

Corso Integrato di Statistica Informatica e Analisi dei Dati Sperimentali

A.A 2012-2013

Esercitazione 1

1 istruzioni

Preparare una cartella di lavoro che contiene: i files dei dati (.txt), i files di lavoro (fogli elettronici Office o OpenOffice) e i files della relazione. Eventualmente utilizzate lo schema proposto sulla pagina web del corso per la relazione. Nominare i files in modo chiaro. Ad esempio: **EsA_Es01.txt**, **EsA_Es01.xls**, **EsA_Es01.doc** e riportare nei files il cognome e n. di matricola di tutti i componenti del gruppo.

Finita l'esercitazione salvare il proprio lavoro, chiudere i files, comprimere la cartella di lavoro (zip, rar,...) e inviarla via mail a:

CI.biologia@gmail.com

indicando nell'oggetto: i cognomi degli studenti che hanno partecipato.

2 Esercizi ed esperimenti pratici proposti

2.1 Esercizi

Scopo

- Importare i dati in un foglio elettronico.
- Presentare dati nominali e ordinali utilizzando tabelle di frequenza e istogrammi.
- Calcolare i principali indici di posizione e variabilità delle distribuzioni.

Svolgere **un paio di esercizi di diversa tipologia** e preparare una breve relazione per esercitarsi ad utilizzare tabelle, istogrammi e strumenti di calcolo dei fogli elettronici. Si possono usare i files proposti o, in alternativa, quelli riguardanti i dati raccolti in aula **dati(2014)**.

Seguire i link **Dati(20xx)** dove si trovano diversi dati ai cui lavorare.

Esercizi 1: importare dati in un foglio elettronico e sintesi di valori nominali.

- *Importare dati* in un foglio elettronico in formato ITA (virgola come separatore dei decimali) o UK (punto come separatore dei decimali). Dalla directory **dati(2013)** importare in un foglio elettronico i dati dal file **dati_ITA_virgola.txt** e **dati_UK_punto.txt**. Calcolare i principali indici statistici delle distribuzioni. (posizione: Media, moda, mediana; dispersione: deviazione standard, distanza interquartile, etc...)
- *Variabili nominali*: dalla directory **dati(2013)** importare i dati del file **a01_colore_pasticche.2013.txt**.

Calcolare la distribuzione di frequenze assolute e relative per colori. Riportare in un istogramma la distribuzione di frequenze (relativo o assoluta)

- *Variabili nominali*: nella directory **dati(2013)** il file **2013.03.13.Date di Nascita.analisi.xls(txt)** contiene informazioni sulla data/giorno di nascita di un gruppo di studenti, sul tipo di parto e sul genere. Confrontare le frequenze di parti naturali e artificiali di uomini e donne.

- *Variabili nominali*: nella directory **dati(2013)** il file **2013.03.13.Date di Nascita.analisi.xls(txt)** contiene informazioni sulla data/giorno di nascita di un gruppo di studenti, sul tipo di parto e sul genere. Confrontare la distribuzione delle nascite durante la settimana per parti cesareo e naturale.

- *Variabili nominali*: confrontare la distribuzione delle nascite (mese) per gli studenti registrati nel **2013** (**2013.03.13.Date di Nascita.dat.txt**) e nel **2011** (**Es.02.nascite.txt**).

- *Distribuzione di Variabili ordinali discrete*: nella directory **dati(2012)** i files: **a03_distribuzione.studenti.7.3.2012.txt** e **a03_distribuzione.studenti.5.3.2012.txt** riportano la mostrano la distribuzione degli studenti nelle diverse file di banche. Confrontare le due distribuzioni usando istogrammi di frequenze.

- *Distribuzione di Variabili ordinali discrete*: Studiare la distribuzione della lunghezza del cognome nei dati del file: **2013.03.13.Cognome.analisi.xls (dati(2013))** confrontando uomini e donne.

Esercizi 2: Sintesi dati Ordinali discreti I voti degli esami di tre studenti sono riportati nel file **Esami.txt**. Importare i voti di esame di tre studenti in un foglio elettronico.

- Calcolare i principali indici statistici di posizione: media, moda, mediana.
- Calcolare i principali indici statistici di dispersione: range, varianza e deviazione standard.
- Calcolare le tabelle di frequenza
- Rappresentare i dati in un istogramma di frequenza (assoluta e relativa)
- Salvare i risultati.

Una possibile realizzazione si può trovare nel file **Esami.xls**

Esercizi 3: sintesi statistica di dati ordinali

- A lezione (2012) sono stati pesati in aula alcuni bicchieri pieni di fagioli. I dati sono disponibili nella directory **dati(2012)** disponibili sul file **a05_PesoBicchiere2012.xls**. Calcolare le distribuzioni di frequenze (assolute/relative integrate/differenziali), riportare i dato su istogrammi, calcolare i principali indici di sintesi statistica. Il file **a05_PesoBicchiere2012.Analisi.xls** mostra un possibile uso delle funzioni del foglio elettronico.

- Nei dati **dati(2011)** nel file **Es.04.liposoma.txt** sono stati registrate le dimensioni di alcuni liposomi. Sintetizzare le informazioni utilizzando istogrammi di frequenza e indici di sintesi statistica.
- Sono stati registrati i ritmi cardiaci di

un certo numero di studenti **dati(2012)**, **a04_FerquenzaCardiaca.2012.xls**. Sintetizzare la distribuzione dei risultati utilizzando indici statistici, tabelle e istogrammi di frequenza.

• Studiare la distribuzione dell'altezza nel gruppo di studenti (**dati(2013)**: **2013.03.13.Cognome.analisi.xls**) confrontando uomini e donne.

2.2 Esperimenti

Scopo: acquisire dati sperimentali e presentare una breve sintesi delle informazioni contenute. Effettuare almeno un'esperimento e presentare una breve relazione.



Figura 1: Esempio di stima statistica della superficie di un oggetto complesso. I puntini sul bordo possono essere contati 1/2. Effettuare diverse misure sovrapponendo l'oggetto in modo diverso

Esperimento 1 Si prepara un sacchetto mischiando una quantità B di fagioli bianchi e una quantità N di fagioli neri. Effettuare una serie di misure pescando una manciata di fagioli dal sacchetto e registrare il numero di fagioli bianchi, il numero di fagioli neri (ogni volta rimettere i fagioli pescati nel sacchetto). Dati analoghi già registrati si trovano in **a02_FagioliBN_2012.txt (dati(2012))** e **Fagioli.BN.2013.dati.txt (dati(2013))**.

Sintetizzare le informazioni statistiche riportando:

- la distribuzione della variabile X=numero di fagioli per estrazione;
- la distribuzione delle variabili Y_B =frazione di fagioli bianchi per estrazione.

Riassumere le osservazioni in una breve relazione con tabelle e grafici.

Esperimento 2 su un foglio A4 disegnare un certo numero di puntini uniformemente distribuiti (almeno 100, massimo 200) (fig. 1). Sovrapporre una forma irregolare, es ritagliata, o una foglia e contare il numero di puntini coperti (o esterni). Effettuare diverse prove (almeno 20

e registrare i risultati su un file. Un foglio A4 ha superficie $S_{A4} = 21.0 \times 29.7 = 623.7\text{cm}^2$. E' lecito aspettarsi che il valore atteso della variabile $X = N_x/N_t$, data rapporto tra il numero di puntini coperti dall'oggetto N_x e il numero di puntini uniformemente distribuiti sul foglio N_t sia eguale al rapporto tra la superficie dell'oggetto S_x e la superficie del foglio S_{A4} . Quindi $S_x \approx \bar{X} * S_{A4}$ - Studiare la distribuzione della variabile X. - Studiare la distribuzione della variabile S_x .

Discutere i risultati in una breve relazione

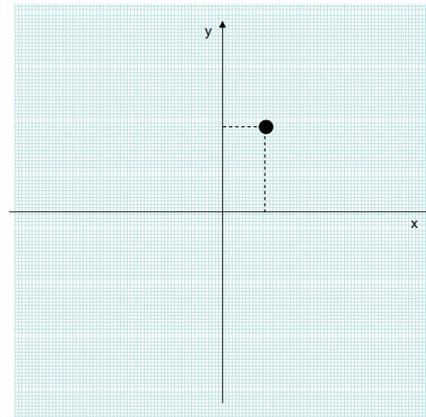


Figura 2: Esercizio di lancio. Utilizzare un foglio di carta millimetrata per determinare le posizioni oppure tracciare due assi ortogonali su un foglio A4 e usare un righello graduato.

Esperimento 3 su un foglio di carta millimetrata si tracciano gli assi di un riferimento cartesiano x,y (fig. 2). Da una certa distanza lanciare una pallina di carta cercando di mandarla nel centro. Registrare su un file ASCII le coordinate (x,y in cm) delle posizioni ottenute ad ogni lancio. Effettuare almeno 60 lanci.

#	i	x[cm]	y[cm]
1		4.5	2.3
2		-3.7	2.6
...			

- importare i dati in un foglio elettronico e salvare il foglio di lavoro in formato opportuno (.xls) per evitare di perdere il lavoro fatto alla chiusura del programma.
- Per ogni punto calcolare la distanza media dall'origine $d = \sqrt{x^2 + y^2}$.
- Riportare le coordinate in un grafico xy.
- Calcolare i principali indici statistici per le distribuzioni delle variabili X,Y (coordinate) e D (distanza).
- Utilizzare istogrammi e/o tabelle di frequenza per confrontare qualitativamente le distribuzioni.
- Riassumere l'esperienza in una breve relazione.