

1 – Analisi statistica bivariata

- Lo studio di un fenomeno di interesse, generalmente, si svolge rilevando contemporaneamente più caratteri su ciascuna unità statistica per spiegare il fenomeno attraverso la relazione tra esso ed altre variabili.

La statistica bivariata consiste nello studio del comportamento di due caratteri osservati congiuntamente sulle stesse unità statistiche

- Si indaga su come si manifesta un carattere al variare dell'altro mediante
 - la costruzione delle distribuzioni condizionate
 - grafici
 - le misure di associazione

2 – Analisi statistica bivariata

Matrice di dati

Cartel13 - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	studente	sexso	età	residenza	maturità	corso di laurea	num.esami sostenuti	sport praticati	tempo libero	conoscenza lingue	fumo	soddisfazione per la città
2	1	F	18	cosenza	scientifica	economia	2	tennis	libri	si	si	bassa
3	2	M	19	catanzaro	classica	ingegneria	4	calcio	musica	si	no	media
4	3	M	20	cosenza	classica	farmacia	6	calcio	teatro	si	si	media
5	4	M	21	cosenza	tecnica	ingegneria	7	calcio	cinema	no	si	alta
6	5	F	18	cosenza	tecnica	scienze politiche	0	nuoto	cinema	no	no	alta
7	6	F	18	cosenza	tecnica	economia	0	nessuno	amici	no	no	alta
8	7	F	19	cosenza	tecnica	farmacia	3	nuoto	sport	no	si	bassa
9	8	M	26	cosenza	classica	matematica	15	box	sport	no	si	bassa
10	9	F	19	cosenza	classica	matematica	4	pilates	amici	no	no	media
11	10	M	19	cosenza	classica	scienze politiche	5	calcio	amici	si	no	media
12	11	F	19	catanzaro	classica	scienze politiche	2	pilates	cinema	si	no	alta
13	12	F	18	catanzaro	scientifica	biologia	1	pilates	cinema	no	si	media
14	13	M	19	catanzaro	scientifica	lettere	2	spinning	teatro	si	si	bassa
15	14	F	19	catanzaro	scientifica	storia	2	spinning	libri	no	si	media
16	15	M	18	crotone	scientifica	economia aziendale	0	nessuno	musica	si	si	media
17	16	M	20	reggio calabria	scientifica	giurisprudenza	3	nessuno	musica	no	si	bassa
18	17	M	20	napoli	scientifica	giurisprudenza	3	nessuno	sport	no	si	alta
19												
20												
21												
22												

Foglio1 | Foglio2 | Foglio3

Pronto | 100% | 19:29 | lunedì | 14/02/2011

3 – Analisi statistica bivariata

Due caratteri rilevati su un collettivo di studenti:

$X = \text{'sesso'}$ e $Y = \text{'comportamento rispetto al fumo'}$

su $n=77$ studenti si rilevano le coppie di dati:
 (F, si) (M, si) (M, si) (F, no) (F, si) (M, si)

e poi si costruisce la tabella doppia che classifica le unità per ciascuna coppia di modalità osservate

X\Y	F	M
si	12	15
no	18	32

Numero di volte in cui compare (F, si)



Dati sul disastro del Titanic: lista passeggeri

- 1, **I**, **1**, Miss Elisabeth Walton, Southampton, St Louis, MO, B-5, 24160, female
- 2, **I**, **0**, Miss Helen Loraine, Southampton, Montreal, PQ / Chesterville, female
- 3, **I**, **0**, Mr Hudson Joshua, Creighton, Southampton, Montreal, PQ / Chesterville, male

Rosso: classe (I, II, III, personale di bordo)

Blu: esito 1 = sopravvissuto, 0 = morto

5 – Analisi statistica bivariata

Dati sul disastro del Titanic (tabella di contingenza)

Classe	Sopravvissuti		Totale
	No	Si	
I	122	203	325
II	167	118	285
III	528	178	706
Personale	673	212	885
Totale	1490	711	2201

n_{ij} frequenze nella cella di riga i e colonna j

$n_{i.}$ totale frequenze riga i (marginali riga)

$n_{.j}$ totale frequenze colonna j (marginali colonna)

N totale frequenze di tabella (2201)

Simbologia delle tabelle doppie

		Y							
		y_1	y_2	y_3	...	y_j	...	y_s	
X	x_1	n_{11}	n_{12}	n_{13}	...	n_{1j}	...	n_{1s}	$n_{1.}$
	x_2	n_{21}	n_{22}	n_{23}	...	n_{2j}	...	n_{2s}	$n_{2.}$

	x_i	n_{i1}	n_{i2}	n_{i3}	...	n_{ij}	...	n_{is}	$n_{i.}$

	x_r	n_{r1}	n_{r2}	n_{r3}	...	n_{rj}	...	n_{rs}	$n_{r.}$
		$n_{.1}$	$n_{.2}$	$n_{.3}$...	$n_{.j}$...	$n_{.s}$	n

7 – Analisi statistica bivariata

n_{ij} frequenze congiunte

n_{ij} / N frequenze congiunte relative

$100 \times n_{ij} / N$ frequenze congiunte %

Classe	Sopravvissuti		Totale
	No	Si	
I	5.5	9.2	14.8
II	7.6	5.4	12.9
III	24.0	8.1	32.1
Personale	30.6	9.6	40.2
Totale	67.7	32.3	100.0

il 9.2% stavano
nella I classe e sono
sopravvissuti

8 – Analisi statistica bivariata

n_{ij} frequenze congiunte

n_{ij} / n_i frequenze relative condizionate di riga

$100 \times n_{ij} / n_i$ frequenze % condizionate di riga

Classe	Sopravvissuti		Totale
	No	Si	
I	37.5	62.5	100.0
II	58.6	41.4	100.0
III	74.8	25.2	100.0
Personale	76.0	24.0	100.0
Totale	67.7	32.3	100.0

il 62.5% **di coloro che stavano nella I classe** (condizione) sono sopravvissuti

9 – Analisi statistica bivariata

n_{ij} frequenze congiunte

$n_{ij} / n_{.j}$ frequenze relative condizionate di colonna

$100 \times n_{ij} / n_{.j}$ frequenze % condizionate di colonna

Classe	Sopravvissuti		Totale
	No	Si	
I	8.2	28.6	14.8
II	11.2	16.6	12.9
III	35.4	25.0	32.1
Personale	45.2	29.8	40.2
Totale	100.0	100.0	100.0

il 28,6% **di coloro
che sono
sopravvissuti**
(condizione) stavano
nella I classe

11 –Analisi statistica bivariata -- esempi

$X = \text{tipo di coltura}; Y = \text{residui di pesticidi}$

X\Y	presenti	assenti	tot
biologico	29	98	127
convenzionale	19485	7086	26571
tot	19514	7196	26698

Quale frequenza è corretto interpretare per capire se i prodotti biologici contengono meno pesticidi?

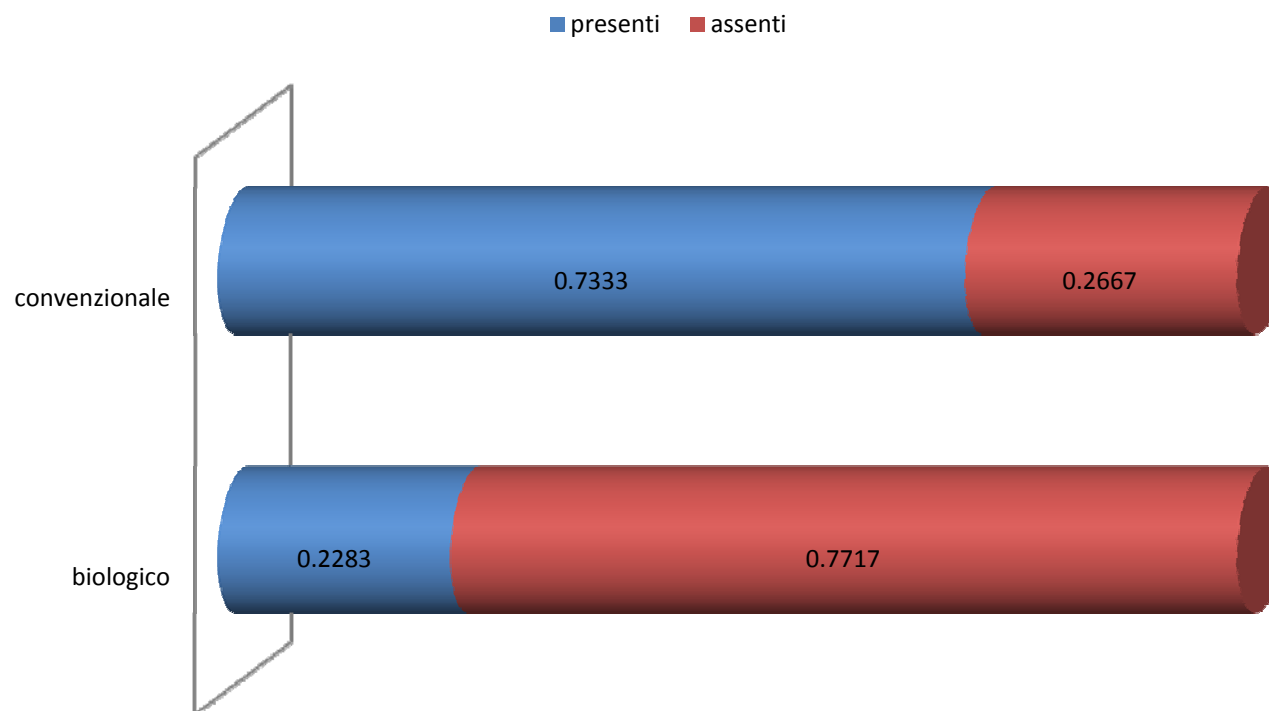
X\Y	presenti	assenti	tot
biologico	0.0011	0.0037	
convenzionale	0.7298	0.2654	
tot			1

X\Y	presenti	assenti	tot
biologico	0.2283	0.7717	1
convenzionale	0.7333	0.2667	1
tot			

12 – Analisi statistica bivariata

graficamente

presenza di pesticidi in prodotti alimentari



13 – Analisi statistica bivariata - esempi

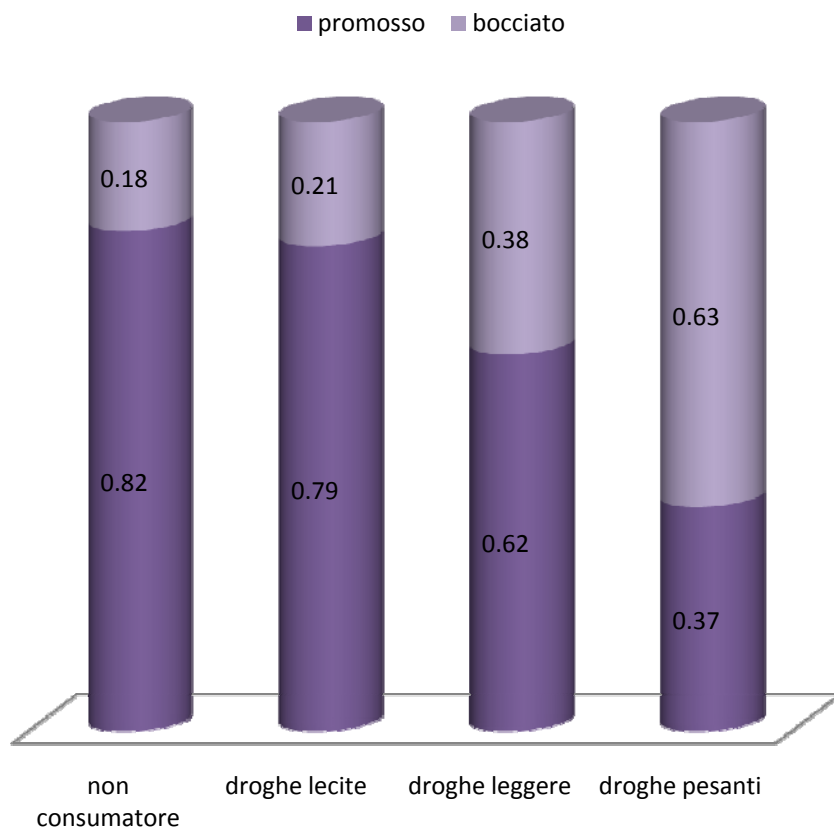
X= carriera scolastica; Y = consumo di droghe

	non consum	droghe lecite	droghe leggere	droghe pesanti			non consum	droghe lecite	droghe leggere	droghe pesanti	
promosso	50	186	34	11	281	promosso	0.132	0.489	0.089	0.029	0.739
bocciato	11	48	21	19	99	bocciato	0.029	0.126	0.055	0.050	0.261
	61	234	55	30	380		0.161	0.616	0.145	0.079	1.000

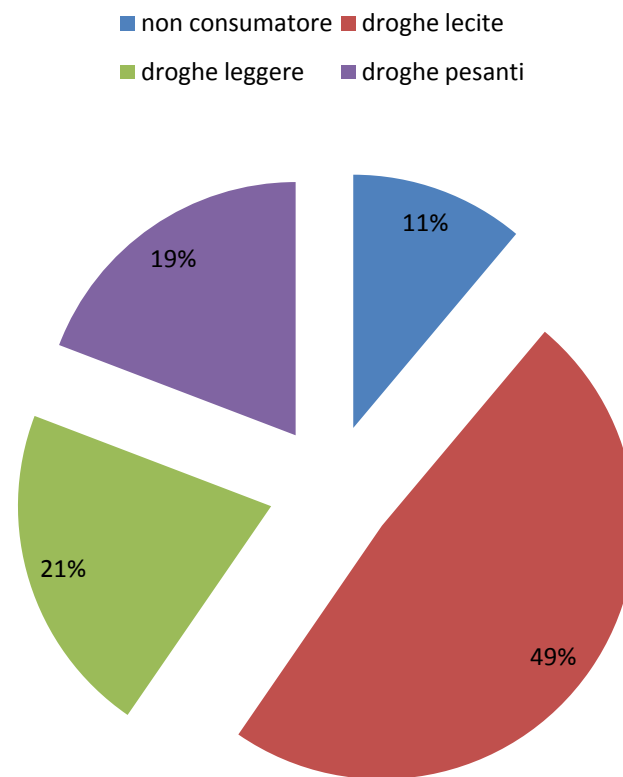
	non consum	droghe lecite	droghe leggere	droghe pesanti			non consum	droghe lecite	droghe leggere	droghe pesanti
promosso	0.18	0.66	0.12	0.04	1.00	promosso	0.82	0.79	0.62	0.37
bocciato	0.11	0.49	0.21	0.19	1.00	bocciato	0.18	0.21	0.38	0.63
							1.00	1.00	1.00	1.00

14 – Analisi statistica bivariata - esempi

consumo di droga e carriera scolastica



bocciato



15 – Analisi statistica bivariata - esempi

livello di alcool nel sangue del
conducente

		basso	medio	alto		
conseguenze incidente	gravi	2	52	158	212	Freq. Assolute
	non gravi	115	48	30	193	
		117	100	188	405	

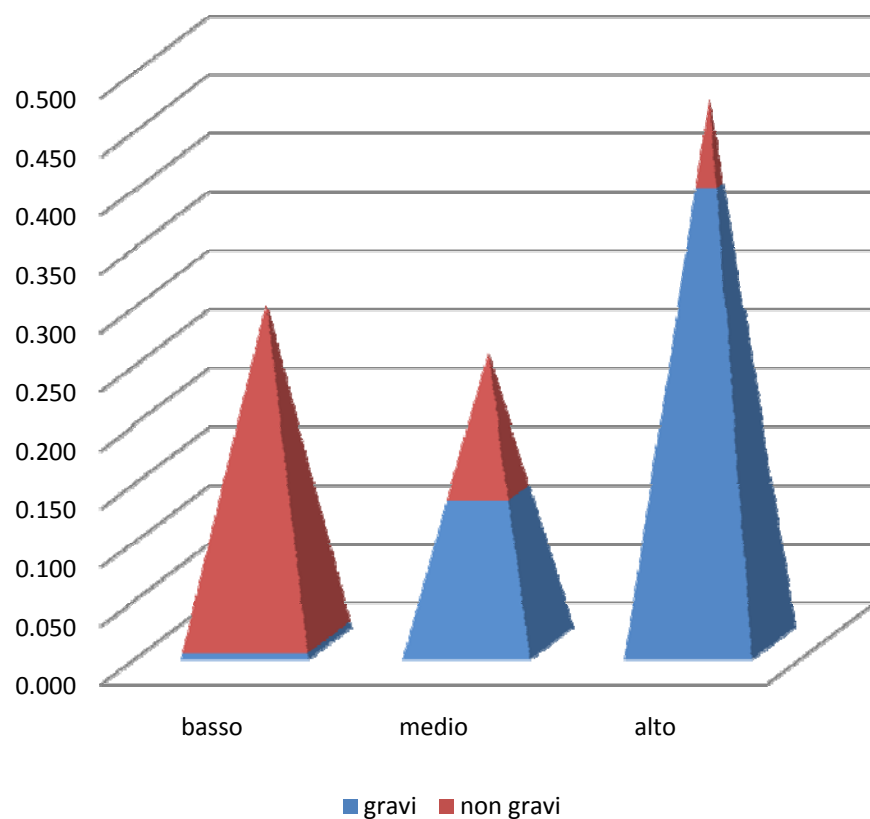
		basso	medio	alto		
conseguenze incidente	gravi	0.005	0.128	0.390	0.523	Freq. Relative
	non gravi	0.284	0.119	0.074	0.477	
		0.289	0.247	0.464	1	

		basso	medio	alto		
conseguenze incidente	gravi	0.009	0.245	0.745	1	Distr. Condizionate Y X
	non gravi	0.596	0.249	0.155	1	

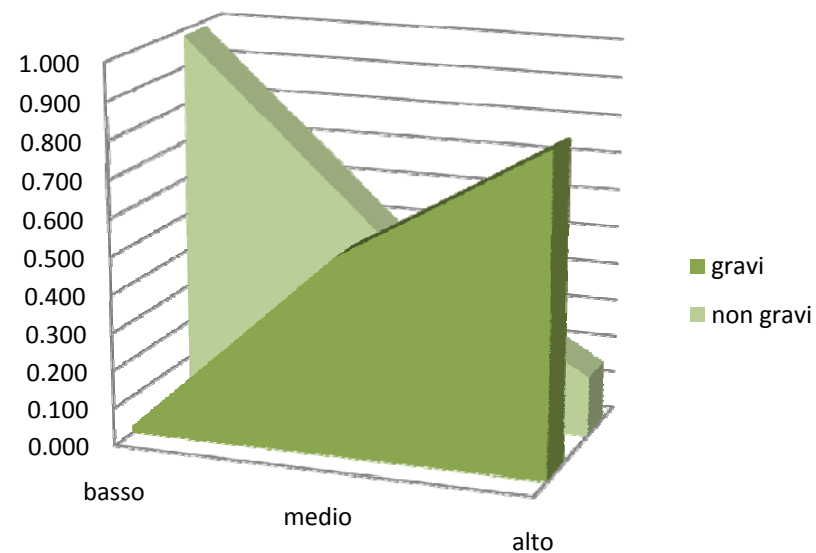
		basso	medio	alto		
conseguenze incidente	gravi	0.017	0.520	0.840		Distr. Condizionate X Y
	non gravi	0.983	0.480	0.160		
		1	1	1		

16 – Analisi statistica bivariata - esempi

livello di alcool nel sangue e incidenti



conseguenze incidenti per tasso alcolemico



promozioni

operatori

	0	1-2	3	>3	
vodafone	120	40	260	12	432
tim	85	316	226	456	1083
wind	7	40	10	28	85
	212	396	496	496	1600

Freq.
Assolute

	0	1-2	3	>3	
vodafone	0.075	0.025	0.163	0.008	0.270
tim	0.053	0.198	0.141	0.285	0.677
wind	0.004	0.025	0.006	0.018	0.053
	0.133	0.248	0.310	0.310	1.000

Freq. Relative

	0	1-2	3	>3	
vodafone	0.28	0.09	0.60	0.03	1.00
tim	0.08	0.29	0.21	0.42	1.00
wind	0.08	0.47	0.12	0.33	1.00

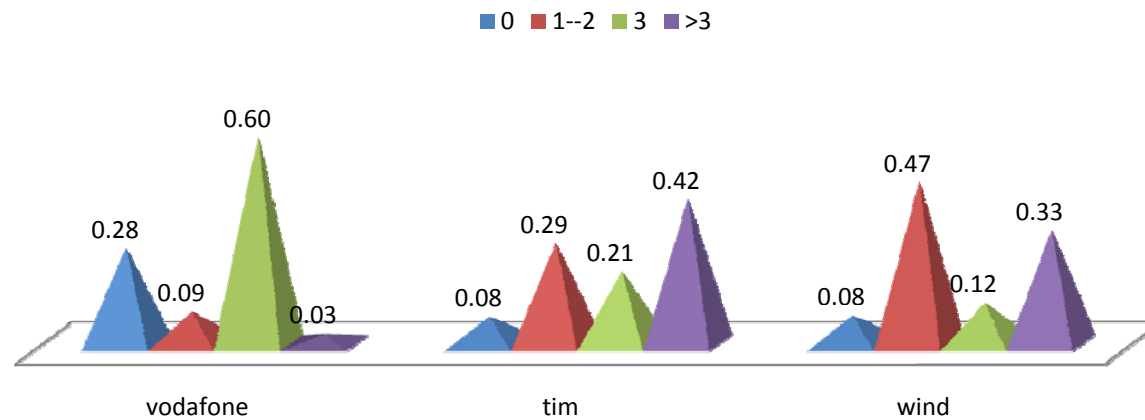
Distr. Condizionate Y|X

	0	1-2	3	>3	
vodafone	0.57	0.10	0.52	0.02	
tim	0.40	0.80	0.46	0.92	
wind	0.03	0.10	0.02	0.06	
	1.00	1.00	1.00	1.00	

Distr. Condizionate X|Y

18 – Analisi statistica bivariata

numero promozioni tra operatori diversi



operatori di telefonia date le promozioni offerte

