INTRODUZIONE A STATA

Settembre 23, 2024

Emma Manneschi e Maria Paola Pagliarulo

emanneschi@luiss.it
mpagliarulo@luiss.it

Introduzione

Questi appunti e il “do-file” delle lezioni vi forniranno il materiale necessario per avere una conoscenza di base di Stata (versione 18 o qualsiasi altra versione abbastanza recente) per la gestione dei dati e l'analisi statistica. Potete utilizzarli in combinazione con le funzioni che Stata vi mette a disposizione per approfondire la conoscenza di comandi specifici:

- Per ottenere informazioni su un comando o una procedura in Stata, digita “help nome_comando” nella finestra di comando oppure accedi al menu "Help" dall'interfaccia.
- Un altro comando utile è “findit”. Se cerchi informazioni su "y", digita “findit y” nella finestra dei comandi. Questo comando esegue una ricerca estesa nella documentazione di Stata, nelle FAQ, nei pacchetti aggiuntivi e online per trovare tutte le risorse disponibili relative a "y".

Questo materiale sarà sufficiente per questo corso. Però, se desideri avere più informazioni su Stata, ecco alcune risorse utili:

- UCLA computing service website: https://stats.oarc.ucla.edu/stata/ . Qui si trova un'ampia gamma di consigli su diversi aspetti di Stata, comprese le guide introduttive che possono essere scaricate.
- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics Using Stata* (Vol. 2). College Station, TX: Stata press.

Cos'è Stata?

- Stata è un software di analisi dei dati versatile, ampiamente utilizzato nella ricerca nelle scienze sociali ed economiche, che consente di esplorare, gestire, riassumere e analizzare dataset in modo efficiente.

Interfaccia di Stata

- **(A) Finestra dei comandi**: è il campo in cui si digitano i comandi che possono essere lanciati premendo invio.
- **(B) Finestra delle variabili**: elenca le variabili (e le loro etichette) in memoria. Facendo clic sul nome di una variabile, la sua descrizione apparirà nella Finestra delle proprietà, mentre facendo doppio clic su di essa apparirà nella Finestra dei comandi.
- **(C) Finestra delle proprietà**: mostra le informazioni sulle variabili e sul dataset.
- **(D) Finestra dei risultati**: visualizza i risultati dei comandi lanciati nella finestra dei comandi (o attraverso lo script/dofile).
- **(E) Finestra di revisione**: elenca i comandi lanciati dalla finestra dei comandi. I comandi andati a buon fine appaiono in nero, mentre quelli non andati a buon fine appaiono in rosso. È possibile lanciare nuovamente un comando facendo doppio clic sulla finestra di revisione.

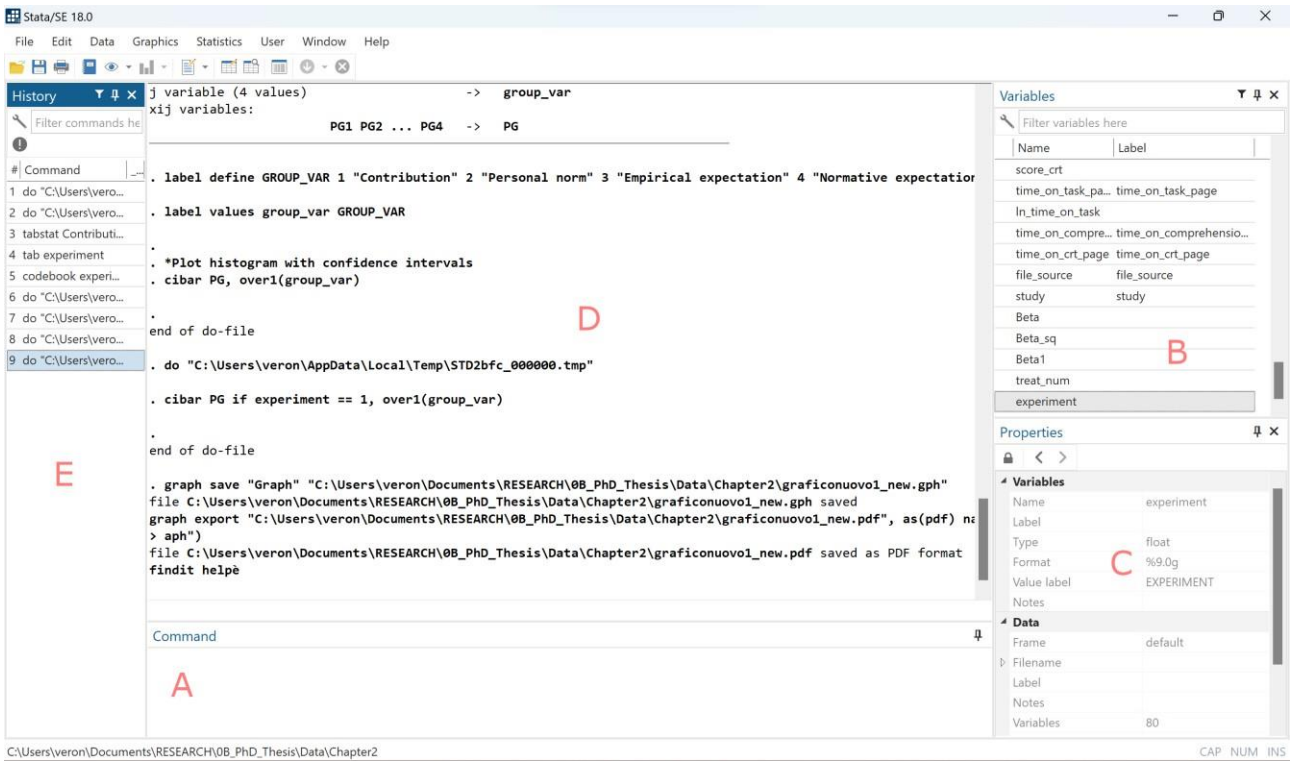

Figura 1: Schermata di Stata.

I menu di Stata

- Per eseguire i comandi si utilizzerà principalmente la sintassi (come fa la maggior parte degli utenti di Stata).
- La sintassi si riferisce all'insieme di regole che definiscono la struttura e il format dei comandi in un linguaggio di programmazione.
- Stata, tuttavia, consente di eseguire i comandi in modo “point-and-click” attraverso i suoi menu.
- I menu di Stata permettono di eseguire la maggior parte dei comandi per la gestione dei dati, la creazione di grafici e l'analisi statistica.

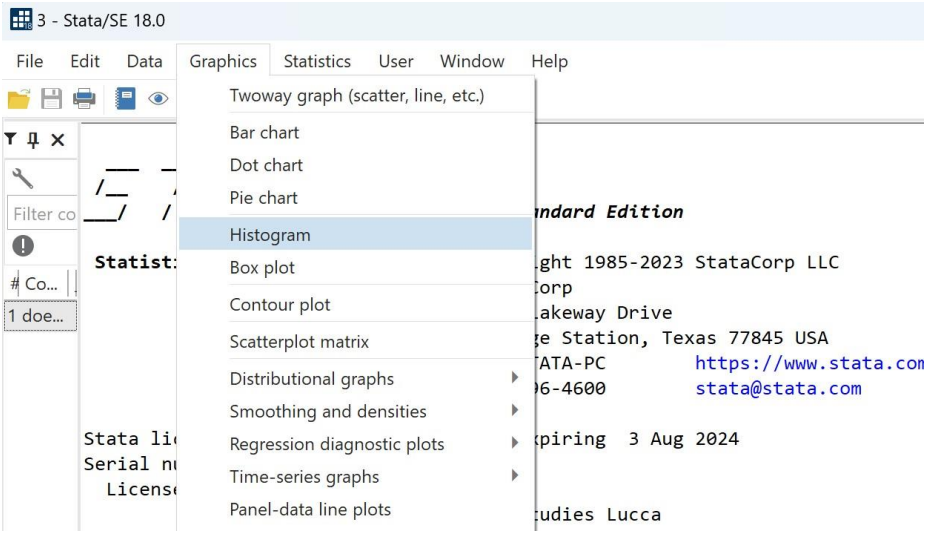

Figura 2: Creare un istogramma attraverso il menu di Stata “Graphics”.

File principali di Stata

- **.dta**: contiene dati che possono essere caricati direttamente in Stata senza utilizzare le funzioni di importazione.
- **.do**: i file do sono file di script contenenti un elenco di comandi.

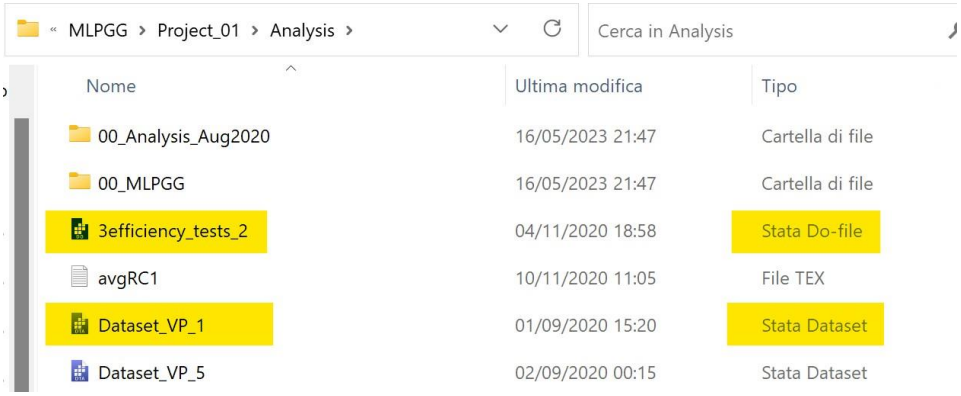

Figura 3: Come appaiono i file “.dta” e “.do” in una cartella.

Do-file

- I do-files sono file di testo in cui è possibile **salvare ed eseguire i comandi per riutilizzarli**, anziché riscriverli ogni volta nella finestra dei comandi.

- Per aprire un nuovo editor di do-file, ci sono due modi:

    - Scrivi il comando *doedit* nella Finestra dei comandi.
    - Fare clic sull'icona “foglio e matita” nella barra degli strumenti (vedere la Figura 4 qui sotto).

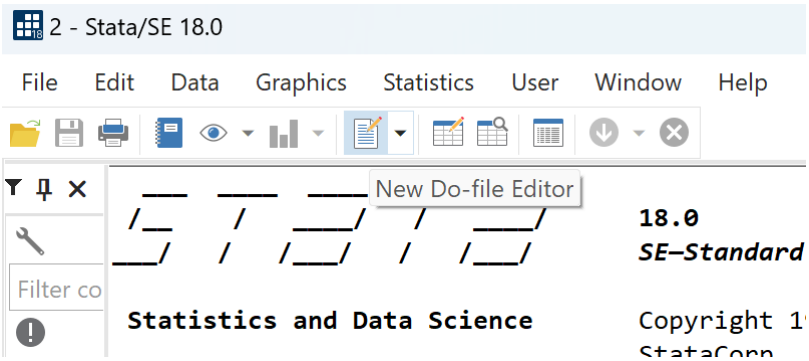

Figura 4: Aprire un nuovo do-file editor.

- Per eseguire i comandi, è necessario evidenziare la parte del codice che si desidera eseguire, quindi premere Ctrl+D (o, se avete un Mac: Shift+Cmd+D) o fare clic sull'icona “Esegui” (come nella Figura 5 qui sotto). È possibile selezionare ed eseguire più comandi!

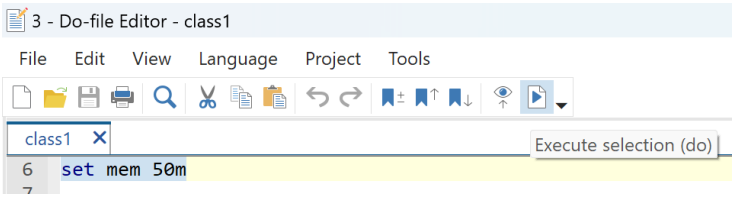

Figura 5: L’icona “Do”.

- È possibile commentare il codice utilizzando i caratteri “/*” prima e “*/” dopo il testo. Si possono anche usare gli asterischi o “//” per commentare una singola riga. I commenti vengono visualizzati in verde. Non sono destinati a essere letti dal software (NON sono comandi), ma dalle persone. Aggiungere commenti al codice aiuterà gli altri (e il “futuro te”) a interpretarlo!

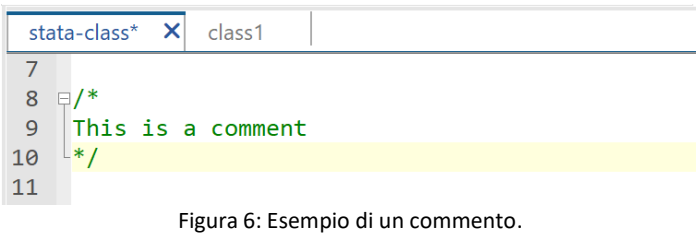

Figura 6: Esempio di un commento.

- Dopo aver utilizzato il file do, ricordarsi di salvarlo facendo clic sull'icona del floppy disk. L'asterisco

nel nome del file significa che ci sono modifiche non salvate!

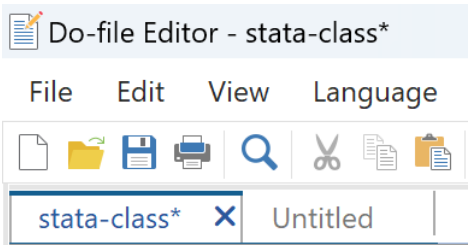

Figura 7: Come salvare le modifiche sul file do.

Caricare i dati

- Per caricare un dataset in formato ".dta" agevolmente, potete seguire i seguenti passaggi: 1) aprire un file do e salvarlo nella stessa cartella dove si trova il dataset da utilizzare. 2) fare doppio clic sul file ".dta" 3) copiare questa riga di codice e incollarla nel vostro do file:

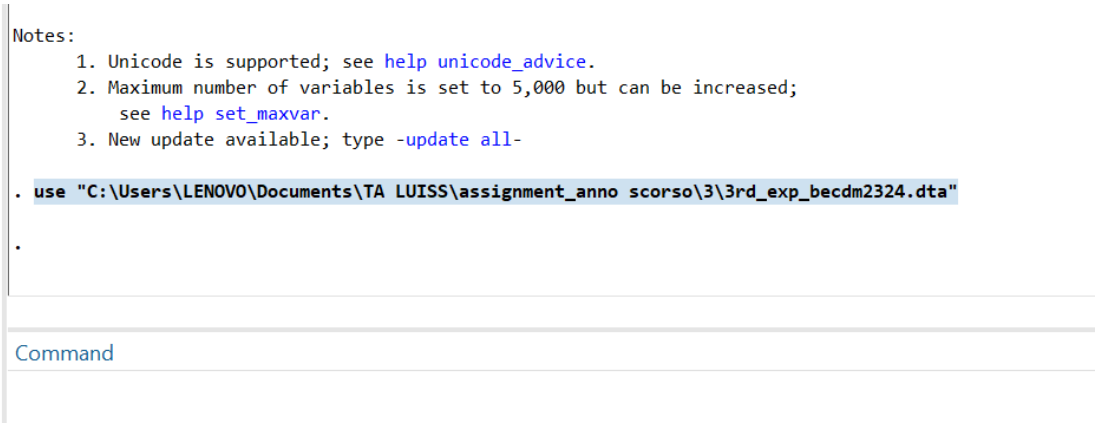

Figura 8: Caricare il dataset.

- Tra virgolette viene specificato il **percorso file** o path, cioè la sequenza di cartelle che specifica *l'esatta posizione di un file* all'interno del sistema di archiviazione del computer. Quando importi un dataset in Stata, è necessario fornire il percorso completo affinché il programma possa localizzare e aprire correttamente il file.

- Se si vuole cancellare tutto dalla memoria (e questo è SEMPRE necessario sia quando si avvia un nuovo file do, sia se si carica un nuovo dataset diverso da quello in memoria nello stesso file do), bisogna specificare l'opzione **clear** dopo una virgola, sulla stessa riga del comando use. La riga del comando apparirà quindi come segue: *use "percorso file", clear*

- Con il comando *import* è possibile caricare anche dataset in formato excel e csv (Figura 9).

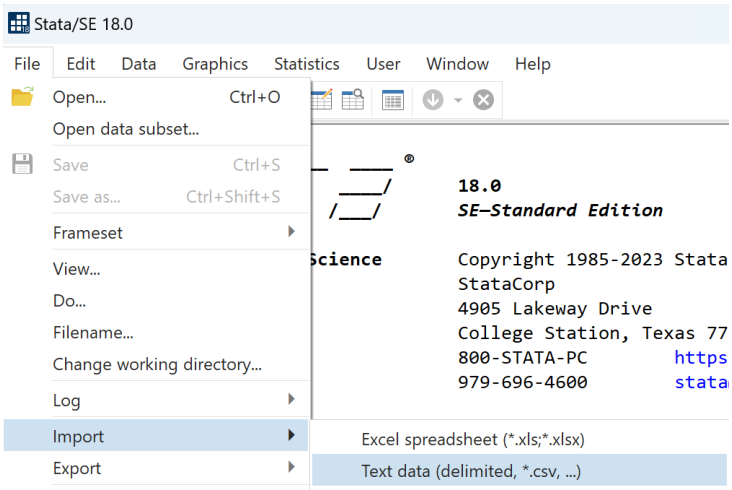

Figura 9: Importazione di file da altri formati tramite il menu “File” dell'interfaccia di Stata.

- In questa lezione utilizzeremo un dataset “.dta” che Stata può recuperare via Internet, quindi non è necessario che sia memorizzato sul PC. Ci sono due comandi che possono farlo. Si tratta dei comandi sysuse e webuse.

Esplorare i tuoi dati

- Per aprire e visualizzare il dataset, è possibile:
  - Scrivere il comando *browse*
  - Fare clic sull'icona “foglio di calcolo con lente di ingrandimento” nella barra degli strumenti.

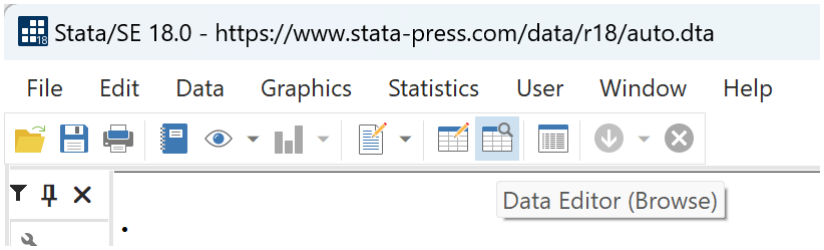

Figura 10: Esplorare il dataset.

- Si può notare che i colori sono diversi per i vari tipi di dati. Le colonne nere sono numeriche, quelle rosse sono stringhe e quelle blu sono numeriche con etichette di stringa.

Figura 11: Diversi tipi di dati.

- I tipi di dati standard sono:
    - Byte (valori interi compresi tra -127 e 100)
    - Int (valori interi compresi tra -32.767 e 32.740)
    - Long (valori interi compresi tra -2.147.483.647 e 2.147.483.620)
    - Float (numeri con decimali con circa 8 cifre di precisione)
    - Double (numeri con decimali con circa 16 cifre di precisione)

o Ora siete pronti per passare a Stata! Creiamo il do-file "**stata-class.do**".

Analisi dei dati con Stata

Statistiche descrittive

In un articolo è sempre utile iniziare la propria analisi fornendo al lettore alcune statistiche descrittive dei dati. Per esempio, dai dati che abbiamo utilizzato finora, possiamo dire che il dataset "auto2" comprende 74 osservazioni di 52 automobili di origine statunitense e 22 automobili di origine straniera, e che il campione presenta un prezzo medio di circa 6.165 dollari, che varia da un minimo di 3.291 a un massimo di 15.906, ecc. ecc. È possibile includere tabelle (come quelle riprodotte di seguito) e una descrizione delle stesse. Attenzione: le tabelle devono essere autoesplicative, quindi si consiglia di aggiungere alcune informazioni nelle "Note sulla tabella", come nell'esempio seguente.

Tabella 1: Statistiche descrittive delle variabili principali.

Variable	Obs.	Media	Dev. Std.	Min	Max
price	72	5914.208	2562.398	3291	13594
mpg	72	21.403	5.801	12	41
rep	67	3.433	.988	1	5
weight	72	2989.583	765.822	1760	4840

Note: Price è misurato in dollari. Mpg rappresenta il chilometraggio (miglia). Rep rappresenta il numero di riparazioni effettuate nel 1978 ed è codificato come 1 "Scarso", 2 "Discreto", 3 "Medio", 4 "Buono", 5 "Eccellente". Weight è misurato in libbre (lb).

A volte è utile visualizzare le statistiche delle stesse variabili ma per gruppi diversi. Ad esempio, nella Tabella 2 abbiamo diviso tra auto nazionali e auto straniere. Possiamo notare che il prezzo medio delle auto straniere è di circa 6.400 dollari e che il prezzo medio delle auto nazionali è di circa 5.700 dollari. Ecc. ecc.

Tabella 2: Statistiche descrittive delle variabili principali, per provenienza.

	N	Media	Dev.Std.	Min	Max
Car origin: Domestic					
price	50	5707.2	2534.666	3291	13594
mpg	50	19.92	4.763	12	34
rep	46	3.043	0.842	1	5
weight	50	3286	689.948	1800	4840
Car origin: Foreign					
price	22	6384.682	2621.915	3748	12990
mpg	22	24.773	6.611	14	41
rep	21	4.286	0.717	3	5
weight	22	2315.909	433.003	1760	3420

Grafici

È sempre consigliabile fornire rappresentazioni visive dei dati e/o delle informazioni che si ritengono importanti per l'analisi.

Figura 1: Distribuzione del prezzo.

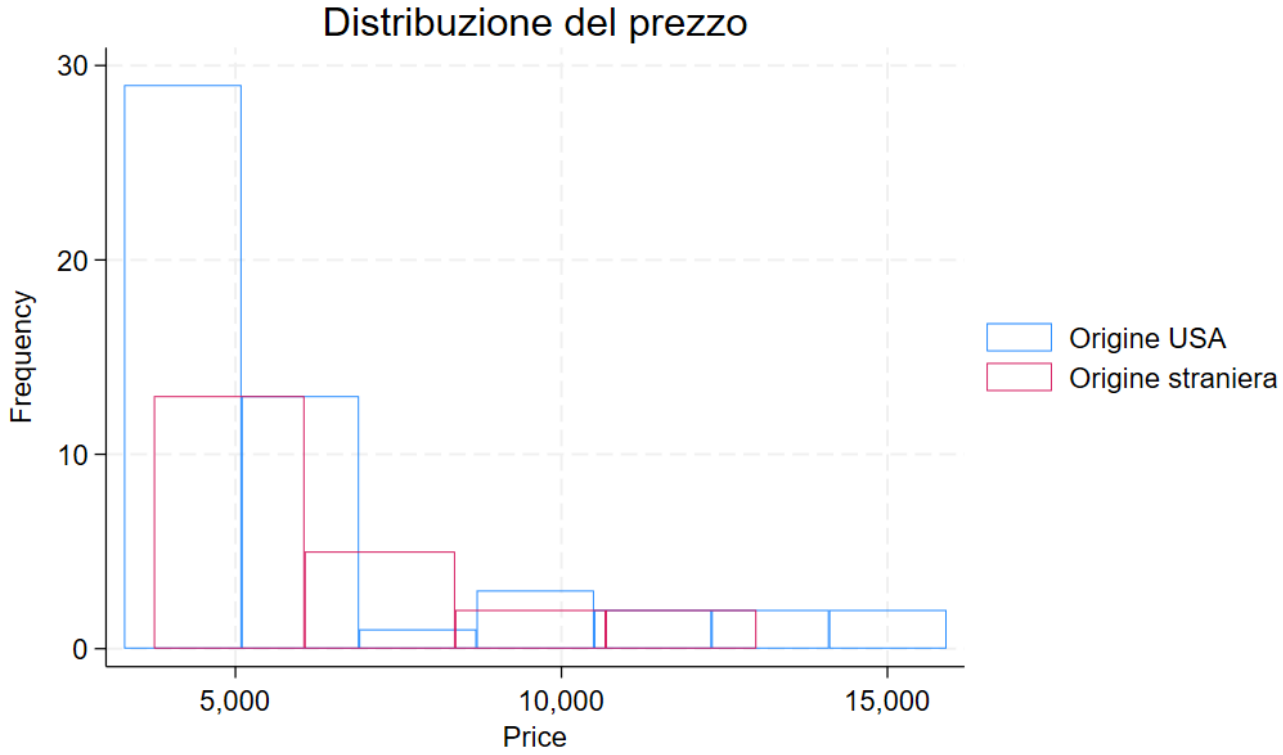

Figura 2: Distribuzione del prezzo,
per provenienza.

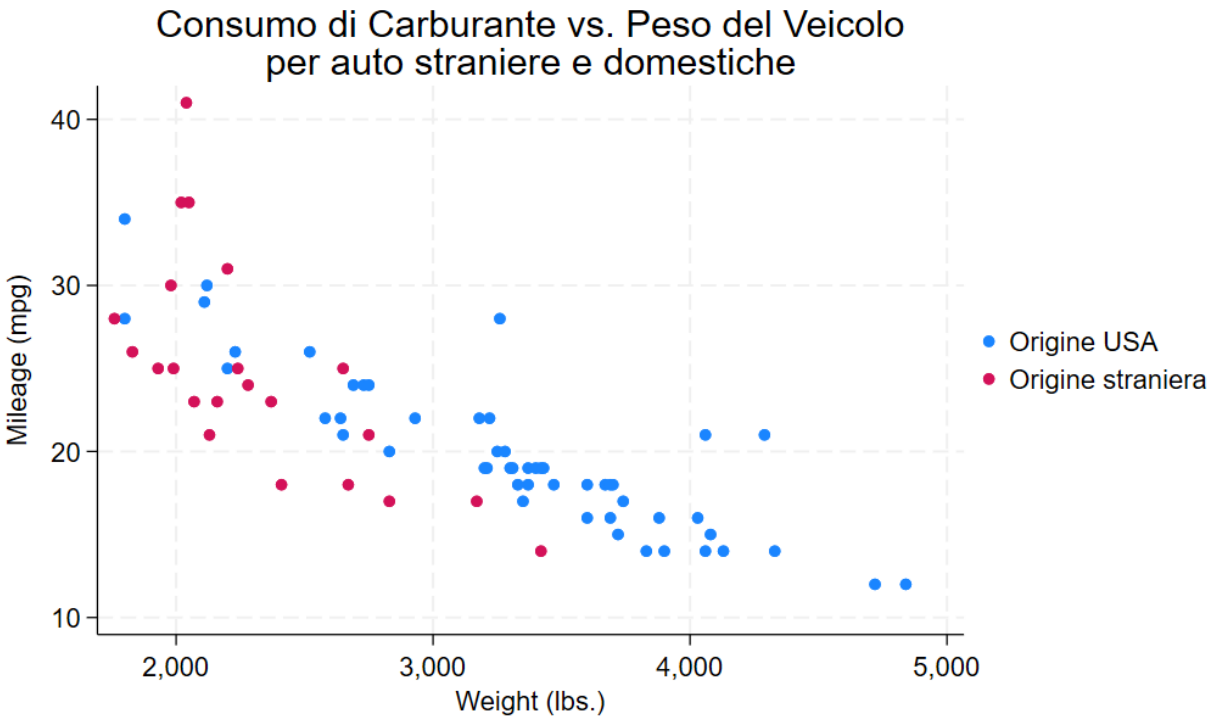

Figura 3: Relazione tra chilometraggio e peso.

È possibile aggiungere informazioni sul coefficiente di correlazione di Pearson quando si commenta un grafico a dispersione. Ad esempio, descrivendo l'andamento della Figura 3, si può dire che il coefficiente di correlazione tra le variabili “mpg” e “peso” è forte e negativo (-0,8177) e che, poiché il relativo valore p è inferiore a 0,05, la correlazione tra queste due variabili è statisticamente significativa.

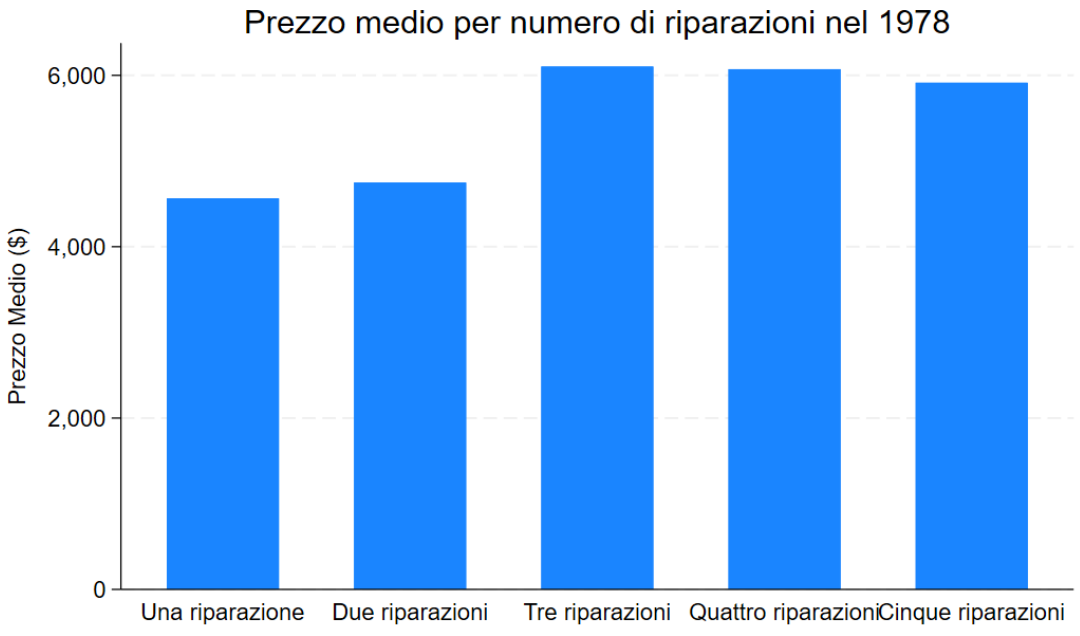

Figura 4: Prezzo medio per numero di riparazioni.

Questo è un esempio di descrizione per la Figura 4:

Dal grafico a barre qui sopra, possiamo notare che, in media, i prezzi sembrano aumentare man mano che lo stato dell'auto migliora in termini di riparazione. Se l'auto ha ricevuto solo una o due riparazioni (cioè è classificata come “Scarsa” o “Discreta”), il prezzo medio è compreso tra 4.500 e 5.000 $; mentre se l'auto ha ricevuto almeno tre riparazioni (cioè è “Media”, “Buona” o “Eccellente”) il prezzo medio è di circa 6.000 $ o leggermente superiore.

Esiste un altro tipo di grafico che potrebbe essere utile (anche per i vostri assignments).

Si tratta di un grafico a linee per dati time-series (ad esempio, se dovete visualizzare la variabile PIL negli ultimi 20 anni). Può essere lanciato tramite il comando *tsline*. Digitare “help tsline” nella finestra dei comandi per saperne di più!

Verifica di ipotesi

La *verifica di ipotesi* è una metodologia fondamentale nel campo della statistica inferenziale, utilizzata per **testare la validità di un'affermazione** o ipotesi **riguardante una popolazione**, basandosi su dati campionari. Questa tecnica è essenziale per determinare se i risultati osservati in un campione possono essere generalizzati a una popolazione più ampia.

Ipotesi nulle e ipotesi alternativa:

Il punto di partenza della verifica di ipotesi statistiche è la specificazione dell'ipotesi da verificare, detta *ipotesi nulla*. La verifica di ipotesi richiede l'uso dei dati al fine di confrontare l'ipotesi nulla con una seconda ipotesi, detta *ipotesi alternativa*, che è valida se la nulla non lo è.

- L'**ipotesi nulla** (H_0) prevede che la media della variabile d'interesse, che chiameremo y, nella popolazione, E(y), assuma un valore specifico, indicato con $\mu_{y,0}$.

$$H_0: E(y) = \mu_{y,0}$$

- L'**ipotesi alternativa** specifica che cosa è vero se l'ipotesi nulla non lo è. L'ipotesi alternativa più generale è che $E(y) \neq \mu_{y,0}$; è detta ipotesi alternativa bilaterale perché prevede che E(y) possa essere sia minore sia maggiore di $\mu_{y,0}$.

$$H_1: E(y) \neq \mu_{y,0}$$

Il problema che gli statistici affrontano è quello di utilizzare l'evidenza empirica, fornita da un campione selezionato casualmente, per stabilire se accettare l'ipotesi nulla oppure rifiutarla in favore dell'ipotesi alternativa.

Valore p dei test

Il **valore p** dei test è la probabilità di ottenere una statistica sfavorevole all'ipotesi nulla, almeno quanto la statistica calcolata effettivamente nel campione, assumendo che l'ipotesi nulla sia vera.

Immagina di voler sapere se la media di un campione di dati che hai raccolto (ad esempio, l'altezza media di un gruppo di persone) è significativamente diversa dalla media della popolazione generale (che potrebbe essere l'altezza media di tutta la popolazione di un paese). Per fare ciò, esegui un **test di ipotesi**. Il **p-value** è il **valore che ti dice quanto è probabile ottenere un risultato come quello del tuo campione** (o più estremo), **se l'ipotesi nulla fosse vera**. In altre parole, il p-value ti indica la probabilità che la differenza tra la media campionaria e la media della popolazione sia dovuta al caso.

Come interpretarlo:

- Se il **p-value** è **sufficientemente piccolo** (ad esempio, inferiore a una soglia comune come 0,05), significa che è **molto improbabile** che la differenza osservata nel campione sia dovuta al caso, e quindi si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa. In altre parole, si conclude che c'è una differenza significativa tra la media del campione e quella della popolazione. Quando il valore p è inferiore a 0,05 diciamo che l'ipotesi nulla è rifiutata a un livello di significatività del 5% (ossia c'è una probabilità inferiore al 5% che l'ipotesi nulla sia vera) Quando il valore p è compreso tra 0,1 e 0,05 diciamo che l'ipotesi nulla è rifiutata a un livello di significatività del 10% (ma non del 5%)
- Se il **p-value** è **grande** (ad esempio, maggiore di 0,05), significa che la differenza osservata potrebbe essere tranquillamente dovuta al caso, quindi non hai abbastanza prove per rifiutare l'ipotesi nulla. In pratica, non puoi dire che le due medie siano diverse in modo significativo.

Un esempio pratico:

- Supponiamo che la media dell'altezza della popolazione sia 170 cm e tu voglia sapere se il tuo campione di 100 persone, con una media di 172 cm, è significativamente diverso da quella media.

Applichi un test statistico e ottieni un **p-value di 0,03**.

- Cosa significa? Un p-value di 0,03 indica che, se la media della popolazione fosse davvero 170 cm, c'è solo il 3% di probabilità di ottenere un campione con una media di 172 cm (o più estremo) **per puro caso**. Poiché 0,03 è inferiore a 0,05, puoi rifiutare l'ipotesi nulla e concludere che la media del tuo campione è significativamente diversa da quella della popolazione.

T test

Il test t è un test statistico per verificare se il valore medio di una distribuzione si discosta significativamente da un certo valore di riferimento.

Il t test può essere usato in tre modi:

1) **T-test a campione singolo (confronta la media di un campione con un valore specifico)**

```
. ttest mpg == 30

One-sample t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
mpg	74	21.2973	.6725511	5.785503	19.9569	22.63769

```
    mean = mean(mpg)                                  t = -12.9398
H0: mean = 30                      Degrees of freedom =       73

   Ha: mean < 30                 Ha: mean != 30                 Ha: mean > 30
Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

I valori p in Stata sono espressi dal simbolo "Pr (...)".
Possiamo accettare la nostra ipotesi nulla (quella scritta nel comando e che appare etichettata come "H_0") se nessuno dei valori p è inferiore a 0,05. Se il valore p "centrale" è inferiore a 0,05, dobbiamo rifiutare la nostra ipotesi nulla.

I valori p "esterni" ci danno la direzione della disuguaglianza (se "mpg" è diverso da 30, vogliamo sapere se è maggiore o minore di 30). Troveremo la nostra soluzione, ancora una volta, con il valore p inferiore a 0,05.

In questo caso, la media di "mpg" è inferiore a 30.

Di conseguenza, possiamo dire che: "**In questo caso, rifiutiamo l'ipotesi nulla. La media del consumo di carburante è statisticamente diversa da 30. Più precisamente, la media di mpg è statisticamente inferiore a 30**".

2) Paired t-test (testa l'uguaglianza tra la media di due variabili)

```
. ttest price_1 == price_2

Paired t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
price_1	74	.4980805	.0322738	.2776299	.4337589	.5624022
price_2	74	.4837835	.0350746	.3017227	.41388	.553687
diff	74	.014297	.0528334	.4544901	-.0909998	.1195938

```
    mean(diff) = mean(price_1 - price_2)                      t =   0.2706
H0: mean(diff) = 0                     Degrees of freedom =       73

Ha: mean(diff) < 0          Ha: mean(diff) != 0          Ha: mean(diff) > 0
Pr(T < t) = 0.6063    Pr(|T| > |t|) = 0.7875          Pr(T > t) = 0.3937
```

La nostra ipotesi nulla è che la media del price_1 sia uguale a quella del price_2. O, in altre parole, che la media della differenza tra le due variabili sia uguale a zero. Per questo Stata ci mostra anche l'ipotesi scritta in questo modo: H_0 media(diff) = 0.

Leggiamo il valore p "centrale". Non è inferiore a 0,05 (infatti è pari a 0,7875). Non rifiutiamo la nostra ipotesi nulla: le medie delle due variabili ("price_1" e "price_2") sono statisticamente uguali. In questo caso, quindi, non è necessario esaminare i valori p "esterni".

In questo caso, le medie di price_1 e price_2 sono statisticamente uguali.

Pertanto, possiamo dire che: "**In questo caso, non rifiutiamo l'ipotesi nulla. La media del price_1 è statisticamente uguale alla media del price_2**".

3) Test t per due campioni (testa che la media di una variabile sia uguale tra gruppi diversi)

```
. ttest mpg, by(foreign)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Domestic	52	19.82692	.657777	4.743297	18.50638	21.14747
Foreign	22	24.77273	1.40951	6.611187	21.84149	27.70396
Combined	74	21.2973	.6725511	5.785503	19.9569	22.63769
diff		-4.945804	1.362162		-7.661225	-2.230384

```
    diff = mean(Domestic) - mean(Foreign)                     t =  -3.6308
H0: diff = 0                             Degrees of freedom =       72

 Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
Pr(T < t) = 0.0003    Pr(|T| > |t|) = 0.0005          Pr(T > t) = 0.9997
```

Si tratta di verificare se la media di una variabile è uguale tra due gruppi. In questo caso, stiamo verificando se la media di “mpg” è uguale tra auto straniere e auto nazionali (o, in alternativa, se la differenza delle due medie è uguale a zero).
Quest'ultimo test deve essere interpretato seguendo le stesse regole dei due test precedenti.

Come sempre, osserviamo innanzitutto il valore p “centrale”. È inferiore a 0,05. Possiamo rifiutare la nostra ipotesi nulla. Ciò significa che la media di “mpg” è diversa tra le auto nazionali e quelle straniere.

Osserviamo i valori p “esterni” e notiamo che la differenza tra le medie è inferiore a 0. Se la differenza è inferiore a 0, significa che la prima media è più piccola della seconda. Quindi, la media di “mpg” per le auto nazionali è più bassa rispetto alla media di “mpg” per le auto straniere.

Di conseguenza, possiamo affermare che: “**Rifiutiamo l'ipotesi nulla che “mpg” sia in media uguale tra le auto nazionali e straniere. La variabile “mpg” è statisticamente diversa tra i due gruppi. In particolare, la media di “mpg” è più alta tra le auto straniere rispetto a quelle nazionali**”.

Il modello di regressione lineare (opzionale)

La regressione lineare è una tecnica di analisi statistica utilizzata per valutare la correlazione tra una variabile dipendente e un insieme di variabili esplicative. Essa consente di misurare l'effetto che una variazione unitaria di ciascuna variabile esplicativa ha sulla variabile dipendente, mantenendo costanti tutte le altre variabili, e di quantificare il potere esplicativo di ogni variabile rispetto alla dipendente.

1) Con un singolo regressore

Il modello di regressione lineare con un singolo regressore mette in relazione una variabile X con un'altra variabile Y. Tale modello postula una relazione lineare tra X e Y; la pendenza della retta che mette in relazione le due variabili è l'effetto di una variazione unitaria di X su Y.

Il modello di **regressione lineare** descrive la relazione tra una variabile dipendente (**Y**) e una variabile indipendente (**X**) attraverso una funzione lineare.

L'obiettivo della regressione lineare è stimare una retta che meglio approssima i dati osservati, secondo l'equazione

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, \ldots, n$$

Dove:

- $\mathbf{Y_i}$ è la variabile dipendente,
- $\mathbf{X_i}$ è la variabile indipendente,
- $\boldsymbol{\beta_0}$ è l'**intercetta** (il valore di Y quando X = 0),
- $\boldsymbol{\beta_1}$ è il **coefficiente di regressione** (l'effetto di una variazione unitaria di X su Y, quindi la pendenza della retta),
- $\mathbf{u_i}$ rappresenta l'errore residuo, cioè la differenza tra i valori osservati e quelli stimati.

Il metodo dei **minimi quadrati** viene utilizzato per trovare i valori di $\boldsymbol{\beta_0}$ e $\boldsymbol{\beta_1}$ che minimizzano la somma dei quadrati degli errori, fornendo così la migliore stima lineare della relazione tra X e Y.
La regressione lineare è utile per predire il comportamento di Y in funzione di X, testare ipotesi sulle relazioni tra variabili e determinare l'intensità dell'associazione.

2) Con regressori multipli

Il modello di regressione multipla estende il modello di regressione con un singolo regressore includendo come regressori una serie di variabili addizionali. Quando è utilizzato per l'inferenza causale, questo modello permette di stimare l'effetto su Y_i di una variazione in un regressore (X_{1i}), tenendo costanti gli altri (X_{2i}, X_{3i}, ecc.).
Il modello di regressione multipla è:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i, \qquad i = 1,\ldots,n$$

Dove:

- $\mathbf{Y_i}$ è l'i-esima osservazione della variabile dipendente,
- $\mathbf{X_{1i}, X_{2i}, \ldots, X_{ki}}$ sono le i-esime osservazioni di ciascuno dei k regressori,
- $\boldsymbol{\beta_0}$ è l'**intercetta** (il valore di Y quando tutte le X sono uguali a 0),
- $\boldsymbol{\beta_1}$ è il **coefficiente associato a $\mathbf{X_1}$** , $\boldsymbol{\beta_2}$ è il **coefficiente associato a $\mathbf{X_2}$** , ecc. Il coefficiente $\boldsymbol{\beta_1}$ rappresenta la variazione attesa di $\mathbf{Y_i}$ che risulta da una variazione unitaria di X_{1i}, tenendo costanti tutti gli altri regressori ($X_{2i}, \ldots, X_{ki}$).
- $\mathbf{u_i}$ rappresenta l'errore

Esempio di regressione dal paper "A field study on cooperativeness and impatience in the Tragedy of the Commons"

Table 1
Determinants of size of holes in shrimp traps (*OLS*).

Dependent variable	Average size of holes in shrimp trap in cm				
Model	1	2	3	4	5
Contribution in PGE (in MUs)	0.0105**		0.0094**	0.0103**	0.0088**
	(0.0041)		(0.0038)	(0.0041)	(0.0040)
Impatience (praline dummy)		−0.0504**		−0.0539**	−0.0467**
		(0.0230)		(0.0238)	(0.0217)
Belief in PGE			−0.0011	0.0003	−0.0001
			(0.0051)	(0.0055)	(0.0053)
Preference for praline				−0.1962	−0.1593
				(0.1425)	(0.1228)
Age			−0.0011	−0.0003	−0.0009
			(0.0011)	(0.0012)	(0.0012)
Gender (male dummy)			0.0893***	0.0663***	0.0515
			(0.0232)	(0.0232)	(0.0332)
Children			0.0121**	0.0137**	0.0109*
			(0.0056)	(0.0062)	(0.0058)
Centrality			−0.0001	−0.0003	−0.0005
			(0.0003)	(0.0004)	(0.0004)
Years of schooling			−0.0034	−0.0026	−0.0004
			(0.0039)	(0.0046)	(0.0045)
Years in occupation			−0.0008	−0.0024**	−0.0022**
			(0.0010)	(0.0011)	(0.0011)
Field perception shrimpers			0.0195	0.0437**	0.0490**
			(0.0195)	(0.0203)	(0.0206)
Field belief shrimpers			−0.0140	−0.0286**	−0.0229*
			(0.0120)	(0.0134)	(0.0137)
Daily hours fishing			0.0050	0.0091	0.0039
			(0.0065)	(0.0060)	(0.0062)
Quantity of shrimp traps			−0.0001**	−0.0001*	−0.0000
			(0.0000)	(0.0000)	(0.0000)
Income			−0.0000	−0.0001*	−0.0001
			(0.0000)	(0.0000)	(0.0000)
Constant	0.4122***	0.4061***	0.3115***	0.4432**	0.4435**
	(0.0174)	(0.0164)	(0.0850)	(0.1700)	(0.1705)
Village fixed effects?	no	no	no	no	yes
Observations	114	83	112	83	83
R^2	0.064	0.051	0.272	0.400	0.484

Notes: ***99% significance, **95% significance; *90% significance. Robust standard errors in parentheses.

Il paper esamina il comportamento dei pescatori di gamberi attraverso un esperimento che misura la cooperazione e la pazienza, osservando come queste caratteristiche influiscano sulle pratiche di pesca sostenibile. In particolare, la dimensione dei fori nelle trappole per gamberi è considerata un indicatore della sostenibilità delle pratiche di pesca: fori più piccoli catturano gamberi non fertili, riducendo le risorse future.

La tabella 1 riassume i determinanti delle dimensioni dei fori nelle trappole per gamberi, utilizzando un'analisi di regressione OLS.

Diverse colonne rappresentano diverse regressioni con diversi insiemi di variabili esplicative, noi ci concentreremo sull'ultima colonna.

Qui, ogni variabile indipendente viene associata a un coefficiente che indica quanto la variabile influenza la dimensione dei fori nelle trappole. La significatività di questi coefficienti viene misurata tramite i **p-value** e rappresentata con asterischi:

- **Un asterisco (*):** significatività al 10% ($p < 0,10$)
- **Due asterischi (**)**: significatività al 5% ($p < 0,05$)
- **Tre asterischi (***)**: significatività al 1% ($p < 0,01$)

Dettagli sui coefficienti statisticamente significativi:

1. **Children:**
    - Coefficiente: **0,0109**
    - Significativo al **10%** (**$p < 0,10$**), indicato da **un asterisco**.
    - I risultati mostrano che il numero di figli è positivamente correlato alla dimensione media dei fori. Ogni incremento unitario di questa variabile (un figlio in più) aumenta la dimensione dei fori di circa 0,01 cm, suggerendo che i pescatori con figli tendono a utilizzare fori più grandi, lasciando così sfuggire i gamberi piccoli e non fertili.
2. **Years in occupation:**
    - Coefficiente: **-0,0022**
    - Significativo al **5%** (**$p < 0,05$**), indicato da **due asterischi**.
    - I risultati mostrano che il numero di anni trascorso a fare il pescatore è negativamente correlato alla dimensione media dei fori. Ogni incremento unitario di questa variabile (un anno in più) diminuisce la dimensione dei fori di circa 0,002 cm, suggerendo che i pescatori con più esperienza tendono a utilizzare fori più piccoli, dunque sono più inclini a sfruttare maggiormente le risorse attuali.

Dettagli sui coefficienti non significativi:

3. **Gender (male dummy):**
    - Coefficiente: **0,0515**
    - **Non significativo** ($p > 0,1$), senza asterischi.
    - Questo indica che il genere non è un fattore che influenza la scelta di usare trappole più o meno sostenibili in termini di dimensioni dei fori.

4. **Years of schooling:**
    - Coefficiente: **-0,0004**
    - **Non significativo** ($p > 0,1$), senza asterischi.
    - Questo indica che il livello di istruzione non è un fattore che influenza la scelta di usare trappole più o meno sostenibili in termini di dimensioni dei fori.