

Achegas estatísticas aos métodos cuantitativos en estudos perceptivos

Simposio Dialectoloxía Perceptiva

M^a José Ginzo Villamayor

Departamento de Estadística e Investigación Operativa
Universidad de Santiago de Compostela



- 1 *Conxunto de técnicas que estudan os feitos para chegar a unha avaliación numérica dos mesmos.*
- 2 *Datos numéricos que concirnen a unha categoría de feitos.*

- 1 A Estatística é a *rama das Matemáticas* que $[\dots]$ trata coa aleatoriedade dos resultados da experiencia cotián $[\dots]$.
- 2 A estatística *axuda a todas as demais ciencias* a xerar modelos matemáticos "xerais" onde se teña considerado un *compoñente aleatorio*.

- 1 A Estatística é a *rama das Matemáticas* que $[\dots]$ trata coa aleatoriedade dos resultados da experiencia cotián $[\dots]$.
- 2 A estatística *axuda a todas as demais ciencias* a xerar modelos matemáticos "xerais" onde se teña considerado un *compoñente aleatorio*.

- *A Estatística é a primeira das ciencias inexactas.*
(Edmont Hour De Goncourt)

- *A Estatística é a primeira das ciencias inexactas.*
(Edmont Hour De Goncourt)
- *There are lies, dammed lies and statistics.*
(Mark Twain? Disraeli?)

- *A Estatística é a primeira das ciencias inexactas.*
(Edmont Hour De Goncourt)
- *There are lies, dammed lies and statistics.*
(Mark Twain? Disraeli?)
- *Eu non creo nas estatísticas.*
(Anónimo)

Estatística

A estatística é a arte de aprender a partir dos datos.

- A Estatística é unha ciencia que estuda a recolección, análise e interpretación de datos.
- Utilízase como axuda á toma de decisións ou para explicar condicións regulares ou irregulares dalgún fenómeno.
- É transversal a unha ampla variedade de disciplinas.

- A estatística agrupa un conxunto de técnicas mediante as que se recopilan, agrupan, estruturan, e finalmente, analízanse os conxuntos de datos.
- O propósito da estatística é darlles sentido a estes datos, é dicir, que nos poidan dar unha idea da situación que refrexan para tomar decisións.

As técnicas estatísticas

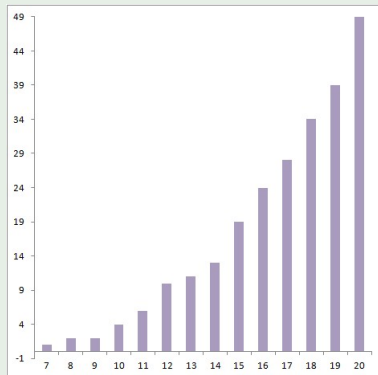
- Descritivas e inferenciais.
- Permiten a obtención de resumos numéricos e gráficos, predición, descrición de patróns subxacentes aos datos (series de tempo, procesos espaciais e espazo-temporais).
- Necesidade! Sempre que teñamos medicións suxeitas a erros (ou variabilidade) precisamos de técnicas estadísticas.

Exemplos de técnicas estatísticas no ámbito lingüístico

- Representación gráficas: histograma, diagramas de caixa, diagramas de barras, gráficos de dispersión, mapas, ...
- Contrastes para a media, mediana, varianza, ...
- Análise de regresión (relación entre dúas ou máis variables cuantitativas), ...
- Técnicas multivariantes: análise de compoñentes principais, cluster, de correspondencias, ...
- ...

- Este test realizárono 49 xuíces, do municipio de Lalín.
- Cada persoa tivo que responder a 20 preguntas.
- Polo tanto, 980 respostas.
- O número de acertos foi de 787, é dicir, acertaron o 81.31 %.

- A figura mostra o gráfico de frecuencias acumuladas de acertos na mostra.
- Non hai ningún xuíz na mostra que acertase menos de 7 preguntas.
- O 20.48 % da mostra acertaron as 20 preguntas.



- 1 A taxa de acertos é significativa?**. No seguinte sentido, **Pode considerarse que a porcentaxe de acertos é maior do 75 %?**
- Para iso realizouse un contraste de hipótese para a proporción.
 - Concluíuse que a porcentaxe de acertos é igual ou maior ao 75 %.

- 2 **Os acertos débense ao azar?** Nas 20 preguntas que evalua cada xuíz, 4 delas eran repetidas, **preguntas de control**.
- Compárase de forma individual pregunta por pregunta para ver si os xuíces responderon do mesmo xeito ou non, en cada pregunta de control.
 - Fanse **test de proporcións** para ver si contestaron da mesma forma.
 - Resulta que as proporcións de acertos nos dous grupos son iguais. \implies **Os acertos non se deben ao azar.**

- 3 Efecto da lingua por grupos.** Houbo que comparar os resultados segundo uns códigos de lingua: GC (bilingües), MC (máis castelán), MG (máis galego), SC (só castelán) e SG (só galego), e comprobar si acertaron na mesma proporción.

	Acertos	Casos	Proporción (%)	IC (95%)
GC	71	80	88.75 %	[81.201 – 96.299]
MC	68	80	85.00 %	[76.550 – 93.450]
MG	110	140	78.57 %	[71.417 – 85.726]
SC	43	60	71.66 %	[59.431 – 83.902]
SG	495	620	79.83 %	[76.600 – 83.077]

Realizáronse **test de proporcións 2 a 2**.

Test 1: Resultados

Comparación	as proporcions de acertos son iguais?.
GC con MC	Si
GC con MG	Si
GC con SC	No, (maior no grupo GC)
GC con SG	Si
MC con MG	Si
MC con SC	Si
MC con SG	Si
MG con SC	Si
MG con SG	Si
SC con SG	Si

Técnicas empregadas



Comparación de proporcións.



Contrastes de hipóteses.



Intervalos de confianza.

- Este test realizárono 49 xuíces, do municipio de Lalín.
- Cada xuíz tivo que responder a 20 preguntas.
- Polo tanto, 980 respostas.
- O número de acertos foi de 673, é dicir, acertaron o 68.67 %.

1 Tasa de acertos.

- Calculouse o **intervalo de confianza para a proporción de acertos**.

- 2 Os acertos débense ao azar?** (Nas 20 preguntas que realiza cada xuíz, 3 delas son repetidas.)
 - 1 Compárase de forma individual pregunta por pregunta para ver se os xuíces responderon do mesmo xeito ou non, en cada pregunta de control.
 - 2 Faise **test de proporcións** para ver si contestaron da mesma forma.
 - 3 As proporcións de acertos nos dous grupos son iguais. \implies **Os acertos non se deben ao azar.**

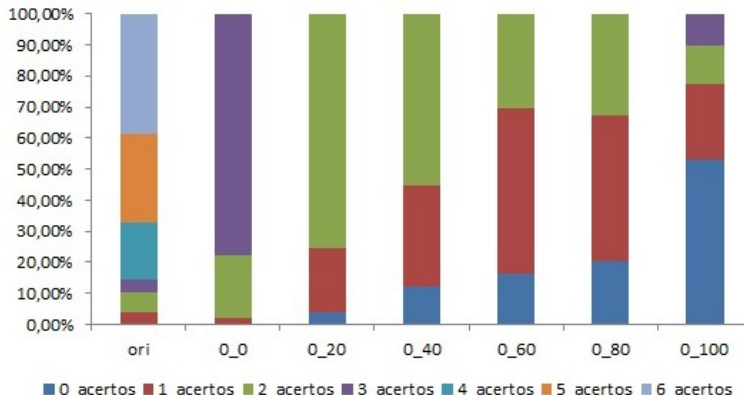
- 3 Efecto da pendente.** Compáranse os resultados nos cambios das pendentes: orixinais, 0.0, 0.20, 0.40, 0.60, 0.80 y 0.100.

	Acertos	Casos	Proporción (%)	IC (95%)
Originales	234	294	79.59 %	[74.815 – 84.369]
0	135	147	91.83 %	[87.070 – 96.603]
20	84	98	85.71 %	[78.276 – 93.153]
40	70	98	71.42 %	[61.974 – 80.883]
60	56	98	57.14 %	[46.835 – 67.451]
80	55	98	56.12 %	[45.787 – 66.457]
100	39	147	26.53 %	[19.053 – 34.008]

- Cálculase a **proporción de acertos en cada grupo**.
- A medida que aumenta a pendente o individuo non percibe si a frase é interrogativa ou non.
- Para valores baixos de pendente acertan máis que coa locución sen alterar.
- **Tablas de frecuencias de acertos.** Ao ir aumentando a pendente o número de persoas que acerta menos aumenta.
- Comparacións 2 a 2 para ver si acertaron na mesma proporción, (comparamos 2 a 2 pendentes diferentes, mediante **test de proporcións**).

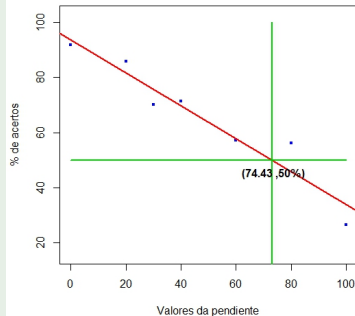
Test 3: Resultados

Na figura recóllense as frecuencias de acertos por individuos en función da pendente.



Test 3: Resultados

- 1 A recta vermella representa a recta que considera no eixo X os valores da pendente e no Y a taxa de acertos.
- 2 A recta verde horizontal marca o 50% dos acertos.
- 3 O punto de corte destas rectas é $(74.43, 50\%)$, e dá lugar á recta verde vertical $y = 74.43$.
- 4 Conclúese que pendentes por encima de 74.43 a taxa de acertos será inferior ao 50% e viceversa, pendente máis pequenas de 74.13 a taxa de acertos será superior do 50%.



- 8 Se a pendente é 74.13 o que se obtén é o 50% de acertos, o mesmo que cando se tira unha moeda ao aire.

Técnicas empregadas



Comparación de proporcións.



Contrastes de hipóteses.



Intervalos de confianza.

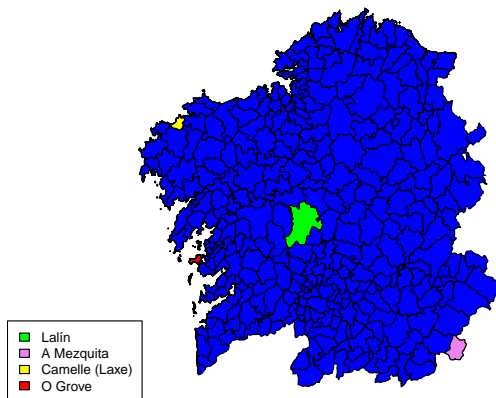
- Este test realizárono 49 xuíces, do municipio de Lalín.
- Cada xuíz tivo que responder a 22 preguntas.
- Polo tanto, 980 respostas.
- Obxetivo: medir se as locucións se parecen . . .

Distancias por estrada entre os catro puntos

O máis próximo a Lalín é Camelle.

	Lalín	Camelle	O Grove	A Mezquita
Lalín		127	130	167
Camelle	127		149	179
O Grove	130	149		181
A Mezquita	167	179	181	

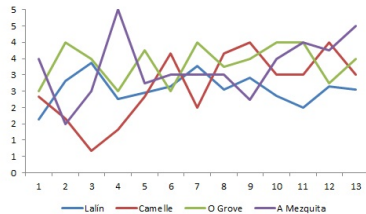
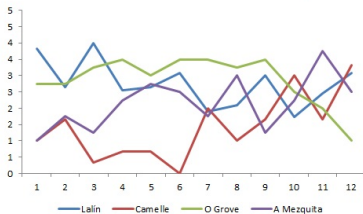
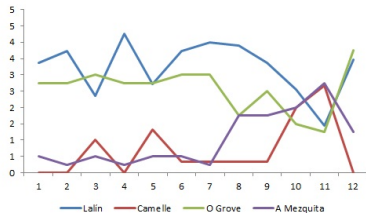
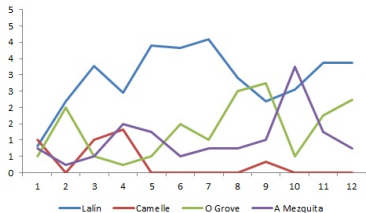
Xuíces: Lalín



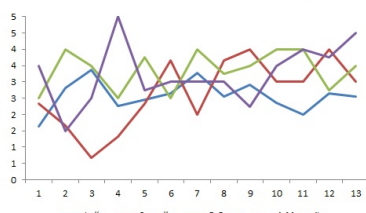
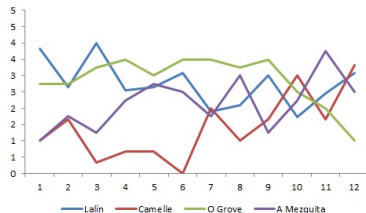
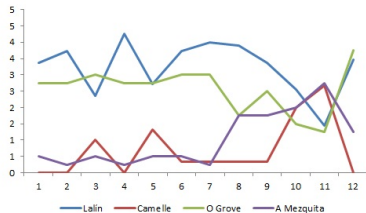
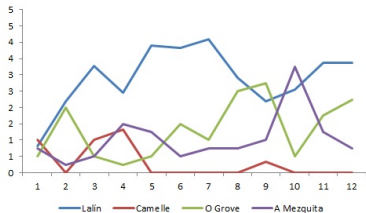
- Agrupáronse as locucións que son do mesmo lugar de orixe e
- se compararon 2 a 2 as 4 locucións.

Comparación	se valoran do mesmo modo as locucións?.
Lalín con Camelle	No
Lalín con O Grove	Si
Lalín con A Mezquita	No
Camelle con O Grove	No
Camelle con A Mezquita	No
O Grove con A Mezquita	No

Test 2: Resultados



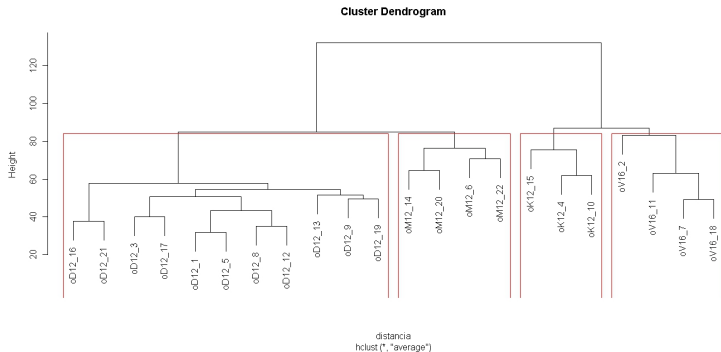
Test 2: Resultados



- Trátase de mostrar se hai grupos de individuos que puntúan de modo diferente ao resto.
- Promediamos por tipo de locución segundo sexa de Lalín, Camelle, O Grove o A Mezquita, para cada xuíz.
- Logo promediáronse todas as respostas por xuíz e repartiuse a mostra de 49 xuíces en 4 partes ordenada de menor a maior.
- Obsérvase que aqueles individuos que teñen a media total máis baixa (primeiro á esquerda) separan mellor as locucións de Lalín do resto.
- De modo análogo os que teñen total máis alto (cuarto á dereita) distinguen peor as catro locucións, é dicir puntúan todas de modo máis ou menos parecido.

Test 2: Resultados

- Por último aplicamos **análise clúster** para ver se as diferentes locucións son ben identificadas polos xuíces.



Técnicas empregadas



Comparación de proporcións.



Contrastes de hipóteses.



Intervalos de confianza.

- 1 Desenvolvemento de aplicación web para realizar os test.
- 2 Ter en conta a capacidade do xuíz.
- 3 Análise estatística dos datos:
 - na liña actual,
 - propiedades de dependencia e independencia dos datos,
 - modelos loxísticos con efectos aleatorios,
 - corrección efectos múltiples (Bonferroni, Benjamini and Hochberg, ...),
 - análise de datos funcionais,
 - realización de mapas de percepción,
 - ...

Achegas estatísticas aos métodos cuantitativos en estudos perceptivos

Simposio Dialectoloxía Perceptiva

M^a José Ginzo Villamayor

Departamento de Estadística e Investigación Operativa
Universidad de Santiago de Compostela

