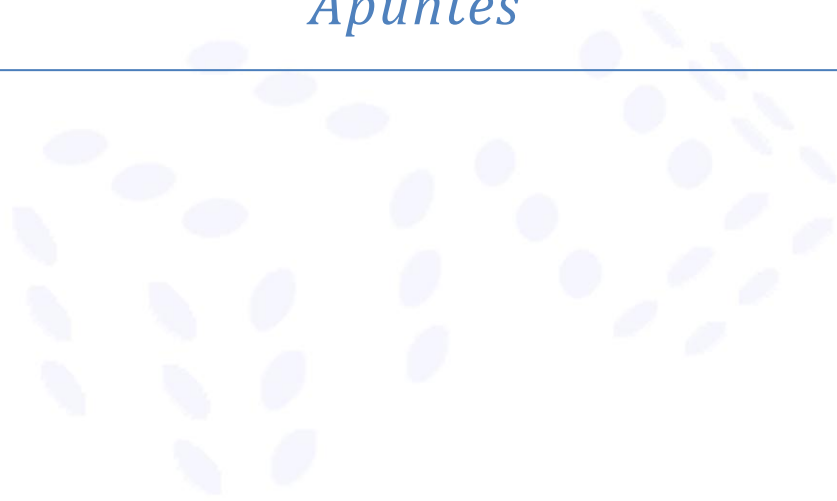


Εstadística I

Grado en ADE/Eco/A+D/CiF/EiT UAB 1/2

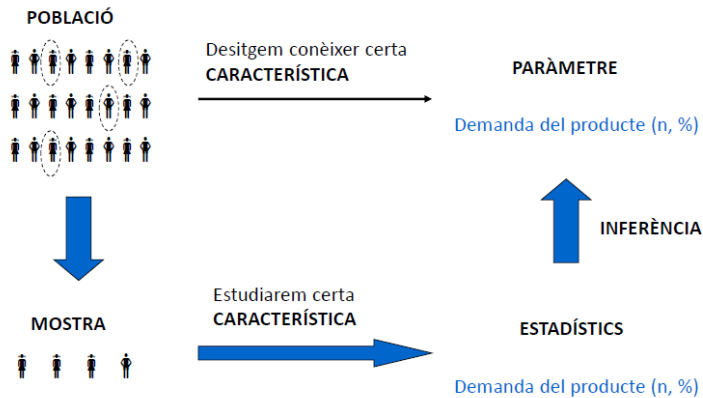
estadistix

Apuntes



1. ANÁLISIS DE DATOS

OBTENCIÓN DE LOS DATOS



Una **población** es el conjunto de todos los elementos que cumplen una o varias características.

Un **parámetro** es una propiedad descriptiva de la población que generalmente es desconocida.

Una **muestra** es un subconjunto de los elementos de una población.

Un **estadístico** es una propiedad descriptiva de la muestra.

La **inferencia estadística** se encarga de extraer conclusiones sobre una población a partir del estudio de una muestra mediante técnicas probabilísticas.

La **estadística descriptiva** se encarga de resumir y describir un conjunto de datos para su comprensión.

Tipos de muestreos representativos:

- **Muestreo aleatorio simple:** totalmente al azar.
- **Muestreo sistemático:** mediante el cálculo de una frecuencia de extracción $k=N/n$.
- **Muestreo aleatorio estratificado:** si conservamos la proporcionalidad de las partes que componen la población antes de hacer la extracción aleatoria.
- **Muestreo por conglomerados y por etapas:** si la población se divide en partes equivalentes (conglomerados) y elegimos algunas por azar para ahorrar costes. Si lo hacemos varias veces: por etapas.

TIPOS DE VARIABLES

Variable: característica observable que cambia entre los elementos de una población. Característica que puede ser medida y que puede adoptar más de un valor.

Variables cualitativas o categóricas: son aquellas que recogen una característica que no se puede expresar mediante una cantidad, aunque sí con una categoría.

- **Nominales:** tienen un conjunto de categorías sin ningún tipo de jerarquía.
- **Ordinales:** tienen un conjunto de categorías con una jerarquía u orden.

Variables cuantitativas o numéricas: son aquellas variables que recogen como información una cantidad numérica de lo que se está observando.

- **Discretas:** tienen un conjunto finito de valores por ejemplo si únicamente toman números enteros.
- **Continuas:** el conjunto de posibles valores entre dos números fijos es infinito.

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

Frecuencia absoluta (n_i): número de veces que se repite un determinado valor. Puede ser individual o acumulada (N_i). En las variables nominales no tiene sentido acumular.

Frecuencia relativa (f_i): proporción que representa las apariciones de ese valor respecto al total $f_r = \frac{n_i}{n}$

También puede ser individual o acumulada (F_i). Si se multiplica por 100 nos dará un porcentaje.

Para datos cuantitativos agrupados en intervalos:

Amplitud del intervalo: $a_i = L_i - L_{i-1}$

Marca de clase: $c_i = \frac{L_i + L_{i-1}}{2}$

Altura del intervalo: $h_i = \frac{n_i}{a_i}$

Ejemplo numérico discreto, número de hijos por familia: 2, 1, 1, 0, 1, 2, 2, 0, 1, 1.

x_i	n_i	N_i	f_i	F_i
0				
1				
2				

Ejemplo: numérico continuo, notas de un examen: 3,4; 5,3; 6,8; 9,1; 7,2; 8,3; 1,9; 5,1; 4,4; 3,2.

x_i	n_i	N_i	f_i	F_i	a_i	c_i	h_i
[0-2)							
[2-4)							
[4-6)							
[6-8)							
[8-10)							

Ejemplo R Studio:

```
# Mostra la taula de freqüències
DiscFreqTable

##          n          %      val%    %cum val%cum
## 1         46      24.5      28.6    24.5     28.6
## 2         36      19.1      22.4    43.6    50.9
## 3         24      12.8      14.9    56.4    65.8
## 4         12       6.4       7.5    62.8    73.3
## 5         43      22.9      26.7    85.6   100.0
## NA         27      14.4       NA   100.0     NA
## Total    188     100.0     100.0   100.0   100.0
```

ContFreqTable								
##	Lower	Upper	Main	Frequency	Percentage	CF	CPF	
## 1	0.0	0.1	0.05	0	0.0	0	0.0	
## 2	0.1	0.2	0.15	0	0.0	0	0.0	
## 3	0.2	0.3	0.25	0	0.0	0	0.0	
## 4	0.3	0.4	0.35	5	2.7	5	2.7	
## 5	0.4	0.5	0.45	24	12.8	29	15.4	
## 6	0.5	0.6	0.55	27	14.4	56	29.8	
## 7	0.6	0.7	0.65	27	14.4	83	44.1	
## 8	0.7	0.8	0.75	56	29.8	139	73.9	
## 9	0.8	0.9	0.85	34	18.1	173	92.0	
## 10	0.9	1.0	0.95	15	8.0	188	100.0	

GRÁFICOS

Diagrama de sectores

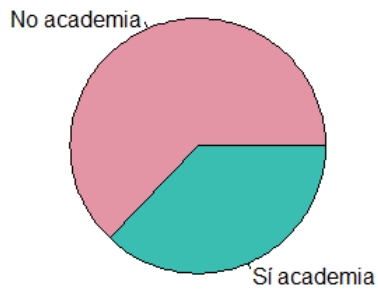
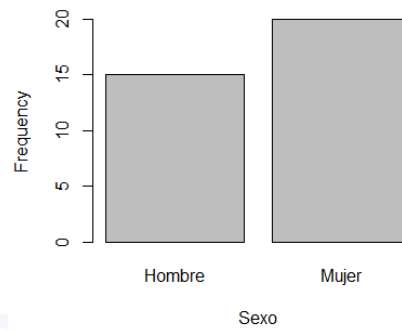
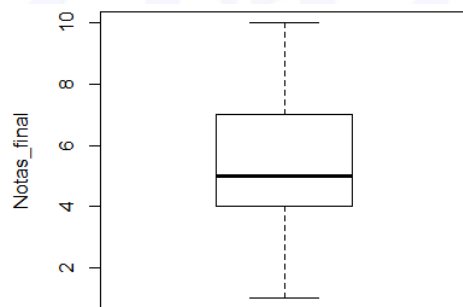


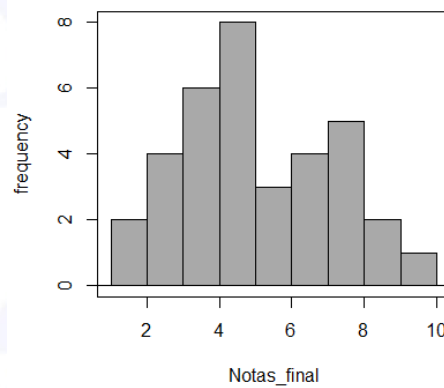
Diagrama de barras



Boxplot



Histograma



MEDIDAS DE POSICIÓN

Media aritmética:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum x_i \cdot n_i}{n} = \sum x_i \cdot f_i$$

Mediana:

$$Pos_{Md} = \frac{n + 1}{2}$$

Cuartiles y percentiles:

Ejemplo R Studio:

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.000   6.000   7.000   7.311   9.000  10.000     10
```

MEDIDAS DE DISPERSIÓN**Varianza muestral (S^2):**

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n} = \frac{\sum x_i^2 \cdot n_i}{n} - \bar{x}^2$$

Varianza corregida o casi-varianza (S^2_{n-1}):

$$S^2_{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1} = \left(\frac{\sum x_i^2}{n} - \bar{x}^2 \right) \cdot \frac{n}{n-1} = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n-1} = \left(\frac{\sum x_i^2 \cdot n_i}{n} - \bar{x}^2 \right) \cdot \frac{n}{n-1}$$

La desviación estándar/típica (S):

$$S = DE = \sqrt{S^2}$$

Coefficiente de Variación de Pearson (CV):

$$CV = V = \frac{S}{\bar{x}}$$

Rango, Recorrido o Amplitud (R):

$$R = X_{max} - X_{min}$$

Rango intercuartílico (Rq):

$$R_I = Q_3 - Q_1$$

MEDIDAS DE FORMA**Moda:**

Asimetría: $CA > 0$ *asimetría positiva* $CA = 0$ *simetría* $CA < 0$ *asimetría negativa***Curtosis:** $K > 0$ *leptocúrtica* $K = 0$ *mesocúrtica* $K < 0$ *platicúrtica**Ejemplo R Studio:*

```
##      var      sd      IQR      cv      skew kurtosis
##    4.000    2.00    3.00    0.274   -0,854    0.021
```

COMBINACIONES LINEALES DE VARIABLES

$$Si \ x = a_1 \cdot x_1 + a_2 \cdot x_2 \quad \begin{cases} \bar{x} = a_1 \cdot \bar{x}_1 + a_2 \cdot \bar{x}_2 \\ S^2 = a_1^2 \cdot S_1^2 + a_2^2 \cdot S_2^2 \end{cases}$$

TABLAS DE FRECUENCIAS DE DOBLE ENTRADA O BIVARIANTES

	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj.8	Suj. 9	Suj.10
x	1	2	0	3	1	3	2	0	1	2
y	0	1	1	1	2	0	2	1	1	1

Frecuencias absolutas conjuntas:

x\y	0	1	2	
0	0	2	0	2
1	1	1	1	3
2	0	2	1	3
3	1	1	0	2
	2	6	2	10

Frecuencias relativas conjuntas:

x\y	0	1	2	
0	0	0,2	0	0,2
1	0,1	0,1	0,1	0,3
2	0	0,2	0,1	0,3
3	0,1	0,1	0	0,2
	0,2	0,6	0,2	1

Tabla de frecuencias condicionales

 $X_i/(Y = 2)$ $n(X_i/(Y = 2))$ $f(X_i/(Y = 2))$

0	0	0
1	1	0,5
2	1	0,5
3	0	0
	2	1

 $Y_i/(X = 0)$ $n(Y_i/(X = 0))$ $f(Y_i/(X = 0))$

0	
1	
2	

Ejemplo R Studio:

Cell Contents					

N					
N / Table Total					

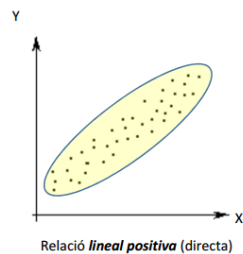
Total Observations in Table: 161					
HDI_Data\$GNI_GROUP	HDI_Data\$HDI_GROUP				
	High	Low	Medium	Very High	Row Total
1	19	0	5	22	46
	0.118	0.000	0.031	0.137	
2	11	2	4	19	36
	0.068	0.012	0.025	0.118	
3	10	2	7	5	24
	0.062	0.012	0.043	0.031	
4	2	5	4	1	12
	0.012	0.031	0.025	0.006	
5	5	26	12	0	43
	0.031	0.161	0.075	0.000	
Column Total	47	35	32	47	161

DESCRIPCIÓN CONJUNTA DE DOS VARIABLES CUANTITATIVAS

La **relación lineal** entre dos variables puede ser:

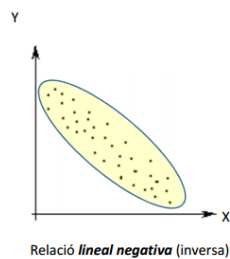
- **Positiva** (+ de una variable, + del otro o al revés), Ej : Horas de estudio (X) y nota en el examen (Y).

Horas de estudio	nota
3	2
8	4
12	7
16	9

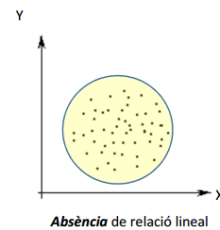


- **Negativa** (+ de una, - del otro) Ej.: ausencias a clase (X) y nota en el examen (Y).

Ausencias	nota
4	9
5	7
9	4
12	2



- **Nula**. Sin relación. Ex: altura (X) y nota (Y).



También nos podemos encontrar con relaciones no lineales

Medidas de asociación lineal

Además del análisis gráfico, podemos calcular una serie de indicadores que nos ayudarán a decidir si las variables están relacionadas (de forma lineal positiva o negativa) y con **qué intensidad**.

La **covarianza** (S_{xy})

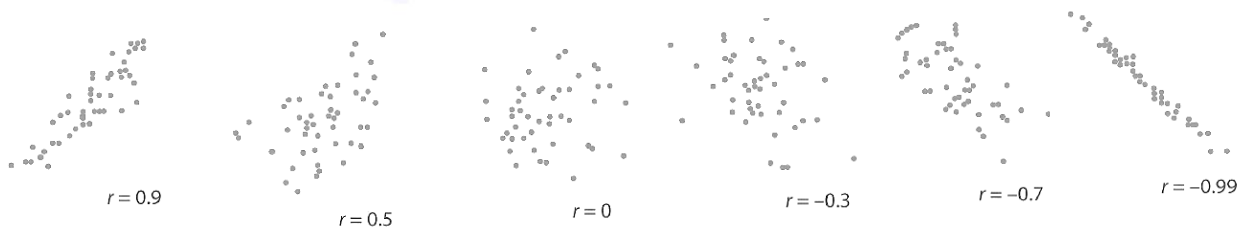
$$S_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_j - \bar{y})}{n} = \frac{\sum x_i y_j}{n} - \bar{x} \bar{y} = \frac{\sum (x_i - \bar{x}) \cdot (y_j - \bar{y}) \cdot n_{ij}}{n} = \frac{\sum x_i y_j \cdot n_{ij}}{n} - \bar{x} \bar{y}$$

La **covarianza corregida o casi-covarianza** ($S_{xy \ n-1}$)

$$S_{xy \ n-1} = \frac{\sum (x_i - \bar{x}) \cdot (y_j - \bar{y})}{n-1} = \left(\frac{\sum x_i y_j}{n} - \bar{x} \bar{y} \right) \cdot \frac{n}{n-1} = \frac{\sum (x_i - \bar{x}) \cdot (y_j - \bar{y}) \cdot n_{ij}}{n-1} = \left(\frac{\sum x_i y_j \cdot n_{ij}}{n} - \bar{x} \bar{y} \right) \cdot \frac{n}{n-1}$$

El **coeficiente de correlación** (r_{xy})

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$



El **coeficiente de determinación** (R^2)

$$R^2 = r_{xy}^2$$

También se puede expresar las correlaciones o las covarianzas en matrices:

$$S = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \dots & \dots & \dots & \dots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & 1 & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

Ejemplo: calcula los indicadores de asociación lineal sabiendo que $\bar{x} = 1,5$ $\bar{y} = 1$ $S_x^2 = 1,05$ $S_y^2 = 0,4$

	Suj. 1	Suj. 2	Suj. 3	Suj. 4	Suj. 5	Suj. 6	Suj. 7	Suj. 8	Suj. 9	Suj. 10
x	1	2	0	3	1	3	2	0	1	2
y	0	1	1	1	2	0	2	1	1	1

x\y	0	1	2	
0	0	2	0	2
1	1	1	1	3
2	0	2	1	3
3	1	1	0	2
	2	6	2	10

Ejemplo R Studio:

```
cor(HDI_Data[2:5])

##           HDI           LEB           EYS           MYS
## HDI 1.0000000 0.8998000 0.9123647 0.8989656
## LEB 0.8998000 1.0000000 0.7866404 0.7347878
## EYS 0.9123647 0.7866404 1.0000000 0.8151615
## MYS 0.8989656 0.7347878 0.8151615 1.0000000

cov(HDI_Data[2:5])

##           HDI           LEB           EYS           MYS
## HDI 0.02393283 1.166926 0.4048757 0.4301132
## LEB 1.16692572 70.274936 18.9161489 19.0504347
## EYS 0.40487569 18.916149 8.2283442 7.2317371
## MYS 0.43011324 19.050435 7.2317371 9.5650113
```

2. TEORÍA DE LA PROBABILIDAD

SUCESOS ALEATORIOS Y ESPACIO MUESTRAL

Experimento aleatorio:

$$\Omega = \{0, X\} \quad \Omega = \{1, 2, 3, 4, 5, 6\}$$

Suceso:

Operaciones con sucesos

Unión: $A \cup B = \{1, 2, 3, 5\}$

Intersección: $A \cap B = \{1, 3\}$

Complementario: $\bar{A} = \{2, 4, 6\}$

Diferencia: $A - B = \{5\}$

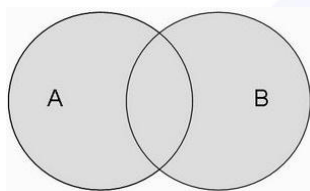
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{1, 3, 5\}$$

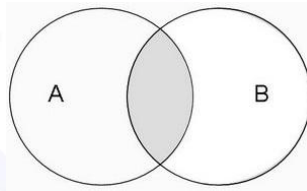
$$B = \{1, 2, 3\}$$

$$C = \{4, 6\}$$

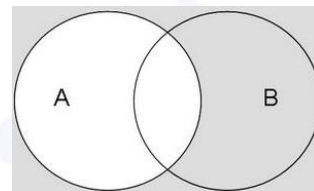
Representación gráfica (Diagramas de Venn):



$$A \cup B$$



$$A \cap B$$



$$\bar{A}$$

AXIOMAS DE PROBABILIDAD

- 1.- La probabilidad nunca puede ser negativa.
- 2.- La probabilidad de que ocurra algún suceso del espacio muestral es uno.
3. Si tenemos dos sucesos disjuntos (que no se pueden dar a la vez) su unión es, la suma de las probabilidades individuales

INTERPRETACIONES DE LA PROBABILIDAD

Concepto clásico de la probabilidad:

Concepto experimental de la probabilidad:

Diagrama del árbol:

OPERACIONES CON PROBABILIDAD

$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A - B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

$$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$$

$$P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B}) = 1 - P(A \cap B)$$

Ejemplo: si A i B son dos sucesos con $P(A)=3/8$, $P(B)=4/8$ i $P(A \cap B)=2/8$

a) $P(A \cup B)$

b) $P(\bar{A})$

c) $P(\bar{A} \cap \bar{B})$

d) $P(A \cap \bar{B})$

e) $P(A \cup \bar{B})$

PROBABILIDAD CONDICIONADA

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

Independencia si: $P(A/B) = P(A)$ o $P(B/A) = P(B)$ o $P(A \cap B) = P(A) \cdot P(B)$

Ejemplo: En una clase el 40% tienen el pelo rubio, el 25% los ojos azules y el 15% las dos cosas.

a) Si el pelo es rubio, ¿qué probabilidad hay de que también tenga los ojos azules?

b) Si tiene los ojos azules, ¿qué probabilidad hay de que no tenga el pelo rubio?

Ejemplo: Con los siguientes datos, calcula:

	EC	EU	
Hombre	11	4	15
Mujer	11	9	20
	22	13	35

a) $P(M)$

b) $P(EC)$

c) $P(H \cap EC)$

d) $P(EC/M)$

e) $P(EU \cup M)$

Ejemplo: Una enfermedad tiene dos síntomas diferentes, A y B. Un 20% de los casos no presentan síntomas, 50% el síntoma A y el 30% a más del A, el B. ¿Son independientes A y B?

TEOREMA DE LA PROBABILIDAD TOTAL

$$P(B) = P\left(\frac{B}{A_1}\right) \cdot P(A_1) + P\left(\frac{B}{A_2}\right) \cdot P(A_2) \dots + P\left(\frac{B}{A_n}\right) \cdot P(A_n)$$

Ejemplo: En un examen de Estadística I el 50% no van a ninguna academia, el 30% van a la academia Estadistix y el 20% a otras academias. La probabilidad de aprobar de los que no van a ninguna academia es de un 60%, de los que van a Estadistix es del 95% y de los que van a otras academias es del 70%. ¿Qué probabilidad tienen los alumnos de aprobar?

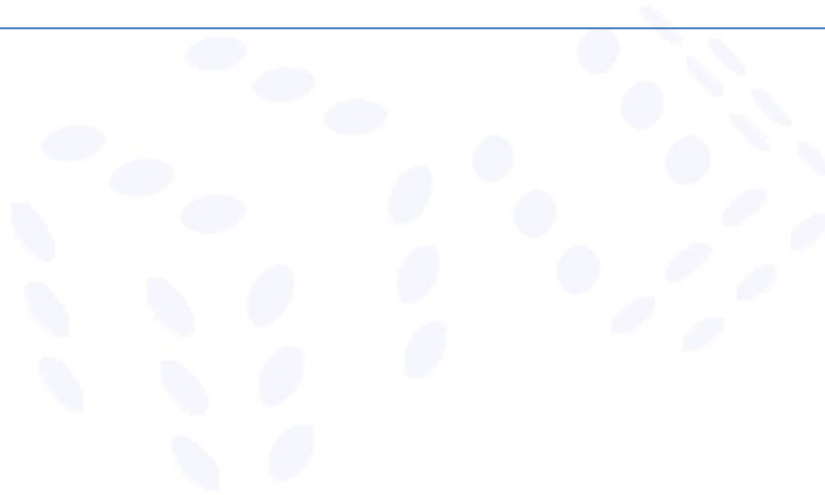
TEOREMA DE BAYES

$$P\left(\frac{A_1}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{P\left(\frac{B}{A_1}\right) \cdot P(A_1)}{P(B)} = \frac{P\left(\frac{B}{A_1}\right) \cdot P(A_1)}{P\left(\frac{B}{A_1}\right) \cdot P(A_1) + P\left(\frac{B}{A_2}\right) \cdot P(A_2) \dots + P\left(\frac{B}{A_n}\right) \cdot P(A_n)}$$

Ejemplo: en caso anterior, ¿cuál es la probabilidad de que si has aprobado, hayas ido a Estadistix?

estadistix

Exámenes parciales



ESTADÍSTICA I

TIPUS A

Examen Parcial

1. Es disposa d'informació sobre la nota d'accés a la universitat (NOTA_ACCES) i la nota en el curs de Matemàtiques I (NOTA_MAT) de 500 estudiants de la facultat. Per cada estudiant es disposa, a més, del seu gènere (CODI_GENERE) i el grau en el que s'ha matriculat (CODI_GRAU) que prenen valors d'acord amb la següent taula de codis:

CODI_GENERE		CODI_GRAU	
Valor	Gènere	Valor	Grau
0	Home	1	ADE
1	Dona	2	ADE+Dret
		3	ECO
		4	CiF
		5	EiT

La informació recollida es resumeix en les sortides GRETL que trobareu a l'altra cara d'aquest full. En base a aquests resultats es demana:

- (a) En aquesta mostra, quin percentatge hi ha de dones ? [0.5 PUNTS]
- (b) Considerant el 50% dels estudiants amb millor nota d'accés, hi ha algú amb una nota d'accés per sota de la mitjana ? Justificar la resposta. [1 PUNT]
- (c) Com s'interpreta el signe negatiu del coeficient d'asimetria de la nota de Matemàtiques I ? [0.5 PUNTS]
- (d) Com s'interpreta el signe negatiu del coeficient de correlació entre el gènere i el grau ? [1 PUNT]
- (e) Existeix una relació clara entre la nota d'accés a la universitat i la nota de Matemàtiques I ? Justificar la resposta. [1 PUNT]

(f) Quin ha estat el grau preferit entre els homes a l'hora de matricular-se ? I entre les dones ? Justificar la resposta. [0.5 PUNTS]

(g) On s'observa més variabilitat en la nota d'accés, entre els homes o entre les dones ? Justificar la resposta. [0.5 PUNTS]

2. La probabilitat que una empresa rebi un ajut a la innovació i a la internacionalització (*AII*) depèn, entre altres factors, del sector. Segons un estudi, si classifiquem les empreses en: baixa tecnologia manufacturera (*BTM*), alta tecnologia manufacturera (*ATM*) i serveis intensius de coneixement (*SIC*), les probabilitats d'obtenir un ajut són respectivament: 0.2, 0.25 i 0.35. En un país hi ha un 60% d'empreses de baixa tecnologia manufacturera, un 25% d'alta tecnologia manufacturera i un 15% del sector de serveis intensius de coneixement. Trobeu:

(a) La probabilitat que una empresa pertanyi a les *BTM* i rebi un ajut a la innovació i internacionalització (*AII*). [1 PUNT]

(b) Quina és la probabilitat que una empresa rebi un ajut *AII*? [1.5 PUNTS]

=

(c) Quina és la probabilitat que una empresa no rebi un ajut *AII*? [0.5 PUNTS]

- (d) Si una empresa ha rebut un ajut *AII*, a quin dels tres sectors és més probable que [2 PUNTS] pertanyi? Justifica la teva resposta a partir de les probabilitats calculades.

Estadístics principals, usant les observacions 1 – 500

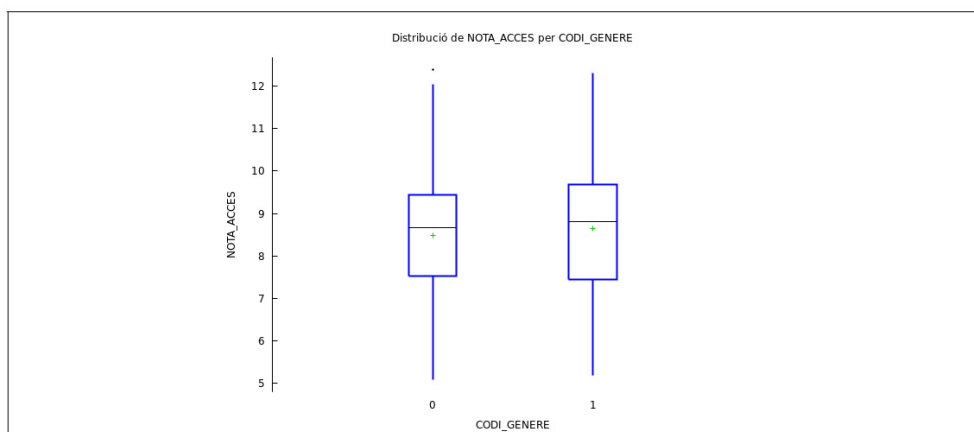
Variable	Mitjana	Mediana	Desv. Estàndard	Asimetria
CODLGENERE	0.38200	0.0000	0.48636	0.48572
NOTA_ACCES	8.5477	8.7330	1.4444	−0.031823
NOTA_MAT	4.4622	5.0000	2.2302	−0.24974
CODLGRAU	2.7860	3.0000	1.3682	0.093670

Coefficients de correlació, usant les observacions 1 – 500

CODLGENERE	NOTA_ACCES	NOTA_MAT	CODLGRAU	
1.0000	0.0574	0.0396	−0.1329	CODLGENERE
	1.0000	0.3639	−0.4641	NOTA_ACCES
		1.0000	−0.1487	NOTA_MAT
			1.0000	CODLGRAU

Tabulació creuada de CODLGENERE (files) contra CODLGRAU (columnes)

	[1]	[2]	[3]	[4]	[5]	TOTAL
[0]	22.30%	13.60%	31.10%	14.90%	18.10%	309
[1]	31.40%	19.40%	16.80%	27.20%	5.20%	191
TOTAL	25.80%	15.80%	25.60%	19.60%	13.20%	500



ESTADÍSTICA I

Examen Parcial

TIPUS A

6 d'Abril de 2021

1. Es realitzen 2000 entrevistes a persones en situació d'atur al Vallès Occidental sobre els temps que porten desocupades (DAYS) i els ingressos mensuals que tenien abans de perdre la feina (INCOME). Per a cada una d'aquestes persones es disposa, a més, del seu gènere (GENDER_CODE) i de l'ocupació que tenien immediatament abans de quedar-se en atur (OCCUPATION_CODE) que prenen valors d'acord amb la següent taula de codis:

GENDER_CODE	
Valor	Gènere
0	Home
1	Dona

OCCUPATION_CODE			
Valor	Ocupació	Valor	Ocupació
1	Enginyers i graduats	6	Subalterns
2	Enginyers tècnics i pèrits	7	Auxiliars administratius
3	Caps d'administració i taller	8	Oficials de primera i de segona
4	Ajudants no titulats	9	Oficials de tercera i especialistes
5	Oficials administratius	10	Mà d'obra no qualificada

La informació recollida es resumeix en les sortides d'**RStudio** que trobareu a l'altra cara d'aquest full. En base a aquests resultats es demana:

- (a) En aquesta mostra, quin percentatge dels entrevistats eren *Auxiliars Administratius* abans de quedar-se en atur? Justificar la resposta. [0,5 PUNTS]
- (b) Calcula l'ingrés mitjà de les persones entrevistades a partir de la taula **IncomeTable**. Expliciteu els càlculs. [1 PUNT]
- (c) Del total d'entrevistats que treballaven d'*Oficials administratius*, quin percentatge són dones? [1 PUNT]
- (d) Es pot interpretar el signe del coeficient de correlació entre el gènere i l'ocupació? Per què? [0,5 PUNTS]
- (e) Existeix una relació clara entre les variables DAYS i INCOME? De quin tipus i per què? [1 PUNT]
- (f) Entre quins valors (interval) trobaríem la *mediana* mostral dels ingressos? Justificar la resposta. [0,5 PUNTS]
- (g) Quina és la moda mostral (o modes mostrals, si fos el cas) de la variable DAYS? Interpreta aquest resultat. [0,5 PUNTS]

2. Un servei de càterring a domicili està estudiant la fiabilitat de les comandes que reben durant la setmana, de les quals un 40% són a l'hora de dinar i el 60% restant pel sopar. De les comandes que reben a l'hora de dinar només 1 de cada 4 és falsa, mentre que de les comandes per sopar hi ha 1 que és falsa de cada 5.

(a) Quina és la probabilitat que una comanda sigui falsa? [1,25 PUNTS]

(b) Si una comanda no és falsa, quina és la probabilitat que sigui per sopar? [1,25 PUNTS]

(c) Si una comanda és falsa, quina és la probabilitat que sigui per dinar? [1,25 PUNTS]

(d) Quina és la probabilitat que una comanda no sigui falsa o que no sigui per sopar? [1,25 PUNTS]

-
- Per poder fer l'examen cal tenir un document vàlid d'identitat i tenir-lo a la vista en tot moment.
 - L'examen no es pot fer amb llapis.
 - **TOTES LES RESPOSTES HAN D'ESTAR DEGUDAMENT JUSTIFICADES.**
 - Només es pot utilitzar una calculadora. No es pot tenir a l'abast, cap mena d'apunts, llibres, ni qualsevol altre tipus de material amb continguts relacionats amb l'assignatura.
 - Els telèfons mòbils i altres dispositius han d'estar desconnectats i guardats. En cap cas es podran fer servir com a rellotge, com a calculadora, o per qualsevol altre funció.
-

Nom:

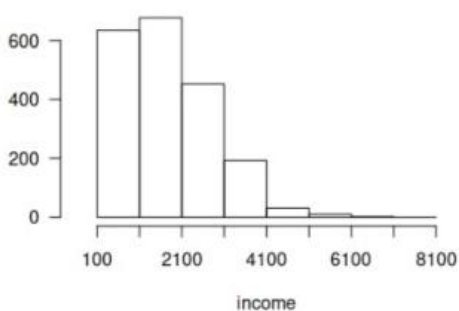
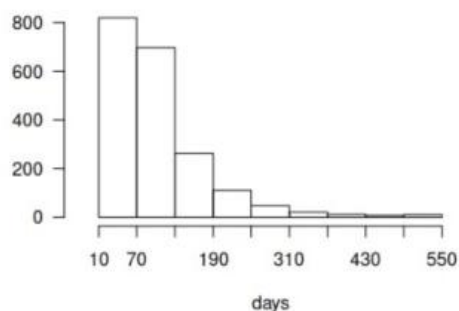
Grup:

Taules de freqüències bivariants

Freqüències conjuntes (absolutes i relatives) i freqüències absolutes marginals

<pre>## ## Cell Contents ## ----- ## N ## N / Table Total ## ----- ## ## Total Observations in Table: 2000 ## ## ## occupation_code gender_code ## ----- ## 0 1 Row Total ## ----- ## 1 182 108 290 ## 0.091 0.054 ## ----- ## 2 521 286 807 ## 0.261 0.143 ## ----- ## 3 153 74 227 ## 0.076 0.037 ## ----- ## 4 37 19 47 ## 0.018 0.005 ## ----- ## 5 30 12 42 ## 0.015 0.006 ## ----- ## 6 6 2 8 ## 0.003 0.001 ## ----- ## 7 8 7 15 ## 0.004 0.004 ## ----- ## 8 139 72 211 ## 0.070 0.036 ## ----- ## 9 172 100 272 ## 0.086 0.050 ## ----- ## 10 44 37 81 ## 0.022 0.018 ## ----- ## Column Total 1292 708 2000 ## ----- ##</pre>							
<pre>## ## 100 1100 600 ## 1100 2100 1600 ## 2100 3100 2600 ## 3100 4100 3600 ## 4100 5100 4600 ## 5100 6100 5600 ## 6100 7100 6600 ## 7100 8100 7600 ## ## 635 ## 677 ## 452 ## 193 ## 31 ## 10 ## 2 ## 0 ## ## 31.8 ## 33.9 ## 22.6 ## 9.7 ## 1.6 ## 0.5 ## 0.1 ## 0.0 ## ## 635 ## 1312 ## 1764 ## 1957 ## 1988 ## 1998 ## 2000 ## 2000 ## ## 31.8 ## 65.6 ## 88.2 ## 97.8 ## 99.4 ## 99.9 ## 100.0 ## 100.0 ##</pre>							

Histogrammes



Coefficients de correlació entre variables

```
## [1] "cor(occupation_code,gender_code)= 0.8161202834579708"
## [1] "cor(days,income)= -0.9350916158004891"
## [1] "cor(days,gender_code)= -0.3323032368920747"
## [1] "cor(occupation_code,income)= -0.70617361123571627"
```

ESTADÍSTICA I

TIPUS A

Examen Parcial

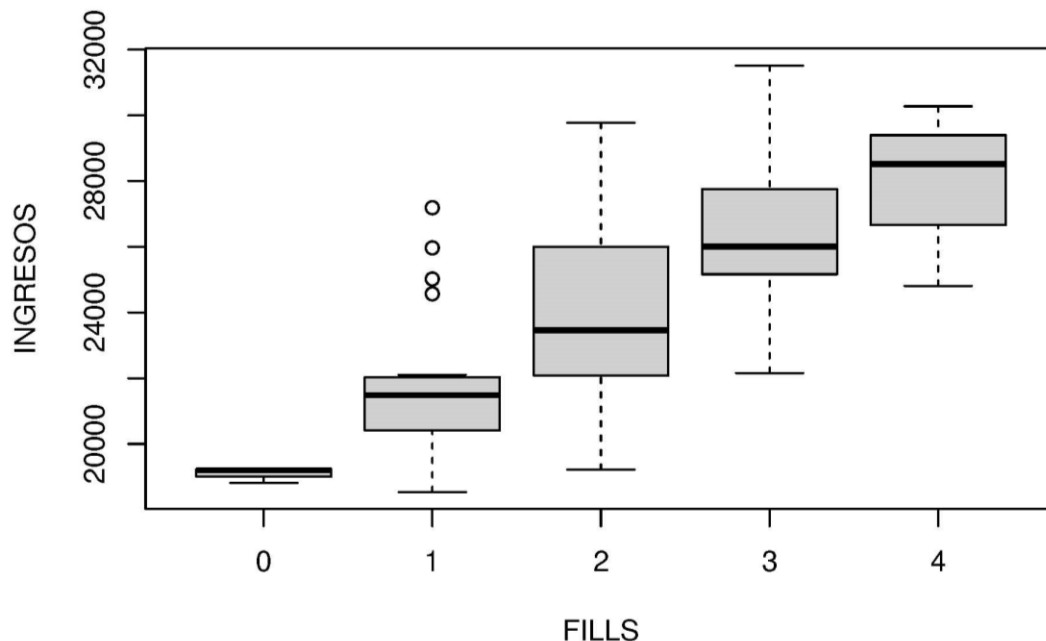
24 de març de 2022

Pregunta 1

Es disposa d'una base de dades obtinguda mitjançant entrevistes a 90 famílies de Sabadell. La informació recollida correspon a les variables **FILLS** (número de fills de la família), **TVs** (número d'aparells de televisió a la casa) i **INGRESOS** (ingressos nets anuals de la família). Utilitzant aquesta informació es construeix la taula bidimensional de freqüències (taula de contingència) entre les variables **FILLS** i **TVs** que hi ha a continuació

```
##
##
## Cell Contents
## |-----|
## | N |
## |-----|
##
##
## Total Observations in Table: 90
##
##
##      | FILLS
##      | 0 | 1 | 2 | 3 | 4 | Row Total |
##-----|-----|
## 0 | 0 | 5 | 8 | 6 | 0 | 19 |
##-----|-----|
## 1 | 0 | 5 | 12 | 3 | 3 | 23 |
##-----|-----|
## 2 | 3 | 6 | 10 | 10 | 0 | 29 |
##-----|-----|
## 3 | 0 | 3 | 9 | 7 | 0 | 19 |
##-----|-----|
## Column Total | 3 | 19 | 39 | 26 | 3 | 90 |
##-----|-----|
##
##
```

Obtenim també aquesta representació gràfica de les observacions de la variable **INGRESOS**.



En base a aquests resultats,

- Quin és el número mig de fills de les famílies a la mostra? Justificar la resposta. [1 punt]
- Quin percentatge de famílies tenen 2 fills? Justificar la resposta. [0,5 punts]
- Quin percentatge de famílies tenen 2 fills i 3 televisors? Justificar la resposta. [0,5 punts]
- De les famílies que tenen 2 fills, quin percentatge té 3 televisors? Justificar la resposta. [1 punt]
- Compareu la distribució mostral del nombre de televisors condicionada al fet que família té 2 fills amb la distribució marginal del nombre de televisors. Són iguals? Perquè? [1 punt]
- Diries que són independents les variables **FILLS** i **INGRESOS** ? Perquè? [1 punt]

Pregunta 2

La probabilitat que l'índex IBEX pugi un dia qualsevol és 0,52 i la probabilitat que pugi un dia que el president del govern apareix en televisió és 0,49. Si el president del govern apareix en televisió el 32,5% dels dies,

- quina és la probabilitat que en un mateix dia l'IBEX hagi pujat i el president del govern hagi aparegut en televisió? [1 punt]
- Si ens informen que un dia l'IBEX ha pujat, quina és la probabilitat que el president del govern hagi aparegut en televisió? [2 punts]
- Un dia que el president del govern NO apareix per televisió, quina és la probabilitat que l'IBEX pugi? [2 punts]

-
- Per poder fer l'examen cal tenir un document vàlid d'identitat i tenir-lo a la vista en tot moment.
 - L'examen no es pot fer amb llapis.
 - Totes les respostes han d'estar degudament justificades.
 - Només es pot utilitzar la calculadora. No es pot tenir a l'abast, cap mena d'apunts, llibres, ni qualsevol altre tipus de material amb continguts relacionats amb l'assignatura.
 - Els telèfons mòbils, smartwatches, i qualsevol altre dispositiu han d'estar desconnectats i guardats. En cap cas es podran fer servir com a rellotge, com a calculadora, o per qualsevol altra funció.

ESTADÍSTICA I

TIPUS A

Examen Parcial

27 de març de 2023

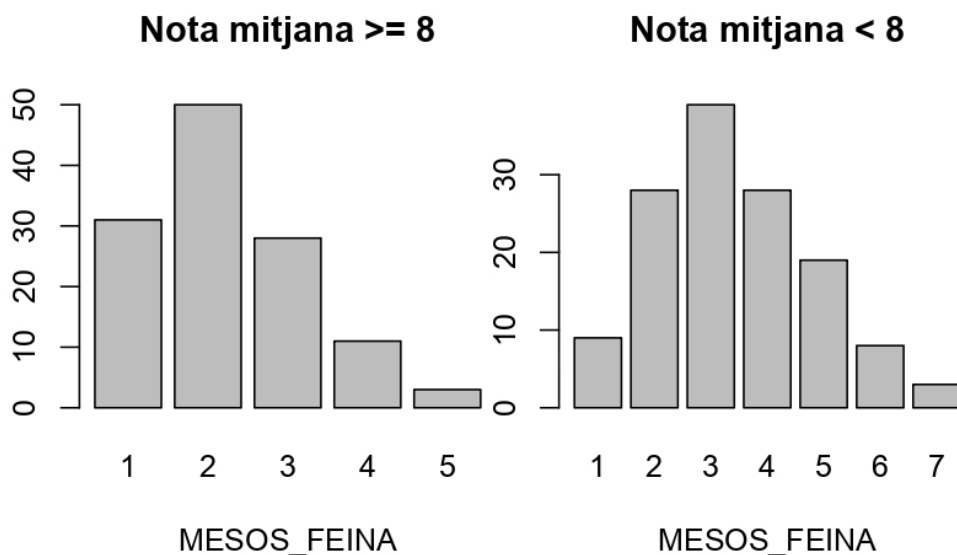
Pregunta 1

Un estudi realitzat amb 257 estudiants que van acabar els seus estudis universitaris l'any passat va analitzar conjuntament les variables ANYS_ESTUDIS (anys que s'ha trigat en completar els estudis), MESOS_FEINA (mesos que s'ha trigat en trobar una feina) i NOTA_MITJANA (Nota mitjana de l'expedient).

Les dades recollides es resumeixen en la següent taula bidimensional de freqüències (taula de contingència) entre les variables ANYS_ESTUDIS i MESOS_FEINA

```
##
##
##      Cell Contents
## |-----|
## |               N |
## |-----|
##
##
## Total Observations in Table:  257
##
##
##      | MESOS_FEINA
## ANYS_ESTUDIS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Row Total |
##-----|-----|-----|-----|-----|-----|-----|-----|
##      4 | 40 | 57 | 31 | 18 | 9 | 5 | 0 | 160 |
##-----|-----|-----|-----|-----|-----|-----|
##      5 | 0 | 21 | 29 | 14 | 9 | 2 | 1 | 76 |
##-----|-----|-----|-----|-----|-----|-----|
##      6 | 0 | 0 | 7 | 6 | 3 | 1 | 1 | 18 |
##-----|-----|-----|-----|-----|-----|-----|
##      7 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 3 |
##-----|-----|-----|-----|-----|-----|-----|
## Column Total | 40 | 78 | 67 | 39 | 22 | 8 | 3 | 257 |
##-----|-----|-----|-----|-----|-----|-----|
##
##
```

Obtenim també les següents representacions gràfiques que mostren la distribució de freqüències de la variable MESOS_FEINA pels estudiants que es van graduar amb una nota mitjana de l'expedient igual o per sobre de 8 i els que es van graduar amb una nota mitjana per sota de 8.



En base a aquests resultats,

- a. Quina és la mitjana del nombre d'anys que s'ha trigat en completar els estudis? Justificar la resposta. [1]
- b. Quin percentatge d'estudiants a la mostra va trigar 5 o més mesos en trobar feina? Justificar la resposta. [1]
- c. Quants estudiants a la mostra van trigar 5 mesos en trobar feina i 5 anys en acabar el seus estudis? Justificar la resposta. [1]
- d. Dels estudiants a la mostra que van trigar 4 anys en acabar els seus estudis, quin percentatge va trobar feina en 3 o menys mesos? Justificar la resposta. [1]
- e. Quina és la moda de la distribució de freqüències de la variable MESOS_FEINA. I la mediana ? Justificar les respostes. [1]
- f. Observant les representacions gràfiques de la distribució de freqüències de la variable MESOS_FEINA segons la nota mitjana de l'expedient, diries que existeix una relació entre aquestes variables? Per què? [1]

ESTADISTIX



Pregunta 2

Del total d'ordinadors que es compren a la Unió Europea un 70% són fabricats a un país asiàtic, mentre que el 30% restant es produeixen a la mateixa UE. Se sap que 8 de cada mil ordinadors fabricats a l'Àsia resulten defectuosos, mentre que només 4 de cada mil ordinadors produïts a Europa són defectuosos.

- Quina és la probabilitat que un ordinador qualsevol comprat a la Unió Europea resulti defectuós?
- Comprem un ordinador i resulta defectuós. Quina és la probabilitat de que hagi estat produït a la Unió Europea?
- La meitat dels ordinadors d'una determinada botiga són d'origen asiàtic, i l'altra meitat són europeus. Si comprem un ordinador triat a l'atzar d'aquesta botiga, quina és la probabilitat de que sigui defectuós?

-
- Per poder fer l'examen cal tenir un document vàlid d'identitat i tenir-lo a la vista en tot moment.
 - L'examen no es pot fer amb llapis.
 - **Totes les respostes han d'estar degudament justificades.**
 - Només es pot utilitzar la calculadora. No es pot tenir a l'abast, cap mena d'apunts, llibres, ni qualsevol altre tipus de material amb continguts relacionats amb l'assignatura.
 - Els telèfons mòbils, smartwatches, i qualsevol altre dispositiu han d'estar desconnectats i guardats. En cap cas es podran fer servir com a rellotge, com a calculadora, o per qualsevol altra funció.

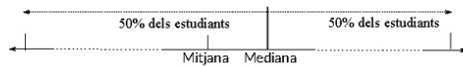


SOLUCIONES EXÁMENES PARCIALES EXAMEN 2019

1.

a) 38,2%

b) No



c) Asimetría por la izquierda, más dispersión por la cola izquierda que por la derecha

d) No se puede interpretar ya que las dos variables son cualitativas nominales

e) Relación positiva de poca intensidad $r=0,364$

f) Los hombres Economía, las mujeres ADE

g) Mirando el Boxplot, es un poco mayor la dispersión en las mujeres

2.

$$a) p(AII \cap BTM) = p(AII/BTM) \cdot p(BTM) = 0,12$$

$$b) p(AII) = p(AII/BTM) \cdot p(BTM) + p(AII/ATM) \cdot p(ATM) + p(AII/SIC) \cdot p(SIC) = 0,235$$

$$c) p(\overline{AII}) = 1 - p(AII) = 0,765$$

d) A BTM

$$p(BTM/AII) = \frac{p(AII/BTM) \cdot p(BTM)}{p(AII)} = 0,511 \quad \uparrow\uparrow$$

$$p(ATM/AII) = \frac{p(AII/ATM) \cdot p(ATM)}{p(AII)} = 0,266$$

$$p(SIC/AII) = \frac{p(AII/SIC) \cdot p(SIC)}{p(AII)} = 0,223$$

EXAMEN 2021

1.

a) 0,75%

$$b) \bar{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{600 \cdot 635 + 1600 \cdot 677 + \dots}{2000} = 1773$$

c) 28,57%

d) No se puede interpretar ya que las dos variables son cualitativas nominales

e) Relación negativa de alta intensidad

f) 1100-2100

g) El intervalo modal es 10-70, no podemos decir cuál es la moda exacta con un histograma

2.

$$a) p(F) = p(F/D) \cdot p(D) + p(F/S) \cdot p(S) = 0,017$$

$$b) p(S/\bar{F}) = \frac{p(\bar{F}/S) \cdot p(S)}{p(\bar{F})} = \frac{(1-p(F/S)) \cdot p(S)}{(1-p(F))} = 0,6$$

$$c) p(D/F) = \frac{p(F/D) \cdot p(D)}{p(F)} = 0,368$$

$$d) p(\bar{F} \cup \bar{S}) = \overline{p(F \cap S)} = 1 - p(F \cap S) = 1 - p(F/S) \cdot p(S) = 0,989$$

EXAMEN 2022**1.**

$$a) \bar{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{0 \cdot 3 + 1 \cdot 19 + \dots}{90} = 2,078$$

b) 43,3%

c) 10%

d) 23,08%

e) No son iguales ni las frecuencias absolutas (las condicionadas siempre serán menos, eso no nos dice nada) ni las frecuencias relativas (lo que significa que las dos variables están relacionadas).

f) Relación directa o positiva entre número de hijos y renta

2.

$$a) p(IS \cap PT) = p(IS/PT) \cdot p(PT) = 0,159$$

$$b) p(PT/IS) = \frac{p(PT \cap IS)}{p(IS)} = 0,306$$

$$c) p(IS/\overline{PT}) = \frac{p(\overline{PT}/IS) \cdot p(IS)}{p(\overline{PT})} = \frac{(1-p(PT/IS)) \cdot p(IS)}{(1-p(PT))} = 0,535$$

EXAMEN 2023**1.**

$$a) \bar{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{4 \cdot 160 + 5 \cdot 76 + \dots}{257} = 4,471$$

b) 12,84%

c) 9

d) 80%

e) Mo=2 Md=3

f) Sí, los que tuvieron mejor nota encontraron trabajo antes

2.

$$a) p(D) = p(D/A) \cdot p(A) + p(D/UE) \cdot p(UE) = 0,007$$

$$b) p(UE/D) = \frac{p(D/UE) \cdot p(UE)}{p(D)} = 0,176$$

$$c) p(D) = p(D/A) \cdot p(A) + p(D/UE) \cdot p(UE) = 0,006$$

Consulta todos nuestros cursos actualizados en:

www.estadistix.com

**Y si tienes cualquier duda,
escribenos un whatsapp al 644310902**

