

# Metodi Quantitativi per Economia, Finanza e Management

*Esempi di Analisi Univariata*



# GRUPPO QUATTRO RUOTE



Alessandro Tondo 19632  
Laura Lavazza 19758  
Matteo Scordo 19813  
Alessandro Giosa 19894

## 2.4. CAMPIONE INTERVISTATO

Il campione intervistato è rappresentato dagli abitanti delle province di Milano e di Varese di età compresa tra i 25 e i 64 anni.

### CONFRONTO ETA' POPOLAZIONE E CAMPIONE

Dopo aver eseguito la **proc freq**, utilizzando come variabile l'**età del campione**, abbiamo aggregato i valori ottenuti in base alle diverse fasce d'età che desideriamo analizzare (25-34 anni, 35-44 anni, 45-54 anni e 55-64 anni). Il nostro campione non rappresenta in modo fedele la popolazione di riferimento, in quanto il numero delle osservazioni relative alle diverse fasce d'età differisce dai dati Istat relativi alla popolazione al 1/1/2014, elaborati dal sito di Tuttitalia\*. In particolare si evidenzia un eccesso di osservazioni nella fascia di età tra i 25 e i 34 anni a discapito delle fasce successive, in particolare quelle tra i 45-54 e 55-64 anni (vedi slide successive). Questo scostamento è causato dall'aver sottoposto un numero eccessivo di questionari a soggetti appartenenti alla nostra personale cerchia di conoscenti. Nella nostra analisi terremo in considerazione questo aspetto, dato che i risultati saranno in parte alterati rispetto alla realtà.

### CONFRONTO PROVINCIA DI RESIDENZA POPOLAZIONE E CAMPIONE

Eseguendo una **proc freq** e utilizzando come variabile la **provincia di residenza del campione**, si evince che la distribuzione del campione tra le province rispecchia la distribuzione della popolazione (vedi slide successive). I dati della popolazione si basano sui dati Istat al primo gennaio 2014.

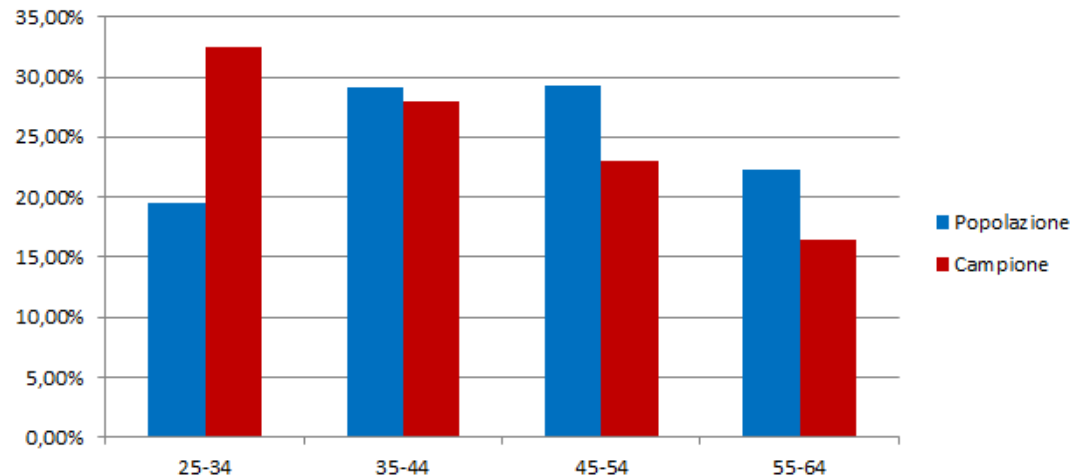
\*Fonti dati sulla popolazione: (<http://www.tuttitalia.it/lombardia/provincia-di-varese/statistiche/popolazione-eta-sesso-stato-civile-2014/> e <http://www.tuttitalia.it/lombardia/18-milano/statistiche/popolazione-eta-sesso-stato-civile-2014/>).

## 2.4. CAMPIONE INTERVISTATO (continua)

Confrontando la distribuzione per età delle osservazioni relative al campione, analizzate attraverso una proc freq e in seguito aggregate per fasce d'età, è possibile riscontrare uno scostamento rispetto alla popolazione totale delle due province considerate.

	Popolazione	% Popolazione	Campione	% Campione
<b>25-34</b>	434.428	19,45%	65	32,50%
<b>35-44</b>	649.729	29,09%	56	28,00%
<b>45-54</b>	652.647	29,22%	46	23,00%
<b>55-64</b>	496.978	22,25%	33	16,50%
	2.233.782		200	

**Tabella 4: confronto età popolazione e campione**

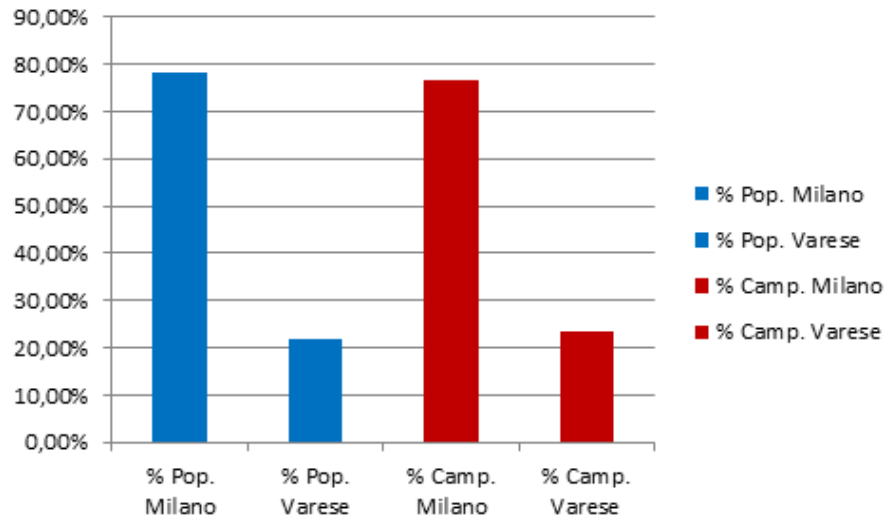


**Grafico 1: confronto età popolazione e campione**

## 2.4. CAMPIONE INTERVISTATO (continua)

Pop. Milano	% Pop. Milano	Pop. Varese	% Pop. Varese	Camp. Milano	% Camp. Milano	Camp. Varese	% Camp. Varese
1.746.350	78,18%	487.432	21,82%	153	76,50%	47	23,50%

**Tabella 5: confronto tra provincia di residenza della popolazione e del campione (non stratificato)**



**Grafico 2: confronto tra provincia di residenza della popolazione e del campione (non stratificato)**

La procedura FREQ

prov				
prov	Frequenza	Percentuale	Frequenza cumulativa	Percentuale cumulativa
Milano	153	76.50	153	76.50
Varese	47	23.50	200	100.00

Confrontando la popolazione delle province di Varese e Milano, con i dati relativi alla distribuzione percentuale del campione tra le due province (ottenuti attraverso la proc freq sopra indicata), risulta che il campione rispetta la distribuzione percentuale della popolazione. (78,18%-21,82% e 76,50%-23,50%)

I dati si riferiscono al campione non stratificato e alla popolazione totale delle province considerate.

## 2.4. CAMPIONE INTERVISTATO (continua)

	Pop. Milano	% Pop. Milano	Pop. Varese	% Pop. Varese	Camp. Milano	% Camp. Milano	Camp. Varese	% Camp. Varese
<b>25-34</b>	336.783	19,28%	97.645	20,03%	53	34,64%	12	25,53%
<b>35-44</b>	510.626	29,24%	139.103	28,54%	43	28,10%	13	27,66%
<b>45-54</b>	512.198	29,33%	140.449	28,81%	32	20,92%	14	29,79%
<b>55-64</b>	386.743	22,15%	110.235	22,62%	25	16,34%	8	17,02%
<b>Tot.</b>	<b>1.746.350</b>	<b>78,18%</b>	<b>487.432</b>	<b>21,82%</b>	<b>153</b>	<b>76,50%</b>	<b>47</b>	<b>23,50%</b>

Tabella 6: confronto tra provincia di residenza/età della popolazione e campione (stratificato)

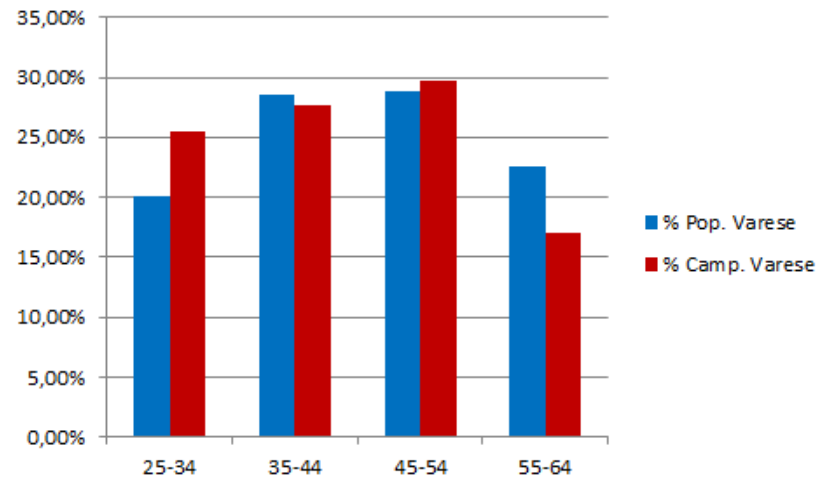
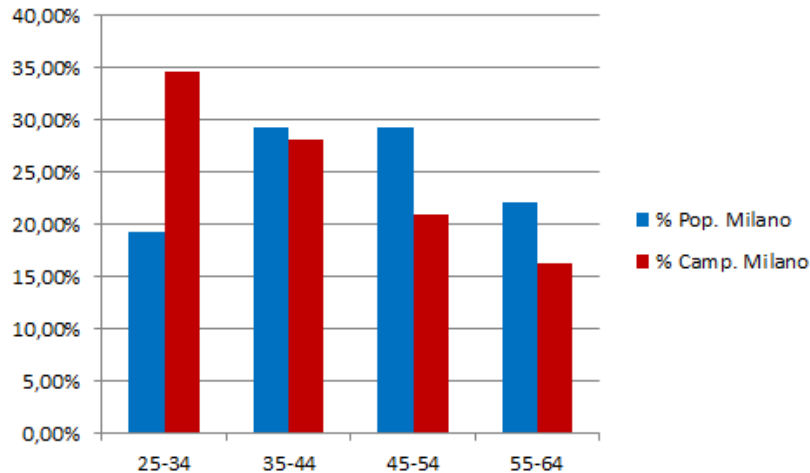


Grafico 3: confronto tra provincia di residenza/età della popolazione e campione (stratificato)

Confrontando il campione stratificato per fasce d'età e provincia di residenza, rispetto alla popolazione, è evidente come lo scostamento precedentemente rilevato sulle fasce d'età del campione rispetto alla popolazione target, si ripresenta in questa analisi.

Il campione è stato analizzato attraverso una proc freq, usando come variabile di classificazione la provincia di residenza. I dati sulla popolazione sono stati aggregati in fasce d'età a partire dai dati Istat.

## 2.6. OBIETTIVI SPECIFICI

Nel determinare le autovetture da inserire nel nostro autosalone, svolgeremo le seguenti analisi:

- **Analisi univariata:** attraverso questa analisi cercheremo di comprendere le distribuzioni di frequenza e le misure di sintesi relative ad alcune variabili come la willingness to pay e la preferenza verso una marca, per spiegare questi fenomeni.
- **Analisi bivariata:** questa analisi è essenziale per comprendere l'entità delle caratteristiche fondamentali in relazione alle variabili chiave come la marca e la willingness to pay.
- **Regressione lineare:** l'obiettivo è comprendere quali relazioni sussistono tra la willingness to pay (variabile dipendente) e le diverse variabili rilevanti per il business.
- **Regressione logistica:** l'obiettivo è quello di capire quali sono gli elementi fondamentali (variabili indipendenti) che influenzano la preferenza verso il brand che sceglieremo.

# 3. PARTE SECONDA

RISULTATI





# 3.1. RISULTATI

## 1. ANALISI UNIVARIATA

Abbiamo iniziato la nostra analisi effettuando un’analisi univariata utilizzando le procedure **proc freq** e **proc univariate**. Le variabili usate sono state: l’età, il sesso, la provincia di residenza (già utilizzate per l’analisi del campione), la marca e la spesa massima (come analisi preliminare).

### 1.1 Marca preferita

Per quanto riguarda la marca abbiamo effettuato una **proc freq** e abbiamo rilevato che i brand più richiesti sono la Fiat (frequenza assoluta pari a 60) e la Volkswagen (frequenza assoluta pari a 59). Gli altri marchi hanno una frequenza assoluta meno significativa, quindi dopo questa prima analisi abbiamo deciso di tenere in considerazione solo la Fiat e la Volkswagen come possibili automobili da possedere nel nostro autosalone. I dati raccolti relativi alle province di Milano e di Varese, non rispecchiano perfettamente i dati relativi al mercato italiano delle automobili, infatti non vi è una netta prevalenza delle Fiat rispetto alle Volkswagen, pur sempre rimanendo i due marchi più ricercati. Le vetture più vendute in Italia nel 2014 sono state la Fiat (281.582) e la Volkswagen (110.227). Questi sono dati di fondamentale importanza da tenere in considerazione nelle nostre decisioni.

La procedura FREQ

marca_fav				
marca_fav	Frequenza	Percentuale	Frequenza cumulativa	Percentuale cumulativa
Citroen	10	5.00	10	5.00
<b>Fiat</b>	<b>60</b>	<b>30.00</b>	70	35.00
Ford	10	5.00	80	40.00
Hyundai	1	0.50	81	40.50
Mini	21	10.50	102	51.00
Peugeot	5	2.50	107	53.50
Renault	9	4.50	116	58.00
Smart	7	3.50	123	61.50
Suzuki	1	0.50	124	62.00
Toyota	17	8.50	141	70.50
<b>Volkswagen</b>	<b>59</b>	<b>29.50</b>	200	100.00

# 3.1. RISULTATI

## 1.2 willingness to pay (spesa massima)

Per quanto riguarda la **spesa massima** abbiamo effettuato una **proc univariate** per iniziare a capire quali sono effettivamente le richieste dei consumatori. Essendo noi un rivenditore di city car, vorremmo possedere nel nostro autosalone solamente vetture che siano in linea con le disponibilità economiche dei nostri clienti. Eseguendo questa procedura abbiamo verificato che la willingness to pay **media** è 12.277,50€, quindi i potenziali clienti sono mediamente disposti a spendere questa cifra. La **mediana** e la **moda** sono pari a 10.000€. Analizzando ulteriori valori possiamo verificare che la deviazione standard (scarto quadratico medio) è di 6.935€. Questo dato ci fa capire quanto i valori sono distanti dal valore centrale. Il **range** in cui varia la spesa massima va da 500€ a 60.000 € . L'indice di **Skewness** > 0, pari a 2,946, indica che esiste asimmetria positiva (la media 12.277,5 è maggiore della mediana 10.000). L'indice di **Kurtosis** > 0, pari a 14,851, indica che la distribuzione è ipernormale quindi la curva di frequenza ha una coda più lunga a destra, come dimostrano i valori estremi superiori 60.000€ e 50.000€.

Poiché la distribuzione di frequenza è caratterizzata da asimmetria positiva con valori estremi a destra elevati, possiamo osservare come la maggior parte delle osservazioni si attestino attorno al valore 12.277,5€ o appena inferiore, mentre vi sono valori molto alti (60.000€ e 50.000€) che spostano la coda della distribuzione verso destra.

La procedura UNIVARIATE  
Variabile: spesa\_max (spesa\_max)

Momenti			
N	200	Somma dei pesi	200
Media	12277.5	Somma delle osservazioni	2455500
Deviazione std	6934.67319	Varianza	48089692.2
Skewness	2.94585315	Curtosi	14.8509582
SS non corretta	3.97173E10	SS corretta	9569848750
Coeff variaz	56.482779	Errore std media	490.355444

Misure statistiche di base			
Posizione		Variabilità	
Media	12277.50	Deviazione std	6935
Mediana	10000.00	Varianza	48089692
Moda	10000.00	Range	59500
		Range interquartile	8000

Test di posizione: Mu0=0			
Test	Statistica	P-value	
T di Student	t 25.03796	Pr >  t	<.0001
Segno	M 100	Pr >=  M	<.0001
Rango con segno	S 10050	Pr >=  S	<.0001

Osservazioni estreme			
Inferiori		Superiori	
Valore	Oss	Valore	Oss
500	112	30000	85
1000	2	30000	121
3000	181	30000	149
3000	25	50000	114
3000	23	60000	89

Distribuzione e grafico delle probabilità per spesa\_max

