

UNIVERSITÀ DEGLI STUDI DI  
MILANO-BICOCCA

FOUNDATIONS OF PROBABILITY AND STATISTICS  
PROGETTO FINALE

---

# Analisi Viaggi e Vacanze, Residenti in Italia 2018

---

*Autore:*

Stefano Biondi - 839358 - s.biondi7@campus.unimib.it

September 7, 2020



## Abstract

In questo elaborato si è analizzato il viaggio dei residenti in Italia nel 2018. Il viaggio è un indicatore di salute della popolazione, per questo è interessante analizzare quanto viaggiano e quanto spendono in viaggio gli Italiani. Partendo da un campione statistico fornito dall'Istat, si sono identificate le medie nazionali di Durata e Spesa Media. Si sono inoltre analizzate le variabili descrittive del campione estraendo informazioni interessanti sulle abitudini degli Italiani, come ad esempio che il mezzo di trasporto maggiormente utilizzato è l'auto di proprietà. Ed informazioni utili per capire il mercato dei viaggi, come ad esempio che la spesa è significativamente diversa se organizziamo trasporto ed alloggio con un tour operator piuttosto che da soli.

## 1 Introduzione

“Viaggi e vacanze” è un focus dell'Istituto Nazionale di Statistica incluso nell'indagine sulle spese delle famiglie, per rilevare informazioni sui movimenti turistici dei residenti in Italia. L'importanza dell'analisi deriva dalla possibile comparazione a livello internazionale per analizzare l'evoluzione dei comportamenti turistici degli individui. Viaggiare è universalmente e scientificamente riconosciuto come una attività che migliora la salute mentale. In letteratura si possono trovare diversi studi a riguardo e concordano tutti che il viaggio apporta benefici alla persona come riduzione dello stress, miglioramento della capacità di problem solving, aumento della creatività, etc. In generale c'è un arricchimento del bagaglio culturale e dell'autostima. Quindi il viaggio può essere considerato un indicatore dello stato di salute della popolazione. Lo scopo dell'elaborato è, dopo una attenta e approfondita analisi delle variabili che compongono il dataset, stimare la spesa media e la durata media dei viaggi compiuti dai residenti in Italia tramite un'inferenza campionaria. Inoltre si è cercato di trovare un modello lineare che spiegasse queste due variabili. Il modello, nonostante sia risultato significativo, non è stato in grado di spiegare a sufficienza il fenomeno quindi l'analisi è conclusa con la dimostrazione, tramite l'anova ad una via, della differenza in media della Spesa nei gruppi delle variabili trovate più promettenti in fase di analisi grafica.

## 2 Dataset e Pre-Processing

Il dataset analizzato è il risultato dell'ultimo aggiornamento dell'indagine dell'Istat sulle spese delle famiglie con focus "Viaggi e Vacanze".

I dati sono reperibili al sito web <https://www.istat.it/it/archivio/178695>. Il dataset fa parte della categoria microdati, file ad uso pubblico.

Il campione utilizzato per le rilevazioni è stato progettato per fornire stime su diversi riferimenti temporali e territoriali. Infatti le stime si riferiscono sia all'intero territorio nazionale che alle regioni, che alle 5 grandi ripartizioni geografiche: Nord-Ovest, Nord-Est, Centro, Sud e Isole.

La base di campionamento adottata, ossia la lista di selezione delle unità campionarie, è costituita dalle LAC, un archivio unico delle famiglie residenti nei comuni italiani, proveniente dalle liste anagrafiche comunali. Il disegno campionario adottato è a due strati di selezione: comuni e famiglie. La necessità che il campione sia concentrato in un numero limitato di comuni è dovuto a ragioni di costo e organizzative. Le unità di primo stadio, i comuni, sono stratificati all'interno di ciascuna regione geografica secondo le due variabili tipologia comunale e dimensione demografica. Da ogni comune campione si estrae un campione di famiglie anagrafiche che costituiscono le unità finali di campionamento.

In particolare il dataset è composto da 4705 rilevazioni. Sono state eliminate le colonne con la presenza di valori NA e quelle che non portavano nessun tipo di informazione, come la colonna "anno di riferimento dei dati". Le proprietà rimanenti sono quindi le seguenti:

- **Tipo Record:** Categorica Nominale
  - 02: Record Individuale
  - 03: Record Viaggi
  - 04: Record Escursioni
- **Sesso:** Categorica Nominale
  - 1: Maschi
  - 2: Femmine
- **Classe di età:** Categorica Ordinale
  - 01:  $\leq 14$  anni

- 02: 15-24 anni
- 03: 25-34 anni
- 04: 35-44 anni
- 05: 45-54 anni
- 06: 55-64 anni
- 07: 65-74 anni
- 08:  $\geq 75$  anni
- **Stato Civile:** Categorica Nominale
  - 1: Celibe/Nubile
  - 2: Coniugato
  - 3: Separato/Divorziato
  - 4: Vedovo
- **Paese di Nascita:** Categorica Nominale
  - 1: Italia
  - 2: Estero
- **Titolo di studio piu alto conseguito:**Categorica Ordinale
  - 1: Nessun titolo/Licenza elementare
  - 2: Licenza di scuola media (o avviamento professionale)
  - 3: Diploma di scuola secondaria superiore o qualifica professionale
  - 4: Diploma universitario, ex scuole parauniversitarie, laurea, post laurea
- **Destinazione principale: italia/estero:** Categorica Nominale
  - 1: Italia
  - 2: Estero
- **Durata del Viaggio:** Quantitativa Discreta
- **Principale mezzo di Trasporto:** Categorica Nominale

- 1: Aereo
- 2: Treno
- 3: Nave, battello, motoscafo
- 4: Auto a noleggio
- 5: Auto propria, di parenti/amici
- 6: Pullman turistico
- 7: Pullman di linea
- 8: Camper, autocaravan
- 9: Moto, Motoscooter
- 10: Altro

• **Principale tipo di Alloggio:**Categorica Nominale

- 1: Albergo/motel/pensione
- 2: Residenza per cure fisiche ed estetiche
- 3: Campo lavoro e vacanza
- 4: Mezzo pubblico di trasporto
- 5: Centro congressi e conferenze
- 6: Villaggio vacanza
- 7: Campeggio
- 8: Marina (Barca in porto, darsena)
- 9: Istituto religioso
- 10: Altra struttura collettiva
- 11: Stanza in affitto
- 12: Abitazione in affitto
- 13: Abitazione in proprietà o multiprop.
- 14: Abitazione di parenti o amici gratuita
- 15: Barca in sito non organizzato
- 16: Altro tipo di sistemazione privata
- 17: Agriturismo

- 18: Bed and breakfast
- **Organizzazione Alloggio:** Categorica Nominale
  - 1: Prenot. diretta
  - 2: Prenot. in agenzia/tour operator
  - 3: Nessuna prenotazione
  - 4: Non sa/non risponde
- **Organizzazione Trasporto:** Categorica Nominale
  - 1: Prenot. diretta
  - 2: Prenot. in agenzia/tour operator
  - 3: Nessuna prenotazione
  - 4: Non sa/non risponde
- **Numero Partecipanti:** Categorica Ordinale
  - 1: 1
  - 2: 2
  - 3: 3
  - 4: 4
  - 5: 5 o più
- **Spesa Media del Viaggio:** Quantitativa Discreta

### 3 Analisi Univariata e Bivariata

L'approccio utilizzato è un'analisi esplorativa mediante gli indici statistici più comuni per quanto riguarda le variabili quantitative e mediante le frequenze assolute per le variabili categoriche. Infine sono stati rappresentati i grafici più esplicativi del fenomeno "Viaggi dei Residenti Italiani nel 2018".

DURATA	ESPE_CO
Min. : 1.000	Min. : 10.0
1st Qu.: 2.000	1st Qu.: 113.0
Median : 3.000	Median : 240.0
Mean : 5.361	Mean : 361.7
3rd Qu.: 7.000	3rd Qu.: 451.0
Max. : 90.000	Max. : 4400.0

Figure 1: Indici di posizione Variabili Quantitative

### 3.1 Univariata Variabili Quantitative

Iniziando dalle variabili quantitative possiamo osservare le loro distribuzioni mediante gli indicatori di posizione principali, figura 1.

Si può subito notare la presenza di outliers dato che la distanza tra il 3° quartile e il massimo sono rispettivamente quasi 14 e 9 volte l'ampiezza tra il minimo e il 3° quartile. Questo viene infatti confermato dal boxplot.

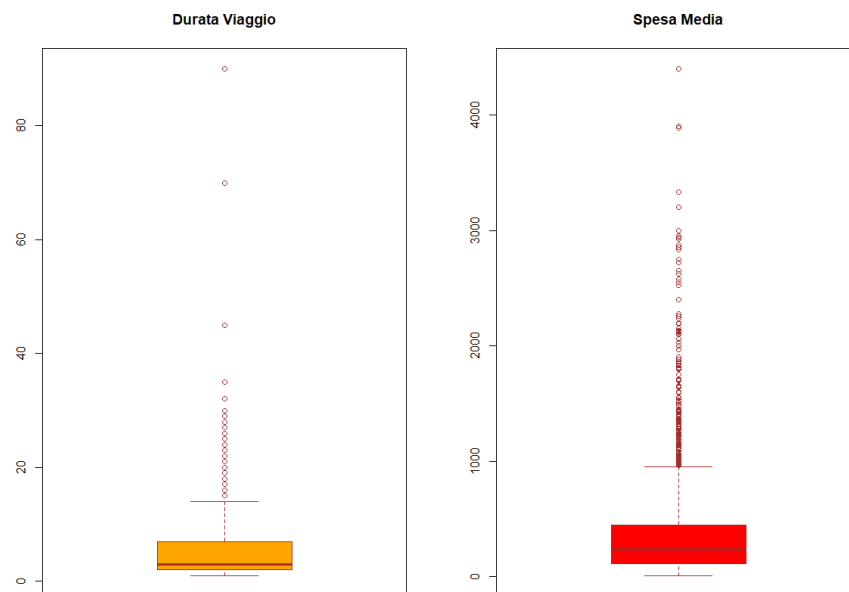


Figure 2: Boxplot Variabili Quantitative

Questo significa che c'è una piccola parte del campione che spende e rimane in viaggio più del resto dei concittadini analizzati.

Analizzando gli indici di dispersione possiamo notare che hanno quasi lo stesso coefficiente di variazione. Questo significa che hanno quasi la stessa variabilità interna. Cioè la diversità del campione riguardo alla durata è simile a quella della spesa. Il che ha senso anche considerando che le due variabili possono essere logicamente correlate, all'aumentare della durata della vacanza aumenta la spesa per la stessa.

```
> var_durata > var_spesa
[1] 30.12134 [1] 162982.9
> sd_durata > sd_spesa
[1] 5.488875 [1] 403.7543
> cv_durata > cv_spesa
[1] 1.023805 [1] 1.116108
```

Figure 3: Indici di Dispersione Variabili Quantitative

Infine, grazie alla suddivisione in classi tramite la varianza empirica del campione, notiamo una similitudine tra le distribuzioni di frequenza. Entrambe mostrano una decrescita esponenziale all'aumentare del valore della variabile, il che conferma la presenza degli outliers nella rappresentazione mediante boxplot. La Durata del Viaggio sembra avere una discesa più rapida all'asintoto orizzontale.

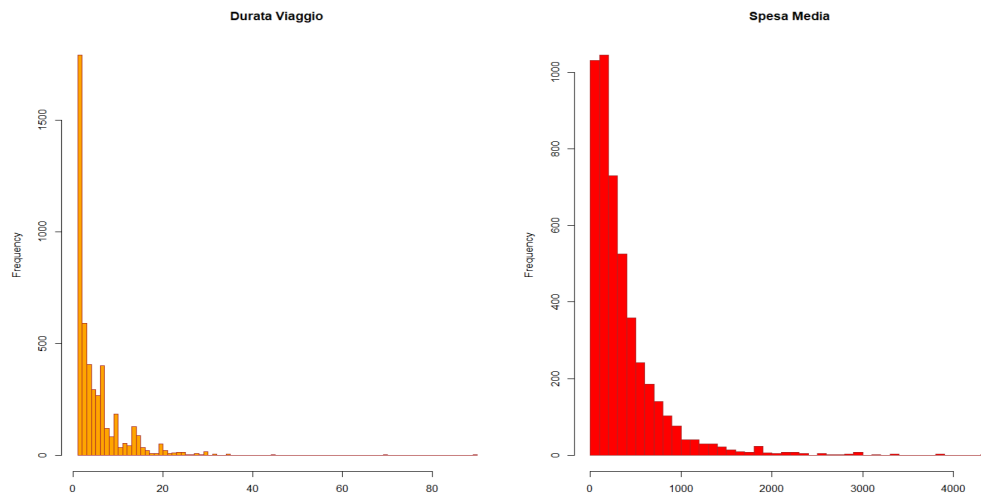


Figure 4: Istogrammi Variabili Quantitative



### 3.2 Univariata Variabili Qualitative

Iniziando dall'analisi delle variabili Categorie Ordinali, grazie alle distribuzioni di frequenza, possiamo analizzare il grafico delle frequenze relative.

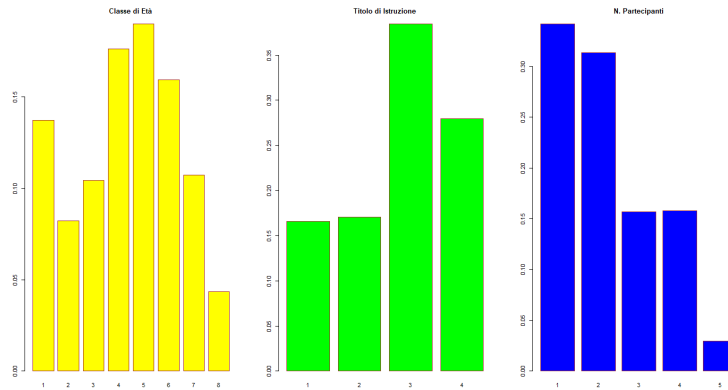


Figure 5: Barplot Variabili Qualitative Ordinali

L'età dei viaggiatori del campione, a parte un picco nella prima classe, sembra seguire una distribuzione normale con picco nella classe 5. Quindi abbiamo un aumento dei viaggi da parte delle unità statistiche all'aumentare dell'età fino ad un massimo nei 45-54 anni, dopodichè c'è una progressiva diminuzione. In generale la coda iniziale è più pesante dell'ultima quindi la distribuzione non sembra del tutto centrata.

Per quanto riguarda il titolo di istruzione possiamo notare che viaggia maggiormente chi ha un diploma di scuola secondaria o superiore. Questo potrebbe essere legato allo stipendio medio, ci aspettiamo che un titolo di studio maggiore sia relazionato positivamente ad un aumento dello stipendio e quindi del numero di viaggi.

Infine il numero di partecipanti sembra diviso in 3 macro gruppi. Le unità statistiche viaggiano maggiormente da sole o in coppia, c'è poi una simile percentuale di persone che viaggiano in 3 o 4 ed infine il gruppo minore che viaggia in 5 o più. Quest'ultimo risultato potrebbe dipendere dalla grandezza delle famiglie se esiste una correlazione tra i viaggiatori e il viaggiare con la propria famiglia. Più numerosa è la famiglia e meno si riuscirà a viaggiare perchè costoso oppure esistono meno famiglie numerose di quelle più piccole. Tutte osservazioni che sarebbe interessante approfondire in futuro.

Concludiamo l'analisi univariata con le variabili Categorie Nominali, quelle in cui non esiste nessun ordinamento naturale tra le classi. Come pos-

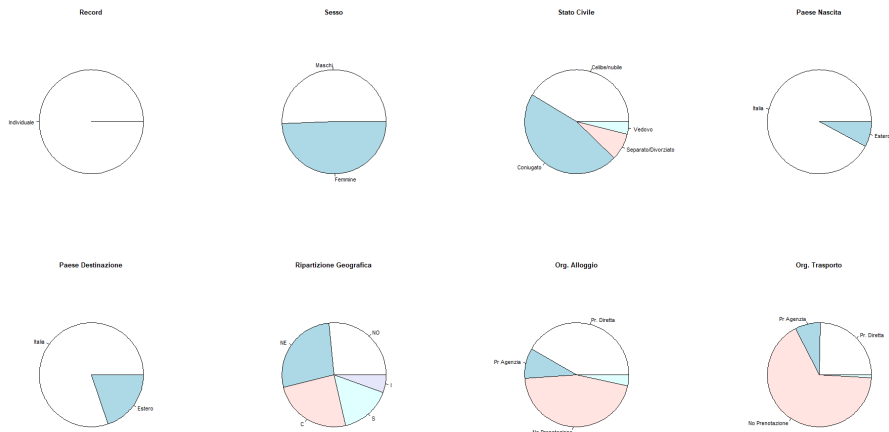


Figure 6: Pieplot Variabili Qualitative Nominali

siamo notare il campione si equidistribuisce rispetto al sesso. Il tipo di record è pressochè sempre individuale e lo stato civile è nella maggioranza dei casi Celibe/nubile o Coniugato. Paese di nascita e di destinazione è prevalentemente l'Italia anche se l'estero è maggiore nella destinazione rispetto alla nascita. Ci potremmo aspettare di vedere una correlazione tra chi viaggia all'estero ed è nato all'estero a causa del ritorno al luogo di origine durante le vacanze. Infine più della metà del campione statistico proviene dal Nord Italia, una bassa percentuale organizza l'alloggio con un'agenzia di viaggi o un tour operator e più del 60% non prenota il trasporto.

Infine vediamo che sia per la tipologia di alloggio che per la tipologia di mezzo di trasporto abbiamo delle categorie decisamente più utilizzate delle altre:

- **Alloggio:**
  - Albergo/motel/pensione
  - Abitazione di parenti o amici gratuita
- **Mezzo di trasporto:**
  - Auto propria, di parenti/amici

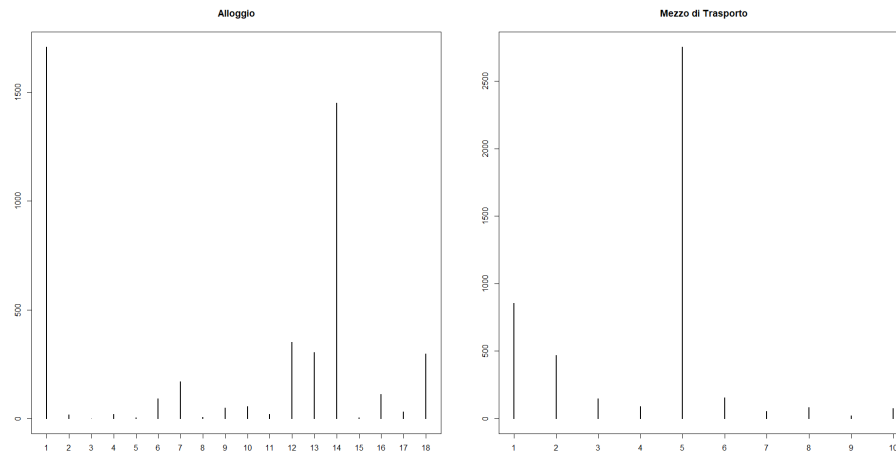


Figure 7: Plot Variabili Qualitative Nominali

Il fatto che l'auto propria o di parenti o di amici sia il mezzo preferito del campione statistico è congruente con il fatto che più del 60% non prenota il trasporto.

### 3.3 Analisi Bivariata

Iniziamo con la visualizzazione dei boxplot bivariati per cercare diverse distribuzioni della spesa media data dalle diverse altre proprietà, figura 8.

Lo scatterplot con la Durata non sembra mostrare chiare relazioni, questo in disaccordo con le supposizioni fatte durante l'analisi univariata, osserveremo successivamente se esiste effettivamente una relazione o no.

Sembra esistere una differenza nella spesa media del viaggio quando il paese di destinazione è l'estero. Aumenta la spesa inoltre quando l'organizzazione dell'alloggio e del trasporto sono fatti in agenzia o tramite tour operator. In figura 9 possiamo osservare che la spesa aumenta quando l'alloggio è di tipo Mezzo pubblico di Trasporto o Villaggio Vacanza. E' difficile che mezzo di trasporto pubblico sia un reale tipo di alloggio per viaggiatori, quindi si dovrebbe chiedere ad Istat di rivedere la categoria 4 della proprietà alloggio. Si possono inoltre notare molti outliers per la categoria Alberghi. Non c'è sorpresa se consideriamo che anche gli hotel di lusso rientrano in questa categoria e che questa è una categoria molto utilizzata dal nostro campione statistico. Per quanto riguarda i mezzi di trasporto notiamo che l'utilizzo dell'aereo comporta una spesa media maggiore. Infine l'auto di proprietà ha una dis-

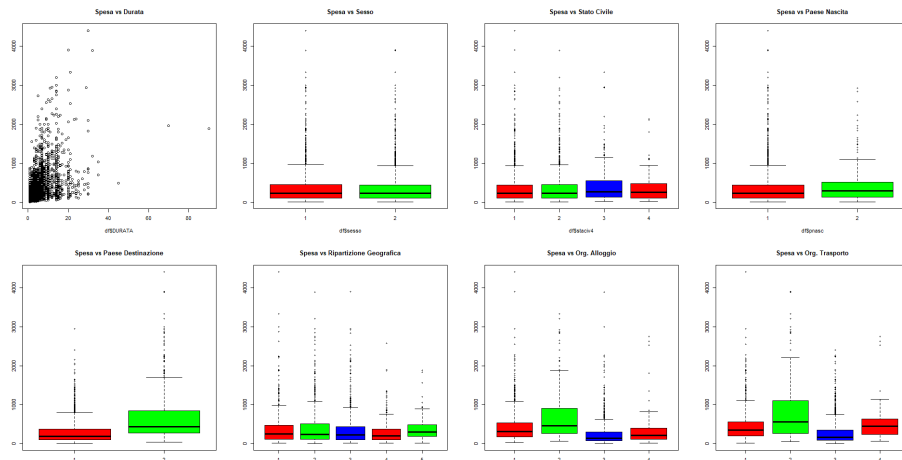


Figure 8: Rappresentazione Grafica Analisi Bivariata della Variabile Spesa Media

tribuzione di spesa medio bassa ma con molti outliers verso l'alto. Questo perchè è la categoria in assoluto più utilizzata e la spesa è molto variabile in base alla tipologia dell'auto. In figura 10, infine, possiamo vedere come

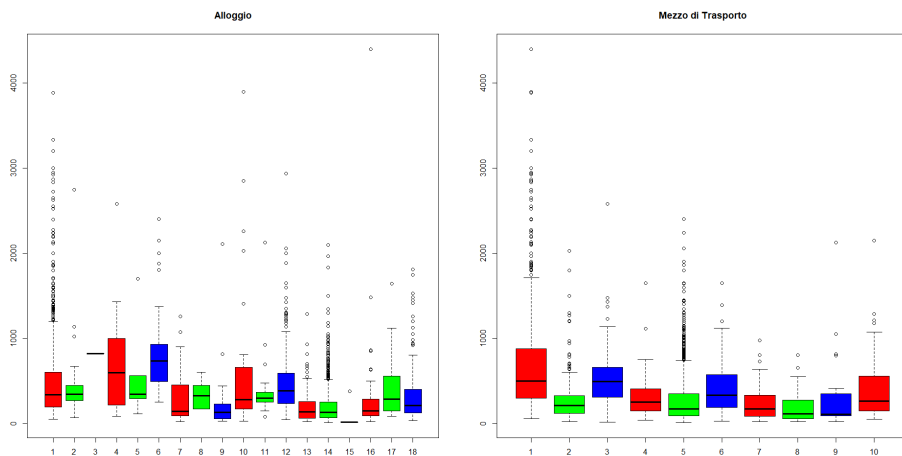


Figure 9: Rappresentazione Grafica Analisi Bivariata della Variabile Spesa Media

si distribuiscono le variabili categoriche ordinali rispetto alla spesa media. L'età e titolo di studio, in modo anti intuitivo, sembrano essere distribuite

uniformemente. Invece il numero di partecipanti aumenta la spesa media. In particolare la spesa media decresce all'aumentare del numero di componenti.

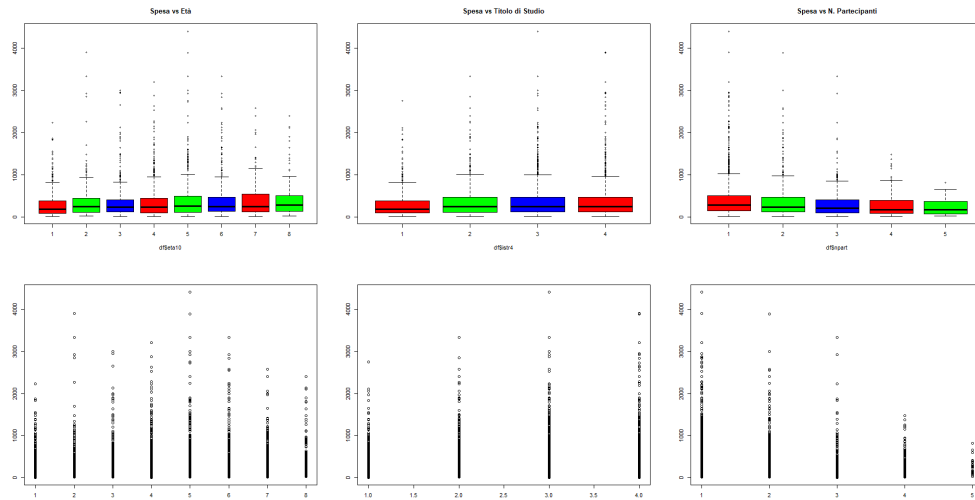


Figure 10: Rappresentazione Grafica Analisi Bivariata della Variabile Spesa Media

Analizziamo quindi analiticamente la dipendenza della spesa media dalle altre variabili. Innanzitutto la matrice di correlazione tra le variabili quantitative restituisce un indice di correlazione di 0.468 che quindi conferma una correlazione positiva tra la spesa media e la durata del viaggio.

Nella tabella seguente sono riportati i p-value del test chi-quadro e l'eta quadro per le variabili categoriche.

Variabile	Chi2	GL	p-value Chi2
Sesso	537.47	802	1
Stato Civico	3004.3	2406	5.826e−16
Paese di Nascita	1081.1	802	1.407e−10
Paese di Destinazione	2169.3	802	2.2e−16
Ripartizione Geografica	6050.4	3208	2.2e−16
Org. Alloggio	5747.1	2406	2.2e−16
Org. Mezzo di Trasporto	6359.5	2406	2.2e−16
Alloggio	29194	13634	2.2e−16
Mezzo di Trasporto	13459	7218	2.2e−16
Età	6866.3	5614	2.2e−16
Titolo di Studio	2783	2406	1.091e−07
N. Partecipanti	7453.8	3208	2.2e−16

Come si vede dalla tabella, a causa di un p-value sotto la soglia di significatività dello 0.01 non possiamo rifiutare l'ipotesi nulla di indipendenza con la spesa media per tutte le variabili tranne la variabile Sesso. Quest'ultima è l'unica che presenta una dipendenza con la spesa media.

## 4 Regressione Lineare

Nonostante non possiamo rifiutare l'ipotesi nulla di indipendenza data dal chi-quadro si è provato ad analizzare il modello di regressione lineare con i minimi quadrati. Il risultato è riportato in figura 11. Come si può notare il valore del test è confermato dal p-value  $< 2.2e-16$ . Le uniche variabili non significative sono il sesso, lo stato civile e l'età. Tutte le altre variabili sono significative per il modello lineare con una significatività dello 0.99. Il modello però ha R-quadro pari a 0.4422. Quindi, anche se significativo, il modello non può essere preso in considerazione per regredire la spesa media.

Dato che Durata è anch'essa una variabile quantitativa è stato creato un modello di regressione lineare per provare a spiegare la durata dei viaggi dei residenti italiani date le restanti variabili. Il risultato in figura 12. Anche in questo caso, nonostante il modello sia significativo e la maggioranza delle variabili siano significative al 99%, il modello non è sufficientemente capace di descrivere la durata dei viaggi. Infatti l'R-quadro vale 0.3331, quindi solo 1/3 dell'informazione è spiegata dal modello.

```

call:
lm(formula = ESPE_CO ~ sesso + staciv4 + pnasc + DEST_IE + rip +
    ORGALL + ORGTRA + ALLOG + MEZZO + eta10 + istr4 + npart +
    DURATA, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-1618.41  -148.36   -27.94    89.13   2879.89

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  367.4766    39.2175   9.370 < 2e-16 ***
sesso        -13.5573     8.9036  -1.523 0.127906
staciv4      -11.7329     7.5586  -1.552 0.120668
pnasc        -64.3441    17.0824  -3.767 0.000167 ***
DEST_IE      224.2050    12.8523  17.445 < 2e-16 ***
rip          -7.3837     3.7200  -1.985 0.047217 *
ORGALL       -50.5573     5.2680  -9.597 < 2e-16 ***
ORGTRA       15.5752     6.8166   2.285 0.022363 *
ALLOG        -13.2133     0.8212 -16.090 < 2e-16 ***
MEZZO        -29.9848     3.0594  -9.801 < 2e-16 ***
eta10         4.7637     3.1408   1.517 0.129408
istr4        16.4065     4.7344   3.465 0.000534 ***
npart       -41.4148     4.3022  -9.626 < 2e-16 ***
DURATA       34.8666     0.8415  41.432 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 302 on 4691 degrees of freedom
Multiple R-squared:  0.4422,    Adjusted R-squared:  0.4407
F-statistic: 286.1 on 13 and 4691 DF,  p-value: < 2.2e-16

```

Figure 11: Regressione Lineare della Variabile Spesa Media

## 5 Inferenza sulla popolazione Italiana

Come descritto nella nota metodologica il disegno campionario è stato progettato per fornire stime a livello anche nazionale. Quindi procediamo con la stima della spesa media e con la stima della durata media nazionale dei viaggi dei residenti in Italia nel 2018.

Essendo ignote sia la media che la varianza della popolazione dovremo utilizzare il T-test per determinare gli intervalli di confidenza delle stime. Utilizzando  $\alpha = 0.01$  e considerando che per l'intervallo di confidenza voglio stimare il valore sulle due code troverò  $t_{\frac{\alpha}{2}} = \pm 2.58$ . Considerando inoltre, per la spesa media, che la media campionaria è  $\bar{x} = 361.7135$ , la deviazione

```

call:
lm(formula = DURATA ~ sesso + staciv4 + pnasc + DEST_IE + rip +
    ORGALL + ORGTRA + ALLOG + MEZZO + eta10 + istr4 + npart +
    ESPE_CO, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-15.547  -2.414  -0.787   1.108  72.615

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.2311776   0.5857018  -5.517 3.64e-08 ***
sesso        0.3290470   0.1321174   2.491  0.0128 *
staciv4      0.0769911   0.1122295   0.686  0.4927
pnasc        1.3567180   0.2531944   5.358 8.80e-08 ***
DEST_IE     -0.0293876   0.1968810  -0.149  0.8814
rip          0.0159154   0.0552450   0.288  0.7733
ORGALL       0.6841711   0.0783316   8.734 < 2e-16 ***
ORGTRA      -0.2002719   0.1012045  -1.979  0.0479 *
ALLOG        0.1589321   0.0123058  12.915 < 2e-16 ***
MEZZO        0.0966239   0.0458575   2.107  0.0352 *
eta10        0.2158046   0.0465299   4.638 3.61e-06 ***
istr4       -0.6032666   0.0698171  -8.641 < 2e-16 ***
npart        0.7752916   0.0634922  12.211 < 2e-16 ***
ESPE_CO      0.0076835   0.0001855  41.432 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.482 on 4691 degrees of freedom
Multiple R-squared:  0.3349,    Adjusted R-squared:  0.3331
F-statistic: 181.7 on 13 and 4691 DF,  p-value: < 2.2e-16

```

Figure 12: Regressione Lineare della Variabile Durata

standard campionaria è  $s = 403.7114$  e  $n = 4705$ , abbiamo che l'intervallo per la spesa media sarà

$$\left( sp\bar{e}sa - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, sp\bar{e}sa + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right) = (346.547, 376.88)$$

L'intervallo è sufficientemente piccolo quindi possiamo concludere che  $n$ , numerosità del campione è sufficiente ad avere un intervallo di confidenza con un range non troppo esteso.

Allo stesso modo è possibile calcolare la durata media dei viaggi compiuti da tutti i residenti in Italia nel 2018.

$$\left( dur\bar{a}ta - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, dur\bar{a}ta + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right) = (5.1544, 5.5669)$$



Questi sono gli intervalli in cui le due medie nazionali cadranno rispettivamente con una probabilità dello 0.99.

Infine sono state analizzate le differenze della spesa in media rispetto alle classi delle variabili che, in fase grafica, identificavano maggiore differenza.

Come mostrato in figura 13 dal p-value sempre inferiore della soglia di significatività 0.01, c'è una significativa differenza nella media tra i gruppi. In particolare c'è almeno un gruppo con media differente dagli altri per ogni variabile analizzata.

DEST_IE	Df	Sum Sq	Mean Sq	F value	Pr(>F)		ORGALL	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Residuals	4703	648791411	1.38e+05	855.7	<2e-16	***	Residuals	4703	735922725	156479	197.5	<2e-16	***
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
ORGTRA	Df	Sum Sq	Mean Sq	F value	Pr(>F)		ALLOG	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Residuals	4703	726691939	154517	259.8	<2e-16	***	Residuals	4703	715200154	152073	339.5	<2e-16	***
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
MEZZO	Df	Sum Sq	Mean Sq	F value	Pr(>F)		npart	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Residuals	4703	696287532	148052	476.5	<2e-16	***	Residuals	4703	755311903	160602	71.75	<2e-16	***
--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Figure 13: Anova Univariata per Differenza in media intra gruppo

## 6 Conclusioni

Concludiamo l'analisi osservando che c'è una differenza in media sulla spesa media dei viaggi nelle variabili

- Paese di Destinazione
- Organizzazione Alloggio
- Organizzazione Mezzo di Trasporto
- Alloggio
- Trasporto
- Numero di Partecipanti

La regressione lineare però non è sufficiente come modello a spiegare la Spesa Media per viaggio, confermato anche dall'analisi del chi-quadro sulla dipendenza tra variabili.

Infine il campione utilizzato è risultato sufficientemente numeroso per garantire un'inferenza adeguata sulla popolazione da cui il campione proviene. Infatti per l'intera popolazione dei residenti in Italia nel 2018 si ha, con una probabilità dello 0.99

$$sp\bar{e}sa \in (346.547, 376.88)$$

$$dur\bar{a}ta \in (5.1544, 5.5669)$$