

MODEL 2

1. Cas pràctic 1

Amb les dades associades als passatgers del Titànic es disposa de la següent taula de freqüències creuades entre els passatgers que van sobreviure a l'enfonsament i la classe en la qual viatjaven (primera, segona i tercera classe):

Supervivència \ Classe	1a Classe	2a Classe	3a Classe
NO	80	97	372
SI	136	87	119

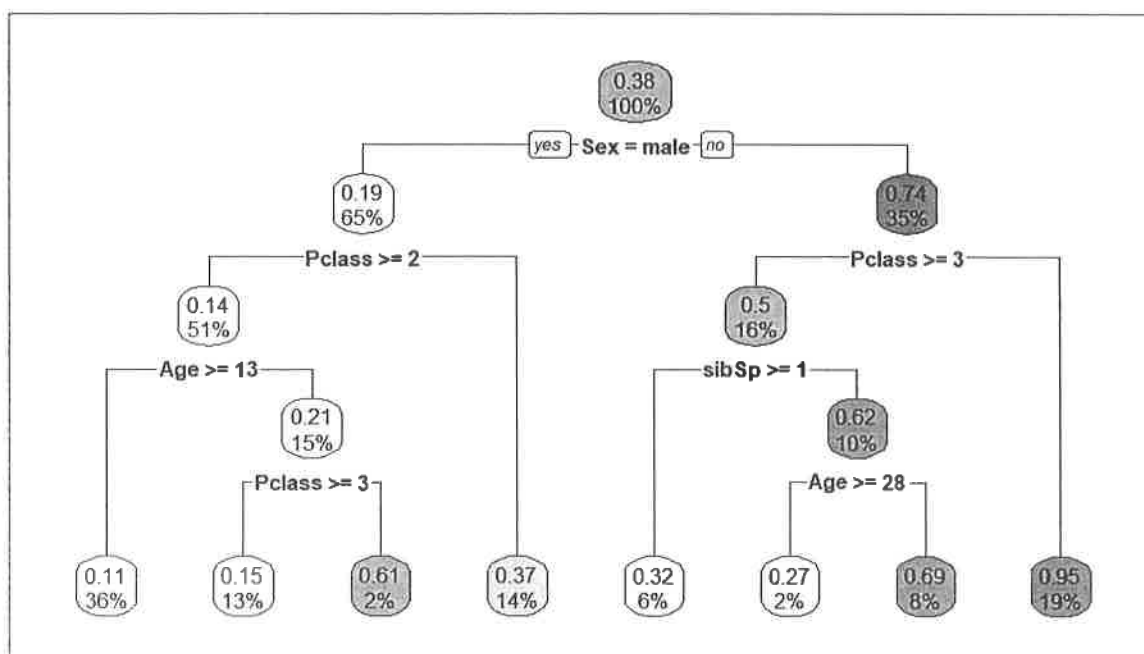
a) Amb la informació de la taula anterior calcula les probabilitats conjuntes i marginals.

b) Calcula les següents probabilitats condicionals: $P(SI | 1a Classe)$, $P(NO | 2a Classe)$, $P(3a Classe | SI)$ i $P(1a Classe | NO)$.

c) Obtengui la taula de freqüències creuades que tindríem en el cas d'independència entre la supervivència i la classe en la qual viatjava en el Titànic.

A més s'ha estimat un arbre (*Decision Tree* o *Regresion Tree*) per a explicar la probabilitat de supervivència (*Survived*, 1 en cas positiu i 0 en cas negatiu) emprant les següents variables explicatives: edat (*Age*, en anys), classe en la qual viatjaven (*Pclass*, 1 per a la 1a classe, 2 per a 2a classe i 3 per a la 3a classe), nombre de familiars amb els quals viatjava (*sibSp*) i finalment el sexe (*sex*, 1 si és home, 0 si és dona).

Els resultats obtinguts han estat els següents:



d) Interpreta els resultats presentats al quadre anterior.

2. Cas pràctic 2

En un determina zona turística se sap que va haver-hi 100 hotels oberts al llarg de 2021. El govern local desitja estimar la xifra de negoci (mesurada en unitats monetàries, um) facturada pels hotels en aquest any i per a això ha seleccionat, per mostreig aleatori simple, a 10 hotels obtenint els resultats següents:

Xifra de negoci (x_i)	Nombre d'hotels (n_i)	$x_i n_i$	$x_i^2 n_i$
400	4	1.600	640.000
500	3	1.500	750.000
600	2	1.200	720.000
1.000	1	1.000	1.000.000
TOTAL	n=10	5.300	3.110.000

- Estimar la xifra de negocis total de la zona turística.
- Estimar l'error de mostreig comès.
- Calcular l'interval de confiança amb un nivell de confiança del 95%, assumint que la població segueix una distribució normal.
- Si s'hagués volgut obtenir un error de mostreig igual o inferior a 3.500 unitats monetàries al mateix nivell de confiança, quants hotels hauria estat necessari enquestar?

Suposem que el govern local coneix per endavant la següent informació relativa a la categoria dels diferents hotels:

Estrat	Nombre d'hotels	Quasi-variància de l'estrat
Hotels de 3 estrelles o menys	80	10.033
Hotels de 4 o més estrelles	20	16.722

- Es vol analitzar la possibilitat de realitzar en estudis futurs un mostreig estratificat de grandària 10 amb afixació uniforme, calcular l'error de mostreig esperat.
- Calcular l'error de mostreig esperat si el futur estudi es realitzés aplicant un mostreig estratificat de grandària 10 amb afixació proporcional.
- Comentar els resultats obtinguts i justificar quin hauria de ser el disseny mostral seleccionat d'entre els analitzats als apartats anteriors.