

LABORATORIO - MINITAB N.1

VARIABILI QUALITATIVE

DESCRIZIONE DEI DATI DA ESAMINARE

A un campione di studenti del Michigan sono state poste una serie di domande per capire quali obiettivi si pongono per il futuro. In particolare, sono interrogati su quale aspetto tra "grado di istruzione raggiunta", "pratica di uno Sport", "Aspetto fisico" e "Situazione Economica" siano più influenti per avere una buona popolarità.

Nel questionario vengono inoltre richiesti genere, livello scolastico e altre informazioni demografiche. Si vogliono analizzare i dati.

Numero dei ragazzi intervistati: 478

Nomi delle variabili:

1. Genere: maschio (M) o femmina (F)
2. Anni: età dello studente
3. Distretto della scuola: rurale (rural) , suburbano (suburban) , urbano (urban)
4. Obiettivi: alto grado di istruzione (1), essere popolare (2), essere un bravo sportivo (3)
5. Istruzione: "raggiungere un buon livello di istruzione" è molto importante(1), importante (2), abbastanza importante (3), poco importante (4) per la popolarità
6. Sport: "pratica dello sport" è molto importante (1), importante (2), abbastanza importante (3), poco importante (4) per la popolarità
7. Aspetto: "essere belli" è molto importante (1), importante (2), abbastanza importante (3), poco importante (4) per la popolarità
8. Soldi: "avere molti soldi" è molto importante (1), importante (2), abbastanza importante (3), poco importante (4) per la popolarità

I dati sono raccolti nel dataset: [Students' Goal](#)

ANALISI STATISTICA IN MINITAB

Iniziamo con l'analisi descrittiva delle variabili considerate singolarmente.

Essendo tutte variabili qualitative, le rappresentazioni usuali sono le tabelle di contingenza o i diagrammi a barre.

Tabelle di contingenza (a una via)

Columns: Genere

	boy	girl	All
Count	227	251	478
% of Total	47,49	52,51	100,00

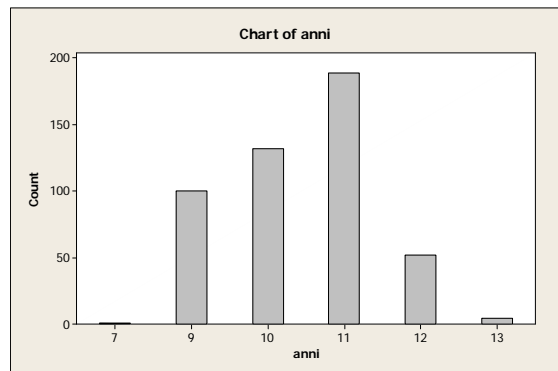
Commento: i maschi e le femmine sono ugualmente rappresentati.

Stat→Table→Descriptive Statistics

Selezionare "Genere" nelle *Categorical variables for columns*.

Nel menu "*Categorical Variables*", scegliere *count* (per i conteggi) e *Total percent* (per le frequenze).

Diagrammi a barre e diagrammi circolari per singole variabili

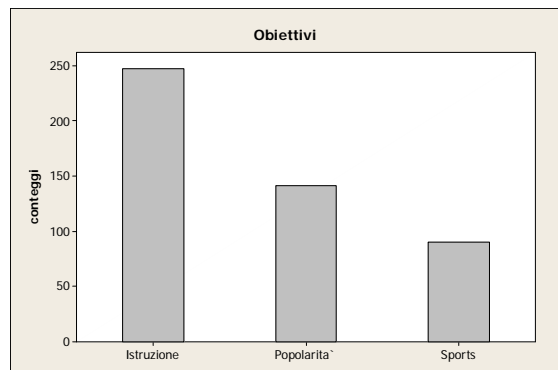


Graph→Bar Chart→Simple

Selezionare "Anni" in *Graph variables*.

Se si "clicca" sulle barre del grafico, si leggono esattamente i conteggi.

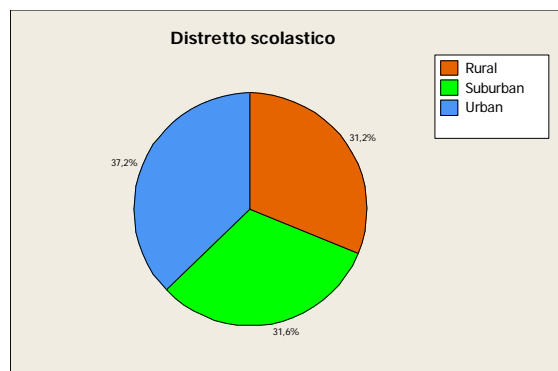
Commenti: gli anni più rappresentati sono i 10 e 11.



Graph→Bar Chart→Simple

Selezionare "Obiettivi" in *Graph variables*.

Commenti: I ragazzi preferiscono raggiungere un buon grado di istruzione.



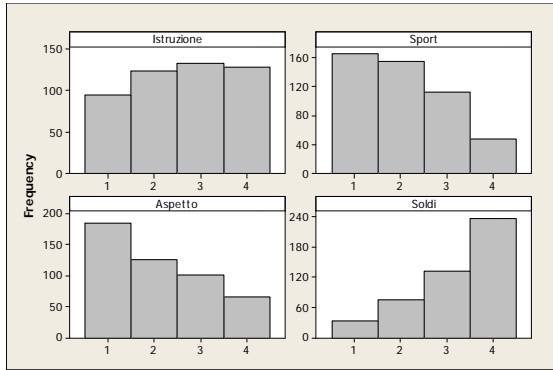
Graph→Pie Chart

Selezionare "*Chart counts of unique values*" e "*Distretto Scolastico*" per *Categorical variables*.

Selezionare nel menu "*Labels*" "*Percent*" per avere le percentuali.

Commento: i distretti scolastici sono ugualmente rappresentati.

Diagrammi a barre di più variabili in un unico "panel"



Graph→Histogram→Simple

Selezionare in "Graph variables" Istruzione, Sport, Aspetto e Soldi e scegliere nel menu "Multiple Graphs" l'opzione "In separate panels of the same graph".

Commento: I ragazzi pensano che per la popolarità sia importante eccellere in uno Sport e avere una bella presenza. Mentre i Soldi non servono per la popolarità.

Dall'analisi appena svolta, si capisce che, per comprendere meglio le opinioni dei ragazzi intervistati, è necessaria un'analisi incrociata delle variabili. Ad esempio può essere interessante conoscere se gli obiettivi sono uguali per Maschi e Femmine? Oppure se il distretto scolastico di appartenenza influenza gli obiettivi dei ragazzi?

Tabelle di contingenza a due vie

- conteggi

Rows: Obiettivi Columns: Genere

	boy	girl	All
1	117	130	247
	51,54	51,79	51,67
2	50	91	141
	22,03	36,25	29,50
3	60	30	90
	26,43	11,95	18,83
All	227	251	478
	100,00	100,00	100,00

Cell Contents: Count
% of Column

Stat→Tables→Descriptive statistics

Selezionare come "Categorical variables"

For rows: Obiettivi

For columns: Genere.

Selezionare nel menu "Categorical variables" l'opzione "Counts" per avere i conteggi e "Columns percent" per avere la tabella profilo colonna.

Commenti: Raggiungere un buon livello di istruzione è il primo obiettivo per Maschi e Femmine. Mentre una percentuale di femmine maggiore rispetto ai maschi preferisce raggiungere una buona popolarità piuttosto che eccellere nello Sport.

- profili riga

Rows: Distretto scolastico Columns: Obiettivi

	1	2	3	All
Rural	38,26	33,56	28,19	100,00
Suburban	57,62	27,81	14,57	100,00
Urban	57,87	27,53	14,61	100,00
All	51,67	29,50	18,83	100,00

Cell Contents: % of Row

Stat→Tables→Descriptive Statistics

Selezionare come "Categorical variables"

For rows: Distretto scolastico

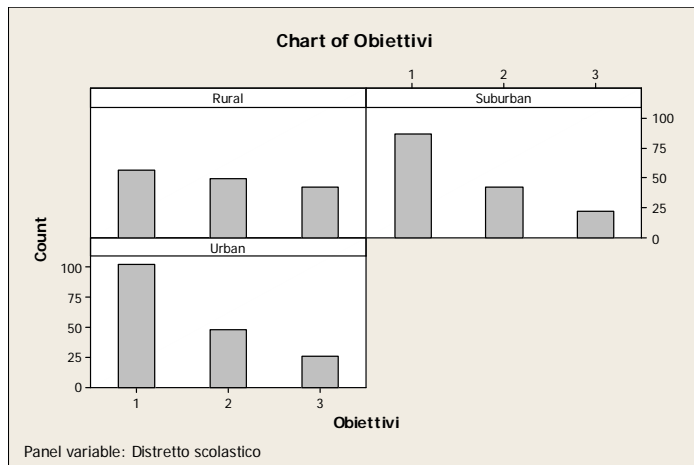
For Columns: Obiettivi

Selezionare nel menu "Categorical variables" l'opzione "Rows percent" per avere la tabella profilo riga.

Commento: il distretto scolastico di appartenenza e gli obiettivi sembrano dipendenti.

Passiamo ora alla rappresentazione grafica di più variabili.

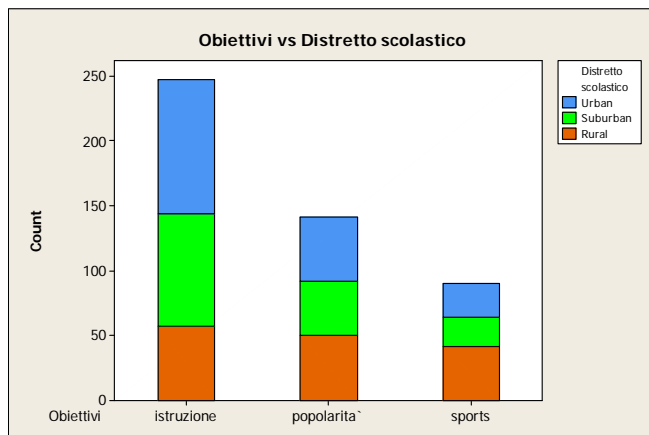
Diagrammi a barre di una variabile suddivisi secondo i livelli di un'altra variabile (profili)



Graph→Bar Chart→Simple

Selezionare "Obiettivi" in *Graph variables* e nel menu "Multiple Graphs", scegliere "Distretto scolastico" in "by variables with groups in separate panels"

Diagrammi a barre di una variabile con evidenziata la distribuzione di un'altra variabile in ciascuna barra (distribuzione congiunta)



Graphs→Bar Charts

Selezionare "Counts of unique values"

→Stack

Selezionare come "Categorical variables" *Obiettivi* e *Distretto scolastico* (in questo ordine!)

Nota: per **modificare i colori** cliccare **sulla parte di barra da modificare** poi

Attributes → Fill pattern

Selezionare *Custom*. In *Background color* scegliere il colore.

Commenti: Per i ragazzi che provengono da un distretto scolastico suburbano e urbano è più importante raggiungere un buon livello di istruzione. Per i ragazzi provenienti da un distretto scolastico rurale i tre obiettivi si equivalgono.

Tabelle di contingenza a due vie con i profili riga e colonna

Rows: Istruzione Columns: Genere

	boy	girl	All
1	41,49	58,51	100,00
2	49,59	50,41	100,00
3	50,38	49,62	100,00
4	46,88	53,13	100,00
All	47,49	52,51	100,00

Cell Contents: % of Row

Rows: Sport Columns: Genere

	boy	girl	All
1	76,97	23,03	100,00
2	38,96	61,04	100,00
3	27,68	72,32	100,00
4	19,15	80,85	100,00
All	47,49	52,51	100,00

Cell Contents: % of Row

Rows: Aspetto Columns: Genere

	boy	girl	All
1	23,78	76,22	100,00
2	58,73	41,27	100,00
3	58,42	41,58	100,00
4	75,76	24,24	100,00
All	47,49	52,51	100,00

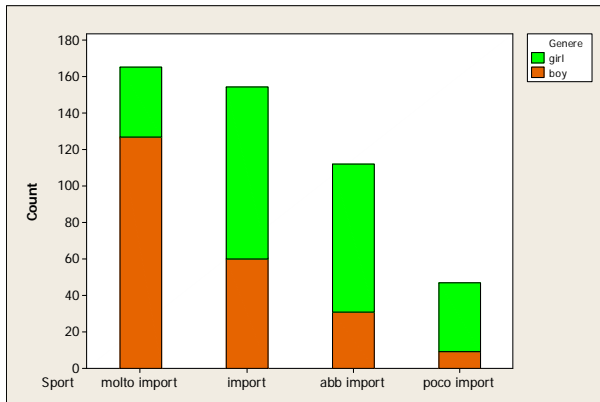
Cell Contents: % of Row

Rows: Soldi Columns: Genere

	boy	girl	All
1	50,00	50,00	100,00
2	42,67	57,33	100,00
3	53,03	46,97	100,00
4	45,57	54,43	100,00
All	47,49	52,51	100,00

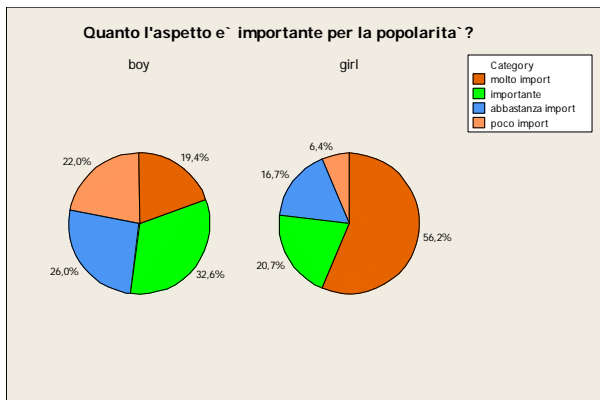
Cell Contents: % of Row

Commenti: Si nota che non c'è indipendenza fra la variabile genere e le variabili aspetto, soldi e sport. Mentre sia maschi che femmine valutano approssimativamente allo stesso modo l'importanza di una buona istruzione per ottenere popolarità.



Ad esempio, eccellere nello sport sembra essere molto importante per i maschi, ma non per le femmine.

Diagrammi circolari di una variabile suddivisi secondo i livelli di un'altra variabile (profili)



Le femmine reputano più importante avere un bell'aspetto.

Graphs→Pie Chart

Selezionare "Chart counts of unique values" e "Aspetto" come "Categorical variables". Nel menu "Multiple Graphs" selezionare "Genere" in "by variables with groups on same graph" nel sotto menu "by variables" e "on the same graph" nel sottomenu "multiple variables".

Nota: per **modificare i colori** procedere come per i diagrammi a barre