

**ESERCIZIO**

La tabella indagine *StudentiFlorida*, disponibile sul sito (in formato pdf e xls), riporta i dati di un'indagine condotta su 60 studenti laureati presso la University of Florida, relativamente alle seguenti variabili:

COLONNA	NOME VARIABILE	DESCRIZIONE
1	id	Codice identificativo dello studente
2	genere	Genere dello studente
3	età	Età (in anni compiuti)
4	punteggioHighSchool	Punteggio ottenuto alla scuola superiore (scala da 1 a 4)
5	punteggioCollege	Punteggio ottenuto al College (scala da 1 a 4)
6	distanzaCampus	Distanza (in miglia) del campus dalla città di provenienza dello studente
7	distanzaAula	Distanza (in miglia) della classe dall'attuale luogo di residenza
8	oreTVsettimana	Numero medio di ore settimanali passate a guardare la TV
9	oreSportSettimana	Numero medio di ore settimanali dedicate allo sport
10	quotidianiPerSettimana	Numero di volte per settimana che si legge un quotidiano
11	AH	Numero di persone conosciute morte per AIDS o sieropositive
12	vegetariano	Regime di alimentazione vegetariano (yes, no)
13	affiliazionePolitica	Affiliazione politica (D = Democratico, R = Repubblicano, I = Indipendente)
14	ideologiaPolitica	Ideologia politica (1 = molto liberale, 2 = liberale, 3 = leggermente liberale, 4 = moderato, 5 = abbastanza conservatore, 6 = conservatore, 7 = molto conservatore)
15	religiosità	Partecipazione ad eventi religiosi (0 = mai, 1 = occasionalmente, 2 = la maggior parte delle settimane, 3 = ogni settimana)
16	opinioneAborto	Opinione circa la legalità dell'aborto nei primi tre mesi di gravidanza (yes, no)
17	azioniDiscriminazione	Impegno in azioni contro le discriminazioni (si, no)
18	vitaMorte	Crede nella vita dopo la morte (si -y, incerto - u, no - n)

Fonte: A. Agresti and B. Finlay, *Statistical Methods for the Social Science*, Prentice Hall, 2008

1) Secondo te i dati sono campionari o censuari? (motiva brevemente la risposta)

--

2) Riporta per ciascuna delle variabili il tipo (distinguendo tra variabili qualitative nominali/ordinali e variabili quantitative) e l'insieme delle possibili modalità/intensità:

NOME VARIABILE	TIPO	MODALITA' / VALORI
genere		
età		
punteggioHighSchool		
punteggioCollege		
distanzaCampus		
distanzaAula		
oreTVsettimana		
oreSport		
quotidianiPerSettimana		
AH		
vegetariano		
affiliazionePolitica		
ideologiaPolitica		
religiosità		
opinioneAborto		
azioniDiscriminazione		
vitaMorte		

3) Costruisci la tabella di frequenza per la variabile **GENERE**, riportando le frequenze assolute, relative e percentuali:

GENERE	$n_i$	$f_i$	$p_i$
F			
M			
<b>Totale</b>	<b>60</b>	<b>1</b>	<b>100</b>

4) Conoscendo la frequenza relativa è sempre possibile risalire alla frequenza assoluta? Spiega brevemente il procedimento.

--

5) Costruisci la tabella di frequenza per la variabile RELIGIOSITA', riportando le frequenze assolute, relative e percentuali:

RELIGIOSITA'	$n_i$	$f_i$	$p_i$
0			
1			
2			
3			
<b>TOTALE</b>	<b>60</b>	<b>1</b>	<b>100</b>

6) Costruisci la tabella a doppia entrata per le variabili GENERE e RELIGIOSITA' riportando le frequenze assolute:

		GENERE		
		F	M	
RELIGIOSITA'	0			
	1			
	2			
	3			
				60

7) Costruisci la tabella a doppia entrata per le variabili GENERE e RELIGIOSITA' riportando le frequenze percentuali:

		GENERE		
		F	M	
RELIGIOSITA'	0			
	1			
	2			
	3			
				100

8) A partire dalla tabella costruita al punto 6, calcola i **profili colonna**, ovvero le distribuzioni condizionate per la variabile RELIGIOSITA' e la distribuzione marginale in frequenze percentuali per la stessa variabile (profilo marginale), utilizzando le frequenze percentuali:

		GENERE		
		F	M	
RELIGIOSITA'	0			
	1			
	2			
	3			
		<b>100</b>	<b>100</b>	<b>100</b>

9) A partire dalla tabella costruita al punto precedente, verifica che il profilo marginale (ultima colonna) può essere ottenuta come media ponderata dei profili colonna, usando come pesi le frequenze osservate di ciascuna colonna (vedi tabella punto 5)

--

SUGGERIMENTO: la media pesata si ottiene come  $\frac{\sum(\text{valori} \times \text{pesi})}{\sum \text{pesi}}$

10) Guardando i profili colonna, ti sembra che ci sia un legame tra la le scelte religiose degli studenti e il loro genere? Motiva brevemente la risposta.

11) A partire dalla tabella costruita al punto 6, calcola i **profili riga**, ovvero le distribuzioni condizionate per la variabile GENERE e la distribuzione marginale in frequenze percentuali per la stessa variabile (profilo marginale), utilizzando le frequenze percentuali:

		GENERE		
		F	M	
RELIGIOSITA'	0			100
	1			100
	2			100
	3			100
				100

12) A partire dalla tabella costruita al punto precedente, verifica che il profilo marginale (ultima riga) può essere ottenuta come media ponderata dei profili riga, usando come pesi le frequenze osservate di ciascuna riga (vedi tabella punto 5)

SUGGERIMENTO: la media pesata si ottiene come  $\frac{\sum(\text{valori} \times \text{pesi})}{\sum \text{pesi}}$