

Laboratorio
Excel

Sessione n. 1

Venerdì 031014 Gruppo PZ

Lunedì 061014 Gruppo AD

Martedì 071014 Gruppo EO

PROGRAMMA SVOLTO NELLA SESSIONE N. 1

(I) Esercizi di raccordo con quanto sinora visto a lezione sulle v.a. (variabili aleatorie):
Esercizi di Calcolo delle Probabilità per la v.a. di Poisson con utilizzo del comando Excel DISTRIB.POISSON.

(II) Esercizio di SIMULAZIONE dei valori osservati di una v.a. (in particolare Poisson) con utilizzo del comando Excel RANDOM NUMBER GENERATION.

(III) Introduzione alla nozione di Data Base aziendale e alla Analisi Statistica dei dati aziendali (o Statistica Descrittiva). Introduzione alle nozioni di:

(a) Dati grezzi o raw data,

(b) Trasformazione dei dati grezzi in informazioni per la gestione aziendale espresse nella forma di Indici statistici, Tabelle e Grafici,

(c) Nozione generale di Indice Statistico. In particolare: frequenza assoluta di un valore osservato, e Min (minimo), Max (massimo), Moda e Media dei valori osservati di una variabile,

(d) TABELLA DELLE FREQUENZE dei valori osservati di una variabile e uso dei comandi Excel TABELLA PIVOT per produrre la Tabella delle frequenze.

(e) Esercizio di produzione delle prime due colonne della Tabella delle frequenze per la variabile $X = \text{'numero di ordini giornalieri che arrivano ad un magazzino in } N \text{ giornate di magazzino aperto'}$,

(f) specificazione dei valori numerici, e del significato, dei principali indici statistici che si leggono direttamente dalla lettura delle prime due colonne Tabella delle frequenze,

(g) calcolo della Media dei valori osservati di una variabile con il comando Excel MEDIA e SOMMA (diviso N); la Media dei valori osservati di una variabile come approssimazione o stima della media dei valori possibili della variabile stessa.

SVOLGIMENTO DETTAGLIATO DELLA SESSIONE N. 1

(I) Esercizi di raccordo con quanto sinora visto a lezione sulle v.a. (variabili aleatorie):
Esercizi di Calcolo delle Probabilità per la v.a. di Poisson con utilizzo del comando Excel DISTRIB.POISSON.

Esercizi sulla v.a. di Poisson con i comandi Excel.

Si consideri la v.a. $X = \text{'n. di ordini di acquisto arrivati ad un magazzino in una giornata di magazzino aperto'}$, e sapendo che il n. medio di ordini che arrivano in una giornata è di 15 ordini in media, si risponda alle seguenti domande (con l'uso del comando Excel per la v.a. di Poisson con $m_i = 15$)

Domanda A1: $P(X=6)=p(6)=?$

Domanda A2: $P(X \leq 6)=?$

Comando Excel per la v.a. di Poisson	DISTRIB.POISSON(*,**,***)
* x=6	** mi della Poisson
	*** 0 o Falso --> p(x)
	1 o Vero --> P(X<=x)

Risposta domanda (A1):

$P(X=6)=p(6)=?$ 0,004839

Risposta domanda (A2):

$P(X \leq 6) = ?$ 0,007632

Domanda (A3):

$P(10 \leq X \leq 20) = ?$ 0,847175

$P(X \leq 20) = ?$ 0,917029

$P(X < 10) = P(X \leq 9) = ?$ -0,06985

0,847175

Domanda (A4):

$P(10 < X < 20) = ?$ 0,756754

$P(X < 20) = P(X \leq 19) = ?$ 0,875219

$P(X \leq 10) = ?$ -0,11846

0,756754

Domanda (A5):

$P(X \geq 12) = 1 - P(X \leq 11) = ?$ 0,815248

$P(X \leq 11) = ?$ 0,184752

0,815248

(II) Esercizi di SIMULAZIONE dei valori osservati di una v.a. (in particolare Poisson) con utilizzo del comando Excel RANDOM NUMBER GENERATION.

SIMULAZIONE DEI VALORI OSSERVATI DI UNA VARIABILE ALEATORIA.

Esercizio. Si simuli il numero di ordini di acquisto che arrivano ad un magazzino in 300 giorni (di magazzino aperto) sapendo che il numero medio di ordini che arrivano al magazzino in una giornata è di 15 ordini (si usi il numero generatore 4).

Comando di simulazione:

_ Barra strumenti DATI

_ Data Analysis

_ Random Number Generation

_ numero delle variabili aleatorie (da cui si vuole simulare)

_ numero dei valori che si vogliono ottenere con la simulazione

_ Distribution: Poisson(*), Binomiale, etc.

_ Parametri della distribuzione

_ Random seed o numero generatore

_ Nuovo foglio di lavoro (per accogliere i valori simulati)

(* per Poisson il parametro è "mi" anche se è scritto "lambda")

Si riportano qui di seguito (nella colonna 'n. ord. g.' = n. di ordini giornaliero) i primi 5 valori e gli ultimi 5 valori dei 300 valori simulati (la colonna 'n. gg.' indica il numero progressivo dei giorni):

n. gg.	n.ord. g.
1	11
2	11
3	24
4	12
5	15

n. gg.	n.ord. g.
296	13
297	16
298	13
299	14
300	14

(III) Introduzione alla nozione di Data Base aziendale e alla Analisi Statistica dei dati aziendali (o Statistica Descrittiva).

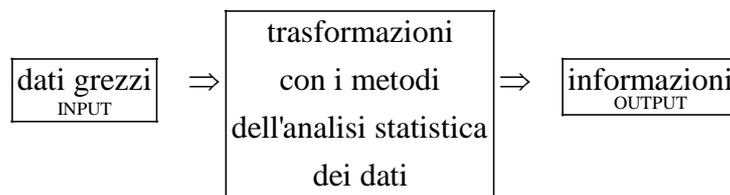
Introduzione alle nozioni di:

- (a) Dati grezzi o raw data,
- (b) Trasformazione dei dati grezzi in informazioni per la gestione aziendale espresse nella forma di Indici statistici, Tabelle e Grafici,
- (c) Nozione generale di Indice Statistico. In particolare (si vedano gli esercizi svolti): frequenza assoluta di un valore osservato, e Min (minimo), Max (massimo), Moda e Media dei valori osservati di una variabile,

IL RUOLO DELL'ANALISI STATISTICA DEI DATI NELLA GESTIONE AZIENDALE

DAI DATI GREZZI ALLE INFORMAZIONI: INDICI STATISTICI, TABELLE E GRAFICI

Un DB aziendale (Base Dati o Data Base) può avere migliaia di record (o righe) e decine o centinaia di variabili (o colonne). Es. 3000 record e 50 colonne, ovvero 150000 campi contenenti altrettanti dati (numeri e/o caratteri alfabetici). E' chiaro che leggere tutti tali 150000 dati, o anche leggere solo pochi di essi, non dà nessuna informazione sull'andamento della gestione aziendale (es. sull'andamento delle vendite nel caso del DB delle fatture attive). Infatti i dati che si trovano nei campi dei DB sono detti DATI GREZZI (o "raw data") perché debbono essere trasformati con i metodi dell'analisi statistica per avere informazioni sull'andamento della gestione aziendale.



Cosa è una informazione?

Una informazione è sempre un dato ma espresso informa tale (e con proprietà tali) da poter essere usato nei processi decisionali aziendali e nelle corrispondenti azioni attuative delle decisioni aziendali (si noti: in-forma-azione). Le informazioni sono espresse da indici statistici, tabelle e grafici.

Un INDICE STATISTICO è un numero che da solo indica una caratteristica di tutti i dati grezzi contenuti in una o più colonne del DB, ovvero di una o più variabili del DB. Tali indici hanno proprietà informative (o "descrittive") circa i dati grezzi che si trovano nel DB e possono avere, con opportune ipotesi, proprietà inferenziali, cioè previsive in senso probabilistico, circa certi dati futuri aziendali che non si trovano nel DB.

Le TABELLE che contengono informazioni su una sola variabile del DB sono dette tabelle uni-dimensionali o uni-variate e fanno parte dell'analisi statistica detta uni-dimensionale o uni-variate perché analizza una sola variabile del DB isolatamente o disgiuntamente dalle altre.

Le TABELLE che contengono informazioni su due variabili del DB sono dette tabelle bi-dimensionali o bi-variate o anche tabelle a doppia entrata e fanno parte dell'analisi statistica detta bi-dimensionale o bi-variate perché analizza congiuntamente due variabili del DB.

Noi produrremo sia le tabelle uni-dimensionali sia le tabelle bi-dimensionali con gli stessi comandi Excel, cioè quelli delle TABELLE PIVOT di Excel.

I GRAFICI evidenziano certe informazioni in modo intuitivo ed immediato. Noi vedremo in particolare, per l'analisi statistica uni-dimensionale, i grafici di sequenza ed i grafici a barre, e per l'analisi statistica bi-dimensionale i grafici che mettono in evidenza il grado di correlazione di due variabili.

(III) Introduzione alla nozione di Data Base aziendale e alla Analisi Statistica dei dati aziendali (o Statistica Descrittiva).

(d) TABELLA DELLE FREQUENZE dei valori osservati di una variabile e uso dei comandi Excel TABELLA PIVOT per produrre la Tabella delle frequenze. Esercizio di produzione delle prime due colonne della Tabella delle frequenze per la variabile X = 'numero di ordini giornalieri che arrivano ad un magazzino in N giornate di magazzino aperto',

(e) Esercizio: specificazione dei valori numerici, e del significato, dei principali indici statistici che si leggono direttamente dalla lettura delle prime due colonne Tabella delle frequenze: min, max, moda.

(f) calcolo della Media dei valori osservati di una variabile con il comando Excel MEDIA e SOMMA (diviso N); la Media dei valori osservati di una variabile come approssimazione o stima della media dei valori possibili della variabile stessa.

(d) **ESERCIZIO.** I 300 valori simulati precedentemente della variabile X = 'n. di ordini giornaliero che arrivano ad un magazzino' vengono considerati in questo esercizio come i valori del n. di ordini giornaliero

effettivamente arrivati ad un magazzino negli ultimi 300 giorni lavorativi. Si richiede di produrre con i comandi Excel Tabella pivot le prime due colonne della Tabella delle frequenze per tali 300 valori osservati della variabile X. **Svolgimento:** anticipiamo qui sotto il risultato ottenuto (i comandi Tabella-pivot sono riportati piu' sotto).

n. ord. g. (numero giornaliero di ordini)	frequenza assoluta
5	1
7	2
8	2
9	7
10	18
11	15
12	30
13	25
14	27
15	37
16	24
17	25
18	19
19	17
20	16
21	12
22	7
23	8
24	3
25	4
31	1
	300

I valori osservati della variabile **n. ord. g.** sono $N = 300$.

La prima colonna riporta solo $n = 21$ valori (dal piu' piccolo al piu' grande) che sono i valori osservati senza le ripetizioni degli stessi che si sono avute nei 300 dati. Tali ripetizioni sono riportate a fianco di ciascun valore x_i ($i=1,2,\dots,21$) nella colonna delle **frequenze assolute** di tali valori x_i

La **frequenza assoluta** di un valore x_i ($i=1,2,\dots,n$) osservato di una variabile e' il numero delle volte che tale valore si e' ripetuto negli N dati osservati.

Simboli: n_i = frequenza assoluta del valore x_i ($i=1,2,\dots,n$).

Domanda: quando si ha che $N = n$? **Risposta:**

quando ogni valore osservato non si e' mai ripetuto, ovvero

quando ogni valore osservato si e' osservato una sola volta e quindi tutti i valori osservati hanno avuto frequenza assoluta = 1), ovvero

quando tutti i valori osservati sono stati diversi tra loro.

Domanda: a quanto e' uguale la somma delle frequenze assolute? **Risposta:**

La somma delle frequenze assolute e' uguale al numero N dei valori osservati. Nel nostro caso degli ordini giornalieri, la somma delle frequenze assolute (nell'ultima riga della seconda colonna) e' uguale a 300 che e' proprio il numero totale dei valori osservati $N = 300$.

Simboli: $\sum_i n_i = N$

Gli N valori osservati di una variabile si dicono Dati grezzi (come gia' visto), ma si dicono anche che sono **N realizzazioni della variabile** perche' di tutti i valori possibili della variabile sono quelli che si sono realizzati nelle N osservazioni della variabile stessa.

COMANDI DI INIZIALIZZAZIONE TABELLA-PIVOT

COMANDI DI INIZIALIZZAZIONE TABELLA-PIVOT:

Barra strumenti: INSERISCI

_ Clik su icona *tabella-pivot* (in alto a sinistra di barra strumenti di *inserisci*)

_ viene richiesto: di selezionare le righe e colonne occupate dal Data Base, e di specificare “nuovo foglio di lavoro” in cui verra’ inizializzata la Tabella-pivot, cio’ fatto:

_ Se compare il *layout-classico* di tabella-pivot (con i suoi campi-riga, campi o area dati, e campi-colonna) l’inizializzazione è terminata.

_ Se non compare il *layout-classico*: fare click-destro sul riquadro che compare, selezionare *opzioni tabella-pivot*, selezionare *visualizza*, selezionare *layout-classico* che a questo punto compare e l’inizializzazione è terminata.

Il *layout-classico* e’ costituito da tre aree: un’area a sinistra detta *campi-riga*, un’area in alto detta *campi-colonna*, ed un’area centrale detta *campi-valore*.

COMANDI TABELLA-PIVOT che danno le prime due colonne della Tabella delle frequenze di cui sopra

COMANDI TABELLA-PIVOT che danno le prime due colonne della tabella delle frequenze di cui sopra

Trascinamento nei *campi-riga* del nome **n. ord. g.** della variabile

Click destro su una riga qualsiasi (nei *campi-riga*) della tabella-pivot

Nel menù che compare, click su **raggruppa**

Nella finestra che compare, in “**raggruppa per**” mettere “1” (se non c’è già)

Trascinamento nei *campi-valore* del nome **n. ord. g.** della variabile

Click destro su una riga qualsiasi (nei *campi-riga*) della tabella-pivot

Nel menù che compare, click su “**separa**”

RISULTATO:una tabella con la prima colonna contenente i valori della variabile ordinati dal più piccolo al più grande senza le ripetizioni, e con la seconda colonna contenente le frequenze assolute di detti valori, che sono cioè le prime due colonne della tabella delle frequenze riprodotta sopra.

(e) **ESERCIZIO:** Si specifichino i valori numerici, ed il significato, dei principali indici statistici che si leggono direttamente dalla lettura delle prime due colonne Tabella delle frequenze: min, max, moda.

Svolgimento.

Min = e’ il minimo valore dei dati osservati della variabile **n. ord. g.** → **Min = 5**

Max = e’ il massimo valore dei dati osservati della variabile **n. ord. g.** → **Max = 31**

Moda = e’ quel valore dei dati osservati che si e’ ripetuto piu’ volte di tutti, cioe’ il valore che ha avuto la massima frequenza assoluta → Moda = **15** (perche’ la sua frequenza assoluta 37 e’ la piu’ grande di tutte le frequenze assolute).

(f) **ESERCIZIO:**

(1) Si determini il numero di ordini medio giornaliero.

(2) Si confronti il valore medio ottenuto con la media (=15) della variabile di Poisson che ha generato i 300 dati grezzi. **Svolgimento:**

(1) Comandi Excel:

$\text{= MEDIA(selezione delle celle che contengono i dati)}$ → media = 15,49333

$\text{= SOMMA(selezione delle celle che contengono i dati)/N}$ → media = $4648/300 = 15,49333$

(2) Confronto fra le due medie: si nota che la media dei dati grezzi e’ una approssimazione della media della variabile di Poisson che ha generato i dati stessi. Sotto condizioni molto generali, e’ molto alta la probabilita’ che tanto piu’ grande e’ il numero **N** dei dati grezzi tanto meglio la media dei dati grezzi approssima la media della variabile che ha generato i dati stessi.