundamental para abordar a seguinte parte da unidade, polo que trataremos de apoiarnos en simulacións para que o alumnado se familiarice coas distribucións na mostraxe.

2. Estimación de parámetros

Concepto de estimador.

Estimación de media, varianza e proporción. Propiedades.

Estimación por intervalos de confianza.

Intervalos de confianza para unha poboación.

Estimación no modelo de regresión lineal.

Unha vez introducidos os conceptos xerais, abordaranse os problemas clásicos na Inferencia Estatística: a estimación de parámetros, puntual e por intervalos de confianza, e o contraste de hipóteses, paramétricas e non paramétricas.

O problema da estimación de parámetros comezarase coa introdución do estimador puntual e as propiedades desexables en termos de sesgo e varianza, con especial énfase no erro cadrático medio e a eficiencia para a comparación de estimadores. En particular, estudaranse as propiedades da proporción mostral, a media mostral e a varianza mostral e cuasivarianza como estimadores da proporción poboacional, esperanza (media poboacional) e varianza poboacional, respectivamente.

Para a ilustración da distribución da proporción mostral e as súas propiedades, poden utilizarse gráficas como as amosadas na Figura 3. Nestas represéntanse os histogramas de cincocentas proporcións calculadas a partir de mostras xeradas dunha distribución de Bernoulli con probabilidade de éxito 0.4, con distintos tamaños mostrais: dez (esquerda), cincuenta (centro) e cen (dereita). As tres distribucións están centradas arredor do verdadeiro valor do parámetro (ilustración da ausencia de sesgo), mentres que a medida que o tamaño mostral aumenta, a variabilidade diminúe, consecuencia de que a proporción mostral é un estimador consistente da probabilidade de éxito neste tipo de distribucións.

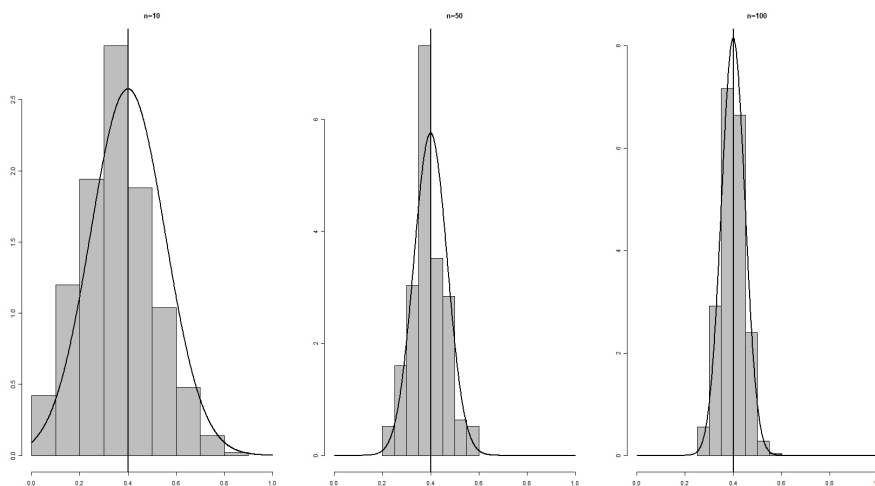


Figura 3. Ilustración da distribución da proporción mostral para distintos tamaños de mostra.

Tamén se introducirán o método dos momentos e o método de máxima verosimilitude (de maneira intuitiva) como procedementos xerais para a obtención de estimadores.

As distribucións na mostraxe que se estudaron previamente serán a base para a construción de intervalos de confianza para a proporción, a media e varianza (en poboacións normais) a través de estatísticos pivote, cuxa distribución non depende do parámetro de interese. Tamén se estudará como determinar o tamaño mostral necesario para estimar os parámetros poboacionais mediante un nivel de confianza e un erro máximo admisible prefixados.

Para finalizar este segundo bloque da UD, considerarase como proceso de xeración de datos o modelo de regresión lineal simple, que xa se tería introducido na UD II da materia, e na UD I de maneira descritiva. Os parámetros da función de regresión realizarase a través do método de mínimos cadrados, obtendo asemade a súa distribución na mostraxe e os intervalos de confianza correspondentes. Tamén se considerará a estimación da varianza do erro, a través da varianza residual.

3. Contraste de hipóteses

Introdución ao contraste de hipóteses.
Procedemento de contraste.
Contrastes nunha e en dúas poboacións.
Contrastes no modelo de regresión lineal simple.
Contrastes de independencia e normalidade.

Finalmente, tratarase o problema do contraste de hipóteses, técnica estatística cuxa finalidade é decidir se unha determinada hipótese, afirmación ou conxectura sobre a distribución poboacional estudada é confirmada ou invalidada estatisticamente a partir das observacións obtidas na mostra.

Un aspecto distintivo do contraste de hipóteses fronte ás técnicas de estimación é a existencia dunha hipóteses sobre a distribución poboación, que se denomina hipótese nula, a cal se quere contrastar fronte a unha hipótese alternativa. Comezarase introducindo os conceptos xerais (tipos de hipóteses, tipos de erro, tamaño e potencia dun contraste, nivel de significación, rexión crítica,...) e proseguirase co procedemento de contraste. Este procedemento detallarase en contrastes paramétricos sobre a media e a varianza (en poboacións normais) e sobre a proporción, e estenderase ao caso de comparación de medias, varianzas e proporcións en poboacións independentes. O contraste de comparación de medias tamén se estudará para o caso no que se teñan mostrax relacionadas.

Ao igual que no bloque temático anterior, e dado que tamén se obtiveron as distribucións dos estimadores no modelo de regresión lineal simple, traballarase os contrastes de significación sobre os parámetros da función de regresión e o test F.

Finalmente, este bloque complétase con algúns contrastes baseados na distribución chi-cadrado. Baixo unha mesma filosofía, pódense introducir as

adaptacións deste tipo de contrastes ao caso de bondade de axuste, homoxeneidade de poboacións e independencia, con aplicación a táboas de continxencia. Dentro dos contrastes de bondade de axuste, introducirase tamén o contraste de Kolmogorov-Smirnov, e dado que a distribución normal xoga un papel importante ao longo de toda a materia, tamén se considerará o contraste de Shapiro-Wilk. No estudo dos contrastes non paramétricos, non se pretende que o alumnado coñeza as propiedades teóricas dos mesmos, senón que comprendan o procedemento de construción e saiban identificar en que situacións a súa aplicación é posible, entendendo o risco inherente a calquera toma de decisións baseadas neste tipo de técnicas.

METODOLOXÍA E ACTIVIDADES

A metodoloxía docente prevista nesta UD, entendendo como tal a maneira na que docentes e alumnado confrontan e organizan o traballo co propósito de que o alumnado acade os obxectivos marcados, contempla actividades presenciais e non presenciais. Dentro das actividades presenciais, inclúense as sesións expositivas e interactivas (de resolución de problemas e en aula de informática). Como actividades non presenciais, considéranse a resolución dun caso práctico e algunhas lecturas ou actividades recomendadas, así como as horas de estudo.

Na Táboa 1 inclúese a distribución horaria das actividades previstas para o ensino/aprendizaxe dos contidos desta UD. Considéranse necesarias un total de vinte e unha horas presenciais, distribuídas en dez horas de sesións expositivas, outras dez horas de sesións interactivas (seminarios de resolución de problemas e prácticas na aula de informática) e unha hora de titorías.

| Actividades presenciais | | Actividades non presenciais | |
|---|----------|-------------------------------|----------|
| Sesións expositivas | 10 horas | Estudo individual | 15 horas |
| Sesións interactivas (seminarios) | 4 horas | Caso práctico | 4 horas |
| Sesións interactivas (aula informática) | 6 horas | Lecturas e outras actividades | 2 horas |
| Titorías | 1 horas | | |

Táboa 1. Distribución de actividades formativas presenciais e non presenciais (en horas de traballo do alumnado) para a UD de Inferencia Estatística.

Actividades presenciais

Sesións expositivas (10 horas): nas sesión de docencia expositiva explicaranse os conceptos teórico-prácticos dos contidos, apoiándose en presentacións multimedia, que se facilitarán ao alumnado. Tamén se resolverán algúns problemas tipo, de maneira que o alumnado poida traballar sobre os boletíns de exercicios propostos de maneira independente. Respecto ao material para o seguimento das sesións expositivas, ademais dos temas da materia, os boletíns de exercicios e a bibliografía recomendada, o alumnado disporá de material docente complementario (exercicios resoltos, exemplos de casos, enderezos na rede, bibliografía na rede, etc.).

Sesións interactivas de resolución de exercicios (4 horas): para as sesións interactivas de resolución de problemas, facilitaráselle ao alumnado una colección de boletíns temáticos dos que se resolverán problemas

significativos, deixando para o traballo persoal outros similares. Os problemas serán resolto polo profesorado e polo alumnado, de maneira individual ou en grupo, coa posterior exposición dos resultados dependendo da temática e/ou da necesidade de clarificar ou incidir nalgún dos contidos tratados. A planificación de actividades realizouse considerando sesións de dúas horas, se ben podería adaptarse para sesións dunha hora. Sería recomendable que o número de estudantes por grupo nas sesións interactivas de seminarios non excedese os vinte.

Dado que un dos obxectivos é que o alumnado saiba formular problemas reais en termos estatísticos, unha actividade que pode axudarlles a esta tarefa pode levarse a cabo nas sesións interactivas, ou nas expositivas, dependendo do calendario do curso, traballando os conceptos do contraste de hipóteses dunha forma máis activa, involucrando ao alumnado. Por exemplo, para introducir os contrastes sobre a media ou a proporción, é doado atopar nos medios de comunicación algún titular que resulte interesante ou sospeitoso. Elixido un ou varios titulares, pode solicitarse ao alumnado que faga unha breve enquisa sobre o tema, considerando algunhas cotas por sexo e idade, e cos datos reais recollidos, levaríase a cabo na clase o contraste do titular. Este tipo de enfoque fai que os estudantes amosen unha actitude máis activa, e que exerciten a súa capacidade crítica.

Sesións interactivas en aula de informática (6 horas): parte das sesións de docencia interactiva levaranse a cabo na aula de informática onde se introducirá ao alumnado no manexo do software R para a análise estatística de datos. Para cada unha destas sesións o alumnado disporá dun guión da práctica, da presentación utilizada como fio condutor e dun boletín de exercicios. Na aula de informática, o alumnado resolverá problemas preparados ao efecto. Ao igual que para as sesións de resolución de problemas, a planificación realizouse considerando sesións de dúas horas para grupos de vinte estudantes.

Titorías (1 hora): as titorías están destinadas ao seguimento da aprendizaxe do alumnado, considerando grupos reducidos (aproximadamente dez alumnas/os).

Actividades non presenciais

Caso práctico: o obxectivo principal das sesións interactivas na aula de informática é que o alumnado sexa quen de realizar unha análise de datos utilizando as técnicas básicas explicadas. Organizados por parellas ou grupos de traballo, esta actividade consistirá na elaboración dun informe utilizando técnicas de estatística descritiva e de inferencia estatística (construción de intervalos de confianza, contrastes de hipóteses e modelado da relación entre variables mediante regresión lineal simple), cuxa formulación e aplicación práctica se traballará nas sesións interactivas en aula de informática. A resolución dos casos prácticos farase empregando o software R. Cabe salientar que esta actividade pode ser transversal a varias

UD, dado que previo á análise inferencial resulta necesario realizar unha análise descritiva dos datos.

Resolución de exercicios: proporase a resolución de exercicios que o alumnado, de maneira individual, deberá entregar para a súa avaliación.

Outras actividades: outras actividades de aprendizaxe non presencial son as horas de estudo individual e de consulta bibliográfica e/ou recursos na rede. Nesta UD proponse a consulta de applets. Un applet é unha compoñente dunha aplicación que se executa no contexto doutro programa, por exemplo, dun navegador web. Moitos dos applets que se poden atopar na rede serven para ilustrar conceptos estatísticos básicos, por exemplo, sobre estimación no modelo de regresión lineal simple, cobertura de intervalos de confianza ou elementos dos contrastes de hipóteses.

AVALIACIÓN DA UNIDADE DIDÁCTICA

A continuación detállanse os procedementos de avaliación de cada un dos obxectivos da UD, presentados na segunda sección desta memoria.

[OB1]--[OB4] Mediante a resolución de exercicios, avaliarase se o alumnado é quen de discriminar entre unha análise estatística descritiva ou inferencial. Seguindo as indicacións que se darán nas sesións expositivas, previa a resolución dos exercicios formulados, o alumnado realizará unha análise da información da que se dispón, identificando os factores mostrais e poboacionais, e determinando claramente os elementos a calcular. A través dos exercicios propostos, tamén se poderá avaliar se o alumnado coñece os principios e aplicación dos contrastes de hipóteses, tanto para unha poboación como para dúas poboacións, e se saben estimar parámetros.

O exame, que será de tipo práctico, incluírá a resolución dalgúns exercicios. Ademais, tamén se formularán algunhas cuestións breves que permitan avaliar a comprensión dos contidos teóricos relacionados fundamentalmente coa estimación de parámetros e co contraste de hipóteses. Por exemplo, poden suxerirse como cuestións curtas a avaliación das propiedades dun determinado estimador (en termos de sesgo e erro cadrático medio) ou o razoamento sobre a decisión a tomar fronte a un contraste de hipóteses dependendo do nivel de significación prefixado.

[OB5] A través do caso práctico, preténdese que os estudantes se familiaricen co procedemento de análise de datos reais, formulando e resolvendo cuestións inferenciais sobre os mesmos.

As actividades de avaliación dos obxectivos [OB1]—[OB4] terían un peso asociado do 80%, mentres que o 20% restante sería para a avaliación do caso práctico.

ANEXO

A comprensión dalgúns conceptos estatísticos pode simplificarse a través de resultados de simulación. No apartado de contidos, amosáronse algunhas gráficas elaboradas con R. Este software proporciona ademais algunhas librarías que poden resultar especialmente útiles na actividade docente. En concreto, a librería *rpanel*, dispón dunha utilidade (*rpcartoons*, véxase a Figura 4) que permite a ilustración dalgúns conceptos relacionados co cálculo de intervalos de confianza e cos contrastes de hipóteses, contidos centrais do segundo e terceiro bloque desta UD

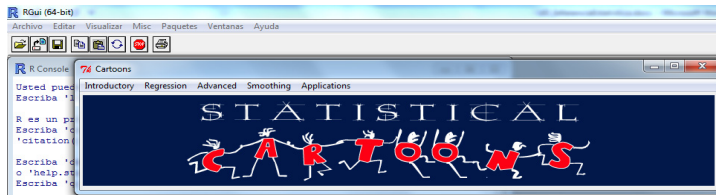


Figura 4. Menú da utilidade *rpcartoons* da librería *rpanel*.

Como pode verse na Figura 5, a utilización dos exemplos da utilidade *rpcartoons* pode facerse de maneira sinxela a través dos menús despregables.

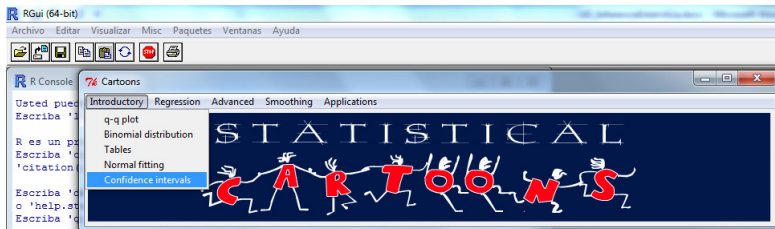


Figura 5. Exemplo de menú despregable da utilidade *rpcartoons* para cálculo de intervalos de confianza.

Para afianzar os contidos do segundo bloque temático desta UD, é importante que o alumnado comprenda a interpretación da cobertura dun intervalo e que sexa consciente do efecto que ten o tamaño da mostra empregada para o cálculo de intervalos sobre a lonxitude dos mesmos. Así, por exemplo, para o cálculo de intervalos de confianza para a media, a utilidade *rpcartoons* dispón dunha opción na que permite fixar os parámetros da distribución normal teórica considerada (media e varianza), e variar o tamaño da mostra. Na Figura 6, preséntanse dous exemplos de cálculos de intervalos de confianza para a media (valor teórico nulo, neste caso), con distintos tamaños de mostra. Estes parámetros poden ser modificados polo usuario. Por defecto, a aplicación xera cen mostras da distribución elixida. No exemplo, probáronse dous tamaños de mostra (cincuenta e cincocentos datos), observándose claramente unha diminución

da lonxitude dos intervalos. Sinálanse en vermello aqueles intervalos que non inclúen ao verdadeiro valor do parámetro. Do mesmo xeito, tamén podería ilustrarse o efecto de aumentar ou diminuír a varianza poboacional.

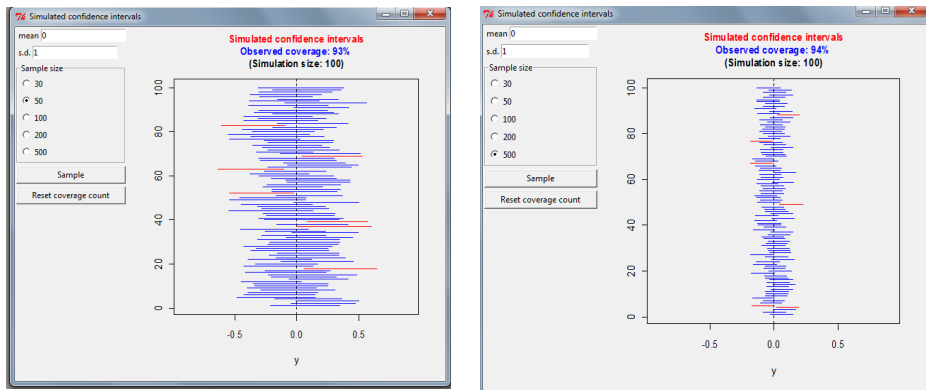
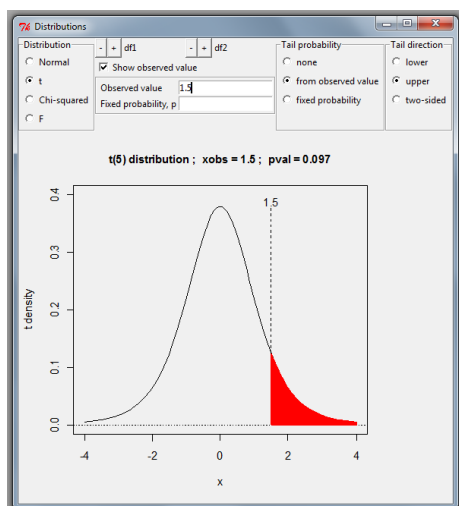


Figura 6. Ilustración da lonxitude e cobertura de intervalos de confianza para a media, nunha poboación normal, para distintos tamaños de mostra.

Para a realización de contrastes de hipóteses, unha vez calculado o valor do estatístico de contraste coa mostra observada, é necesario cuantificar o p-valor asociado (probabilidade de obter un valor máis extremo baixo a hipótese nula) para poder decidir se a hipótese nula debe ser rexeitada ou se non existen evidencias suficientes en contra da mesma.



Como se pode ver na Figura 7, a utilidade *rpartoons* tamén permite calcular p-valores para distintas distribucións, amosando graficamente a probabilidade asociada. Esta representación, simplemente indicando que se desexa representar a cola inferior e non a superior, pode utilizarse ao introducir os cuantiles na UD II da materia.

Figura 7. Cálculo de p-valores para a distribución normal, t-Student, chi-cadrado e F de Snedecor.

BIBLIOGRAFÍA

- CAO, R. E OUTROS (2005) *Introducción a la Estadística y sus Aplicaciones*. Pirámide.
- CRUJEIRAS, R. E FARALDO, P. (2010) *Estadística Básica para Ciencias de la Salud*. Unidixital
- FEBRERO, M. E OUTROS (2008) *Estadística: Ingeniería Técnica en Informática de Sistemas*. Universidade de Santiago de Compostela.
- DEVORE, J.L. (2005) *Probabilidad y Estadística para Ingeniería y Ciencias*. Paraninfo.
- MONTGOMERY, D. E RUNGER, G. (1996) *Probabilidad y Estadística aplicadas a la Ingeniería*. McGraw-Hill.

Recursos en internet

Ligazóns con applets:

http://www.math.hope.edu/swanson/statlabs/stat_applets.html

http://www.stat.tamu.edu/_west/.

R: A Language and Environment for Statistical Computing. (2012) R Development Core Team. <http://www.R-project.org>.



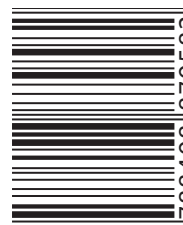
Unha colección orientada a editar materiais docentes de calidade e pensada para apoiar o traballo do profesorado e do alumnado de todas as materias e titulacións da universidade



Impreso en papel 100% reciclado e libre de cloro



SERVIZO DE NORMALIZACIÓN
LINGÜÍSTICA



9 788498 879582