

ECOCAPSULA

**METODI QUANTITATIVI PER ECONOMIA, FINANZA E MANAGEMENT
2017/2018**



0023057 Bacchi Francesca
0023059 Criscenti Salvatore
0023120 Giambrone Alberto

INDICE

► INTRODUZIONE

- Profilo aziendale
- Obiettivo della ricerca
- Descrizione del mercato
- Analisi competitiva
- Definizione della popolazione
- Disegno del campione
- Controllo della rappresentatività del campione

► DEFINIZIONE QUESTIONARIO

- Canale di somministrazione e Data audit
- Struttura del questionario
- Variabili utilizzate

► ANALISI UNIVARIATA

- Obiettivi dell'analisi univariata
- 1° ANALISI UNIVARIATA: Caratteristiche del campione target
- 2° ANALISI UNIVARIATA: Conoscenza delle capsule ricaricabili
- 3° ANALISI UNIVARIATA: Possesso delle capsule ricaricabili
- 4° ANALISI UNIVARIATA: Spesa mensile
- 5° ANALISI UNIVARIATA: Willingness to pay
- 6° ANALISI UNIVARIATA: Disponibilità ad acquistare una capsula Ecocapsula
- 7° ANALISI UNIVARIATA: Conoscenza delle marche concorrenti
- 8° ANALISI UNIVARIATA: Sostenibilità ambientale
- Conclusioni

INDICE

► ANALISI BIVARIATA

- Obiettivi dell'analisi bivariata
- Analisi di connessione (1)
- Analisi di connessione (2)
- Analisi di correlazione (1)
- Analisi di correlazione (2)
- Analisi di correlazione (3)
- Anova (1)
- Anova (2)
- Anova (3)
- Conclusioni

► ANALISI FATTORIALE

- Obiettivi dell'analisi fattoriale
- Scelta delle variabili
- Componenti principali
 - Scelta del numero delle componenti
 - Scree plot
 - Comunalità
 - Interpretabilità delle componenti
- Soluzioni
- Conclusioni

INDICE

► **REGRESSIONE LINEARE**

- Obiettivi della regressione lineare
- Valutazione del modello
- Variabili dummy
 - Valutazione del nuovo modello
 - Metodo Stepwise
 - Calcolo della Vif
- Interpretazione dei coefficienti
- Conclusioni

► **REGRESSIONE LOGISTICA**

- Obiettivi della regressione logistica
- Scelta delle variabili
- Disponibilità all'acquisto di Ecocapsula

- Valutazione della bontà del modello
 - Wald test
 - Percentuale di Concordant
 - Wald Chi Squared test
- Stima del nuovo modello con variabili impattanti
 - Wald test
 - Percentuale di Concordant
 - Wald Chi Squared test
 - Calcolo della Vif
- Interpretazione dei coefficienti
- Conclusioni

► **CONCLUSIONI BUSINESS CASE**

INTRODUZIONE

Profilo aziendale



Centocaffè S.r.l.

È un'azienda italiana che vuole misurarsi nei confronti di un mercato sempre più esigente e alla ricerca di novità che apportino un effettivo aumento qualitativo della vita.

Le caratteristiche peculiari proprie di Centocaffè S.r.l. sono:

- ❖ sviluppo;
- ❖ innovazione;
- ❖ coerenza verso gli alti standard qualitativi tipici di una azienda Italiana.

La forte esperienza maturata in campo commerciale e industriale della Centocaffè S.r.l. si avvale inoltre della collaborazione in sinergia di realtà industriali leader a livello nazionale, in un'ottica di espansione rapida ma controllata, avente come comune denominatore il binomio Efficienza e Qualità 100% Italiana.

Da questo pensiero nasce Ecocapsula, una linea di capsule/cialde ricaricabili e riutilizzabili per macchine da caffè domestiche, disponibile per tutti i sistemi di macchine. Ecocapsula è un prodotto brevettato, ecologico e 100% Made in Italy.

Profilo aziendale



Ecocapsula regala tutta la qualità dei prodotti più genuini in pochi semplici gesti.



Essendo Ecocapsula composta da materiali riciclabili, terminato il suo ciclo di 300 erogazioni, si può smaltire negli appositi contenitori per la plastica.



La linea di prodotti Ecocapsula permette un risparmio fino all'80% rispetto alle capsule originali monouso.



Ecocapsula è un brevetto 100% Made in Italy, dalla progettazione alla realizzazione, utilizzando risorse lavorative del territorio e materiali idonei e certificati per l'uso, garanzia di un prodotto sicuro e di Alta Qualità Italiana.

Obiettivo della ricerca

Nella divisione marketing di Centocaffè S.r.l. è stato formato un team di ricerca con il compito di fornire all'azienda la consulenza necessaria per la vendita del prodotto Ecocapsula.

Questa capsula ricaricabile e riutilizzabile è una tipologia di capsula ecologica in acciaio. Il prodotto, grazie al materiale usato, è riutilizzabile e dura moltissimo tempo (fino a 300 utilizzi), risolvendo quindi il problema dei rifiuti.



In particolare, l'intento dell'analisi è quello di comprendere le opinioni e le abitudini comportamentali dei consumatori per quanto riguarda il possesso delle macchine del caffè in capsule e il relativo consumo di quest'ultime.

Tramite la survey si misurerà la propensione all'acquisto di Ecocapsula. Il fine ultimo della ricerca è l'individuazione della provincia lombarda che riscontra maggiore interesse, e la successiva attivazione di un nuovo punto vendita aziendale.

Obiettivo della ricerca

Di conseguenza, l'attenzione del team si focalizzerà su una serie di sotto-obiettivi che risultano essere complementari e propedeutici all'obiettivo principale.

Quindi, dalla nostra ricerca di mercato ci aspettiamo di conseguire nel dettaglio i seguenti sotto-obiettivi:

- analizzare la frequenza e le modalità di consumo di capsule tradizionali;
- esaminare la spesa mensile degli intervistati;
- comprendere quali siano le caratteristiche a cui i soggetti target prestano attenzione durante l'acquisto;
- capire la propensione al consumo di capsule ricaricabili;
- percepire la sensibilità che gli individui nutrono verso la sostenibilità ambientale;
- cercare di prevedere la futura predisposizione all'acquisto di capsule ricaricabili in caso di apertura del nuovo punto vendita.



Descrizione del mercato

Il mercato italiano delle cialde e capsule di caffè nel 2016 ha espresso una disponibilità produttiva (produzione interna e import) a valore di 1,094 milioni di euro.

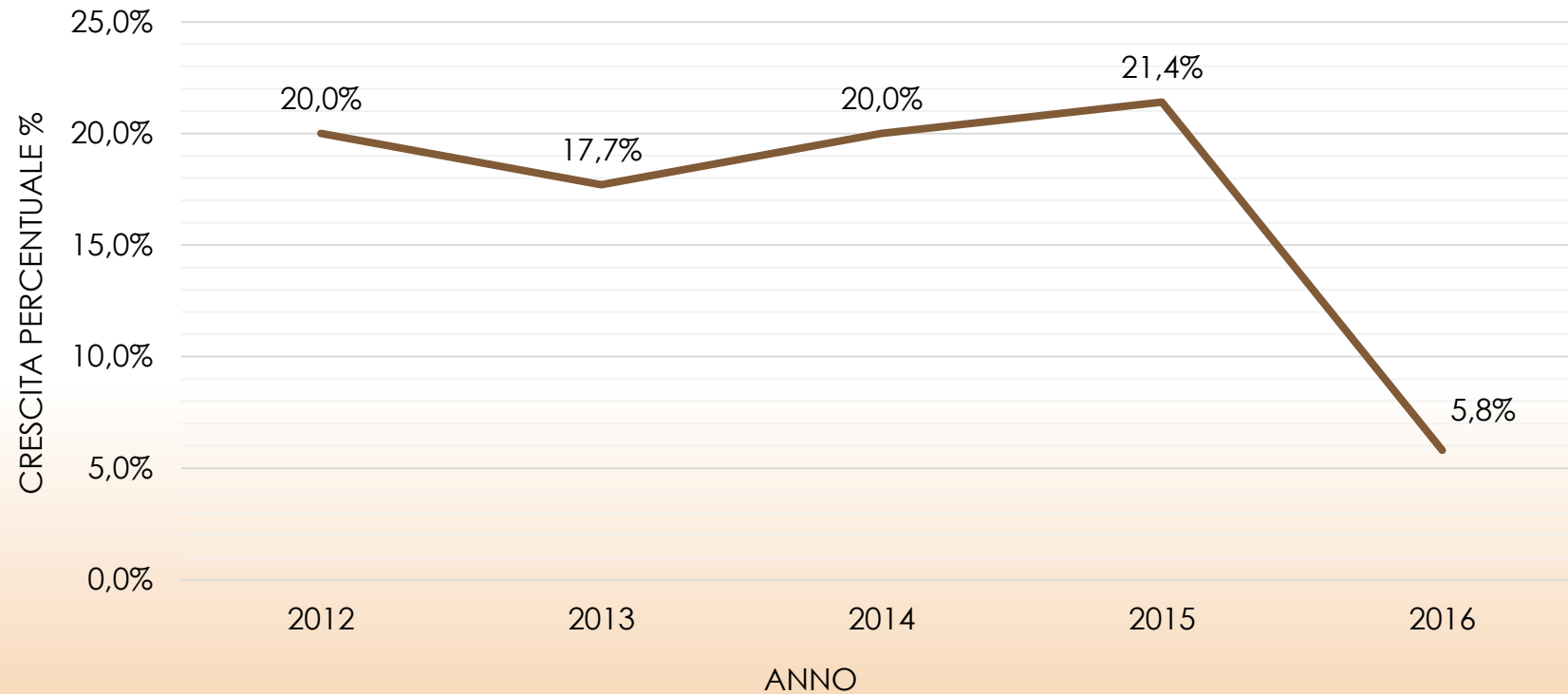
L'export a valore è stato di 306 milioni di euro e la domanda interna è stata pari a 788 milioni di euro.

La crescita del mercato interno a valore è stata del **5,8%**, ma con andamento differenziato nei vari canali:

- + 4,6% nel canale Famiglia;
- - 4,5% nel canale OCS (Office coffee service);
- + 23,8% nel canale Ho.Re.Ca. (canale alberghiero);
- + 53,4% nel canale Internet.



Descrizione del mercato



Nel grafico viene illustrato l'andamento della crescita percentuale del mercato italiano delle capsule. Si nota un continuo aumento, anche se ridotto nell'ultimo anno (2016).

Fonti: Nielsen, Competitive Data, Host.FieraMilano

Descrizione del mercato

Se tutto sommato la differenza dal 2015 al 2016 nell'area d'affari famiglie (Grande Distribuzione, door to door, e-commerce proprietari, ecc.) è nell'ordine del punto percentuale perso dai sistemi proprietari, è invece nell'area d'affari business (OCS tradizionale, e-commerce indipendenti, Ho.Re.Ca.) che assistiamo a flessioni più sensibili, con la quota aggregata a volume dei principali player con sistemi proprietari che passa dal 46,1% del 2015 al 43,8% del 2016.

Questa flessione è legata a doppio filo al calo del 4,5% fatto registrare dal canale OCS tradizionale, e dalla crescita del 52,4% fatta registrare dai rivenditori e-commerce, mentre anche nel canale Ho.Re.Ca., servito con cialde e capsule (la crescita nel 2016 è stata del 23,8% a volume), cominciano a farsi strada le capsule compatibili e quelle ricaricabili.

Le capsule speciali sono infatti il segmento che cresce di più nel 2016 (+11,3%), a fronte di una flessione delle capsule in plastica e delle cialde in carta. Aziende nate invece per sviluppare sistemi proprietari stanno rivedendo la strategia complessiva per sfruttare anche quest'opportunità di business.

Descrizione del mercato

In commercio esistono diverse tipologie di capsule, molte delle quali compatibili, ovvero più economiche rispetto a quelle originali, e possono essere utilizzate sulle macchine dei marchi più famosi.

Oltre al risparmio poi, il caffè prodotto è generalmente di buon livello, sia per via della qualità delle miscele utilizzate che per il modo in cui sono costruite.



Purtroppo il consumo esponenziale di capsule per il caffè, sia per uso casalingo che in ufficio, porta con sé il problema dell'**aumento dei rifiuti** e lo smaltimento delle capsule che sono sia organiche (i residui del caffè) che in plastica o alluminio e quindi non riciclabili e finiscono nel bidone dell'indifferenziato.

Considerando poi, che l'Italia è il sesto paese al mondo per consumo di caffè, questo significa che ogni anno vengono buttate tonnellate di capsule usa e getta.

Descrizione del mercato

In merito a questa situazione, numerose aziende, sensibili al problema ambientale, hanno studiato e realizzato una soluzione alternativa: le **capsule ricaricabili**, che sono identiche a quelle monouso, ma possono essere sciacquate dopo l'utilizzo e riutilizzate per un nuovo caffè.

La durata delle capsule ricaricabili è di circa 250-300 utilizzi.

Sono state brevettate anche delle capsule in acciaio che potrebbero essere usate anche per più tempo.



A fronte di un piccolo sforzo legato al fatto di dover pulire la capsula, quest'idea innovativa porta con se molteplici vantaggi:

- risparmio sui costi di circa 80-85% rispetto alle capsule monouso;
- la possibilità di scegliere la miscela che più si desidera;
- caffè sempre di gran qualità;
- un aiuto concreto all'ambiente con una forte riduzione dei rifiuti.

Analisi competitiva

I **principali competitors** di Ecocapsula che producono capsule ricaricabili e riutilizzabili sono:



Waycap



Emocup



MyCoffeeStar



Quick caffè

Definizione della popolazione

La popolazione target di riferimento della nostra survey è composta da:

- ❖ Residenti nella regione Lombardia
- ❖ Sesso: sia femmine che maschi
- ❖ Età: compresa tra i 18 e 60 anni

Questo target analizzato descrive una popolazione molto ampia e diversificata di consumatori, proprio perché il caffè è una bevanda notoriamente consumata da persone di ogni età.

Attraverso il questionario, con domande mirate, si è voluta fornire una rappresentazione maggiormente fedele alla realtà di riferimento, motivo per cui la popolazione è stata suddivisa in categorie e sono state escluse la fascia inferiore ai 18 e quella superiore ai 60 perché, secondo pareri medici, a queste fasce è sconsigliata la somministrazione di caffè.

Definizione della popolazione

Esaminando la popolazione residente nella regione Lombardia, ne abbiamo rilevato la distribuzione per fasce d'età e sesso.
I dati sono stati raccolti esaminando la ricostruzione intercensuaria della popolazione all'1 gennaio 2017.

FASCE D'ETÀ	MASCHI	% MASCHI	FEMMINE	% FEMMINE	TOTALE POP.	%
18 – 26 anni	434.677	7,691%	405.232	7,171%	839.909	14,862%
27 – 35 anni	506.258	8,958%	495.718	8,771%	1.001.976	17,730%
36 – 44 anni	683.321	12,091%	664.593	11,759%	1.347.914	23,851%
45 – 53 anni	761.289	13,471%	747.271	13,223%	1.508.560	26,693%
54 – 60 anni	468.376	8,287%	484.710	8,576%	953.086	16,864%
TOTALE	2.853.921	50,499%	2.797.524	49,501%	5.651.445	100,000%

Fonte: Istat

Disegno del campione

Il campione in esame risulta essere composto da 200 elementi estratti casualmente dalla popolazione target, ovvero residente nelle province delle Lombardia, di età compresa tra i 18 e i 60 anni.

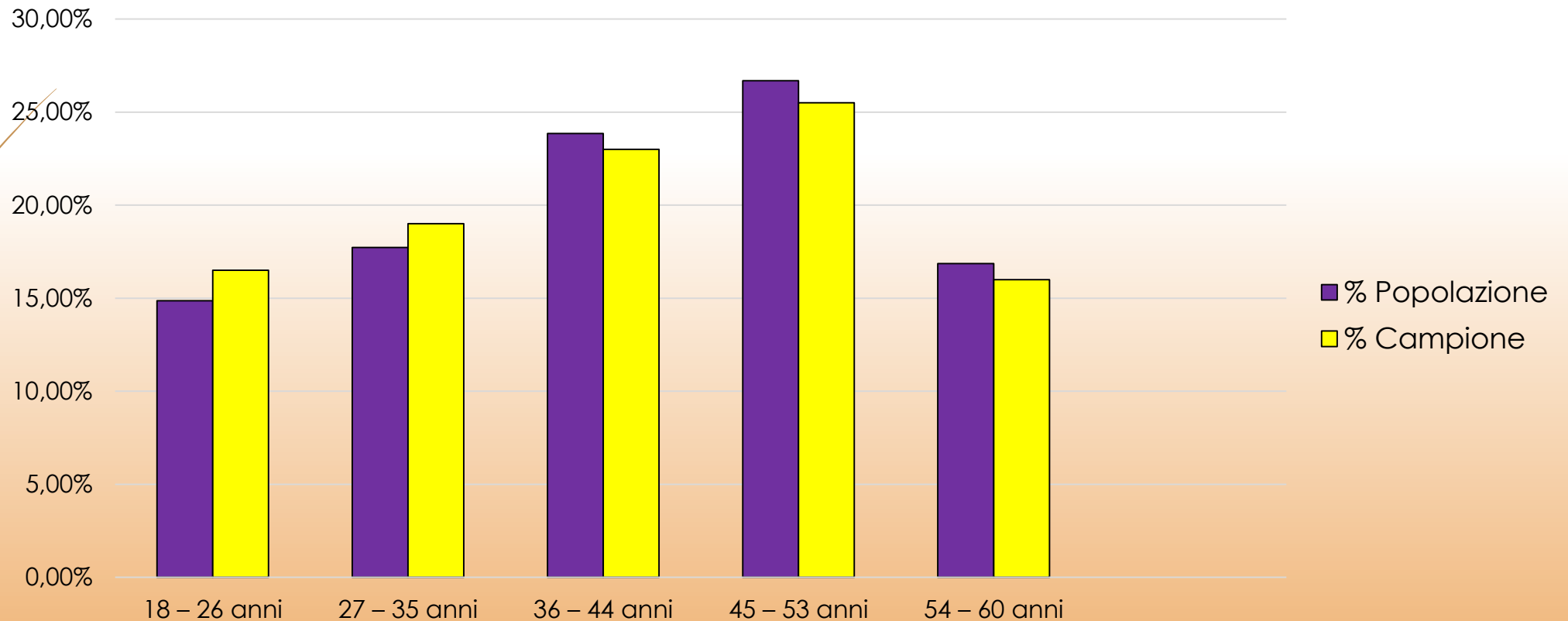
Adottando il metodo di campionamento probabilistico stratificato su due variabili di stratificazione, età e sesso, risulta in linea con la popolazione descritta precedentemente, come dimostra la tabella sottostante.

Ciò permette di effettuare inferenza statistica, estendendo le nostre osservazioni all'intera popolazione.

FASCE D'ETÀ	MASCHI	% MASCHI	FEMMINE	% FEMMINE	TOTALE POP	%
18 – 26 anni	17	8,5%	16	8%	33	16,500%
27 – 35 anni	19	9,5%	19	9,5%	38	19,000%
36 – 44 anni	24	12%	22	11%	46	23,000%
45 – 53 anni	23	11,5%	28	14%	51	25,500%
54 – 60 anni	14	7%	18	9%	32	16,000%
TOTALE	97	48,5%	103	51,5%	200	100,000%

Controllo della rappresentatività del campione

Come è possibile notare anche dal grafico sottostante, il campione individuato è rappresentativo della popolazione di riferimento.





DEFINIZIONE QUESTIONARIO

Canale di somministrazione e Data audit

Il questionario è stato somministrato a soggetti appartenenti al target di riferimento tramite due canali:

- ❖ **Online survey** su piattaforma Google, proposto e compilato solo dal campione preso in considerazione → 114 rispondenti;
- ❖ Questionario cartaceo **face to face**, consegnato personalmente dal team ai soggetti target → 86 rispondenti.



Per tutto il mese di Novembre sono state raccolte le informazioni necessarie per procedere al sondaggio. Sono stati intervistati un totale di 223 persone di differente età, sesso e provincia. Di questi 223 questionari:

- 10 sono stati scartati perché compilati senza seguire alcun criterio logico;
- 7 sono stati scartati perché errati;
- 6 sono stati scartati perché incompleti.

Quindi la valutazione si è concentrata su 200 questionari corretti. Dopo la raccolta dei dati, questi sono stati inseriti in un documento Excel e rielaborati attraverso il software R.

Struttura del questionario

Il questionario elaborato è composto da 32 domande, sia qualitative che quantitative, ed è diviso in 3 sezioni:

- 1) Dati comportamentali** (21 domande): effetti del caffè; fasce orarie in cui si beve il caffè; dove si beve il caffè; possesso della macchina del caffè per capsule; spesa media mensile di capsule; aspetti fondamentali per l'acquisto; soddisfazione della marca acquistata; willingness to pay per capsule ricaricabili e riutilizzabili. In questa sezione l'obiettivo è quello di individuare caratteristiche direttamente funzionali agli obiettivi della nostra ricerca;
- 2) Dati attitudinali** (5 domande): aggettivi che descrivono l'individuo; influenza dei social networks sulle scelte d'acquisto; negozi d'acquisto per elettrodomestici; propensione al rispetto dell'ambiente; significato di consumatore sostenibile. In questa sezione l'obiettivo è quello di identificare il profilo dell'intervistato, studiando i suoi interessi e il suo stile di vita;
- 3) Dati anagrafici/socio-demografici** (6 domande): sesso; età; titolo di studio; professione esercitata; provincia lombarda di residenza; numero di componenti del nucleo familiare. In questa sezione l'obiettivo è quello di inquadrare il rispondente in un contesto spazio-temporale per comprendere la composizione del campione.

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
Caffè	Bere il caffè	Dicotomica
Macchina_Caffè	Possesso della macchina del caffè	Dicotomica
Conos_Capsule	Conoscenza della capsule ricaricabili e riutilizzabili	Dicotomica
Poss_Capsule	Possesso delle capsule ricaricabili e riutilizzabili	Dicotomica
D_5_Attenzione	Di solito il caffè «Migliora l'attenzione»	Quantitativa
D_5_Energia	Di solito il caffè «Dona energia»	Quantitativa
D_5_NoEffetto	Di solito il caffè «Non provoca effetti»	Quantitativa
D_5_Agitazione	Di solito il caffè «Genera Agitazione»	Quantitativa
D_6_PrimaFascia	Di solito beve il caffè «Dalle 6.00 alle 10.00»	Qualitativa
D_6_SecondaFascia	Di solito beve il caffè «Dalle 10.00 alle 14.00»	Qualitativa
D_6_TerzaFascia	Di solito beve il caffè «Dalle 14.00 alle 18.00»	Qualitativa
D_6_QuartaFascia	Di solito beve il caffè «Dalle 18.00 alle 22.00»	Qualitativa
D_6_QuintaFascia	Di solito beve il caffè «Dalle 22.00 alle 02.00»	Qualitativa

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
D_6_SestaFascia	Di solito beve il caffè «Dalle 02.00 alle 06.00»	Qualitativa
D_7_Casa	Di solito beve il caffè a «Casa»	Quantitativa
D_7_Lavoro	Di solito beve il caffè al «Lavoro»	Quantitativa
D_7_ScuolaUniversità	Di solito beve il caffè a «Scuola/università»	Quantitativa
D_7_Bar	Di solito beve il caffè al «Bar»	Quantitativa
Marca_Macchina	Macchina del caffè che si possiede	Qualitativa
Tempo_Preparazione	Tempo dedicato alla preparazione del caffè in minuti	Quantitativa
Volte_Acquisto	Volte al mese in cui si acquistano capsule	Quantitativa
Spesa_Acquisto	Spesa media mensile delle capsule	Quantitativa
Num_Capsule	Numero di capsule che si consumano settimanalmente	Quantitativa
D_13_Prezzo	Importanza del «Prezzo» nell'acquisto delle capsule	Quantitativa
D_13_Qualità	Importanza della «Qualità» nell'acquisto delle capsule	Quantitativa
D_13_Gusto	Importanza del «Gusto» nell'acquisto delle capsule	Quantitativa

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
D_13_Sostenibilità	Importanza della «Sostenibilità» nell'acquisto delle capsule	Quantitativa
D_13_Brand	Importanza del «Brand» nell'acquisto delle capsule	Quantitativa
D_13_Provenienza	Importanza della «Provenienza Made in Italy» nell'acquisto delle capsule	Quantitativa
D_13_ProdottoBio	Importanza del «Prodotto Bio» nell'acquisto delle capsule	Quantitativa
D_13_Packaging	Importanza del «Packaging» nell'acquisto delle capsule	Quantitativa
D_13_RapportoQP	Importanza del «Rapporto qualità/prezzo» nell'acquisto delle capsule	Quantitativa
Soddisfazione_Marca	Soddisfazione della marca	Quantitativa
Disp_Acquisto	Disponibilità ad acquistare una capsula ricaricabile e riutilizzabile	Dicotomica
D_16_QualitàProdotto	Importanza di «Una maggiore qualità del prodotto» nella propensione all'acquisto	Quantitativa
D_16_SceltaCaffè	Importanza della «Possibilità di scelta del caffè da inserire nella capsula» nella propensione all'acquisto	Quantitativa

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
D_16_Convenienza	Importanza di «Una maggiore convenienza» nella propensione all'acquisto	Quantitativa
D_16_RiutilizzoCapsula	Importanza della «Possibilità di riutilizzare la capsula» nella propensione all'acquisto	Quantitativa
D_16_ImpattoAmbiente	Importanza di «Un impatto positivo sull'ambiente» nella propensione all'acquisto	Quantitativa
Wtp	Propensione all'acquisto di una capsula ricaricabile e riutilizzabile	Quantitativa
D_18_Waycap	Conoscenza della marca «Waycap»	Qualitativa
D_18_Mycoffeestar	Conoscenza della marca «Mycoffeestar»	Qualitativa
D_18_Emocup	Conoscenza della marca «Emocup»	Qualitativa
D_18_Quickcaffè	Conoscenza della marca «Quick caffè»	Qualitativa
D_18_Tutte	Conoscenza di «Tutte le precedenti» marche	Qualitativa
D_18_Nessuna	Conoscenza di «Nessuna» marca	Qualitativa

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
D_19_Prezzo	Importanza del «Prezzo» nell'acquisto di capsule ricaricabili e riutilizzabili	Quantitativa
D_19_Durevolezza	Importanza della «Durevolezza del prodotto» nell'acquisto di capsule ricaricabili e riutilizzabili	Quantitativa
D_19_SensAmbiente	Importanza della «Sensibilità ambientale» nell'acquisto di capsule ricaricabili e riutilizzabili	Quantitativa
D_19_QuantitàCaffè	Importanza della «Possibilità di inserire la quantità di caffè desiderata» nell'acquisto di capsule ricaricabili e riutilizzabili	Quantitativa
D_19_Ingombro	Importanza del «Minore spazio di ingombro» nell'acquisto di capsule ricaricabili e riutilizzabili	Quantitativa
Acquisto_EcoCap	Acquisto in passato della capsula ricaricabile e riutilizzabile Ecocapsula	Dicotomica
Wtp_EcoCap	Disponibilità all'acquisto della capsula ricaricabile e riutilizzabile Ecocapsula	Dicotomica
D_22_Energico	Descrizione dell'intervistato come «Energico»	Quantitativa
D_22_Attivo	Descrizione dell'intervistato come «Attivo»	Quantitativa

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
D_22_Sedentario	Descrizione dell'intervistato come «Sedentario»	Quantitativa
D_22_Stanco	Descrizione dell'intervistato come «Stanco»	Quantitativa
Influenza_Social	Influenza dei social sulle scelte d'acquisto	Quantitativa
D_24_GrandiSuper	Acquisto degli elettrodomestici nei «Grandi Supermercati»	Qualitativa
D_24_NegoziDettaglio	Acquisto degli elettrodomestici nei «Negozi al Dettaglio»	Qualitativa
D_24_Online	Acquisto degli elettrodomestici «Online»	Qualitativa
D_24_Tutte	Acquisto degli elettrodomestici in «Tutte» le opzioni	Qualitativa
D_24_Altro	Acquisto degli elettrodomestici in «Altro»	Qualitativa
Rispetto_Ambiente	Propensione dell'intervistato a rispettare l'ambiente	Quantitativa
D_26_BassoImpatto	Consumatore sostenibile significa «Comprare prodotti con un basso impatto ambientale»	Quantitativa
D_26_UsaeGetta	Consumatore sostenibile significa «Non acquistare prodotti usa e getta»	Quantitativa

Variabili utilizzate

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGIA
D_26_CompportDannosi	Consumatore sostenibile significa «Non comprare prodotti di aziende che hanno comportamenti dannosi in ambito sociale e ambientale»	Quantitativa
D_26_NoNecessari	Consumatore sostenibile significa «Limitare l'acquisto dei prodotti non strettamente necessari»	Quantitativa
Sesso	Sesso dell'intervistato	Qualitativa
Età	Età dell'intervistato	Quantitativa
TitoloStudio	Titolo di studio dell'intervistato	Qualitativa
Professione	Professione esercitata dell'intervistato	Qualitativa
Provincia	Provincia di residenza dell'intervistato	Qualitativa
NucleoFam	Numero componenti del nucleo familiare dell'intervistato	Quantitativa



ANALISI UNIVARIATA

Obiettivi dell'analisi univariata

L'analisi univariata, condotta attraverso il software R, ha come obiettivo una riflessione sulle variabili ritenute essenziali per l'individuazione del target di riferimento.

Sono state utilizzate le seguenti funzioni di R: *freq*, *frequencyBy*, *summary*, *basicStats*, *cv* e *getmode*. Tali funzioni permettono l'analisi della distribuzione delle variabili e le misure di posizione, di dispersione e forma della distribuzione.

Le analisi svolte hanno considerato le seguenti variabili:

- Caratteristiche del campione target (età, titolo di studio, professione, provincia di residenza);
- Conoscenza delle capsule ricaricabili e riutilizzabili;
- Possesso delle capsule ricaricabili e riutilizzabili;
- Spesa mensile;
- Willingness to pay per una capsula ricaricabile e riutilizzabile;
- Disponibilità ad acquistare una capsula Ecocapsula;
- Conoscenza delle marche concorrenti;
- Sensibilità ambientale.

1° ANALISI UNIVARIATA: Caratteristiche del campione target

ETÀ

```
> basicStats(capsule$eta)
      X..capsule.eta
nobs      200.000000
NAs        0.000000
Minimum    18.000000
Maximum    60.000000
1. Quartile 30.000000
3. Quartile 50.000000
Mean       40.330000
Median     40.000000
Sum        8066.000000
SE Mean    0.852814
LCL Mean   38.648288
UCL Mean   42.011712
Variance   145.458392
Stdev      12.060613
Skewness   -0.105773
Kurtosis   -1.172655
```

```
> getmode(capsule$eta)
[1] 22 38
```

```
> max(capsule$eta)-min(capsule$eta)
[1] 42
```

L'età media del nostro campione è di 40,33 anni; il valore più frequente, la moda, è sia 22 anni che 38 anni (variabile bimodale).

La mediana, pari a 40, indica che il 50% della popolazione intervistata ha un'età minore di 40,33 anni: vi è quindi un lieve scostamento tra media e mediana.

Il campo di variazione corrisponde a 42 anni: questo valore indica che vi è un range significativo tra il valore massimo di età rilevata (60) e il valore minimo (18).

Da questi dati si evince come l'intervallo di età considerata nel campione è ampio e spazia dai giovani ai più anziani.

1° ANALISI UNIVARIATA: Caratteristiche del campione target

ETÀ

QUANTILE	VALORE
100% Max	60
99%	60
95%	58,05
90%	57
75% Q3	50
50% Mediana	40
25% Q1	30
10%	22,9
5%	21,95
1%	19
0% Min	18

```
> quantile(capsule$eta, c(.01, .05, .10, .25, .50, .75, .90, .95, .99))
1%    5%    10%   25%   50%   75%   90%   95%   99%
19.00 21.95 22.90 30.00 40.00 50.00 57.00 58.05 60.00
```

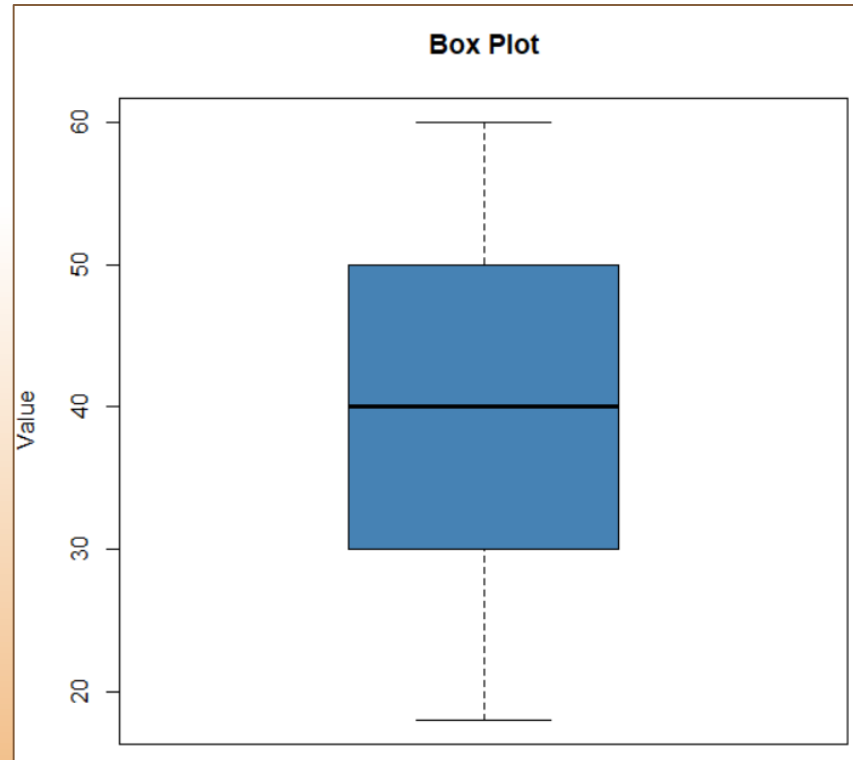
```
> summary(capsule$eta)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 18.00  30.00  40.00  40.33  50.00  60.00
```

```
> IQR(capsule$eta)
[1] 20
```

Osservando i dati ottenuti, la differenza interquartile (Q3-Q1) tra 50 anni e 30 anni risulta pari a 20. Inoltre possiamo capire che il 75% dei rispondenti ha un'età minore o uguale a 50 anni e solo il 25% ha al più 30 anni.

1° ANALISI UNIVARIATA: Caratteristiche del campione target

ETÀ

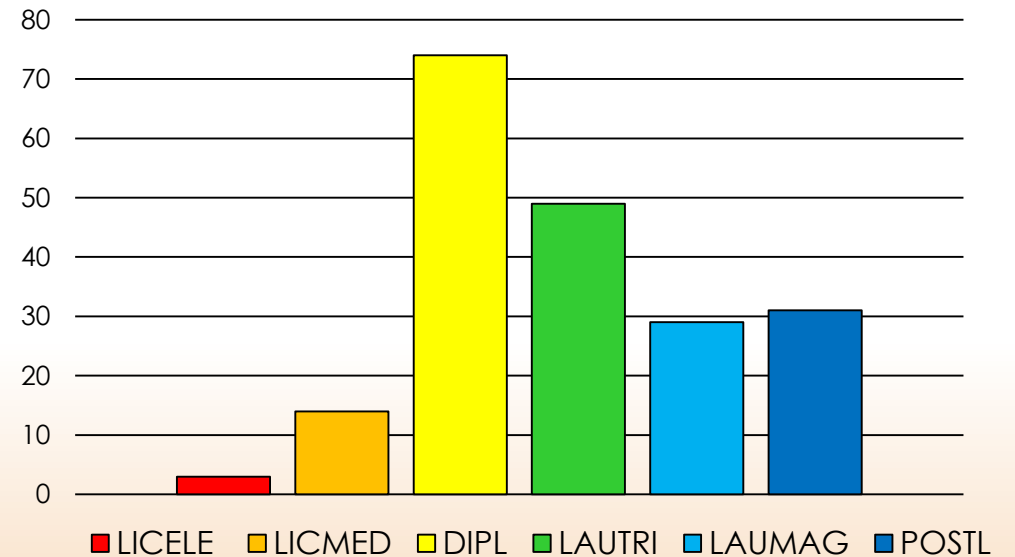


Attraverso il box plot notiamo infatti che la distribuzione è ben distribuita ed omogenea.

1° ANALISI UNIVARIATA: Caratteristiche del campione target

TITOLO DI STUDIO

```
> freq(capsule$titolostudio)
capsule$titolostudio
  Frequency Percent
DIPL          74   37.0
LAUMAG         29   14.5
LAUTRI         49   24.5
LICELE          3    1.5
LICMED          14    7.0
POSTL          31   15.5
Total         200  100.0
```



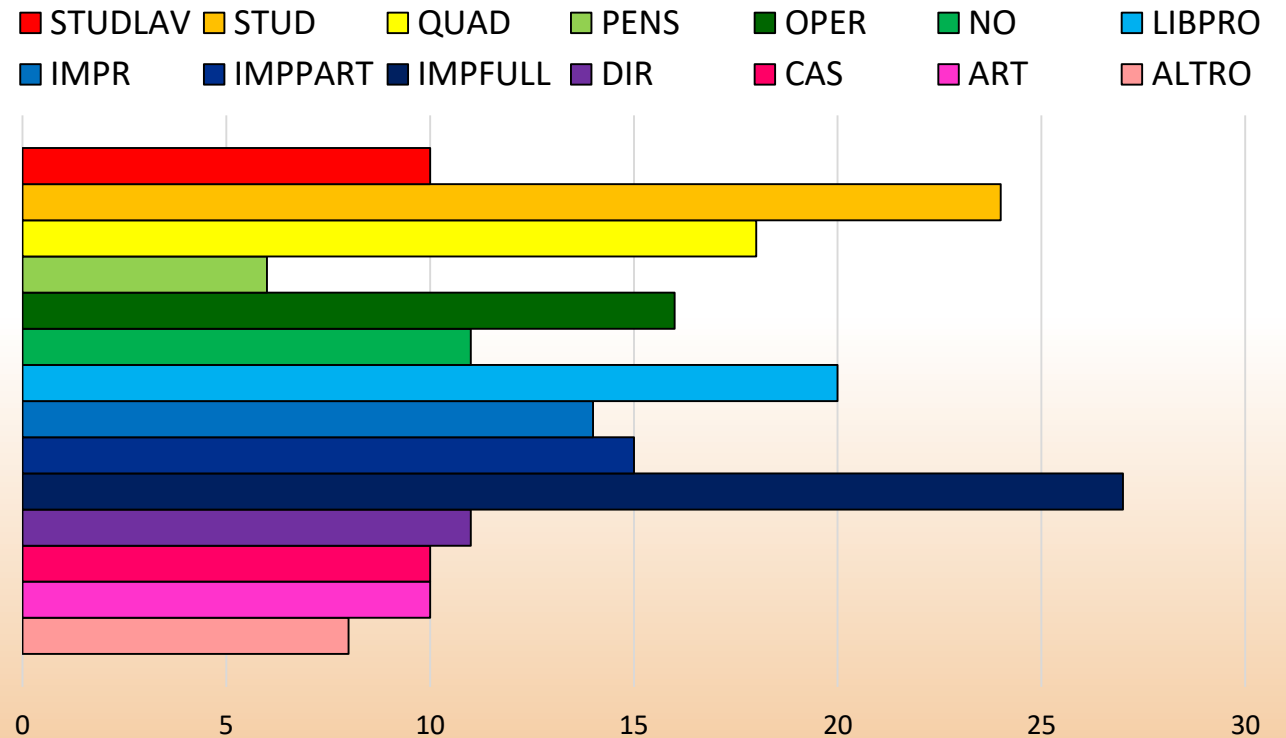
Dai risultati dei questionari, è emerso che solamente 3 intervistati si sono fermati alla Licenza Elementare (1,5%), 14 intervistati la Licenza Media (7%), mentre ben 74 individui hanno ottenuto un Diploma (37%).

Passando agli studi universitari, è emerso che 49 rispondenti hanno conseguito una Laurea Triennale (24,5%), mentre 29 una Laurea Magistrale (14,5%). Infine, 31 soggetti hanno ottenuto una qualifica Post Laurea (15,5%).

1° ANALISI UNIVARIATA: Caratteristiche del campione target

PROFESSIONE

```
> freq(capsule$professione)
capsule$professione
  Frequency Percent
ALTRO          6     3.0
ART            10    5.0
CAS            10    5.0
DIR            11    5.5
IMPFULL       27   13.5
IMPPART       15    7.5
IMPR          14    7.0
LIBPRO        20   10.0
NO            11    5.5
OPER          16    8.0
PENS          6     3.0
QUAD          18    9.0
STUD          26   13.0
STUDLAV       10    5.0
Total        200  100.0
```

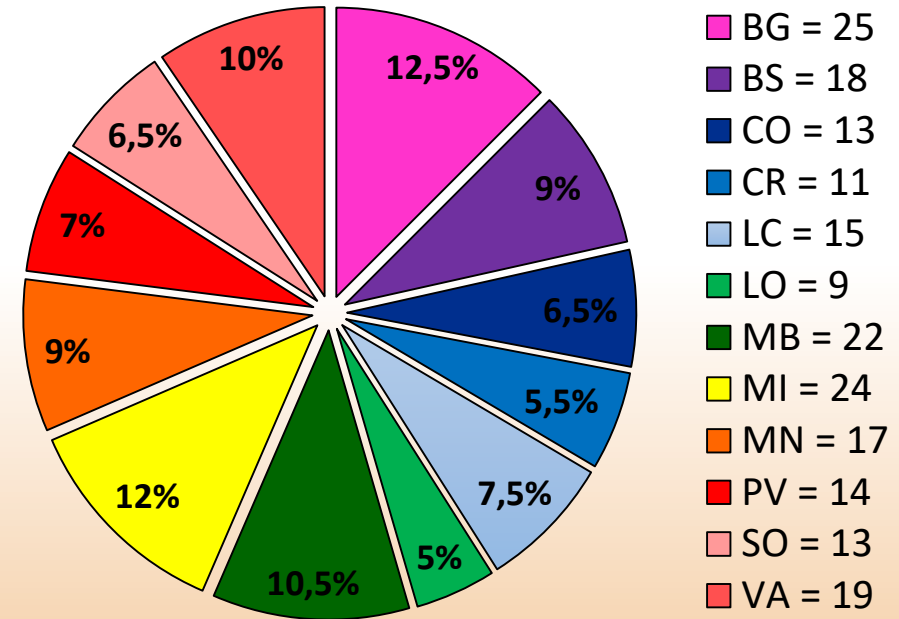


La professione che ha registrato il maggior numero di rilevazioni è quella dell'impiegato full-time (13,5%), seguita dallo studente (13%), dal libero professionista (10%) e dal quadro (9%). Durante il lavoro di analisi, ai fini della ricerca, sarà interessante valutare eventuali relazioni tra la professione degli intervistati e le loro abitudini di acquisto e di consumo.

1° ANALISI UNIVARIATA: Caratteristiche del campione target

PROVINCIA

```
> freq(capsule$provincia)
capsule$provincia
  Frequency Percent
BG          25    12.5
BS          18     9.0
CO          13     6.5
CR          11     5.5
LC          15     7.5
LO           9     4.5
MB          22    11.0
MI          24    12.0
MN          17     8.5
PV          14     7.0
SO          13     6.5
VA          19     9.5
Total       200   100.0
```



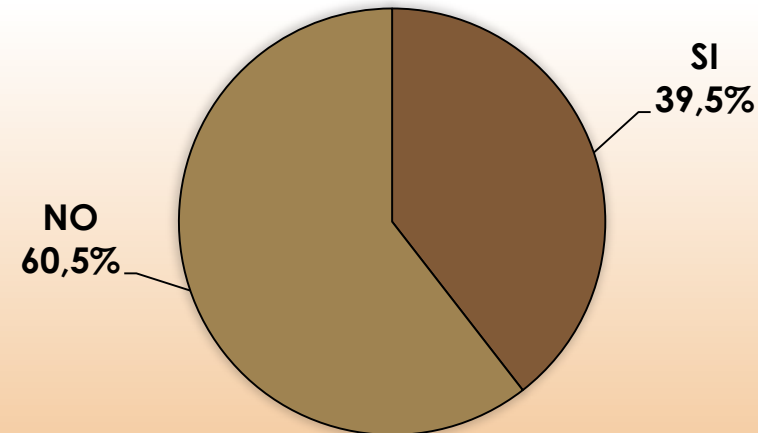
Dai risultati dei questionari, la provincia lombarda che ha riscontrato maggiori dati è stata la provincia di Bergamo, BG, con 25 intervistati (12,5%). Successivamente abbiamo Milano, Monza e Varese, rispettivamente con 24 (12%), 22 (10,5%) e 19 (9,5%) rispondenti.

2° ANALISI UNIVARIATA: Conoscenza delle capsule ricaricabili

La variabile a cui facciamo riferimento ci permette di capire rispetto al nostro campione quanti soggetti conoscono le capsule ricaricabili e riutilizzabili, assumendo la modalità di risposta dicotomica (SI/NO).

Dal grafico a torta emerge che la differenza tra i soggetti che conoscono le capsule ricaricabili e riutilizzabili, rispetto a quelli che non le conoscono, è significativa (60,5% contro 39,5%).

```
> freq(capsule$conos_capsule)
capsule$conos_capsule
  Frequency Percent
0           121    60.5
1           79    39.5
Total        200   100.0
```



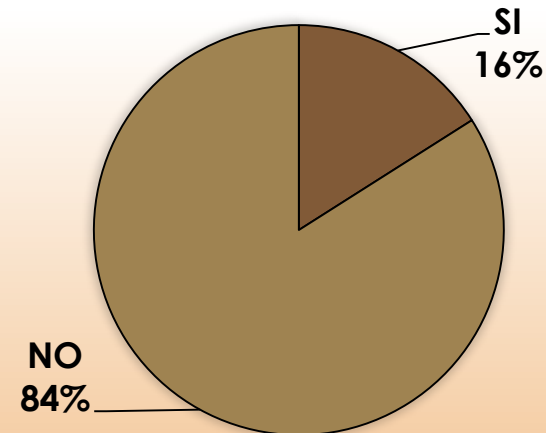
Questi dati sottolineano il fatto che questo prodotto innovativo non è conosciuto a pieno dalla popolazione target.

3° ANALISI UNIVARIATA: Possesso delle capsule ricaricabili

Un'ulteriore variabile posta in esame ci consente di capire rispetto al nostro campione quanti soggetti possiedono le capsule ricaricabili e riutilizzabili, assumendo la modalità di risposta dicotomica (SI/NO).

Dal grafico a torta emerge che il valore di coloro che non possiedono le capsule ricaricabili e riutilizzabili, rispetto a quelli che le possiedono, è molto elevato (84% contro 16%).

```
> freq(capsule$poss_capsule)
capsule$poss_capsule
  Frequency Percent
0           168     84
1           32     16
Total        200    100
```



Questi dati sono interessanti perché ci daranno la possibilità di analizzare, e successivamente confrontare, la disponibilità dei consumatori all'acquisto della nostra particolare capsule.

4° ANALISI UNIVARIATA: Spesa mensile

Per una successiva analisi univariata abbiamo preso in considerazione la variabile spesa media mensile. L'obiettivo di questa indagine è quello di comprendere quanto i nostri individui spendono mediamente nell'arco di un mese per valutare la loro disponibilità economica.

```
> getmode(capsule$spesa_acquisto)
[1] 20
```

```
> basicStats(capsule$spesa_acquisto)
      X..capsule.spesa_acquisto
nobs          200.000000
NAs            0.000000
Minimum         5.000000
Maximum        150.000000
1. Quartile    20.000000
3. Quartile    60.000000
Mean           45.725000
Median         40.000000
Sum            9145.000000
SE Mean        1.888406
LCL Mean       42.001145
UCL Mean       49.448855
Variance       713.215452
Stdev          26.706094
Skewness       0.770400
Kurtosis       0.411582
```

Dall'analisi è risultato che la moda è di 20€ a settimana.

In seguito abbiamo calcolato le seguenti misure di sintesi tramite la funzione *basicStats*.

Dalla tabella si nota che la spesa minima mensile rilevata dal campione è di 5 €, mentre la spesa massima mensile è di 150 €.

Dall'analisi interquartile possiamo capire che il 75% dei rispondenti spende al massimo 60€ al mese e solo il 25% spende al più 20€.

La spesa **media** è pari a 45,725€ al mese e la mediana a 40€.

4° ANALISI UNIVARIATA: Spesa mensile

Per interpretare meglio il valore della media abbiamo analizzato, attraverso un'analisi grafica (**box plot**) la presenza di valori estremi, gli *outliers*, che potrebbero influenzare la nostra media.

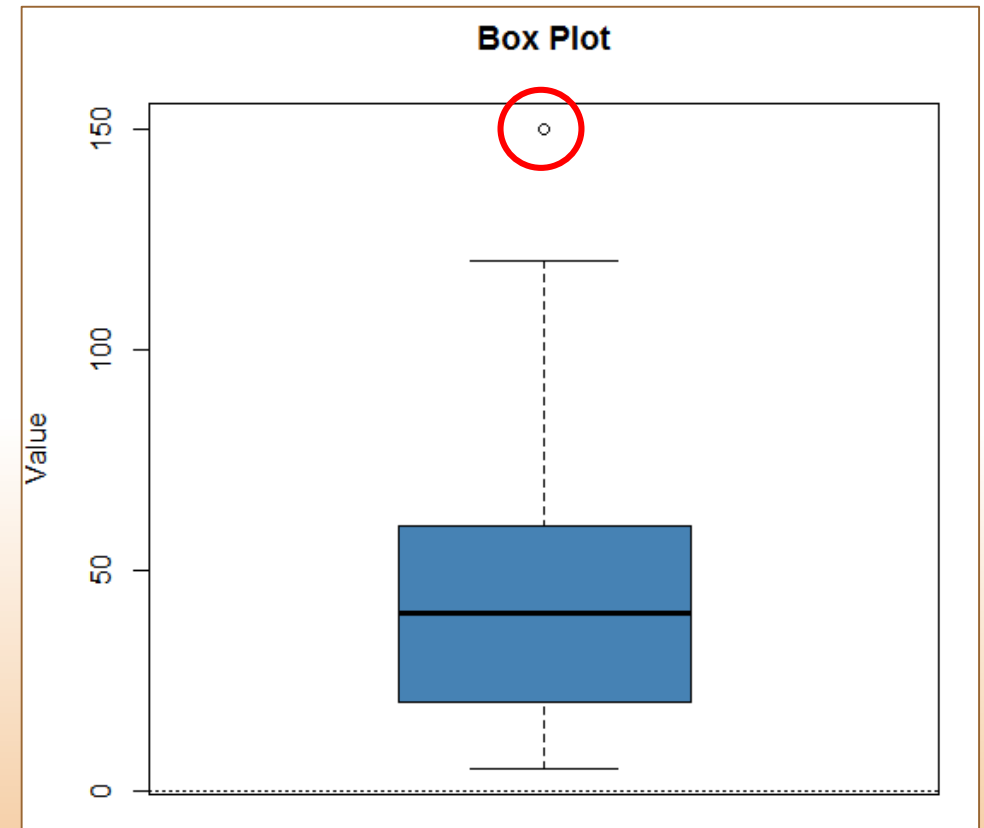
Con la formula $[Q3 + 1,5(Q3 - Q1)]$ abbiamo calcolato il valore massimo: $[60 + 1,5(40)] = 120 \text{ €}$, sopra al quale i valori verranno considerati outliers.

Da quest'analisi si nota che è presente un solo valore outlier che corrisponde a 150€ e che potrebbe influenzare la media.

Quindi abbiamo calcolato nuovamente la media, escludendo questo valore per verificare eventuali cambiamenti significativi.

La nuova media, senza il valore outlier, risulta essere pari a 45,201€. Questo valore si discosta dalla media precedentemente calcolata (45,725 €) solo di 0,524€.

Di conseguenza questo scostamento non è rilevante e quindi manteniamo nella nostra analisi il valore outlier.



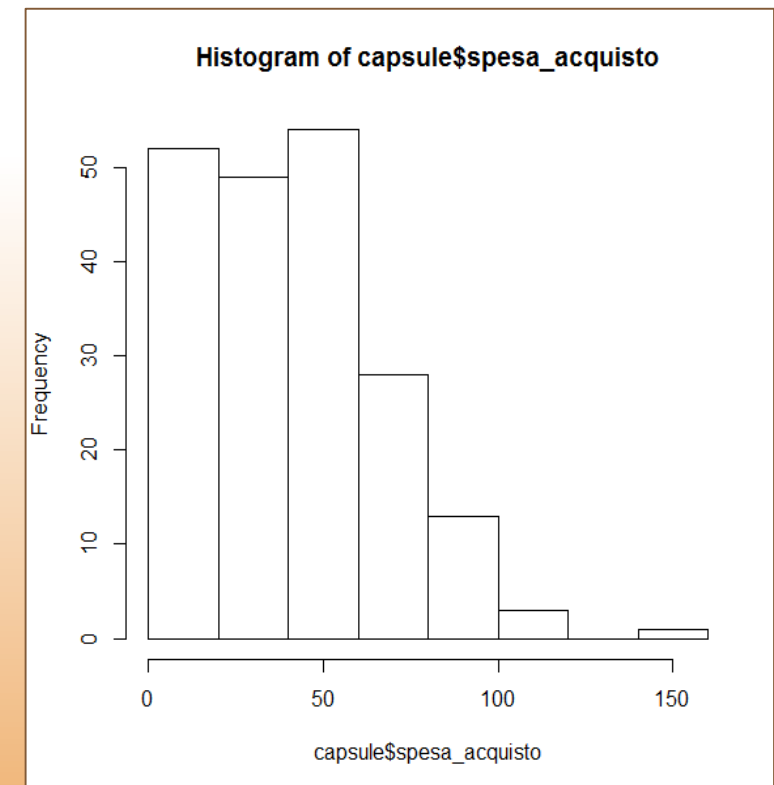
4° ANALISI UNIVARIATA: Spesa mensile

Sempre per la spesa mensile, abbiamo analizzato la distribuzione della spesa attraverso la varianza, in particolare abbiamo utilizzato lo scarto quadratico medio perché ci permette di dare un'interpretazione più accurata in quanto non è sensibile agli scostamenti.

Lo scarto quadratico medio (**Stdev**) mostra la variabilità rispetto alla media ed è pari a 26,706: i nostri dati evidenziano una discreta variabilità strutturale quindi è presente una leggera dispersione.

Variance	713.215452
Stdev	26.706094
Skewness	0.770400
Kurtosis	0.411582

Come si nota dal grafico, la distribuzione non è simmetrica. Infatti il valore dell'**indice di Kurtosis** è pari a 0,411 ($\beta < 3$) e sottolinea che i dati seguono una distribuzione di tipo iponormale; il valore dell'**indice di Skewness** è pari 0,77 ($\gamma > 0$) e evidenzia l'asimmetria positiva (mediana < media). Quindi i dati da noi raccolti indicano una disomogenea distribuzione della spesa tra i consumatori.



5° ANALISI UNIVARIATA: Willingness to pay

Tramite l'analisi di questa variabile vogliamo analizzare quale sia la willingness to pay per una capsula ricaricabile e riutilizzabile.

```
> getmode (capsule$wtp)
[1] 10
```

```
> basicStats (capsule$wtp)
      X..capsule.wtp
nobs      200.000000
NAs        0.000000
Minimum    2.000000
Maximum   100.000000
1. Quartile 10.000000
3. Quartile 33.000000
Mean       24.085000
Median     20.000000
Sum        4817.000000
SE Mean    1.298603
LCL Mean   21.524211
UCL Mean   26.645789
Variance   337.274146
Stdev      18.365025
Skewness   1.428788
Kurtosis   2.547939
```

Dall'analisi è risultato che la moda è pari a 10€. Successivamente abbiamo calcolato le seguenti misure di sintesi tramite la funzione *basicStats*.

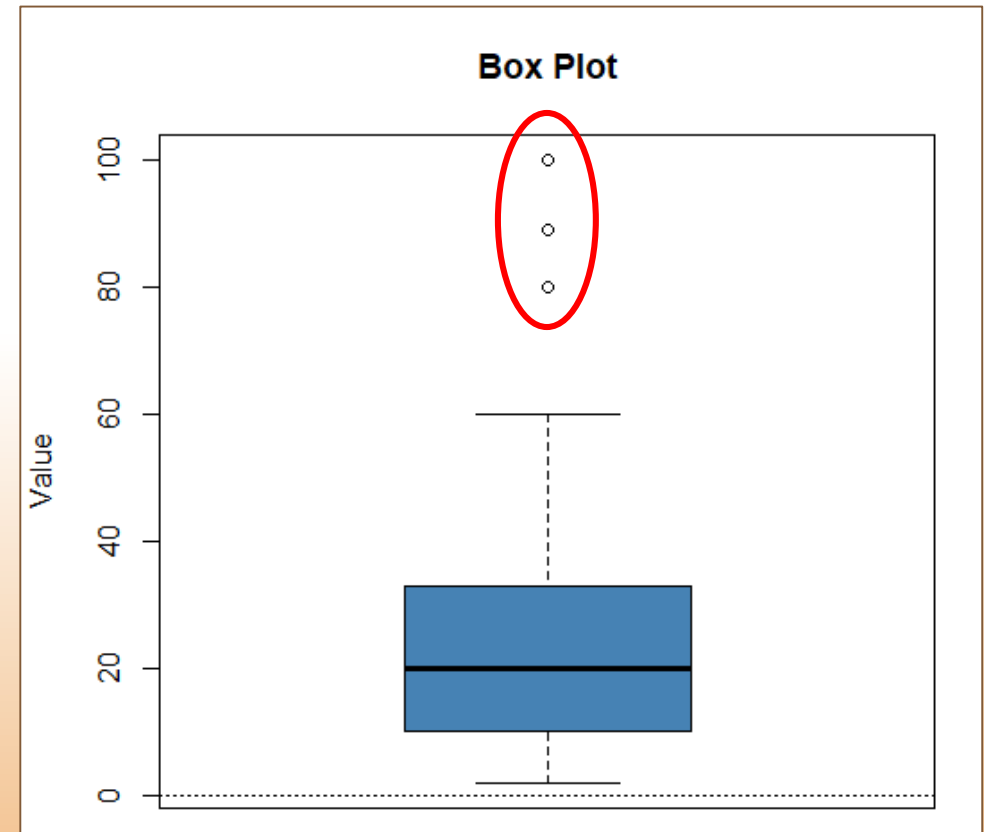
La spesa minima che gli intervistati sarebbero disposti ad affrontare per l'acquisto di 2€, mentre la spesa massima è pari a 100€.

La **media** corrisponde a 24,085€, la **mediana** a 20€. Dall'analisi interquartile si nota che il 75% dei rispondenti spenderebbe al più 33€ e solo il 25% spenderebbe 10€; quindi la differenza interquartile corrisponde a 23€.

5° ANALISI UNIVARIATA: Willingness to pay

Con la formula $[Q3 + 1,5(Q3 - Q1)]$ abbiamo calcolato il valore massimo: $[33 + 1,5(23)] = 67,5€$, sopra al quale i valori verranno considerati outliers. Attraverso il **box plot** si può notare la presenza di *outliers*; da quest'analisi si nota che sono presenti tre valori outliers, relativamente 80€, 89€ e 100€, che corrispondono a 5 questionari. Successivamente abbiamo calcolato di nuovo la media, escludendo questi valori per verificare eventuali cambiamenti significativi. La media calcolata senza gli outliers è pari 22,4€: il valore si discosta dalla media precedentemente calcolata (24,085€) solamente di 1,685€.

Quindi questo scostamento non è rilevante e per questo motivo i valori outliers vengono considerati nella nostra analisi.



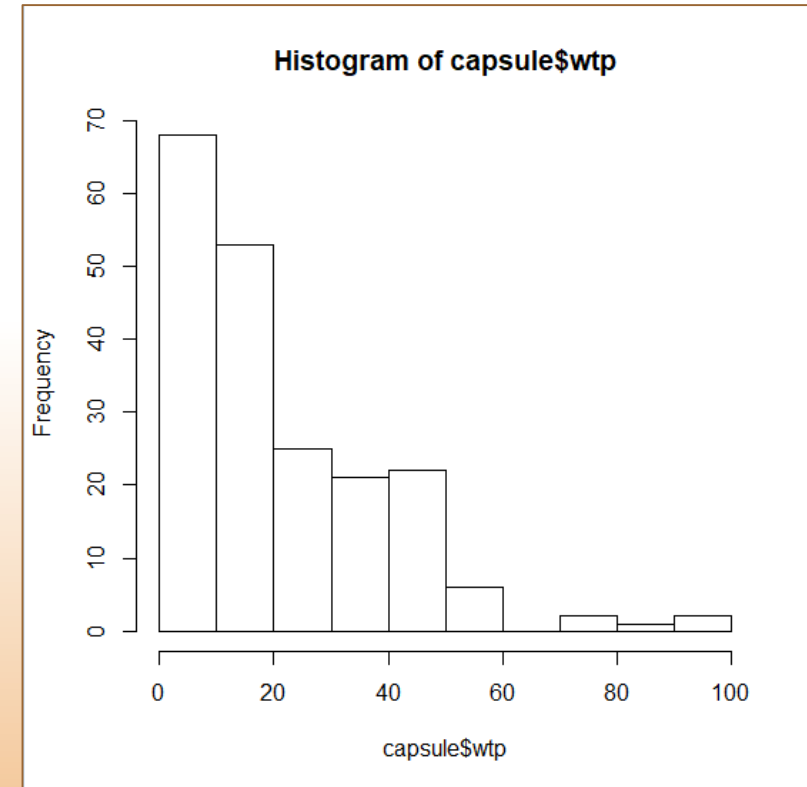
5° ANALISI UNIVARIATA: Willingness to pay

Variance	337.274146
Stdev	18.365025
Skewness	1.428788
Kurtosis	2.547939

Lo **scarto quadratico medio** è pari a 18,365: i nostri dati evidenziano una discreta variabilità strutturale.

Come si può notare dal grafico, la distribuzione è iponormale ($\beta < 3$), in quanto l'**indice di Kurtosis** è pari a circa a 2,547.

Inoltre si riscontra un'asimmetria positiva poiché l'**indice di Skewness** è pari a 1,58 ($\gamma > 0$) e la media è maggiore della mediana.

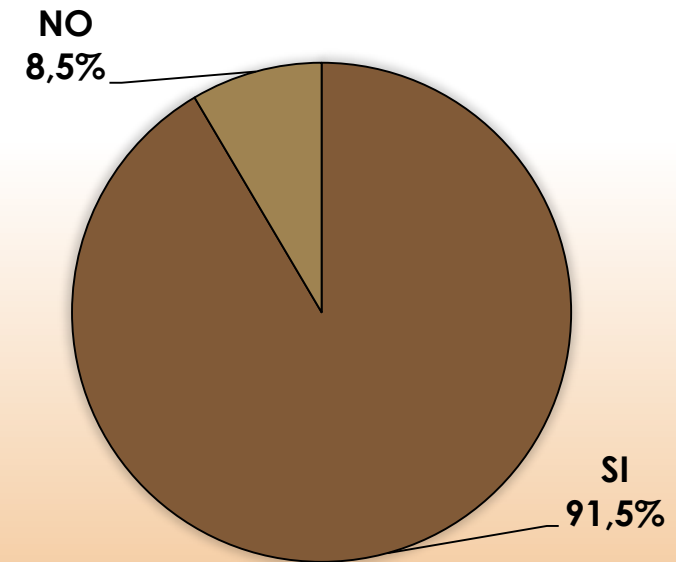


6° ANALISI UNIVARIATA: Disponibilità ad acquistare una capsula Ecocapsula

Un'altra variabile importante ai fini della comprensione del nostro target è la disponibilità ad acquistare la capsula ricaricabile e riutilizzabile Ecocapsula con un riutilizzo di 300 volte.

Attraverso la funzione di R *freq* abbiamo individuato qual è la percentuale di individui che è propenso ad acquistare una capsula ricaricabile e riutilizzabile.

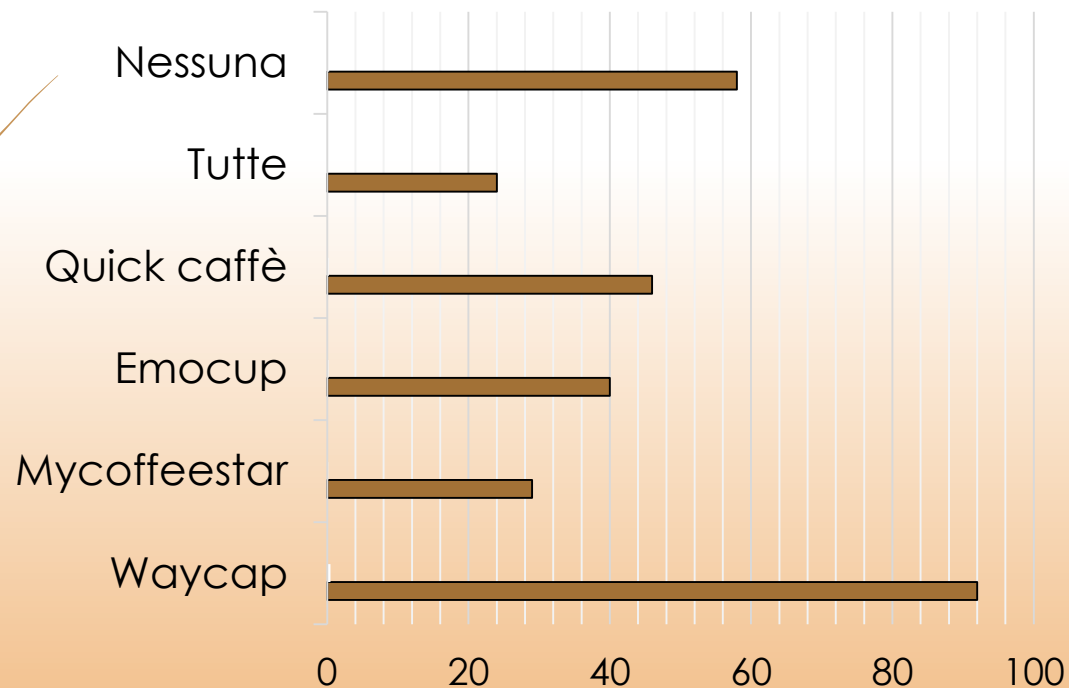
```
> freq(capsule$wtp_ecocap)
capsule$wtp_ecocap
  Frequency Percent
0             17    8.5
1            183   91.5
Total         200  100.0
```



Abbiamo verificato che il 91,5% è propenso ad acquistare una capsula ricaricabile e riutilizzabile, mentre il 8,5% no.

7° ANALISI UNIVARIATA: Conoscenza delle marche concorrenti

Al fine di comprendere il mercato delle capsule ricaricabili e riutilizzabili abbiamo analizzato la conoscenza delle marche concorrenti.



I rispondenti avevano la possibilità di indicare più opzioni (massimo 3), quindi ciascuna marca rappresenta una singola variabile da analizzare usando la funzione **freq** in R. Dal risultato dell'analisi, come è possibile osservare dal grafico, risulta che la marca più conosciuta è «Waycap».

Questo dato può essere rilevante per ulteriori analisi, al fine di comprendere se questo competitor viene preferito o meno a Ecocapsula da parte dei nostri rispondenti.

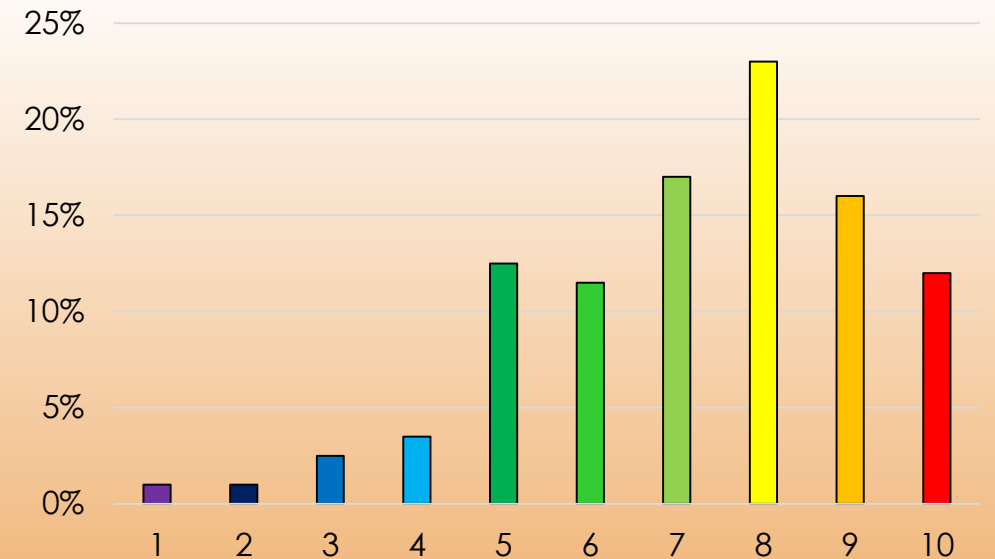
8° ANALISI UNIVARIATA: Sostenibilità ambientale

L'ultima analisi univariata che abbiamo condotto riguarda la sostenibilità ambientale, ovvero quanto gli intervistati ritengono importante il rispetto dell'ambiente, per comprendere se l'ecosostenibilità del nostro prodotto risulta essere un aspetto fondamentale. Da quello che si evince dalla funzione *freq* di R è che il campione risulta molto sensibile all'argomento in quanto 159 rispondenti su 200, ovvero il 79,5% degli intervistati, hanno dato, in una scala da 1 a 10, un voto superiore a 6.

```
> freq(capsule$rispetto_ambiente)
capsule$rispetto_ambiente
  Frequency Percent
1           2     1.0
2           2     1.0
3           5     2.5
4           7     3.5
5          25    12.5
6          23    11.5
7          34    17.0
8          46    23.0
9          32    16.0
10         24    12.0
Total      200   100.0
```

```
> cbind(cumsum(table(capsule$rispetto_ambiente)),
+ cumsum(table(capsule$rispetto_ambiente)/200))
  [,1] [,2]
1     2 0.010
2     4 0.020
3     9 0.045
4    16 0.080
5    41 0.205
6    64 0.320
7    98 0.490
8   144 0.720
9   176 0.880
10  200 1.000
```

Il 20,5% degli intervistati hanno dato un voto inferiore o pari a 5



Conclusioni

L'analisi condotta fino a questo momento ha consentito di comprendere le abitudini d'acquisto del campione, fornendo un quadro di partenza per lo sviluppo del business case.

Sulla base delle analisi univariate svolte risulta che:

- ❖ L'età media del nostro campione è di circa 40 anni. Il 37% degli individui hanno ottenuto almeno un diploma e la professione che risulta più praticata dal nostro campione è l'impiegato full time. Inoltre la provincia lombarda che ha riscontrato maggiori dati è stata la provincia di Bergamo, con il 12,5%;
- ❖ Il 39,5% degli intervistati dichiara di conoscere le capsule ricaricabili e riutilizzabili, e di questi solamente il 16% le possiede. Tali risultati indicano che questo prodotto innovativo non è conosciuto a pieno dalla popolazione target;
- ❖ La spesa media mensile dei nostri intervistati si attesta tra i 5 e 150 euro. Quindi nel nostro negozio ci aspettiamo che, mensilmente, un individuo spenda intorno ai 20 euro (il valore modale dei nostri rispondenti);

Conclusioni

- ❖ Il 91,5% dei rispondenti è propenso ad acquistare una capsula ricaricabile e riutilizzabile e la willingness to pay media è pari a 24,085€. Ciò indica che la nostra capsula viene percepita come un prodotto valido, al quale viene attribuita una percezione di prezzo adeguata;
- ❖ Infine, per quanto riguarda la conoscenza delle marche concorrenti, il più noto competitor risulta essere Waycap. In futuro sarà interessante ottenere un profilo più dettagliato possibile del target al fine di attuare un'efficace strategia marketing non solo nei confronti dei consumatori ma anche dei concorrenti;
- ❖ La sostenibilità è un aspetto fondamentale per le caratteristiche del nostro prodotto e il campione risulta molto sensibile all'argomento, infatti il 79,5% dei consumatori ha indicato un valore di almeno 6 su una scala da 1 a 10;

In conclusione l'offerta di Ecocapsula sembra in linea con le preferenze della popolazione target.



ANALISI BIVARIATA

Obiettivi dell'analisi bivariata

L'analisi bivariata ha come obiettivo l'analisi della relazione di dipendenza o indipendenza tra le variabili.

Sono state utilizzate le seguenti funzioni di R: *CrossTable*, *CramerV*, *cor.test*, *rcorr*, *anova*, *summary(anova)*, *model.tables*, *etaSquared(anova)*.

Le analisi condotte con le relative variabili sono state:

ANALISI DI CONNESSIONE	ANALISI DI CORRELAZIONE LINEARE	ANALISI ANOVA
Provincia e disponibilità ad acquistare una capsula Ecocapsula	Provenienza made in Italy e sostenibilità	Willingness to pay e sesso
Disponibilità ad acquistare una capsula Ecocapsula e la conoscenza della marca Waycap	Spesa mensile e nucleo familiare	Willingness to pay e provincia
	Willingness to pay e età	Spesa mensile e professione

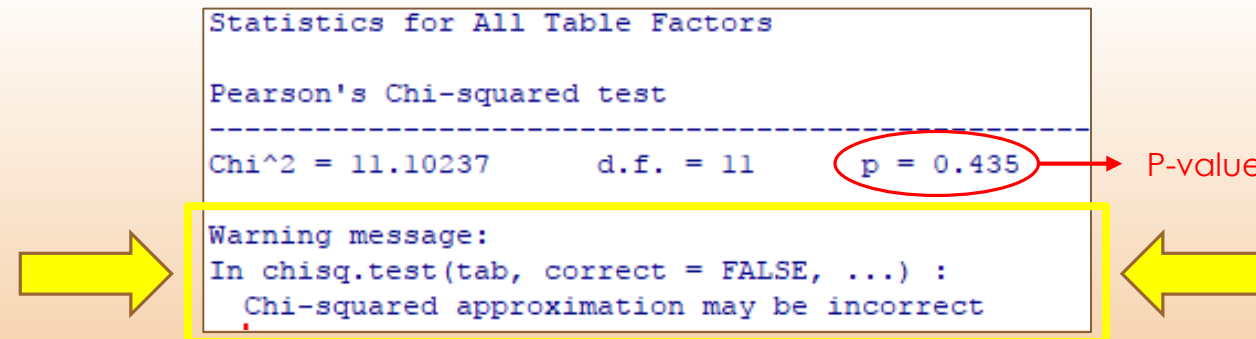
Analisi di connessione (1)

Le variabili utilizzate per l'analisi di connessione sono variabili qualitative. Nella prima analisi di connessione abbiamo verificato se le variabili Provincia e Disponibilità all'acquisto della capsula Ecocapsula sono correlate. Questa verifica è stata effettuata per valutare in quale provincia sarebbe più indicato avviare il nostro punto vendita. Le due variabili analizzate sono entrambe qualitative quindi la funzione utilizzata è *CrossTable*. Per valutare l'esistenza di una relazione statisticamente rilevante tra le due variabili, abbiamo svolto un test di ipotesi **Chi-quadro X^2** per l'indipendenza statistica.

```

Statistics for All Table Factors

Pearson's Chi-squared test
-----
Chi^2 = 11.10237      d.f. = 11      p = 0.435
Warning message:
In chisq.test(tab, correct = FALSE, ...) :
  Chi-squared approximation may be incorrect
  
```



Poniamo H_0 come ipotesi nulla, ovvero di indipendenza statistica e H_1 come ipotesi alternativa. Il p-value corrisponde a 0,435 ed è maggiore del livello di significatività ($\alpha = 0,05$) accettiamo l'ipotesi nulla H_0 , quindi le nostre due variabili sono statisticamente indipendenti.

Analisi di connessione (1)

Come si può notare dalla tabella di contingenza, più del 20% delle celle ha frequenza congiunta assoluta < 5 e quindi R segnala che il test svolto è inaffidabile (slide precedente).

cps\$_	capsule\$provincia												Total
	BG	BS	CO	CR	LC	LO	MB	MI	MN	PV	SO	VA	
0	4	1	0	1	0	1	1	3	0	1	1	4	17
	0.235	0.059	0.000	0.059	0.000	0.059	0.059	0.176	0.000	0.059	0.059	0.235	0.085
	0.160	0.056	0.000	0.091	0.000	0.111	0.045	0.125	0.000	0.071	0.077	0.211	
	0.020	0.005	0.000	0.005	0.000	0.005	0.005	0.015	0.000	0.005	0.005	0.020	
1	21	17	13	10	15	8	21	21	17	13	12	15	183
	0.115	0.093	0.071	0.055	0.082	0.044	0.115	0.115	0.093	0.071	0.066	0.082	0.915
	0.840	0.944	1.000	0.909	1.000	0.889	0.955	0.875	1.000	0.929	0.923	0.789	
	0.105	0.085	0.065	0.050	0.075	0.040	0.105	0.105	0.085	0.065	0.060	0.075	
Total	25	18	13	11	15	9	22	24	17	14	13	19	200
	0.125	0.090	0.065	0.055	0.075	0.045	0.110	0.120	0.085	0.070	0.065	0.095	

Per questo motivo, anche se la prima analisi svolta evidenzia indipendenza tra le variabili, non possiamo stabilire in quale provincia avviare il nostro punto vendita Ecocapsula. A questo punto proviamo a raggruppare le provincia in due macro-aree: la prima categoria «Nord» composta da MB, LC, SO, CO, BS e BG; mentre la seconda categoria «Sud» formata da MI, VA, LO, CR, MN e PV.

Analisi di connessione (1)

Dopo aver raggruppato le province nelle due macro-aree (Nord e Sud) abbiamo svolto nuovamente un test di ipotesi **Chi-quadro χ^2** .

```
Statistics for All Table Factors
-----
Pearson's Chi-squared test
-----
Chi^2 = 1.042674      d.f. = 1      p = 0.307
-----
Pearson's Chi-squared test with Yates' continuity correction
-----
Chi^2 = 0.5884508    d.f. = 1      p = 0.443
```

Poniamo H_0 come ipotesi nulla e H_1 come ipotesi alternativa. Il p-value corrisponde a 0,307 ed è maggiore del livello di significatività ($\alpha = 0,05$) accettiamo l'ipotesi nulla H_0 , quindi le nostre due variabili sono statisticamente indipendenti.

Analisi di connessione (1)

Per verificare l'esito del test abbiamo ricalcolato l'**indicatore V di Cramer** che è pari a 0,07220366: questo risultato è inferiore a 0,2 e quindi ciò conferma la mancaza di connessione.

```
> CramerV(raggruppamenti$provincia,raggruppamenti$wtp_ecocap)
[1] 0.07220366
```

Perciò la provincia di residenza dei rispondenti non influenza la disponibilità di acquisto di Ecocapsula, e viceversa. Di conseguenza possiamo dedurre che non sarà importante in quale provincia avviare il nostro punto vendita.

Il 41,2% che abita al Nord non acquisterebbe Ecocapsula

Il 58,8% che abita al Sud non acquisterebbe Ecocapsula

raggruppamenti\$provincia	raggruppamenti\$wtp_ecocap		Total
	0	1	
NORD	7 0.066 0.412 0.035	99 0.934 0.541 0.495	106 0.530
SUD	10 0.106 0.588 0.050	84 0.894 0.459 0.420	94 0.470
Total	17 0.085	183 0.915	200

Il 54,1% che abita al Nord acquisterebbe Ecocapsula

Il 45,9% che abita al Sud acquisterebbe Ecocapsula

Analisi di connessione (2)

Per la seconda analisi di connessione abbiamo verificato se le variabili Disponibilità ad acquistare una Capsula Ecocapsula e la conoscenza della marca Waycap, considerata il principale competitor, sono correlate.

Questo esame è stato condotto per valutare se Ecocapsula, nonostante non sia il first player sul mercato, abbia comunque possibilità di vendita rispetto al principale concorrente. Le due variabili analizzate sono entrambe qualitative quindi le funzioni utilizzate sono *CrossTable* e *CramerV*.

Statistics for All Table Factors		
Pearson's Chi-squared test		
Chi ² = 0.1740228	d.f. = 1	p = 0.677
Pearson's Chi-squared test with Yates' continuity correction		
Chi ² = 0.02650198	d.f. = 1	p = 0.871

Dall'analisi di **Chi-quadrato X^2** risulta p-value pari a $0.677 > 0,05$ quindi accettiamo l'ipotesi H_0 e possiamo affermare che non vi sia connessione tra le due variabili.

Analisi di connessione (2)

Per verificare l'esito del test abbiamo calcolato l'**indicatore V di Cramer** che risulta essere $0.02949769 < 0,2$: questo ci conferma la manca di connessione.

```
> CramerV(capsule$wtp_ecocap, capsule$d_18_waycap)
[1] 0.02949769
```

Quindi la conoscenza della marca Waycap non influenza la disponibilità di acquisto di Ecocapsula, e viceversa. In termini economici questo comporta che la nostra capsula potrebbe avere successo sul mercato, anche con la presenza dell'attuale concorrenza.

Il 9,3% che non conosce Waycap non acquisterebbe Ecocapsula

Il 90,7% che non conosce Waycap non acquisterebbe Ecocapsula

capsule\$wtp_ecocap	capsule\$d_18_waycap		Total
	0	1	
0	10 0.588 0.093 0.050	7 0.412 0.076 0.035	17 0.085
1	98 0.536 0.907 0.490	85 0.464 0.924 0.425	183 0.915
Total	108 0.54	92 0.46	200

Il 7,6% che conosce Waycap non acquisterebbe Ecocapsula

Il 92,4% che conosce Waycap acquisterebbe Ecocapsula

Analisi di correlazione (1)

Le variabili utilizzate per l'analisi di correlazione sono variabili quantitative. Per la prima analisi di correlazione abbiamo considerato le variabili quantitative influenza della provenienza made in Italy e quella della sostenibilità del prodotto. Abbiamo fatto questa analisi, utilizzando la funzione *Cor.test*, per valutare quanto la propensione dei consumatori ad acquistare prodotti sostenibili aumenti all'aumentare della garanzia dei prodotti di provenienza italiana.

```
> cor.test(capsule$d_13_provenienza, capsule$d_13_sostenibilita)

Pearson's product-moment correlation

data: capsule$d_13_provenienza and capsule$d_13_sostenibilita
t = 8.814, df = 198, p-value = 6.188e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4232733 0.6236501
sample estimates:
cor
0.5308405
```

→ Coefficiente di correlazione lineare ρ

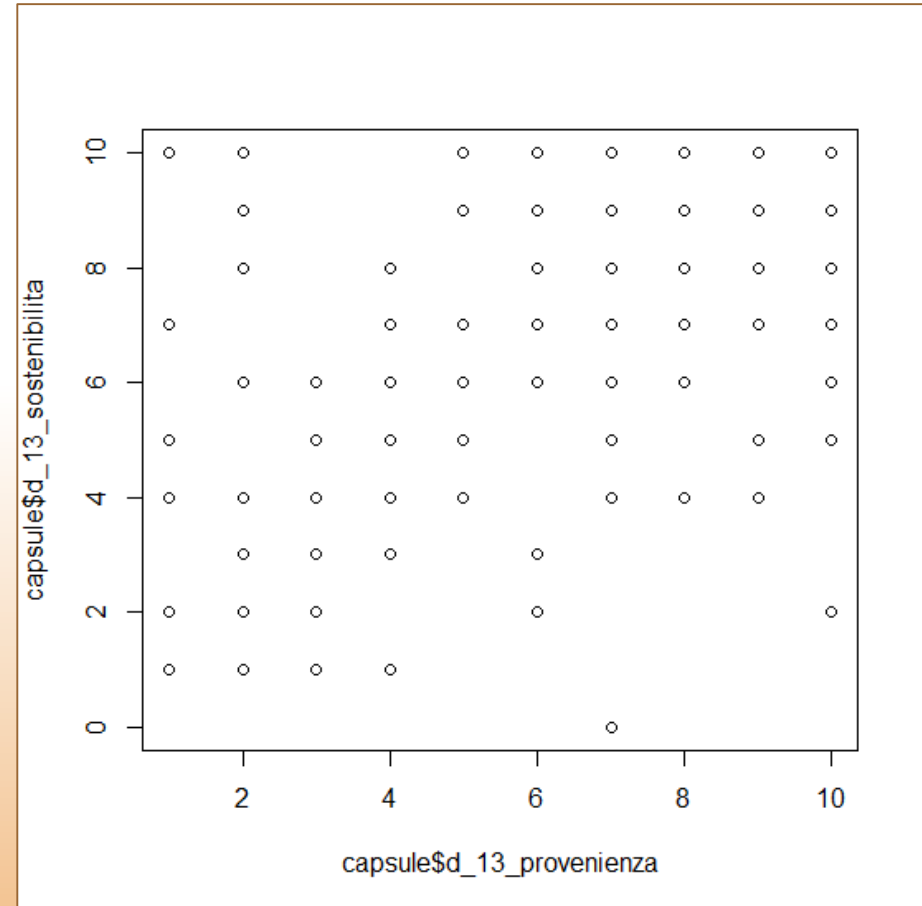
Per verificare se esiste una relazione lineare tra le due variabili è stato condotto un **test d'ipotesi t di Student**. Dall'output ottenuto risulta che le variabili sostenibilità del prodotto e made in Italy sono

linearmente correlate, considerando il livello di significatività α pari a 0,05. Inoltre il coefficiente di correlazione lineare ρ è risultato 0,5308405 maggiore di 0,2. Quindi questo implica che tra sostenibilità del prodotto e made in Italy ci sia una correlazione lineare positiva, essendo l'indice > 0 .

Analisi di correlazione (1)

La relazione lineare viene mostrata nello **scatterplot** seguente.
Questo risultato è in linea con le nostre aspettative perché la propensione all'acquisto di un prodotto sostenibile aumenta al crescere della garanzia della provenienza Made in Italy.

Quindi il nostro obiettivo sarà quello di differenziare la nostra offerta, puntando su entrambe le variabili di sostenibilità e provenienza italiana.



Analisi di correlazione (2)

Per la seconda analisi di correlazione abbiamo considerato le variabili spesa media mensile e quella del nucleo familiare.

Questa analisi attraverso la funzione `Cor.test` è stata eseguita per valutare se all'aumentare del numero di componenti del nucleo familiare aumenta anche la spesa d'acquisto mensile.

Abbiamo effettuato il **test t di Student** per valutare se esiste una relazione lineare tra le due variabili.

```
> cor.test(capsule$spesa_acquisto, capsule$nucleofam)

Pearson's product-moment correlation

data: capsule$spesa_acquisto and capsule$nucleofam
t = 1.8648, df = 198, p-value = 0.06369
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.007501853  0.265281569
sample estimates:
 cor
0.1313759
```

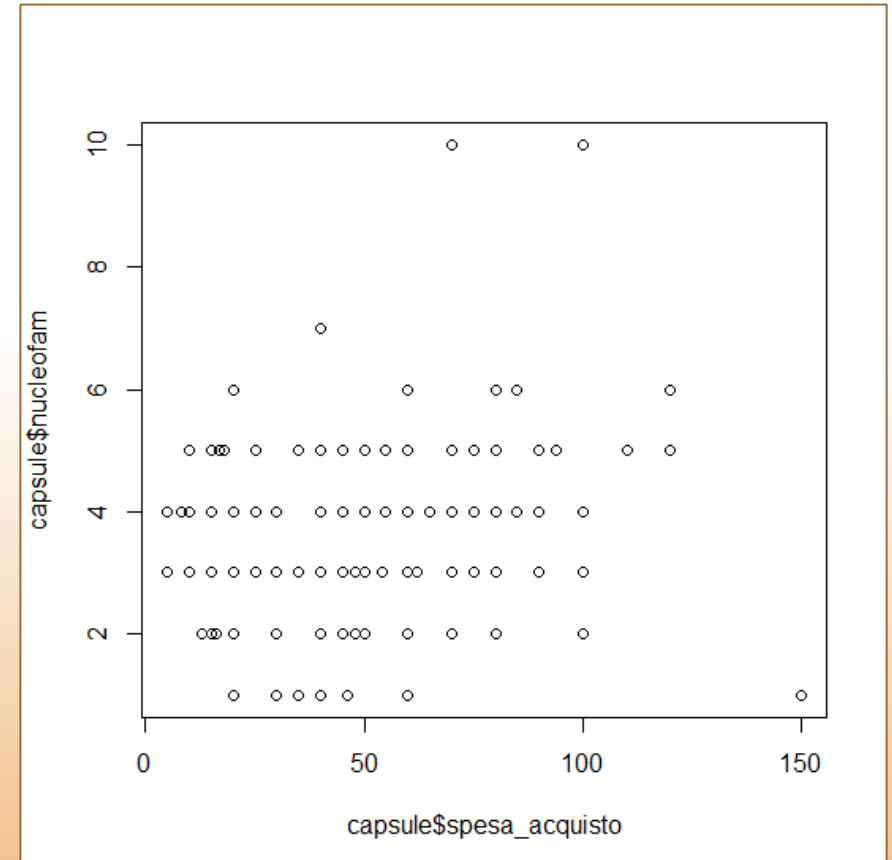
L'esito del test t fornisce un p-value pari a 0,06369 e maggiore di 0,05: quindi accettiamo H_0 e rifiutiamo l'ipotesi H_1 . Le variabili willingness to pay ed età non sono linearmente correlate, avendo considerato il livello di significatività α pari a 0,05.

A conferma dell'indipendenza statistica, il coefficiente di correlazione lineare ρ 0,1313759 è minore di 0,2. Perciò tra le variabili spesa mensile e numero di componenti del nucleo familiare non vi è correlazione lineare.

Analisi di correlazione (2)

Anche lo **scatterplot** mostra che i dati non hanno un legame lineare; questo evidenzia che tra spesa mensile e numero di componenti del nucleo familiare non è presente alcuna correlazione lineare. Contrariamente all'output conseguito, immaginavamo che le due variabili fossero correlate perché normalmente un numero maggiore di componenti del nucleo familiare implica una maggiore spesa mensile, ma dall'evidenza empirica ne deriva che ciò non è riscontrato.

Di conseguenza per l'apertura del nostro punto vendita non sarà rilevante segmentare l'offerta e pubblicizzare il prodotto in base al nucleo familiare, in quanto garantire una prospettiva di risparmio per le famiglie numerose non è significativo.



Analisi di correlazione (3)

Come ultima analisi di correlazione abbiamo considerato le variabili willingness to pay ed età.

La realizzazione della seguente analisi è fatta per determinare se la propensione all'acquisto delle capsule aumenta all'aumentare dell'età del nostro campione.

Con la funzione `Cor.test` abbiamo effettuato un **test d'ipotesi t di Student** per valutare se esiste una relazione lineare tra le due variabili.

```
> cor.test(capsule$wtp, capsule$eta)

      Pearson's product-moment correlation

data:  capsule$wtp and capsule$eta
t = -0.44256, df = 198, p-value = 0.6586
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1694380  0.1077751
sample estimates:
      cor
-0.03143598
```

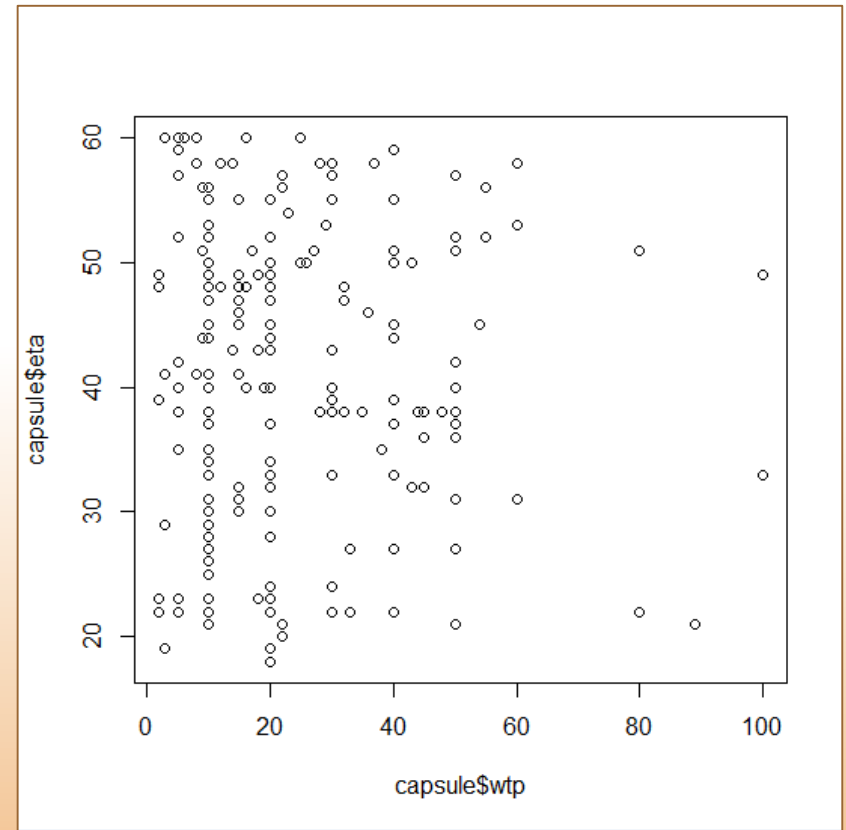
Dal risultato del test t emerge che $P\text{-value} = 0.6586 > 0,05$ quindi accettiamo H_0 e rifiutiamo l'ipotesi H_1 . Le variabili willingness to pay ed età non sono linearmente correlate, avendo considerato il livello di significatività α pari a 0,05.

A conferma dell'indipendenza statistica, il coefficiente di correlazione lineare ρ è $-0,03143598$ e prossimo allo zero. Perciò tra le variabili willingness to pay ed età non è presente nessuna correlazione lineare.

Analisi di correlazione (3)

Dal seguente **scatterplot** si può notare che non esiste un legame lineare tra i dati; ciò significa che tra willingness to pay ed età non vi è correlazione lineare.

Di conseguenza, ai fini della nostra analisi, possiamo affermare che la propensione all'acquisto della nostra capsula non aumenta in modo significativo al crescere dell'età; quindi la distribuzione del nostro target sarà alquanto omogenea nei confronti delle due variabili prese in considerazione.



ANOVA (1)

Le variabili utilizzate per Anova sono miste, ovvero una quantitativa e una qualitativa e le funzioni utilizzate sono *aov*, *model.table*, *summary*, *etaSquared*.
 Abbiamo svolto la prima analisi dell'Anova per capire se c'è dipendenza tra le variabili willingness to pay e sesso dei rispondenti.

Poniamo H_0 come ipotesi nulla, ovvero indipendenza in media tra le variabili e H_1 come ipotesi alternativa, ovvero dipendenza tra esse.

Fissiamo $\alpha = 0,05$ come livello di significatività.

Dall'analisi del **test F di Fisher** emerge che il p-value 0,834 è maggiore di 0,05. Per questo motivo accettiamo l'ipotesi H_0 e rifiutiamo H_1 e quindi non vi è dipendenza in media tra le variabili willingness to pay e sesso.

```
> summary(anova)
      Df Sum Sq Mean Sq F value Pr(>F)
sesso  1     15    14.9    0.044  0.834
Residuals 198 67103   338.9
```

ANOVA (1)

Per confermare l'esito del test e dimostrare che non vi è comunque dipendenza in media, abbiamo calcolato il valore dell'**indice eta quadro** pari a 0,0002213903. Tale valore, essendo molto vicino a 0 e inferiore a 0,2 (indicatore soglia), avvalorata l'ipotesi di indipendenza in media.

```
> anova=aov(wtp~sesso, data=capsule)
> model.tables(anova, type='means')
Tables of means
Grand mean
24.085 → Media totale

sesso
      F      M
24.35 23.8 → Media dei singoli gruppi
rep 103.00 97.0
```

```
> etaSquared(anova)
      eta.sq  eta.sq.part
sesso 0.0002213903 0.0002213903
```

La media della willingness to pay sembra molto vicina tra i diversi gruppi. Rispetto al nostro obiettivo, non vi è una classe dominante verso cui indirizzare l'offerta però è comunque interessante notare che la propensione alla spesa della classe maschile si accosta moltissimo a quella femminile, da cui deduciamo un potenziale interesse per il prodotto.

ANOVA (2)

Abbiamo svolto la prima analisi dell'Anova per capire se c'è dipendenza tra le variabili willingness to pay e provincia di residenza dei rispondenti.

Poniamo H_0 come ipotesi nulla, ovvero indipendenza in media tra le variabili e H_1 come ipotesi alternativa, ovvero dipendenza tra esse.

Fissiamo $\alpha = 0,05$ come livello di significatività.

Dall'analisi del **test F di Fisher** emerge che il p-value 0,425 è maggiore di 0,05. Per questo motivo accettiamo l'ipotesi H_0 e rifiutiamo H_1 e quindi non vi è dipendenza in media tra le variabili willingness to pay e provincia di residenza.

```
> summary(anova)
      Df Sum Sq Mean Sq F value Pr(>F)
provincia  11   3801   345.5   1.026  0.425
Residuals 188  63317   336.8
```

ANOVA (2)

Per confermare l'esito del test e dimostrare che non vi è comunque dipendenza in media, abbiamo calcolato il valore dell'**indice eta quadro** pari a 0,05663165. Tale valore, essendo molto vicino a 0 e inferiore a 0,2 (indicatore soglia), avvalorata l'ipotesi di indipendenza in media.

```
> anova=aov(wtp~provincia, data=capsule)
> model.tables(anova, type='means')
Tables of means
Grand mean
24.085

provincia
  BG  BS  CO CR  LC  LO  MB  MI  MN  PV  SO
  22.12 27.5 15.77 25 22.73 21.78 18.86 22.88 28.06 34.07 24.08
rep 25.00 18.0 13.00 11 15.00  9.00 22.00 24.00 17.00 14.00 13.00
  VA
  27.42
rep 19.00
```

```
> etaSquared(anova)
      eta.sq eta.sq.part
provincia 0.05663165 0.05663165
```

I residenti delle province di Pavia, Mantova, Brescia e Varese hanno la willingness to pay maggiore.

ANOVA (3)

L'ultima analisi dell'Anova è stata condotta per conoscere se c'è dipendenza tra le variabili spesa mensile e professione dei rispondenti.

Poniamo H_0 come ipotesi nulla, ovvero indipendenza in media tra le variabili e H_1 come ipotesi alternativa, ovvero dipendenza tra esse.

Fissiamo $\alpha = 0,05$ come livello di significatività.

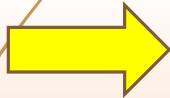
Dall'analisi del **test F di Fisher** emerge che il p-value 0,0132 è minore di 0,05. Per questo motivo rifiutiamo l'ipotesi H_0 e accettiamo H_1 e quindi vi è dipendenza in media tra le variabili spesa mensile e professione.

```
> summary(anova)
              Df Sum Sq Mean Sq F value Pr(>F)
professione  13  18548  1426.8    2.151 0.0132
Residuals   186 123382   663.3
```

ANOVA (3)

Siccome le variabili sono risultate dipendenti in media dal test di ipotesi, ora calcoleremo l'indice η^2 per misurare l'intensità della relazione. Considerando 0,2 come valore soglia, si osserva che il valore dell'indice è pari a 0,1306842, quindi, non c'è dipendenza.

```
> etaSquared(anova)
      eta.sq eta.sq.part
professione 0.1306842 0.1306842
```



Questo esito è opposto a quello precedentemente raggiunto quindi questo può determinare che i valori medi sono differenti ma le strutture di variabilità sottostanti non li rendono statisticamente diversi.

Per risolvere il problema di variabilità, abbiamo deciso di raggruppare le opzioni di risposta: la prima categoria «Reddito alto» è composta da dirigenti, imprenditori, liberi professionisti, quadri e impiegati full-time; la seconda categoria «Reddito medio» racchiude gli impiegati part-time, gli artigiani, gli operai e altre professioni; invece la terza e ultima categoria «Reddito basso» comprende gli studenti, gli studenti lavoratori, le casalinghe, i disoccupati e i pensionati.

ANOVA (3)

Dopo la suddivisione delle 14 opzioni di risposta in 3 nuove categorie ripetiamo l'analisi dell'Anova.

```
> model.tables(anova, type='means')
Tables of means
Grand mean
45.725

professione
  REDALTO REDBASSO REDMEDIO
    48.77   36.51   52.26
rep    90.00   63.00   47.00
```

```
> summary(anova)
      Df Sum Sq Mean Sq F value Pr(>F)
professione  2   8189    4095   6.031 0.00287 **
Residuals 197 133741     679
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Abbiamo effettuato il **test F di Fisher** per la verifica di ipotesi sulla differenza tra medie. L'output ci fornisce un p-value 0,00287 minore del livello di significatività α pari a 0,05. Per il seguente risultato rifiutiamo l'ipotesi H_0 e quindi le due variabili sono dipendenti in media per cui le medie non sono tutte uguali: almeno un valor medio è diverso. A questo punto calcoliamo nuovamente l'indice η^2 .

ANOVA (3)

```
> etaSquared(anova)
           eta.sq eta.sq.part
professione 0.05769816 0.05769816
```

Nonostante l'aggregazione, l'indice assume un valore ancora più basso rispetto a quello precedentemente calcolato. Quindi non essendo riusciti a risolvere i problemi di variabilità non possiamo confermare la relazione tra la spesa mensile e la professione degli intervistati.

L'obiettivo dell'analisi condotta era comprendere quanto i consumatori, in base alla loro professione, spendono mensilmente.

Dai risultati ottenuti, contrariamente a quanto immaginavamo, possiamo dedurre che la professione svolta dagli intervistati non è rilevante per la spesa mensile che essi fanno. Questo ottenuto è fortemente indicativo: infatti, a livello di comunicazione, implica l'applicazione di una strategia omogenea e trasversale.

Conclusioni

Partendo dai risultati dell'analisi univariata, lo studio è proseguito con l'analisi bivariata, suddivisa in analisi di connessione, analisi di correlazione e Anova. Abbiamo ottenuto da queste analisi i seguenti risultati:

- ❖ Per la prima analisi di connessione abbiamo voluto verificare se le variabili Provincia e Disponibilità all'acquisto della capsula Ecocapsula sono correlate per valutare in quale provincia sarebbe più conveniente avviare il punto vendita. È risultato che le variabili non sono connesse quindi desumiamo che non è importante in quale provincia aprire il punto vendita di Ecocapsula;
- ❖ Per la seconda analisi di connessione abbiamo verificato se le variabili Disponibilità ad acquistare una Capsula Ecocapsula e la Conoscenza della marca Waycap. Dall'analisi è risultato che la conoscenza della marca Waycap non influenza la disponibilità di acquisto di Ecocapsula, e viceversa. Per questo motivo, anche se il competitor principale è conosciuto, il nostro prodotto potrebbe avere ugualmente successo;

Conclusioni

- ❖ Per la prima analisi di correlazione abbiamo considerato l'influenza della Provenienza made in Italy e quella della Sostenibilità del prodotto per stimare quanto la propensione dei consumatori ad acquistare prodotti sostenibili aumenti all'aumentare della garanzia dei prodotti di provenienza italiana. Dagli output ottenuti è emerso che le due variabili sono correlate positivamente per cui sarà fondamentale differenziare la nostra offerta, puntando su entrambe le variabili di sostenibilità e provenienza italiana;
- ❖ Per la seconda analisi di correlazione abbiamo valutato il legame tra Spesa mensile e Numero dei componenti del nucleo familiare con l'intento di comprendere se all'aumentare del numero di componenti del nucleo familiare aumentasse anche la spesa d'acquisto mensile. Il risultato ottenuto è che le variabili non sono correlate tra loro e quindi, per l'apertura del nostro punto vendita, non sarà rilevante segmentare l'offerta e pubblicizzare il prodotto in base al nucleo familiare;
- ❖ Per la terza analisi di correlazione abbiamo voluto determinare se la Propensione all'acquisto delle capsule aumenta all'aumentare dell'età del nostro campione. I risultati ottenuti hanno evidenziato che tra willingness to pay ed età non vi è correlazione quindi la distribuzione del nostro target sarà alquanto omogenea nei confronti delle due variabili prese in considerazione;

Conclusioni

- ❖ Per la prima analisi Anova abbiamo esaminato la possibile dipendenza tra le variabili willingness to pay e sesso dei rispondenti. Dai risultati ottenuti possiamo notare che tra gli intervistati non è presente una classe dominante verso cui indirizzare l'offerta. Ne ricaviamo che non dobbiamo rivolgerci, nel proporre la nostra offerta, principalmente a uno o l'altro sesso;
- ❖ Per la seconda analisi Anova volevamo capire se c'è dipendenza tra le variabili willingness to pay e provincia di residenza dei rispondenti. Dai risultati ottenuti non vi è dipendenza in media tra le due variabili e di conseguenza nell'apertura del nostro punto vendita non è rilevante considerare la variabile willingness to pay riferito al campione;
- ❖ Per la terza analisi Anova abbiamo voluto verificare se la Professione degli intervistati influenza la Spesa mensile di capsule di caffè. È emerso che, diversamente da quanto immaginavamo, la professione non incide sulla spesa media mensile poiché, secondo il nostro team, essendo comunque una spesa poco rilevante è accessibile a tutti, indipendentemente dalla professione e di conseguenza dal reddito. Questa considerazione è indispensabile poiché a livello di comunicazione dell'offerta, implica l'implementazione di una strategia omogenea e trasversale.



ANALISI FATTORIALE

Obiettivi dell'analisi fattoriale

L'obiettivo della nostra analisi è quello di spiegare come la spesa media mensile venga influenzata dai comportamenti e dalle esigenze strutturali e funzionali dei consumatori (regressione lineare).

Per fare ciò abbiamo utilizzato i nostri dati comportamentali di tipo quantitativo (in particolare le scale di soddisfazione) trasformandoli in fattori per evitare problemi di multicollinearità, al fine di costruire un modello di regressione lineare solido.

Nell'eventualità in cui il modello spieghi poco, inseriremo variabili socio-demografiche (sesso, provincia, età, titolo di studio e nucleo familiare) per completare la realizzazione del modello.



Scelta delle variabili

Nella seguente tabella sono elencate le variabili quantitative indipendenti che abbiamo selezionato per spiegare la variabile dipendente spesa media mensile.

#	VARIABILE	DESCRIZIONE VARIABILE	#	VARIABILE	DESCRIZIONE VARIABILE
1	D_13_Gusto	Importanza del «Gusto»	10	D_16_SceltaCaffè	Importanza della «Possibilità di scelta del caffè da inserire nella capsula»
2	D_13_Sostenibilità	Importanza del «Sostenibilità»	11	D_16_Convenienza	Importanza di «Una maggiore convenienza»
3	D_13_Brand	Importanza del «Brand»	12	D_16_RiutilizzoCapsula	Importanza della «Possibilità di riutilizzare la capsula»
4	D_13_Provenienza	Importanza del «Provenienza (Made in Italy)»	13	D_16_ImpattoAmbiente	Importanza di «Un impatto positivo sull'ambiente»
5	D_13_ProdottoBio	Importanza del «Prodotto Bio»	14	D_19_Prezzo	Importanza del «Prezzo»
6	D_13_Packaging	Importanza del «Packaging»	15	D_19_Durevolezza	Importanza del «Durevolezza del prodotto»
7	D_13_RapportoQP	Importanza del «Rapporto qualità/prezzo»	16	D_19_SensAmbiente	Importanza del «Sensibilità ambientale»
8	Soddisfazione_marca	Soddisfazione della marca che si acquista generalmente	17	D_19_QuantitàCaffè	Importanza della «Possibilità di inserire la quantità di caffè desiderata»
9	D_16_QualitàProdotto	Importanza di «Una maggiore qualità del prodotto»	18	D_19_Ingombro	Importanza del «Minore spazio di ingombro»

Queste 18 variabili, da sintetizzare in n fattori, sono state scelte per conferire maggiore robustezza al modello di regressione lineare.

Componenti principali

A questo punto procederemo a svolgere l'analisi fattoriale con il metodo delle componenti principali.



Ai fini della scelta del numero dei fattori da utilizzare adopereremo i seguenti metodi:

- Regola empirica (rapporto tra numero di componenti e variabili);
- Regola degli autovalori > 1 ;
- Percentuale Varianza spiegata (compresa tra 60% e 75%);
- Grafico Scree-Plot.

Scelta del numero delle componenti

```

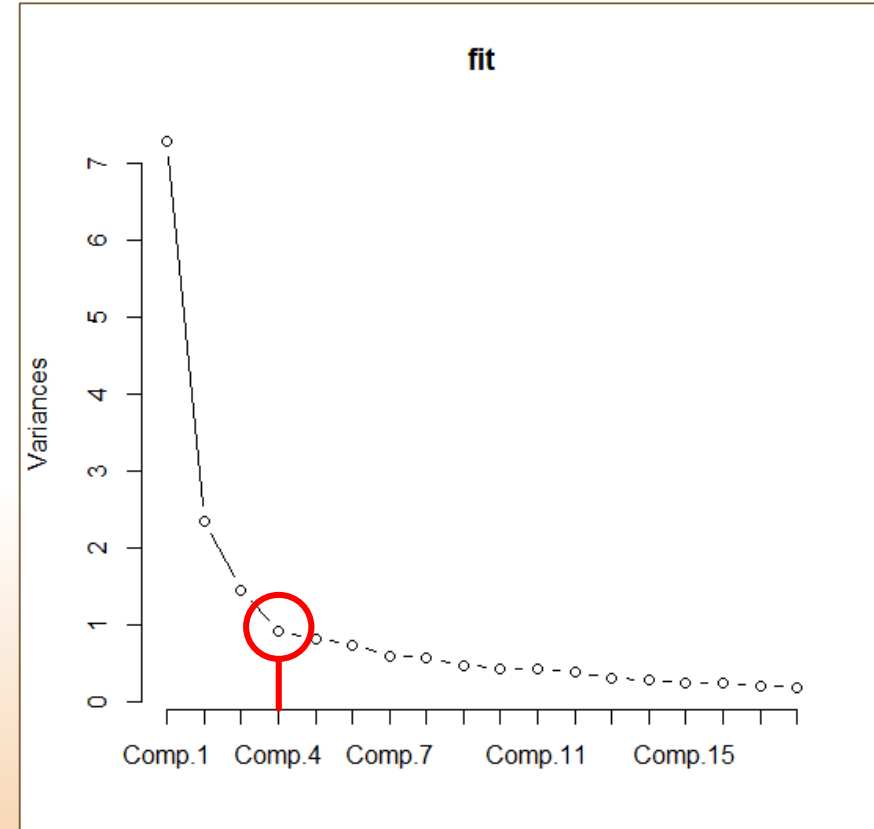
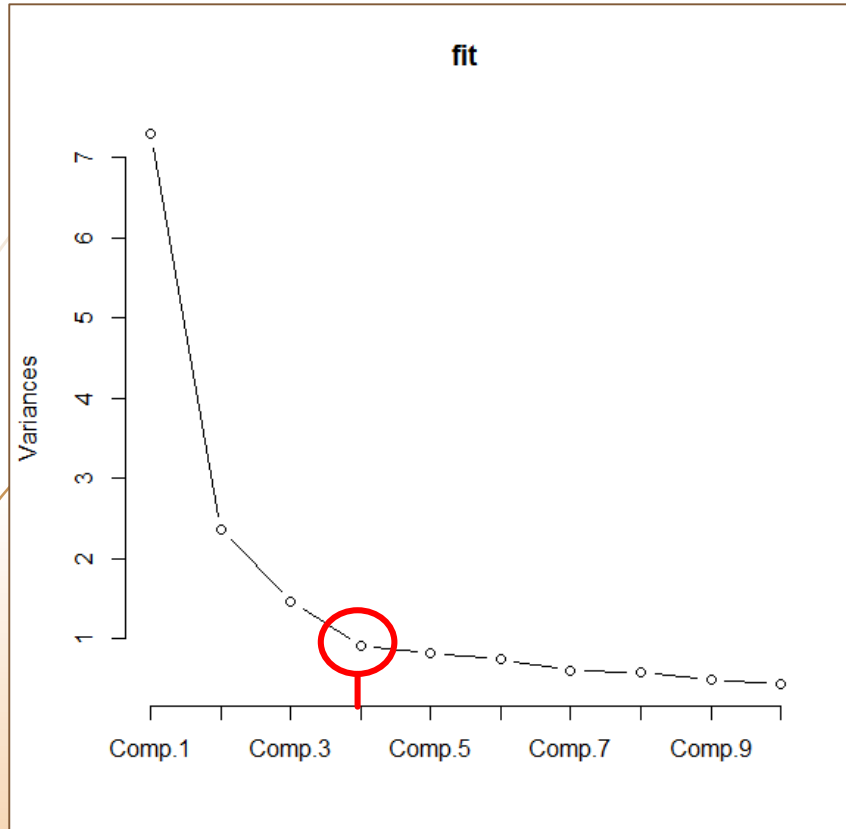
> eig.val<-get_eigenvalue(fit)
> eig.val
      eigenvalue variance.percent cumulative.variance.percent
Dim.1  7.2935763          40.519868          40.51987
Dim.2  2.3557447          13.087470          53.60734
Dim.3  1.4587450           8.104139          61.71148
Dim.4  0.9145998           5.081110          66.79259
Dim.5  0.8244746           4.580414          71.37300
Dim.6  0.7423959           4.124422          75.49742
Dim.7  0.6037747           3.354304          78.85173
Dim.8  0.5817633           3.232018          82.08374
Dim.9  0.4808074           2.671152          84.75490
Dim.10 0.4389411           2.438562          87.19346
Dim.11 0.4227542           2.348634          89.54209
Dim.12 0.3908246           2.171248          91.71334
Dim.13 0.3040022           1.688901          93.40224
Dim.14 0.2917692           1.620940          95.02318
Dim.15 0.2525632           1.403129          96.42631
Dim.16 0.2413696           1.340942          97.76725
Dim.17 0.2100768           1.167093          98.93435
Dim.18 0.1918177           1.065654         100.00000
  
```

Per la **regola empirica**, secondo la quale una buona sintesi è espressa da 1/3 delle variabili originarie, dobbiamo considerare 6 fattori.

Con il **metodo degli autovalori** selezioniamo le componenti principali con varianza maggiore di 1, ovvero gli autovalori (*eigenvalue*) maggiori di 1: consideriamo quindi 3 fattori, che spiegano insieme il 61,7% della varianza totale.

Utilizzando il **metodo della varianza spiegata** potremmo scegliere 3, 4, o 5 fattori perché la loro varianza spiegata totale è compresa tra il 60% e il 75%.

Scree plot



Osservando gli **scree-plot**, notiamo che è presente un gomito in corrispondenza di 4 fattori (la soluzione a 4 fattori presenta una varianza spiegata pari al 66%).

I metodi di analisi utilizzati precedentemente ci hanno fornito 3 possibili soluzioni: 3 o 4 o 5 fattori. Per essere certi del numero di fattori da utilizzare procediamo al confronto delle 3 soluzioni con il metodo delle comunalità.

Comunalità

Variabili	n=3	n=4	n=5
d_13_gusto	0,6727	0,6858	0,6873
d_13_sostenibilità	0,6961	0,7285	0,7292
d_13_brand	0,6165	0,6387	0,6401
d_13_provenienza	0,5579	0,5869	0,5870
d_13_prodotto bio	0,7065	0,7366	0,7383
d_13_packaging	0,6497	0,6500	0,6511
d_13_rapporto qp	0,4833	0,4885	0,7089
soddisfazione_marca	0,4904	0,5208	0,7796
d_16_qualità prodotto	0,6015	0,6034	0,6140
d_16_scelta caffè	0,5947	0,6359	0,6505
d_16_convenienza	0,7153	0,7445	0,7484
d_16_riutilizza capsula	0,6301	0,7911	0,7912
d_16_impatto ambiente	0,7199	0,7911	0,7954
d_19_prezzo	0,5902	0,7167	0,7521
d_19_durevolezza	0,6756	0,7514	0,7922
d_19_sensambiente	0,6800	0,6800	0,6876
d_19_quantità caffè	0,5846	0,6630	0,6707
d_19_ingombro	0,4421	0,6091	0,8228

Abbiamo deciso di confrontare la varianza spiegata rispetto a una soluzione a 3 fattori, a 4 fattori e a 5 fattori.

Le soluzioni evidenziate in giallo sono le celle in cui la comunalità aumenta in maniera sostanziale per effetto dell'estrazione di un altro fattore. Paragonando le comunalità delle tre soluzioni, si può notare come tutte e tre le possibili soluzioni potrebbero essere adeguate (valori maggiori di 0,4).

Siccome preferiamo avere una mediazione tra i criteri di sintesi e di informazione, scegliamo la soluzione a 4 fattori, soddisfacente per entrambi i canoni, e procediamo all'interpretazione.

Interpretabilità delle componenti

Ora andremo a interpretare i fattori al fine di completare l'analisi fattoriale.

```

Loadings:
          PC1    PC2    PC3
d_13_gusto      0.681      0.359
d_13_sostenibilita 0.689  0.460
d_13_rapportoqp  0.606      0.332
d_16_qualitaprodotto 0.646 -0.349
d_16_sceltacaffe  0.716
d_16_convenienza 0.753 -0.384
d_16_riutilizzocapsula 0.755
d_16_impattoambiente 0.746      -0.404
d_19_prezzo      0.699 -0.318
d_19_durevolezza 0.750 -0.337
d_19_sensambiente 0.718      -0.394
d_19_quantitacaffe 0.738
d_19_ingombro    0.516      -0.352
d_13_provenienza 0.523  0.531
d_13_prodotto bio 0.538  0.636
d_13_packaging   0.351  0.705
soddisfazione_marca      0.597
d_13_brand       0.469  0.414  0.474
  
```

A causa della difficile possibilità di comprendere questo output, sfrutteremo la rotazione dei fattori (metodo Varimax) per interpretare al meglio i risultati ottenuti.

Usando il **metodo Varimax** operiamo una rotazione dei fattori in modo che ogni variabile originaria sia molto correlata con al massimo un solo fattore e poco correlata con gli altri.

Così facendo otterremo un output che ci permette di raggruppare più facilmente i fattori mantenendo la stessa variabilità.

Interpretabilità delle componenti

	RC1	RC3	RC2
d_16_scltcaffè	0,564	0,523	
d_16_convenienza	0,632	0,561	
d_16_riutilizzocapsula	0,719		
d_16_impattoambiente	0,814		
d_19_prezzo	0,568	0,517	
d_16_durevolezza	0,591	0,570	
d_19_sensambiente	0,773		
d_19_quantitacaffè	0,664	0,343	
d_19_ingombro	0,576		0,322
d_13_gusto	0,304	0,749	
d_13_rapportoqp		0,519	0,411
soddisfazione_marca		0,679	
d_16_qualitaprodotto	0,359	0,686	
d_13_sostenibilita	0,502		0,662
d_13_brand		0,386	0,683
d_13_provenienza			0,694
d_13_prodotto bio			0,808
d_13_packaging			0,805



Importanza
caratteristiche ?

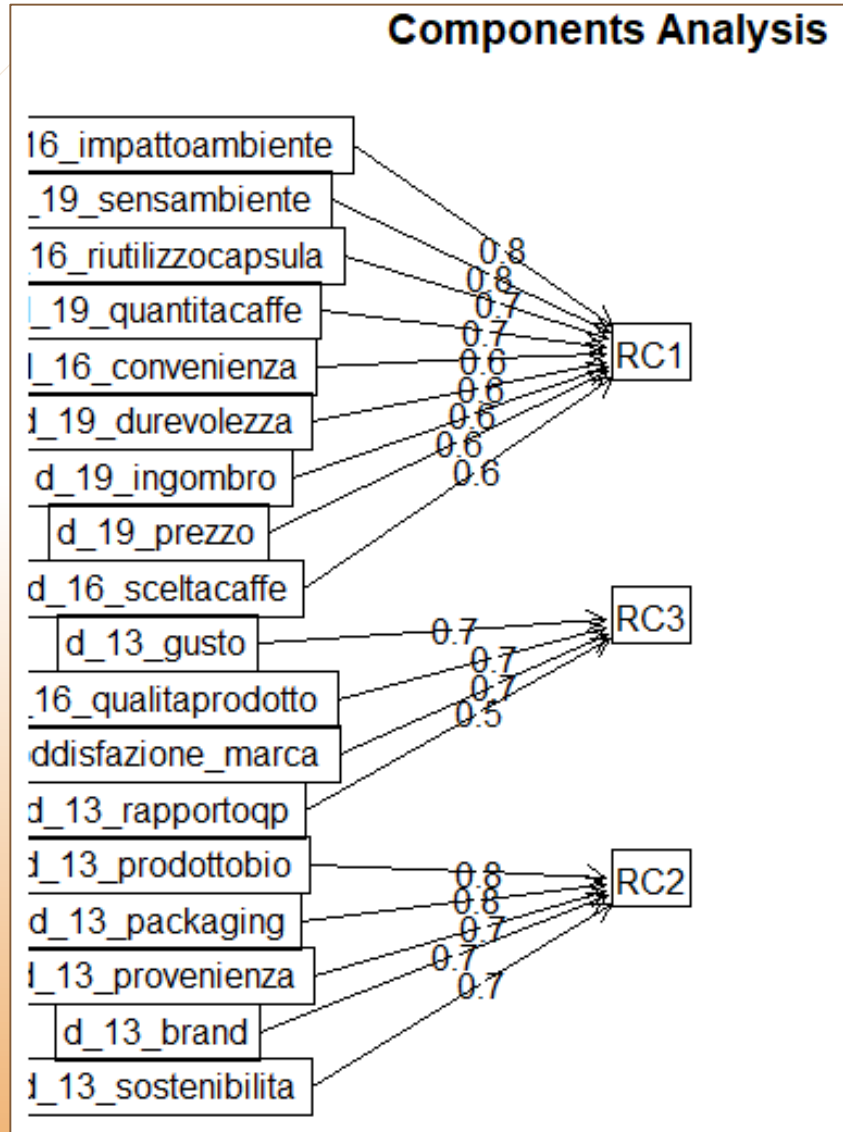


Qualità



?

Interpretabilità delle componenti



Analizziamo i 4 fattori che abbiamo precedentemente selezionato:

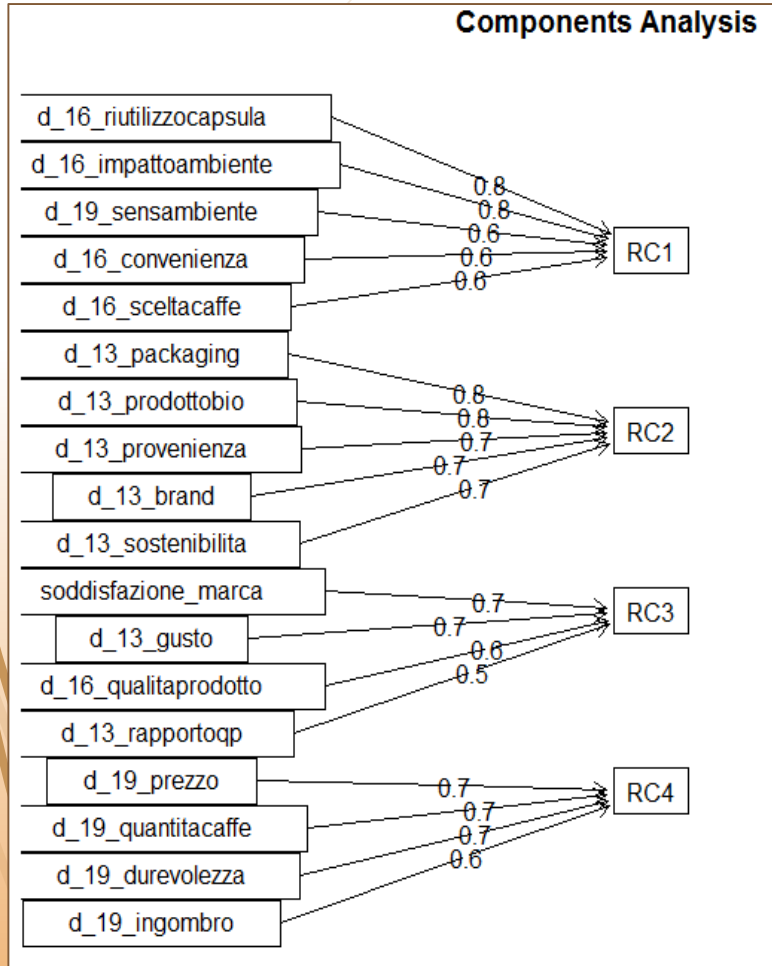
- ❖ Fattore 1 (RC1) = Importanza caratteristiche ?;
- ❖ Fattore 2 (RC2) = ?;
- ❖ Fattore 3 (RC3) = Qualità.

Con la seguente analisi fattoriale con soluzione a 3 fattori, non siamo stati in grado di interpretare in maniera corretta due gruppi.

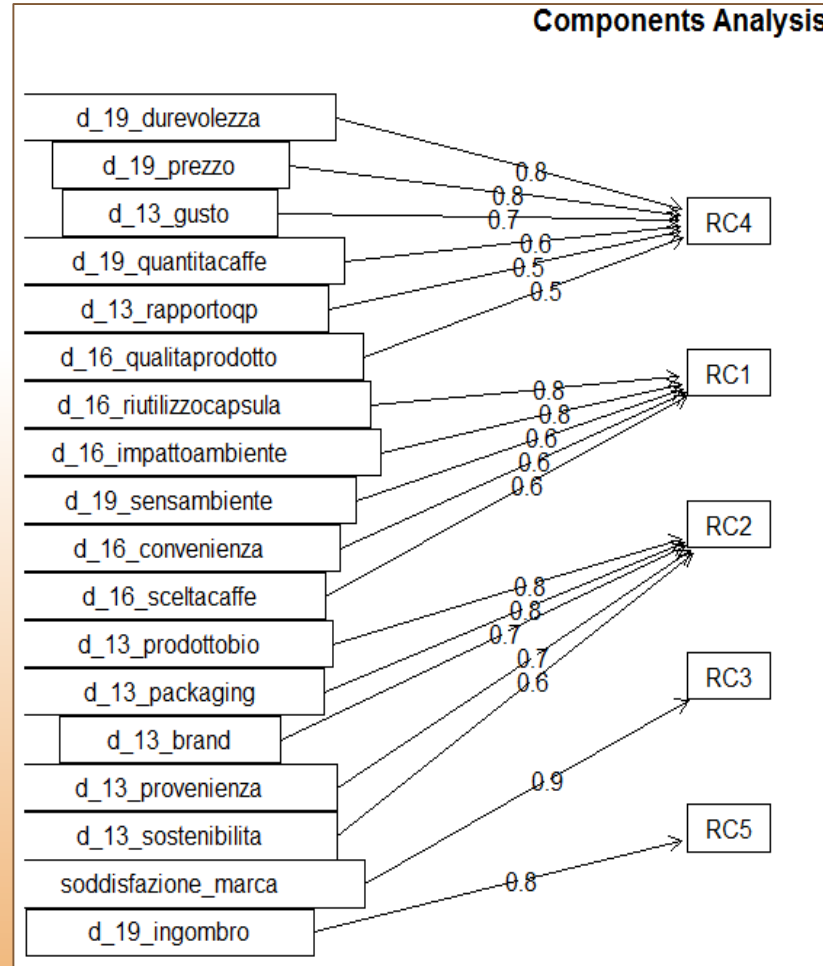
Per questo motivo proseguiamo ad analizzare le soluzioni a 4 e 5 fattori.

Soluzioni

Soluzione a 4 fattori



Soluzione a 5 fattori

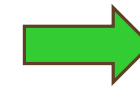


Grazie alla soluzione a 4 fattori siamo riusciti ad interpretare in maniera più consona i fattori, rispetto alla soluzione a 3 o a 5 fattori.

A questo punto procediamo con il processo di rotazione dei 4 fattori (**Varimax**).

Soluzioni

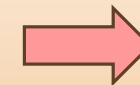
	RC1	RC2	RC3	RC4
d_16_scltcaffè	0,575		0,480	
d_16_convenienza	0,609		0,505	
d_16_riutilizzocapsula	0,818			
d_16_impattoambiente	0,811			
d_19_sensambiente	0,616			0,467
d_13_sostenibilita	0,516	0,658		
d_13_brand		0,680	0,381	
d_13_provenienza		0,700		
d_13_prodotto bio		0,802		
d_13_packaging		0,806		
d_13_gusto			0,693	0,396
d_13_rapporto qp		0,401	0,507	
soddisfazione_marca			0,710	
d_16_qualitaprodotto	0,316		0,643	
d_19_prezzo			0,401	0,707
d_19_durevolezza	0,304		0,462	0,667
d_19_quantitacaffè	0,359			0,672
d_19_ingombro		0,342		0,643



Ambiente



Caratteristiche



Qualità



Vantaggi

Conclusioni

L'obiettivo dell'esecuzione dell'analisi fattoriale era la sintesi delle variabili in modo tale da evitare successivamente problemi di multicollinearità nelle analisi di regressione (lineare e logistica).

Come variabili sono state prese in considerazione scale di punteggio facenti parte della sezione dati comportamentali del questionario d'indagine.

Tramite il metodo delle componenti principali abbiamo determinato che le soluzioni possibili erano a 3, 4 o 5 fattori. Tra queste, abbiamo poi deciso di utilizzare la soluzione a 4 fattori per spiegare il nostro modello, andando quindi ad scartare le altre soluzioni (3 e 5 fattori) per difficoltà interpretative.

Identificati i fattori latenti, li abbiamo successivamente rinominati in base alle variabili originarie spiegate.

I fattori selezionati sono:

- ❖ Ambiente → comprende gli aspetti principali che influenzano l'acquisto del prodotto in riferimento alla sostenibilità ambientale;
- ❖ Caratteristiche → evidenzia le caratteristiche basic del prodotto che lo differenziano da altri;
- ❖ Vantaggi → denota l'importanza che viene attribuita a tutte le caratteristiche che permettono ai consumatori di avere un vantaggio in termini di qualità, prezzo, soddisfazione e spazio;
- ❖ Qualità → indica la soddisfazione dei clienti per quanto riguarda la qualità.



REGRESSIONE LINEARE

Obiettivi della regressione lineare

Ora che abbiamo individuato i fattori che consideriamo utili a spiegare il modello di regressione lineare, siamo in grado di proseguire con questa analisi.

L'obiettivo è quello di spiegare la spesa media mensile utilizzando i fattori precedentemente determinati.

Abbiamo selezionato 18 variabili di partenza poiché volevamo comprendere le abitudini di consumo dei nostri intervistati, al fine di capire se spiegano la spesa media mensile.



Per eludere problemi di multicollinearità abbiamo eseguito l'analisi fattoriale, raggruppando le nostre 18 variabili in 4 fattori. I regressori che abbiamo scelto sono i seguenti:

- 1) Ambiente;
- 2) Vantaggi;
- 3) Caratteristiche;
- 4) Qualità.

Per raggiungere il nostro obiettivo ora andremo a svolgere la regressione lineare multipla.

Valutazione del modello

Per la valutazione del modello abbiamo utilizzato la funzione *linear model (lm)* in modo da valutare la bontà del modello sfruttando i seguenti indicatori:

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	45.725	1.766	25.897	< 2e-16 ***
vantaggi	-5.684	1.770	-3.211	0.00155 **
caratteristiche	8.228	1.770	4.649	6.15e-06 ***
ambiente	-1.450	1.770	-0.819	0.41367
qualita	-0.370	1.770	-0.209	0.83464

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.97 on 195 degrees of freedom
 Multiple R-squared: 0.1434, Adjusted R-squared: 0.1258
 F-statistic: 8.158 on 4 and 195 DF, p-value: 4.2e-06

Annotations:
 - Green box around the Pr(>|t|) column: Test t
 - Blue box around Multiple R-squared: R-quadro
 - Purple box around Adjusted R-squared: R-quadro adjusted
 - Red circle around p-value: Test F

- ❖ **Test F** → considerando H_0 come l'ipotesi in cui tutti i coefficienti sono pari a 0 e H_1 come l'ipotesi alternativa, secondo la quale almeno uno dei coefficienti sia diverso da zero, osserviamo che abbiamo un p-value molto basso, inferiore al livello di significatività $\alpha = 0,05$. Quindi rifiutiamo l'ipotesi nulla e procediamo con la nostra analisi;
- ❖ **R-quadro** → l'output ottenuto è di 0,1434 (spiega il 14,34% della spesa media mensile) ed è un valore molto basso, inferiore al valore soglia (0,2). Quindi, in base all'indice R-squared, il nostro modello ha una bassa capacità esplicativa;
- ❖ **R-quadro adjusted** → tale valore è pari al 12,58%, ed è inferiore a R-quadro, dato che tiene conto del numero dei regressori e dell'ampiezza campionaria;
- ❖ **Test t** → si nota che le uniche variabili impattanti sono i vantaggi e le caratteristiche.

Variabili Dummy

Considerando che i regressori selezionati spiegano poco, uniamo al modello alcune variabili socio-demografiche: sesso, età, provincia, nucleo familiare e titolo di studio. Riteniamo che queste variabili, da inserire nella nostra analisi, possano contribuire a migliorare la spiegazione della variabile dipendente senza condividere informazioni tra loro, e quindi migliorare la capacità esplicativa del modello.

Poiché alcune delle variabili socio-demografiche selezionate sono qualitative e non quantitative, saranno convertite in *variabili dummy*:

- ❖ Sesso → abbiamo attribuito il valore 0 all'opzione «Femmina» e il valore 1 all'opzione «Maschio»;
- ❖ Provincia → è stata suddivisa in 2 categorie in modo da non avere una variabile qualitativa ma una qualitativa ordinale. Alla prima categoria «Sud» (MI, VA, LO, CR, MN e PV) viene attribuito il valore 0, mentre alla seconda categoria «Nord» (MB, LC, SO, CO, BS e BG) viene attribuito il valore 1;
- ❖ Titolo di studio → abbiamo raggruppato in 2 classi le opzioni presenti nel questionario. Alla categoria «Senza laurea» è stato associato il valore 0 e comprende: licenza elementare, licenza media e diploma. Alla categoria «Laurea» è stato associato il valore 1 e comprende laurea triennale, laurea magistrale e post laurea.

Valutazione del nuovo modello

Dopo aver inserito le variabili dummy, svolgeremo nuovamente l'analisi di regressione lineare.

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.21058   8.66228   1.641  0.10255
vantaggi     -5.21044   1.74949  -2.978  0.00328 **
caratteristiche  7.93403   1.79854   4.411  1.72e-05 ***
ambiente     -1.45188   1.75952  -0.825  0.41032
qualita      -0.11681   1.75901  -0.066  0.94712
sex          4.67465   3.48106   1.343  0.18091
prov         0.03363   3.55687   0.009  0.99247
eta          0.38880   0.14984   2.595  0.01020 *
nucleofam    3.00878   1.34106   2.244  0.02601 *
titstudio    5.09693   3.56455   1.430  0.15439
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.38 on 190 degrees of freedom
Multiple R-squared:  0.2043,    Adjusted R-squared:  0.1666
F-statistic: 5.419 on 9 and 190 DF, p-value: 1.302e-06
  
```

R-quadro



Test t



R-quadro adjusted

Test F



Valutazione del nuovo modello

Valutiamo la bontà del nuovo modello sfruttando i seguenti indicatori:

- ❖ **Test F** → osserviamo che abbiamo un p-value molto basso, inferiore al livello di significatività. Perciò rifiutiamo l'ipotesi nulla e procediamo con la nostra analisi;
- ❖ **R-quadro** → l'output fornitoci è di 0,2043 (spiega l'20,43% della spesa media mensile) ed è un valore accettabile, appena superiore al valore soglia (0,2). Si può di conseguenza affermare che il nuovo modello ha capacità esplicativa sufficientemente buona;
- ❖ **R2adj** → dal momento che stiamo valutando molteplici regressori, procediamo ad esaminare anche il valore di R-quadro adjusted; tale valore è pari al 16,66%, che si dimostra superiore all'indice R-quadro adjusted del modello precedente (12,58%). Di conseguenza optiamo di prendere in considerazione questo modello per portare avanti l'analisi;
- ❖ **Test t** → si nota che le variabili che impattano maggiormente sono i vantaggi, le caratteristiche. Altri fattori che impattano in maniera minore sulla spesa sono l'età e il nucleo familiare.

Metodo Stepwise

Per trovare il sottoinsieme di variabili ottimo tra quelli possibili usiamo il Metodo Stepwise.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   21.2947    7.6154   2.796  0.00569 **
vantaggi      -5.3468    1.7284  -3.093  0.00227 **
caratteristiche  7.6953    1.7764   4.332 2.36e-05 ***
eta           0.3419    0.1468   2.329  0.02090 *
nucleofam     2.9727    1.3071   2.274  0.02403 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.34 on 195 degrees of freedom
Multiple R-squared:  0.186,    Adjusted R-squared:  0.1693
F-statistic: 11.14 on 4 and 195 DF, p-value: 3.681e-08
  
```

- ❖ **Test F** → abbiamo un p-value molto basso, inferiore $\alpha = 0,05$. Quindi rifiutiamo l'ipotesi nulla, quindi almeno uno dei coefficienti è diverso da zero, e procediamo con la nostra analisi;
- ❖ **R-quadro** → il valore ottenuto è di 0,186 (spiega il 18,6% della spesa media settimanale) ed è un valore al limite dell'accettabilità, per cui il modello ha una capacità esplicativa accettabile;
- ❖ **R-quadro adjusted** → tale valore è pari al 16,93%, ed è superiore al precedente indice;
- ❖ **Test t** → il p-value relativo ad ogni regressore risulta inferiore al livello di significatività prefissato. Per ogni test t si può rifiutare l'ipotesi nulla e affermare che i regressori che impattano maggiormente sono caratteristiche e vantaggi. Altri fattori che impattano in maniera minore sulla spesa sono età e nucleo familiare.

Calcolo della vif

Verifichiamo ora la presenza di multicollinearità: la funzione che utilizziamo è la *Variance Inflation Factors* .

Dato che abbiamo aggiunto le variabili età e nucleo familiare che, non essendo state sottoposte all'analisi fattoriale, potrebbero avere dei problemi di correlazione con altre variabili, prima di analizzare l'output ottenuto calcoliamo la vif per valutare se esistono problemi di **multicollinearità**.

Se i valori sono superiori a 1,2/1,3 è presente un problema di multicollinearità.

```
> vif(caps_vif)
  Variables      VIF
1   vantaggi 1.003482
2 caratteristiche 1.059923
3          eta 1.053132
4   nucleofam 1.011367
```

Nel nostro caso, dalla vif possiamo vedere che i valori sono prossimi a 1, perciò non siamo in presenza di un problema di multicollinearità.

Nella prossima slide verranno analizzati l'impatto dei regressori e saranno interpretati i coefficienti standardizzati.

Interpretazione dei coefficienti

```
> lm.beta(p)
      vantaggi caratteristiche          eta      nucleofam
-0.2002094    0.2881481    0.1543948    0.1477724
```

Come era prevedibile, data $Y =$ spesa media mensile, il regressore più importante (relativamente agli altri) è **caratteristiche**, ovvero tutte le caratteristiche che sono proprie del prodotto innovativo: all'aumentare della singola unità all'interno di «caratteristiche», la spesa mensile per l'acquisto di capsule aumenta del 28,8%.

All'aumentare della singola unità all'interno di «vantaggi», fissate le altre variabili, la spesa mensile per l'acquisto di capsule diminuisce del 20%.

All'aumentare della singola unità all'interno di «età», fissate le altre variabili, la spesa mensile per l'acquisto di capsule aumenta del 15,4%.

All'aumentare della singola unità all'interno di «nucleofam», fissate le altre variabili, la spesa mensile per l'acquisto di capsule aumenta del 14,7%.

Conclusioni

Attraverso il modello di regressione lineare si è voluto indagare su come la spesa media mensile (variabile dipendente) venisse influenzata dalle variabili indipendenti scelte.

Il primo passo è stato quello di condurre un'analisi fattoriale tra le variabili per risolvere, alla base, il problema di multicollinearità. Quest'analisi ha consentito di ridurre le nostre 18 variabili, scelte in partenza, in 4 fattori finali.

Dopodiché abbiamo proceduto con l'analisi di regressione lineare multipla.

Con la valutazione della bontà del modello attraverso i vari indicatori (R^2 , R^2_{adj} , test F e test T) abbiamo ottenuto un modello che non era in grado di spiegare in maniera adeguata la spesa media mensile.

Per questo motivo abbiamo aggiunto il sesso, l'età, la provincia, il titolo di studio e il nucleo familiare come regressori nel modello.

Valutata la bontà del nuovo modello tramite gli indicatori R^2 , R^2_{adj} , test F e test T, fornendo tutti riscontri positivi, abbiamo controllato se fossero presenti problemi di multicollinearità, non avendo sottoposto all'analisi fattoriale le variabili aggiunte in secondo luogo.

Tramite la Vif, è stata evidenziata assenza di multicollinearità e siamo passati all'interpretazione dei coefficienti.

Conclusioni

Da questa analisi è emerso che:

- ❖ È possibile spiegare la spesa media settimanale attraverso le caratteristiche base ricercate dai consumatori;
- ❖ Sia l'età che il nucleo familiare sono fattori importanti nel determinare la spesa media mensile, infatti, quando sono stati inseriti nel modello, si è verificato un incremento della capacità esplicativa del modello;
- ❖ La spesa media mensile, come era prevedibile, è influenzata dalle caratteristiche proprie del prodotto innovativo;
- ❖ La spesa media mensile, diversamente da quanto potevamo prevedere, è condizionata negativamente dai vantaggi che derivano dal prodotto in termini di prezzo, qualità, sostenibilità e spazio;
- ❖ La spesa media mensile è influenzata positivamente dall'età e dal nucleo familiare.

Perciò, a livello di business, sarà importante:

- enfatizzare le caratteristiche e i punti di forza del prodotto attraverso la pubblicità;
- porre l'attenzione sull'offerta indirizzata a età differenti;
- far intuire nella nostra offerta che un nucleo familiare più ampio otterrebbe un risparmio, in termini economici, maggiore dall'acquisto della capsula Ecocapsula.



REGRESSIONE LOGISTICA

Obiettivi della regressione logistica

Attraverso l'analisi di regressione logistica andremo ad analizzare quanto determinate variabili influiscono sulla disponibilità all'acquisto della capsula Ecocapsula.

L'obiettivo del modello di regressione logistica è quello di prevedere la probabilità che i rispondenti del questionario acquistino la capsula Ecocapsula, partendo da un insieme di variabili indipendenti e sarà rilevante comprendere come i regressori influenzano l'esito della variabile dipendente dicotomica, ovvero la scelta del consumatore.



Adottiamo come variabile dipendente (Y) la disponibilità del campione ad acquistare Ecocapsula, associando il valore 1 alla risposta positiva e il valore 0 alla risposta negativa.

Scelta delle variabili

Dall'analisi fattoriale precedentemente svolta, prendiamo i seguenti regressori: ambiente, vantaggi, caratteristiche e qualità. A questi regressori aggiungiamo le variabili socio-demografiche, età e nucleo familiare, e le variabili dummy, sesso e provincia. Per quanto riguarda le variabili attitudinali andiamo a evidenziare quanta importanza hanno i social nelle scelte d'acquisto, quanto è ritenuto importante il rispetto dell'ambiente, cosa significa essere un consumatore sostenibile e quanto influenza esercita il luogo dove si consuma il caffè. Per fare questa analisi abbiamo utilizzato la funzione *glm(mylogit)*.

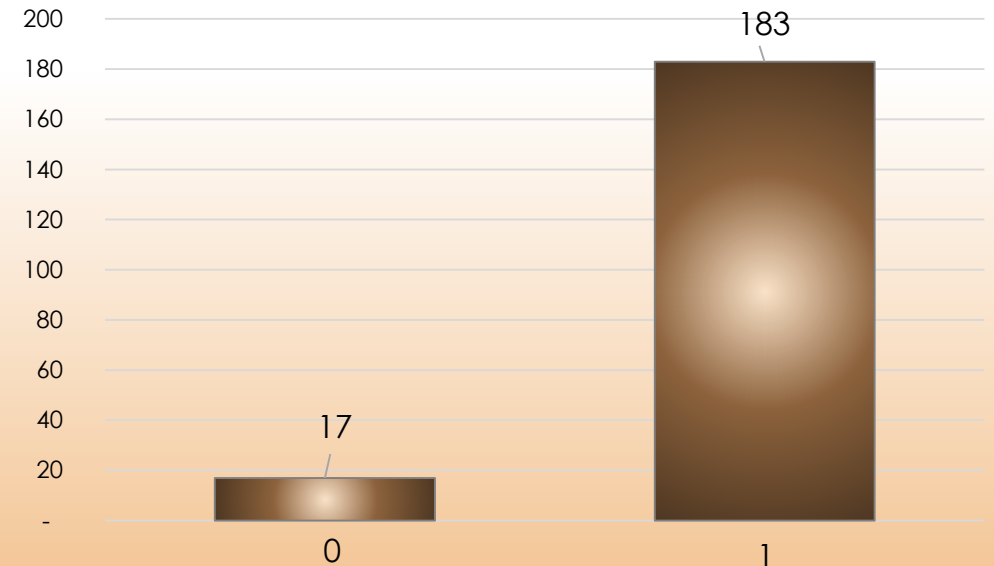
#	VARIABILE	DESCRIZIONE VARIABILE
1	Ambiente	Aspetti che influenzano l'acquisto del prodotto per alla sostenibilità ambientale
2	Vantaggi	Caratteristiche che permettono ai consumatori di avere un vantaggio in termini di qualità, prezzo, soddisfazione e spazio
3	Caratteristiche	Caratteristiche basic del prodotto che lo differenziano da altri
4	Qualità	Soddisfazione dei clienti per quanto riguarda la qualità
5	Sesso	Sesso del campione
6	Età	Età del campione

#	VARIABILE	DESCRIZIONE VARIABILE
7	Provincia	Provincia di residenza del campione
8	Nucleofam	Numero componenti del nucleo familiare del campione
9	Influenza_social	Influenza dei social sulle scelte d'acquisto
10	d_7_casa	Di solito beve il caffè «A casa»
11	d_7_scuolauniversità	Di solito beve il caffè «A scuola/università»
12	d_7_lavoro	Di solito beve il caffè «Al lavoro»
13	Rispetto_ambiente	Propensione al rispetto ambientale
14	d_26_bassoimpatto	Consumatore sostenibile significa «Comprare prodotti con un basso impatto ambientale»

Disponibilità all'acquisto di Ecocapsula

Prima di effettuare la regressione logistica, abbiamo verificato quale fosse la percentuale di intervistati che sarebbe stata favorevole all'acquisto di Ecocapsula. La nostra variabile dipendente assume valore 1 nel caso in cui i soggetti siano disposti ad acquistare una capsula Ecocapsula, assume valore 0 nel caso contrario. Tramite la funzione *freq* abbiamo individuato che la frequenza della modalità 1 è pari al 91,5% e quindi supera la soglia del 5% necessaria per proseguire per l'analisi.

```
> freq(capsule$wtp_ecocap)
capsule$wtp_ecocap
  Frequency Percent
0           17     8.5
1          183    91.5
Total       200   100.0
```



Il risultato ottenuto è un riscontro positivo per la nostra analisi poiché conferma, come viene riportato dalla descrizione iniziale del mercato, che l'acquisto di capsule ricaricabili e riutilizzabili sia in aumento tra i consumatori.

Valutazione della bontà del modello

A questo punto procediamo alla valutazione della bontà del modello, e per fare ciò ricorriamo a tre test:



- Wald Test o Test di significatività congiunta dei coefficienti → valuta la capacità esplicativa del modello (è equivalente al test F nella regressione lineare);
- Percentuale di Concordant → valuta la capacità del modello di stimare la probabilità che il fenomeno si verifichi (quanto più la percentuale è alta tanto migliore è il modello);
- Wald Chi squared test o Test di significatività dei singoli coefficienti → valuta la significatività dei singoli coefficienti, ossia la rilevanza dei corrispondenti regressori nella spiegazione della variabile dipendente (è equivalente al test t nella regressione lineare).

Wald test

```
Wald test

Model 1: wtp_ecocap ~ vantaggi + caratteristiche + ambiente + qualita +
sex + prov + nucleofam + eta + influenza_social + d_7_casa +
d_7_lavoro + d_7_scuolauniversita + rispetto_ambiente + d_26_bassoimpatto
Model 2: wtp_ecocap ~ 1
  Res.Df  Df      F  Pr(>F)
1     185    .    .    .
2     199 -14  1.5683 0.09158 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Per valutare la bontà del modello di regressione logistica usiamo il Wald test. Ponendo H_0 come ipotesi in cui tutti i coefficienti sono pari a 0 e H_1 come ipotesi alternativa, dall'output del Wald test otteniamo un p-value pari a 0,09158 minore del livello di significatività $\alpha = 0,1$.

Rifiutiamo l'ipotesi nulla, quindi esiste almeno un coefficiente diverso da zero e di conseguenza possiamo affermare che il modello ha capacità esplicativa (seppur bassa).

Percentuale di Concordant

```
> CalculateConcordance(mylogit)
$Concordance
[1] 0.8949

$Discordance
[1] 0.1051

$Tied
[1] 0

$Pairs
[1] 3111
```

Proseguiamo a calcolare la percentuale di Concordant.

La percentuale di Concordant è pari a 89,49%: questo è sicuramente un buon esito, il quale supera la soglia per decretare la validità del modello del 70%.

Possiamo affermare che il modello stimato a partire dai fattori estratti è soddisfacente in quanto ha buona capacità esplicativa.

Wald Chi Squared test

Con il test Wald Chi Squared procediamo a valutare quali sono i coefficienti significativi.

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept)   9.16237   3.23818   2.829  0.00466 **
vantaggi     -0.14551   0.28552  -0.510  0.61032
caratteristiche  0.12520   0.36790   0.340  0.73363
ambiente      1.64425   0.40691   4.041  5.33e-05 ***
qualita       0.24955   0.28923   0.863  0.38826
sex          -0.18498   0.64265  -0.288  0.77347
prov         0.66117   0.74960   0.882  0.37776
nucleofam    -0.51874   0.24667  -2.103  0.03547 *
eta          0.01556   0.03596   0.433  0.66518
influenza_social  0.19148   0.14958   1.280  0.20049
d_7_casa     -0.01455   0.01802  -0.808  0.41929
d_7_lavoro   -0.03947   0.01981  -1.993  0.04629 *
d_7_scuolauniversita -0.03935  0.02280  -1.726  0.08438 .
rispetto_ambiente -0.19528  0.18068  -1.081  0.27977
d_26_bassoimpatto -0.27347  0.18503  -1.478  0.13941
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  
```

Dall'output di R si nota che i regressori che meglio spiegano il modello sono: ambiente, nucleofam, d_7_lavoro. Questi regressori spiegano più adeguatamente il modello perché p-value dei singoli coefficienti sono minori del livello di significatività (0,05).

Stima del nuovo modello con variabili impattanti

Siccome non tutti i coefficienti sono rilevanti ai fini della spiegazione della nostra y , andiamo ad utilizzare il metodo della stepwise in modo da trovare il sottoinsieme dei regressori realmente impattanti.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    7.41033    2.10692   3.517 0.000436 ***
ambiente       1.37562    0.33810   4.069 4.73e-05 ***
nucleofam     -0.44646    0.21662  -2.061 0.039301 *
influenza_social  0.17112    0.12043   1.421 0.155345
d_7_lavoro    -0.03127    0.01436  -2.177 0.029478 *
d_7_scuolauniversita -0.03372    0.01470  -2.294 0.021803 *
d_26_bassoimpatto -0.29386    0.15638  -1.879 0.060223 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  
```

L'output ci ha lasciato i regressori più significativi, creando un nuovo modello con 6 regressori. Ora andremo a valutare la bontà del nuovo modello.

Wald test

```

Wald test

Model 1: wtp_ecocap ~ ambiente + nucleofam + influenza_social + d_7_lavoro +
          d_7_scuolauniversita + d_26_bassoimpatto
Model 2: wtp_ecocap ~ 1
  Res.Df Df      F    Pr(>F)
1     193
2     199 -6 3.7638 0.001445 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  
```

Per valutare la bontà del modello di regressione logistica usiamo il Wald test. Ponendo H_0 come ipotesi in cui tutti i coefficienti sono pari a 0 e H_1 come ipotesi alternativa, dall'output del Wald test otteniamo un p-value pari a 0,001445 minore del livello di significatività $\alpha = 0,05$.

Rifiutiamo l'ipotesi nulla, quindi esiste almeno un coefficiente diverso da zero e di conseguenza possiamo affermare che il modello ha una buona capacità esplicativa.

Percentuale di Concordant

```
> CalculateConcordance(a)
$Concordance
[1] 0.8811

$Discordance
[1] 0.1189

$Tied
[1] 0

$Pairs
[1] 3111
```

Proseguiamo a calcolare la percentuale di Concordant.

La percentuale di Concordant è pari a 88,11%: questo è un esito molto buono, che supera la soglia per decretare la validità del modello del 70%.

Possiamo affermare che il modello stimato a partire dai fattori estratti è soddisfacente in quanto ha buona capacità esplicativa.

Wald Chi Squared test

Per valutare quali sono i coefficienti rilevanti utilizziamo il test Wald Chi Squared. I regressori che meglio spiegano il modello sono: ambiente, nucleofam, d_7_lavoro e d_7_scuolauniversità.

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	7.41033	2.10692	3.517	0.000436	***
ambiente	1.37562	0.33810	4.069	4.73e-05	***
nucleofam	-0.44646	0.21662	-2.061	0.039301	*
influenza_social	0.17112	0.12043	1.421	0.155345	
d_7_lavoro	-0.03127	0.01436	-2.177	0.029478	*
d_7_scuolauniversita	-0.03372	0.01470	-2.294	0.021803	*
d_26_bassoimpatto	-0.29386	0.15638	-1.879	0.060223	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Sebbene l'utilizzo della stepwise, si può notare che sono rimasti alcuni coefficienti non molto rilevanti ai fini della spiegazione della disponibilità di acquisto di Ecocapsula. Questo probabilmente perché ai fini della valutazione globale del modello sia influenza_social sia d_26_bassoimpatto sono utili.

Calcolo della Vif

Con la funzione VIF andiamo a valutare la presenza di multicollinearità tra i regressori che sono risultati significativi.

```
> vif(caps2_vif)
      Variables      VIF
1      ambiente 1.192757
2 influenza_social 1.048972
3      d_7_lavoro 1.169159
4 d_7_scuolauniversita 1.160761
5      d_26_bassoimpatto 1.151881
```

Si può notare che tutti i valori del VIF sono inferiori alla soglia 1,2/1,3 e quindi vi è assenza di multicollinearità tra i regressori.

I regressori del nostro modello non sono fortemente correlati tra loro.

Successivamente è possibile procedere con la spiegazione del modello, con l'analisi dei regressori esplicativi, tramite l'interpretazione della stima standardizzata.

Interpretazione dei coefficienti

Per stabilire le variabili più importanti si utilizzano i parametri standardizzati, ordinandoli in valore assoluto dal più grande al più piccolo, attraverso la funzione `lm.beta`.

```
> lm.beta(a)
  ambiente      nucleofam influenza_social
  4.920293     -2.119979         1.649614
d_7_lavoro d_7_scuolauniversita d_26_bassoimpatto
-2.679459     -2.297396         -2.161248
```

Considerando il valore assoluto dei coefficienti standardizzati possiamo osservare che la variabile `ambiente` è l'informazione più rilevante tra i regressori selezionati. Quindi la scelta del caffè da inserire, la convenienza, la possibilità di riutilizzare la capsula, l'impatto ambientale e la sensibilità ambientale hanno un'influenza molto elevata sulla scelta di acquisto del prodotto.

Nel nostro negozio sarà fondamentale sottolineare queste caratteristiche peculiari che garantiscano una soddisfazione delle esigenze dei consumatori.

Interpretazione dei coefficienti

```
> exp(a$coefficient)
      (Intercept)          ambiente          nucleofam
      1652.9787692          3.9575418          0.6398868
influenza_social      d_7_lavoro d_7_scuolauniversita
      1.1866367           0.9692124           0.9668383
d_26_bassoimpatto
      0.7453845
```

- ❖ Al crescere del regressore «ambiente», la probabilità che un consumatore acquisti una capsula ricaricabile e riutilizzabile Ecocapsula aumenta di circa il 295%;
- ❖ Al crescere del numero dei componenti del nucleo familiare, la probabilità che un consumatore acquisti una capsula Ecocapsula diminuisce di circa il 36%;
- ❖ Al crescere dell'influenza che i social networks hanno sulle scelte d'acquisto dei consumatori, la probabilità che uno di essi acquisti una capsula ricaricabile e riutilizzabile Ecocapsula aumenta di circa il 19%;
- ❖ Al crescere del consumo di caffè sul luogo di lavoro, la probabilità che un consumatore acquisti una capsula Ecocapsula diminuisce di circa il 3%;
- ❖ Al crescere del consumo di caffè nella scuola o in università, la probabilità che un consumatore acquisti una capsula Ecocapsula diminuisce di circa il 3,5%;
- ❖ Al crescere della variabile secondo cui consumatore sostenibile significhi comprare prodotti a basso impatto ambientale, la probabilità che un consumatore acquisti una capsula ricaricabile e riutilizzabile Ecocapsula diminuisce di circa il 25%.

Interpretazione dei coefficienti

Così come ci aspettavamo, il regressore «ambiente» (che comprende gli aspetti principali che influenzano l'acquisto del prodotto in riferimento alla sostenibilità ambientale) aumenta in maniera netta e sostanziale la probabilità di acquistare una capsula ricaricabile e riutilizzabile Ecocapsula.

Mentre in riferimento alla variabile «d_26_basso impatto» si verifica una certa ambiguità: i rispondenti, nonostante in precedenti analisi siano risultati molto sensibili al tema della sostenibilità ambientale, non ritengono che «consumatore sostenibile» significhi comprare prodotti con un basso impatto ambientale, ma piuttosto non comprare prodotti usa e getta, non comprare da aziende che hanno comportamenti dannosi in ambito sociale e limitare gli acquisti ai soli prodotti necessari.

Infine possiamo dedurre, come visto già in precedenza, che è controproducente rivolgersi a nuclei familiari estesi e soprattutto la nostra offerta deve essere rivolta a consumatori che non amano e non bevono il caffè nel luogo dove lavorano o studiano. Non bisogna dimenticare, inoltre, di avvicinarci al nostro target, non solamente attraverso il marketing tradizionale ma anche tramite i canali digital e social, in quanto essi hanno una discreta influenza sulle scelte di acquisto del campione.

Conclusioni

Per valutare la possibilità di vendita futura della capsula ricaricabile Ecocapsula, il nostro team ha sviluppato un'analisi che ci permette di analizzare quali variabili influenzano la disponibilità all'acquisto del nostro campione.

Da questa analisi di regressione logistica si può affermare che i fattori che influenzano, in maniera significativa, la disponibilità all'acquisto sono:

- ❖ la variabile ambiente → comprende gli aspetti principali che influenzano l'acquisto del prodotto in riferimento alla sostenibilità ambientale;
- ❖ le variabili lavoro e scuola/università → sono riferite al luogo dove si consuma abitualmente il caffè, in questo caso al lavoro e a scuola/in università;
- ❖ la variabile d_26_bassoimpatto → indica come gli intervistati considerino l'essere un consumatore sostenibile il comprare prodotti con un basso impatto ambientale;
- ❖ La variabile influenza_social → si riferisce all'influenza che hanno i social sulle scelte d'acquisto dei consumatori;
- ❖ La variabile nucleofam → rappresenta il numero dei componenti del nucleo familiare.

Nel proporre la nostra capsula dovremmo prestare attenzione ai fattori ambientali, infatti il profilo del nostro consumatore è quello di una persona attenta agli aspetti ambientali e che non consuma caffè né al lavoro o in università, in quanto Ecocapsula in questi luoghi pecca di praticità.



CONCLUSIONI BUSINESS CASE

Conclusioni

L'obiettivo della nostra analisi consiste nell'individuazione della provincia lombarda che riscontra maggiore interesse nei confronti del nostro prodotto Ecocapsula, e la successiva attivazione di un nuovo punto vendita aziendale.

Nelle prime analisi effettuate sono stati riscontrati molteplici risultati positivi. Nel nostro campione, composto da 200 persone (tutti possessori di macchine di caffè per capsule), il 91,5% risulta propenso ad acquistare una capsula ricaricabile e riutilizzabile Ecocapsula; inoltre, considerando che solamente il 39,5% dei rispondenti era a conoscenza dell'esistenza di questa tipologia di capsule, è facilmente intuibile come questo prodotto possa, in termini di volumi di vendita, inserirsi nel mercato in tempi molto brevi.

La sostenibilità dovrà essere un aspetto fondamentale per le caratteristiche del nostro prodotto poiché il campione risulta molto sensibile all'argomento: il 79,5% dei consumatori ha attribuito alla propensione per il rispetto dell'ambiente un valore di almeno 6, su una scala da 1 a 10. Questo dato ottenuto può risultare importante nella proposta del prodotto, poiché sottolineando gli effetti positivi che nel suo piccolo la nostra capsula può avere sull'ambiente. In questo modo si riesce non solo ad avvicinarsi al target di riferimento, ma anche a proporsi ai prospect come un'azienda ecosostenibile.

Conclusioni

Attraverso l'analisi bivariata siamo riusciti a notare che le variabili provincia e disponibilità all'acquisto della capsula Ecocapsula non sono correlate, in quanto oltre il 90% del campione è propenso all'acquisto, per cui è possibile desumere che non è importante in quale provincia avviare il punto vendita ma riteniamo che sia importante aprirlo in prossimità dei grandi centri in modo da favorire sia la conoscenza del prodotto e sia la facilità di reperimento del suddetto.

Avendo analizzato il mercato dei competitor e avendo individuato in Waycap il principale concorrente, in quanto il più conosciuto dal campione (29%), abbiamo messo in connessione la conoscenza della marca Waycap e la disponibilità ad acquistare Ecocapsula. Da questa analisi è emerso che, nonostante il competitor sia conosciuto in misura notevole, il nostro prodotto potrebbe riscontrare ugualmente successo siccome non vi è connessione tra le due variabili selezionate.

Sempre dall'analisi bivariata svolta, è emerso che per differenziare la nostra offerta bisogna puntare non solo sulla sostenibilità, ma anche sulla provenienza made in Italy del prodotto, in quanto vi è una correlazione positiva tra le due variabili considerate.

Conclusioni

Successivamente, procedendo all'analisi di regressione lineare individuiamo che i regressori che spiegano la spesa media mensile in modo più rilevante sono le caratteristiche, il nucleo familiare e l'età. Questi regressori ci indicano come sostenibilità, brand, provenienza made in Italy, packaging del prodotto, l'età del campione e numero dei componenti del nucleo familiare siano elementi fondamentali che porterebbero a un incremento della spesa mensile delle capsule. A livello di business sarà quindi importante enfatizzare le caratteristiche e i punti di forza del prodotto e porre attenzione alle offerte indirizzate a specifici nuclei familiari.

Invece vi è ambiguità per quanto riguarda il regressore vantaggi, il quale denota l'importanza che viene attribuita a tutte le caratteristiche che permettono ai consumatori di avere un vantaggio in termini di qualità, prezzo, soddisfazione e spazio. Dall'analisi risulta che all'aumentare dei vantaggi, la spesa media mensile per l'acquisto di capsule diminuisce e questa correlazione negativa va contro le nostre aspettative: il nostro team aveva previsto inizialmente che, grazie alla peculiarità della capsula (la durevolezza, la possibilità di inserire la quantità desiderata di caffè nella capsula, il minore spazio di ingombro e la convenienza in termini di prezzo), il nostro prodotto potesse allo stesso tempo differenziarsi e andare a soddisfare le esigenze dei consumatori, che in realtà hanno più a cuore gli aspetti ambientali del prodotto.

Conclusioni

Passando, invece, alla regressione logistica il nostro team ha voluto valutare i fattori che influiscono sulla disponibilità all'acquisto della nostra capsula ricaricabile e riutilizzabile Ecocapsula. Dopo la costruzione del modello che ha una buona capacità esplicativa, gli elementi che si sono dimostrati rilevanti sono l'ambiente e il consumo di caffè a scuola/università e al lavoro: il primo in termini positivi, in quanto l'attenzione all'ambiente da parte dell'azienda e quindi anche da parte del prodotto condiziona positivamente la propensione all'acquisto; mentre gli altri due elementi in maniera negativa, come era prevedibile, in quanto chi consuma principalmente caffè fuori casa non beneficia in maniera esaustiva delle caratteristiche della capsula. Questi elementi ci danno informazioni preziose al fine di indirizzare le nostre campagne di marketing, rivolgendoci in primo luogo ai cosiddetti «consumatori sostenibili», per poi prestare attenzione su un target più ampio.

In conclusione, dai vari risultati che abbiamo ottenuto tramite le nostre analisi, possiamo affermare che nonostante sia abbastanza indifferente in quale provincia lombarda avviare il nostro punto vendita, quest'ultimo ha tutte le carte in regola per avere successo e svilupparsi sempre più.

«Come con arte va preparato, così con arte va bevuto.»