Clase inicial

Curso 2024-2025

Introducción

- Asignatura relacionada con el análisis las pruebas diagnósticas
  - Tema 1: Estudio de la concordancia
  - Tema 2: Validación
- Problema asociado: Clasificar a un individuo en un grupo de categorías

Planificar un problema (ejemplo)

- Objetivo: Regalar a los alumnos del máster un anillo corporativo
- Cuestiones asociadas:
  1. ¿A qué alumnos?
     a. Solo a los del curso actual
     b. A todos los alumnos desde este curso en adelante
  2. Tengo una oferta si hago un pedido grande
  3. ¿El modelo es único? No, hay tres tallas:

  Pequeña, Mediana, Grande

Objetivo: Regalar un anillo a cada alumno

- Cuestiones asociadas:
  1. ¿A qué alumnos? A todos los alumnos
  2. Tengo una oferta si hago un pedido grande
  3. ¿El modelo es único? No, hay tres tallas: P, M y G
  4. Necesito: Medir la circunferencia del dedo de los alumnos. ¿A todos? ¿qué dedo?
  5. ¿Quién aporta la información de 4? Los alumnos del curso actual. ¿Cómo se realiza la medida?
     a. Hay varios procedimientos ¿Por ejemplo?
     b. Los resultados de cada procedimiento
        i. Concuerdan (puedo utilizar cualquiera o mezclarlos) (*1)
        ii. Hay alguno “mejor” (*2)
  6. Con (todas/parte de) las medidas de los alumnos establecemos las tallas P, M y G
  7. Decisión sobre la cantidad a pedir, en total y de cada talla
  8. Poner en práctica → REVISAR: Si hay errores volver a 4

Objetivo: Regalar un anillo a cada alumno

- Conceptos estadísticos que han aparecido:
  - Observaciones muestrales (medidas del dedo de los alumnos elegidos*)
  - ¿Estos alumnos representarán bien a los de las siguientes promociones?
  - Número de observaciones que se van a utilizar
  - Información a recoger (*, edad, sexo, peso, mano dominante, alergias ...)
  - Si utilizamos solo parte de la información muestral ¿cómo decidimos qué datos se reservan para revisar (8)--?
  - Fiabilidad de las observaciones
  - Fiabilidad de los resultados
- El problema no es puntual, queremos tomar decisiones que se van a llevar a la práctica en los próximos cursos sin conocer las características de los futuros alumnos

Planteamiento ‘estadístico’

- Necesitamos apoyarnos en un modelo teórico que podamos aplicar a cualquier promoción de alumnos:

  X=medida del dedo anular de la mano derecha de un alumno

- X es una variable aleatoria y es la base del modelo teórico
- La idea es estudiar X, conocer su comportamiento, tomar decisiones sobre X y aplicarlas sobre los alumnos ‘reales’

Planteamiento ‘estadístico’

- Necesitamos apoyarnos en un modelo teórico que podamos aplicar a … moción de alumnos:

  X=medida del dedo ar… e un alumno

Aleatoria: hay diferencias no controlables entre las medidas, pero tienen un patrón común

- X es una variable aleatoria y es la base del modelo teórico
- La idea es estudiar X, conocer su comportamiento, tomar decisiones sobre X y aplicarl… os ‘reales’

Buscar el patrón: PROBABILIDADES

Planteamiento 'estadístico'

- Modelo teórico que podamos aplicar a cualquier promoción de alumnos:

  X=medida del dedo anular de la mano derecha de un alumno

- ¿Probabilidad de que un alumno tenga un diámetro en su dedo superior a 6,5 cm?
  - P{ X> 6,5}   (A)
- ¿La variable X ha sido estudiada antes?
  – Sí. Puedo utilizar esos resultados y aplicarlos a mi problema. Ejemplo: X tiene distribución Normal(6.5, 1) ¿Cuál es el valor de la probabilidad (A)?
  – No. Debo crear el modelo y estudiarlo teóricamente

Planteamiento ‘estadístico’

- Estudio del modelo teórico que podamos aplicar a cualquier promoción de alumnos:

  X=medida del dedo anular de la mano derecha de un alumno

  Necesitamos suponer que tenemos varias copias de la variable anterior: $(X_1, X_2, ..., X_N)$

  X_1= medida del alumno 1 ...

  N= tamaño de la muestra

  ¿Quién es el alumno 1?

  El primer alumno al que yo entregue su anillo

  No podemos pensar en él como una persona concreta cuando realizamos un estudio teórico

Variables aleatorias más utilizadas

- Binomial (n,p) https://people.hsc.edu/faculty-staff/blins/StatsTools/binomialPlotter.html
- Geométrica (p) https://homepage.divms.uiowa.edu/~mbognar/applets/geo1.html
- Poisson (λ) https://homepage.divms.uiowa.edu/~mbognar/applets/pois.html
- Uniforme (a,b)
- Normal (μ, σ)
- Exponencial (λ) https://statpowers.com/distributionContinuous.html?f=exponential
- Gamma (κ, λ)
- Chi-cuadrado χ^2
- …

Resume en tus apuntes las propiedades de estas variables (p.ej. Interpretación de los parámetros, esperanza, varianza…)

Ejemplo

- Supongamos que en una ciudad se han estudiado episodios de gripe estacional
- El modelo teórico asociado a este tipo de epidemias indica que
  - La probabilidad de que en una familia la madre tenga gripe es 0.1
  - En el 12% de las familias el padre tiene gripe
  - Ambos progenitores tienen gripe en el 2% de las familias (=con probabilidad 0.02)

Distribución conjunta del estado de ambos padres

Sean las variables aleatorias Xm y Xp

Xm=1, si la madre tiene gripe

Xm=0, si la madre no tiene gripe

Xp=1, si el padre tiene gripe

Xp=0, si el padre no tiene gripe

Sabemos que $P(Xm=1)=0.1$ $P(Xp=1)=0.12$

Y que $P(Xm=1,Xp=1)=0.02$

Distribución conjunta del estado de ambos padres

La distribución conjunta satisface

		Xp		
		1	*0*	total
Xm	*1*	0,02		0,1
	0			
total		0,12		

Distribución conjunta del estado de ambos padres

La distribución conjunta satisface

		Xp		
		1	*0*	total
Xm	*1*	0,02	0,08	0,1
	0	0,10	0,80	0,9
total		0,12	0,88	**1**

La probabilidad de que en una familia ambos padres estén sanos es:----

¿Hay independencia en el estado de los padres?

¿P(Xm=1,Xp=1)=P(Xm=1)*P(Xp=1)?

$$0.02 \neq 0.1*0.12$$

Distribución conjunta del estado de ambos padres

		Xp		
		1	*0*	total
Xm	*1*	0,02	0,08	0,1
	0	0,10	0,80	0,9
total		0,12	0,88	1

¿Saber que la madre está enferma aporta información sobre el estado del padre?

$P(Xp=1)=0.12$

$P(Xp=1|Xm=1)=P(Xm=1,Xp=1)/P(Xm=1)$

$=0.02/0.1=0.2$

Simulación del estado de salud de los padres de n=20 familias

		Xp		
		1	*0*	total
Xm	*1*	0,02	0,08	0,1
	0	0,10	0,80	0,9
total		0,12	0,88	**1**

Esquema A:

1.- Simulamos el estado de la madre

2.- Conocido el estado de la madre simulamos el del padre

Simulación del estado de salud de los padres de n=20 familias

Esquema A:

1.- Simulamos el estado de la madre

2.- Conocido el estado de la madre simulamos el del padre

$P(Xm=1)=0,1$ y $P(Xm=0)=0,9$ (*)

$$\left.\begin{array}{l} P(Xp=1|Xm=1)=0,2 \text{ y } P(Xp=0|Xm=1)=0,8 \\ P(Xp=1|Xm=0)=1/9 \text{ y } P(Xp=0|Xm=0)=8/9 \end{array}\right\} \quad (**)$$

Paso 1: Hacer $Xm=0$ y generar $u \rightarrow U(0,1)$

Paso 2: Si $u<0,1$ ➔ hacer $Xm=1$

Paso 3: Hacer $Xp=0$ y generar $v \rightarrow U(0,1)$

Si $Xm=1$ y $v<0,2 \rightarrow$ hacer $Xp=1$

Si $Xm=0$ y $v<1/9 \rightarrow$ hacer $Xp=1$

Simulación del estado de salud de los padres de n=20 familias

Esquema A:

1.- Simulamos el estado de la madre

2.- Conocido el estado de la madre

simulamos el del padre

$P(Xm=1)=0,1$ y $P(Xm=0)=0,9$ (*)

$P(Xp=1|Xm=1)=0,2$ y $P(Xp=0|Xm=1$

$P(Xp=1|Xm=0)=1/9$ y $P(Xp=0|Xm=0)=$

1= Con gripe

0= Sin gripe

familia	Madre	Padre
1	1	1
2	0	0
3	0	0
4	0	0
5	0	0
6	0	1
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	1	1
14	0	0
15	0	0
16	1	0
17	0	0
18	0	0
19	0	0
20	0	0

Tabla de doble entrada de las 20 familias simuladas

	padre enf	padre sano
madre enf	2	1
madre sana	1	16

Simulación del estado de salud de los padres de n=20 familias

Esquema B: Trabajando sobre la distribución conjunta

1.- Creamos una variable auxiliar Y, con puntos de masa

{11, 10, 01, 00} probabilidades respectivas

{0.02, 0.08, 0.10, 0.80}

La equivalencia de sucesos es obvia: {00}={Xm=0,Xp=0}, {01}={Xm=0, Xp=1} …

Paso 1: Y=00

Paso 2: Generar u → U(0,1)

Si u≤0.02 → hacer Y=11, STOP

Si u≤0.1 → hacer Y=10, STOP

Si u≤0.20 → hacer Y=01, STOP

¿Este procedimiento es más rápido que el anterior? ¿es menos costoso?

Simulación del estado de salud de los padres de n=20 familias

Esquema B: Trabajando sobre la distribución conjunta

familia	nº aleatorio	Y
1	0,30996312	00
2	0,23749002	00
3	0,09635188	10
4	0,11609877	01
5	0,77763526	00
6	0,24218958	00
7	0,16113956	01
8	0,56229979	00
9	0,41246499	00
10	0,46146399	00
11	0,1024335	01
12	0,94543511	00
13	0,95840574	00
14	0,30122097	00
15	0,91173938	00
16	0,95087031	00
17	0,05211314	10
18	0,41149933	00
19	0,42981552	00
20	0,31497124	00

Tabla de doble entrada de las 20 familias simuladas

	padre enf	padre sano
madre enf	0	2
madre sana	3	15

Simulación del estado de salud de los padres de n=20 familias

Esquema C: Bootstrap o remuestreo sobre los datos observados

Supongamos que en el centro de salud se tienen datos del estado de salud de los progenitores de 100 familias

observaciones

familia	enfermos
1	MP
2	SP
3	SM
4	SM
5	SM
6	N
7	N
8	SP
9	N
…	…
…	…
99	N
100	SP

La metodología bootstrap proporciona una nueva muestra remuestreando con reposición sobre la información de las 100 familias anteriores

Ejercicio propuesto

Supongamos que en las familias con dos hijos la probabilidad de que cada niño tenga gripe es de 0,4 y que el 50% de las familias tiene a ambos hijos sanos.

1) Determina la distribución conjunta del estado de salud de los niños de la familia.

2) ¿Cuál es la probabilidad de que alguno de los hijos tenga gripe? ¿Cuál es la probabilidad de que ambos tengan gripe si uno de ellos la tiene?

3) Simula el estado de salud de los hijos de 20 familias, tabula los resultados y da estimaciones para las probabilidades calculadas en 2).

4) Recoge los resultados de Ns=1000 simulaciones y da la estimación de las probabilidades calculadas en 2).