

Esercitazione del corso di Informatica per Scienze Naturali Ambientali

Lezione 8 (Laboratorio) – 10 marzo 2010

Le soluzioni agli esercizi vanno messe assieme e salvate nel documento **esercitazione4-nomecognome.odt** (dove al posto di “nomecognome” dovete mettere il vostro vero nome e cognome) ed inviate per posta elettronica all'indirizzo nicotra@di.unipi.it entro il 24 marzo 2010 per essere tenute in considerazione nel punteggio finale dell'esame. Il file con le soluzioni (perlopiu' corrette) valge fino a 2 punti aggiuntivi rispetto ai 30 di base che verranno assegnati durante l'esame finale.

1) Tabelle delle frequenze e istogrammi

1. Scaricate i seguenti file e salvateli nella cartella del corso di informatica:
 - <http://www.di.unipi.it/~nicotra/corsoinformatica/Actors.ods>
 - <http://www.di.unipi.it/~nicotra/corsoinformatica/EnvironmentalPolicy.ods>
 - <http://www.di.unipi.it/~nicotra/corsoinformatica/HomeData.ods>
 - <http://www.di.unipi.it/~nicotra/corsoinformatica/Otis1.ods>
2. Iniziamo a lavorare sul file **EnvironmentalPolicy**. Il file descrive il risultato di un indagine demografica e contiene per ogni persona intervistata (ce ne sono 30) l'età (Age), il sesso (Gender), lo stato di residenza (State), il numero di figli (Children), il salario (Salary) oltre ad un campo su una opinione espressa in merito ad una domanda dell'intervistatore (Opinion).
3. Aprite il programma OpenOffice Calc (lo trovate nel menù dei programmi oppure aprendo una shell e digitando il comando “**oocalc**”) e il file EnvironmentalPolicy.ods.
4. Innanzitutto controlliamo che la distribuzione delle età tra uomini e donne sia simile.
5. Per fare questo decidiamo di dividere le età in tre fasce, quella fino a 34 anni, quella dai 34 ai 59 anni e quelli maggiori di 59 anni. Per fare questo scriviamo in un area vuota del fogli

34	< 34 anni
59	34-59
	>59

dove la prima colonna rappresenta i tre intervalli (notate la casella vuota per rappresentare i maggiori di 59 anni) e la seconda rappresenta invece le etichette che assegnamo agli intervalli (insomma la prima e per il computer la seconda è per noi)

6. Calcoliamo la piccola tabella delle frequenze selezionando le tre caselle vuote accanto a quelle che avete inserito al punto 5 e poi selezionando Insert->Function. Poi selezionate FREQUENCY tra le funzioni (poi Next), e cliccando sull'icona Select corrispondente a Data selezionate tutte le età delle intervistate donne nella colonna Age. Quando li avete selezionati cliccate nuovamente sull'icona. A questo punto cliccate invece sull'icona Select in corrispondenza di Classes e selezionate le tre caselle 34, 59 e quella vuota sottostante. Quindi cliccate su ok.
7. Ora vogliamo inserire un grafico. Per farlo dovete selezionare 6 caselle, quelle delle etichette inserite al punto 5 e quelle delle frequenze calcolate al punto 6. Poi selezionate Insert->Chart, spuntate la casella 3D look e date come titolo al grafico “Female Age”
8. Seguite lo stesso procedimento dei punti 5, 6 e 7 per calcolare il grafico delle età maschili. La distribuzione è la stessa? Scrivetelo nel documento dell'esercitazione, inserendo anche I due

grafici ottenuti (basta copiarli ed incollarli nel documento di testo).

9. Ora invece lavoriamo sul file **Actors** che dovete quindi aprire in Calc. Esso contiene i compensi di diversi attori di Hollywood.
10. Vogliamo anche in questo caso rappresentare un grafico dei guadagni, ma per farlo vogliamo calcolare automaticamente gli intervalli. Come? Dovete usare le funzioni MAX e MIN (si possono inserire come la funzione FREQUENCY al punto 6 per calcolare il massimo e il minimo valore della colonna Salary. A questo punto dividete per tre l'intervallo e trovate le soglie in modo opportuno. Inseritele come al punto 5 nel foglio di calcolo e calcolate il grafico risultante (utilizzando come al punto 5 delle etichette per le diverse colonne). Inseritelo nel documento dell'esercitazione, specificando anche gli intervalli prescelti.
11. Ora invece lavoriamo sul file **Otis1** che contiene il diametro dei cavi da ascensore prodotti da un'azienda (sono 399).
12. Volete dividere I valori in tre intervalli che contengano lo stesso numero di valori. A questo scopo ordinate i valori in senso decrescente. A questo punto la prima soglia sarà il valore alla riga 133, la seconda quella alla riga 266 e l'ultimo valore sarà l'ultima riga.
13. A questo punto, come per I dati precedenti, calcolate una tabella delle frequenze e il corrispondente grafico, da allegare al documento dell'esercitazione.

2) Statistica Descrittiva

1. Riprendiamo il file **EnvironmentalPolicy**. Utilizzando sempre Insert->Function calcolate la media, la mediana e la moda dei redditi di tutte le persone intervistate. Poi fate la stessa cosa per I redditi femminili e maschili separatamente. Quali valori ottenete? Quali considerazioni potete trarne? Provate a tenere in considerazione anche i grafici delle età e riportate tutto nel documento dell'esercitazione.
2. Calcolate anche deviazione standard (STDEV) e varianza (VAR) di tutti gli intervistati, degli uomini e delle donne. Riportate nel file i valori ottenuti e le considerazioni che potete trarne, riportando tutto nel documento dell'esercitazione.

3) Preprocessing dei dati

1. Ora lavoriamo sul file **HomeData** che contiene di una serie di case in vendita il prezzo, la dimensione, l'età e diversi altri attributi. A noi interessa lavorare sui prezzi.
2. Come fatto per i punti precedenti trovare le tre soglie per dividere i dati in tre insiemi della setessa dimensione e inseriamoli nelle celle A122:A124. Nella colonna successiva mettiamo invece le etichette price_low, price_medium, price_high.
3. A questo punto possiamo effettuare il preprocessing utilizzando la funzione IF (SE). Poichè I valori da confrontare sono 3 seguiremo il seguente algoritmo: SE valore < I soglia ALLORA "Price_low" ALTRIMENTI SE valore < II soglia ALLORA "Price_medium" ALTRIMENTI "Price_high". Per fare questo usate una colonna vuota, a cui date nome Price Label scrivete nella prima cella la funzione =IF(A2<=\$A\$123;"Price_low";IF(A2<=\$A\$124;"Price_medium";"Price_high")) e poi copiatela nelle celle successive.
4. Create un istogramma per mostrare la distribuzione dell'ultima colonna ed inseritela nel documento dell'esercitazione.