



## BIBLIOTECA DEI SAPERI

### Modelli per Dati cross-section

#### Introduzione (Slide 2)

Benvenuti!

In questa lezione vedremo che cosa è l'Econometria e perché è fondamentale per la misurazione economica. Tuttavia, la sua importanza si estende ben oltre la disciplina economica.

L'Econometria è anche un insieme di strumenti di ricerca impiegato nelle discipline aziendali di contabilità, finanza, marketing e management. Quindi i metodi di ricerca impiegati dagli economisti, che include il campo dell'Econometria, sono in effetti utili ad un ampio spettro di individui, tra cui anche uno scienziato dei dati.

Infine, tratteremo gli aspetti di base della Regressione semplice e multivariata e come sono implementati in R.

#### Slide 3 Che cosa è l'Econometria?

A questo punto dobbiamo descrivere la natura dell'Econometria.

Letteralmente Econometria significa misurazione dei fenomeni economici e secondo una possibile definizione rappresenta l'applicazione di metodi statistici e matematici per l'analisi dei dati economici, con il fine di dare validazione empirica alle teorie economiche, verificarle, oppure rigettarle. Tuttavia, l'estensione di tali metodi ad altri campi è molto diffusa.

Slide 4 In particolare, secondo una definizione di un altro econometrico (Frisch), ci sono diversi aspetti dell'approccio quantitativo in economia in cui l'approccio della statistica, quello della teoria economica e della matematica, sono necessari, ma non di per sé sufficienti, per una reale comprensione delle relazioni tra le variabili economiche. È l'unificazione di tutti e tre questi aspetti che è potente. Ed è questa unificazione che costituisce l'Econometria.

Slide 5 Inoltre, mentre la statistica si concentra su dati sperimentali, l'Econometria analizza i dati economici che non sono ripetibili e incerti al fine di vagliare la consistenza di una teoria economica con i dati, ricavando così una valutazione quantitativa dell'efficacia delle manovre di politica economica o di quantificare una certa relazione che ha una qualche rilevanza per le decisioni di impresa. Tuttavia, tali metodi possono essere utili nella quarta fase della scienza dei dati, cioè quando vogliamo creare un modello per i dati.

#### Slide 6 La Nozione di modello

Come già detto, il compito principale dell'Econometria è quantificare relazioni tra variabili economiche sulla base dei dati disponibili, usando tecniche statistiche. Quindi il processo che si segue è del tutto simile a quello descritto per la scienza dei dati. Nel caso dell'Econometria, un'analisi empirica usa i dati per sottoporre a verifica (test) una teoria o per stimare una relazione disponendo di un modello economico formale. In generale gli economisti formulano modelli che descrivono le relazioni fra diverse variabili, per esempio la relazione tra reddito e consumo, oppure tra salari e stipendi percepiti e il livello di scolarità, il genere, la regione geografica, ecc.



## Slide 7 Il Modello econometrico

Nella costruzione di un Modello econometrico si possono individuare tre fasi:

- 1) una prima che riguarda la specificazione del modello, ovvero la formalizzazione in termini matematico-statistici delle ipotesi teoriche in base alle informazioni empiriche a disposizione. Quindi si devono individuare le variabili rilevanti per l'analisi, scegliere la forma funzionale della legge sottostante il fenomeno studiato e definire le ipotesi sugli errori commessi
- 2) successivamente si passa alla stima dei parametri del modello, ovvero la quantificazione delle relazioni economiche studiate. In questa fase si cerca di individuare la struttura del modello che meglio approssima la vera struttura del fenomeno in oggetto (incognita)
- 3) una volta stimati i parametri è necessario verificare la validità del modello con una sequenza di operazioni sulla base delle osservazioni disponibili. Quindi è necessario valutare l'attendibilità della specificazione, la capacità descrittiva, la conformità alle aspettative teoriche, la capacità previsionale

## Slide 8 Prima formalizzazione

In economia esprimiamo le nostre idee su relazioni tra variabili economiche utilizzando il concetto matematico di funzione. Ad esempio, per esprimere una relazione tra reddito e consumi, potremmo scrivere la relazione Consumo Reddito che dice che il livello di consumo è una funzione,  $f()$ , del reddito. Oppure se vogliamo indagare la domanda di una singola merce, ad esempio uno specifico modello della Honda, che ci indichi la quantità di automobili Honda richiesta,  $Q_d$ , potremmo scrivere una funzione del prezzo della moto ( $P$ ), del prezzo delle auto sostitutive ( $P_s$ ), del prezzo degli articoli che sono complementari al  $P_c$  (come la benzina) e il livello di reddito  $INC$ .

## Slide 9 Errore nel modello

Ma che cos'è un modello econometrico e da dove viene?

In un modello econometrico dobbiamo prima di tutto renderci conto che le relazioni economiche non sono esatte. La teoria economica non sostiene essere in grado di prevedere il comportamento specifico di qualsiasi individuo o azienda, ma piuttosto descrive il comportamento medio o sistematico di molti individui o aziende. Infatti, quando studiamo le vendite delle auto dobbiamo riconoscere che il numero effettivo di Honda vendute è la somma di questa parte sistematica e di una componente casuale e imprevedibile e che chiameremo errore casuale. Allo stesso modo, un modello econometrico che rappresenta le vendite di Honda si ottiene sommando al precedente modello la componente  $e$ .

Per completare la specificazione del modello econometrico, dobbiamo anche dire qualcosa sulla forma della relazione algebrica tra le nostre variabili economiche. Ad esempio, la quantità domandata è rappresentata come una funzione lineare del prezzo e quindi nel modello per la vendita di Honda introduciamo anche questa specificazione.

## Slide 10 Parametri del modello

I coefficienti  $\beta_1$ ;  $\beta_2$ ; . . . ;  $\beta_5$  sono parametri ignoti del modello che stimiamo utilizzando dati economici e una tecnica econometrica. La forma funzionale rappresenta un'ipotesi sulla relazione tra le variabili.

In ogni problema particolare, la sfida è determinare una forma funzionale che sia compatibile con la teoria economica e i dati. In ogni modello econometrico, che sia un'equazione della domanda, un'equazione dell'offerta o una funzione di produzione, c'è una porzione sistematica e una componente casuale non osservabile. La parte sistematica è la parte che otteniamo dalla teoria economica e comprende un'ipotesi sulla forma funzionale. La componente casuale rappresenta un "rumore", che oscura la nostra comprensione della relazione tra variabili, e che rappresentiamo usando la variabile casuale  $e$ . Usiamo il modello econometrico come base per l'inferenza statistica.



**Slide 11** La teoria economica suggerisce molte relazioni tra variabili economiche. Ad esempio, in microeconomia si studiano i modelli di domanda e offerta in cui le quantità richieste e fornite di un bene dipendono dal suo prezzo. Ciascuno di questi modelli implica una relazione tra variabili economiche. Considereremo quindi come utilizzare un campione di dati economici per quantificare tali relazioni, utilizzando un modello di regressione in cui definiamo una variabile dipendente  $y$  (ad esempio consumi) e la variabile indipendente o esplicativa (ad esempio reddito). Come tutti i modelli anche il modello di regressione si basa su ipotesi.

## **Slide 12** I Dati

Per sviluppare le idee sui modelli di regressione useremo un semplice, ma importante, esempio economico. Supponiamo di essere interessati a studiare la relazione tra reddito familiare e spesa alimentare. Definiamo la popolazione di riferimento (famiglie residenti in una certa città, regione o stato) per poi estrarre dalla popolazione un campione casuale di un certo numero di famiglie con reddito medio di 100 euro al mese a cui chiediamo quale sia stata la loro spesa pro capite in beni alimentari la settimana precedente. La spesa familiare rappresenta la nostra variabile dipendente  $y$ , che è una variabile casuale poiché il valore ci è sconosciuto fino a quando non viene selezionata una famiglia a cui viene posta la domanda sulla spesa familiare. Inoltre, nell'analizzare la relazione economica dobbiamo riconoscere che stiamo descrivendo il comportamento medio o sistematico di molte famiglie, poiché la spesa in alimenti varierà da una famiglia all'altra per svariate ragioni (tipi di cibi che dipendono dai gusti, età media della famiglia, tipo di dieta, ecc.).

## **Slide 13** La Regressione

Per indagare il rapporto tra spesa e reddito dobbiamo costruire un modello economico e quindi un corrispondente modello econometrico, il quale costituisce la base per l'analisi economica quantitativa o empirica. Quindi introduciamo il Modello di regressione. Tuttavia, prima di decidere quale Modello di regressione adottare, è preferibile analizzare i dati a disposizione mediante una rappresentazione grafica per poi passare alla stima econometrica.

## **Slide 14** La Rappresentazione grafica

Come già detto precedentemente, uno scatterplot fornisce molte informazioni sul tipo di relazione che esiste tra le nostre variabili economiche. In questo grafico abbiamo rappresentato sull'asse delle  $x$  il reddito delle famiglie in centinaia di dollari e sull'asse delle  $y$  la loro spesa familiare in beni alimentari. Dall'ispezione grafica possiamo individuare una Relazione lineare tra le due variabili.

## **Slide 15** La Rappresentazione matematica

Nella maggior parte dei libri di testo di economia, le funzioni di "consumo" o "spesa" che mettono in relazione consumo e reddito sono rappresentati come relazioni lineari e inizieremo assumendo la stessa cosa. La rappresentazione matematica del nostro modello economico di spesa alimentare è scritta nell'equazione ed è rappresentata nel grafico. È chiamato un Modello di regressione semplice non perché sia facile, ma perché abbiamo solo una variabile esplicativa sul lato destro dell'equazione. I parametri di regressione incogniti  $\beta_1$  e  $\beta_2$  sono rispettivamente l'intercetta e la pendenza della funzione di regressione.

Nel nostro esempio sulla spesa alimentare l'intercetta  $\beta_1$  rappresenta la media settimanale della spesa per cibo da parte di un nucleo familiare con reddito settimanale nullo. Invece, se il reddito è misurato in dollari, la pendenza  $\beta_2$  ci dice di quanto varia la spesa per una data variazione del reddito disponibile e si potrebbe chiamare la propensione marginale alla spesa per il cibo.

## **Slide 16** La Rappresentazione econometrica



Il modello rappresentato nella precedente equazione descrive il comportamento economico, ma è un'astrazione dalla realtà. Se prendiamo un campione casuale di famiglie con reddito settimanale di 1000 dollari, sappiamo che i valori di spesa effettivi saranno distribuiti attorno al valore medio come mostrato nella figura. Se dovessimo considerare un campione per la spesa delle famiglie ad altri livelli di reddito (ad esempio 2000 dollari), ci aspetteremmo che i valori del campione siano sparsi attorno al loro valore medio. Nella figura abbiamo rappresentato con una distribuzione a forma di campana la spesa alimentare lungo la regressione lineare per ogni livello di reddito. Inoltre, abbiamo assunto che la dispersione dei valori  $y$  rispetto alla loro media sia la stessa per tutti i livelli di reddito  $x$ . Tale ipotesi di varianza costante implica che per ogni livello di reddito  $x$  siamo ugualmente incerti su quanto i valori di  $y$  potrebbero cadere dal loro valore medio e l'incertezza non dipende dal reddito o altro. Dati che soddisfano questa condizione si dicono omoschedastici. Se questa ipotesi viene violata, allora i dati sono detti eteroschedastici.

### Slide 17 Errore

L'essenza dell'analisi di regressione è che qualsiasi osservazione sulla variabile dipendente  $y$  può essere scomposta in due parti:

- 1) una componente sistematica
- 2) e una componente casuale

La componente sistematica di  $y$  è la sua media, che abbiamo rappresentato nella precedente equazione, che di per sé non è casuale poiché è una funzione matematica. La componente casuale di  $y$  è la differenza tra  $y$  e il suo valore medio condizionale. Questa è chiamata **Termine di errore casuale** ed è definito nella prima equazione.

Se riorganizziamo le due equazioni otteniamo il Modello di regressione lineare semplice della seconda equazione della slide, in cui la variabile dipendente  $y$  è spiegata da una componente che varia sistematicamente con la variabile indipendente  $x$  e dal termine di errore casuale  $e$ . Poiché  $y$  è casuale, lo è anche il termine di errore  $e$ . Il valore medio del termine di errore, dato  $x$ , è zero. Quindi sia  $e$  che  $y$  sono variabili casuali, con la differenza che  $y$  è osservabile, mentre  $e$  è non osservabile.

### Slide 18 Errore e Retta di regressione

Se i parametri di regressione  $\beta_1$  e  $\beta_2$  fossero noti, allora per qualsiasi valore di  $y$  si potrebbe calcolare  $e$  come illustrato nella figura. Conoscendo la funzione di regressione potremmo separare  $y$  nelle sue componenti, sistematica e casuale. Tuttavia, i parametri non sono mai noti ed è impossibile calcolare  $e$ .

Che cosa comprende il termine di errore  $e$ ? L'errore casuale  $e$  rappresenta tutti i fattori che influenzano  $y$  diversi da  $x$ . Questi fattori fanno sì che le singole osservazioni  $y$  differiscano dal valore medio. Nell'esempio della spesa alimentare, quali fattori possono comportare una differenza tra la spesa familiare pro capite  $y$  e la sua media,  $E(y)$ ? Abbiamo incluso il reddito come unica variabile esplicativa in questo modello. Qualsiasi altro fattore economico che influisce sulla spesa per il cibo viene "raccolto" nel termine di errore.

### Slide 19 La Spesa alimentare: i dati

Prima di passare alla stima è necessario ottenere i dati che possono essere organizzati in una forma strutturata come nella tabella in questa slide in cui per 40 famiglie disponiamo della spesa alimentare e del reddito. Quindi abbiamo una matrice con 3 variabili numeriche (colonne) e 40 famiglie (righe). L'Econometria affronta il problema di usare i dati campionari  $(x_i, y_i)$  per ottenere le stime numeriche dei parametri incogniti  $\beta_1$  e  $\beta_2$ .

### Slide 20 La Stima dei minimi quadrati

Assumendo che i dati di spesa nella tabella precedente soddisfino la prima ipotesi che il valore atteso della spesa alimentare delle famiglie sia una funzione lineare del reddito, ipotizziamo che l'errore casuale abbia valore atteso zero, ovvero che non abbiamo omissio nessun fattore importante. Inoltre, supponiamo che la varianza di  $y$  sia



costante, ovvero siamo ugualmente incerti riguardo alla relazione tra  $y$  e  $x$  per tutte le osservazioni. Inoltre, presumiamo che i valori di  $y$  per le diverse famiglie non siano correlati tra loro, possiamo ottenere una stima dei parametri utilizzando uno dei possibili metodi, ovvero una regola, o formula, che ci dica come utilizzare le osservazioni nel campione. Sono possibili molte regole, ma quella che useremo si basa sul principio dei minimi quadrati per il quale la somma dei quadrati delle distanze verticali di ciascun punto dalla retta sia più piccola possibile. Le distanze sono al quadrato per evitare che grandi distanze positive vengano cancellate da grandi distanze negative.

### Slide 21 La Stima dei minimi quadrati in R

In R è molto semplice stimare una regressione mediante il metodo dei minimi quadrati. La funzione da utilizzare è `lm` nella quale l'input è espresso da una formula  $y \sim x$  con  $y$  variabile dipendente (nel nostro caso `food_exp`) e  $x$  variabile esplicativa (nel nostro caso `income`) contenuti nell'oggetto `food` che viene passato ad `lm` con l'opzione `data=`

### Slide 22 Spesa alimentare: la Stima della retta di regressione

Nel grafico è rappresentata la nuvola di punti insieme alla retta di regressione stimata con valori per il parametro `beta_1` uguale a 83,42 e `beta_2` uguale a 10,21

### Slide 23 Verifica d'ipotesi: in generale

Dopo la stima dei parametri del modello di regressione lineare la terza fase prevede la verifica della bontà di adattamento del modello. Una prima analisi è la formulazione di ipotesi sul comportamento economico mediante affermazioni sui parametri del modello. In particolare, le verifiche di ipotesi usano le informazioni sui parametri contenuta nel campione di dati, rappresentata dalle stime puntuali dei minimi quadrati e i rispettivi standard error, per giungere ad una conclusione sulla validità delle ipotesi.

### Slide 24 Verifica d'ipotesi: Stima minimi quadrati

Quindi è necessario definire un test di ipotesi che prevede la definizione dell'ipotesi nulla ( $H_0$ ) in cui viene testato se la stima del parametro è uguale ad una costante ( $c$ ) che costituisce un valore importante nel contesto di un modello di regressione specifico (un valore comune per  $c$  è 0). Definita l'ipotesi alternativa ( $H_1$ ) flessibile e che dipende in una certa misura dalla teoria economica (nel nostro caso parametro diverso da 0), sulla base del valore di una statistica test decidiamo di rifiutare l'ipotesi nulla o di non rifiutarla. La distribuzione di probabilità della statistica test è completamente nota quando l'ipotesi nulla è vera. Nel nostro esempio `t` indica la statistica test che ha distribuzione  $t$  di Student.

### Slide 25 Verifica d'ipotesi: p-value in R

In generale nei test di ipotesi è necessario definire una zona di accettazione. Tuttavia, quando si illustrano i risultati di un test statistico di ipotesi, è diventata pratica standard riportare il p-value. Conoscendo il valore del p-value di un test,  $p$ , possiamo determinare l'esito del test confrontando il valore  $p$  con il livello di significatività scelto,  $\alpha$ , senza cercare o calcolare i valori critici. In R utilizzando la funzione `summary` dell'oggetto in cui è stato salvato l'output della stima abbiamo subito il valore del p-value. Nel nostro esempio il p-value è cerchiato e dice che è molto basso e quindi si può rifiutare l'ipotesi che il parametro è uguale a 0.

### Slide 26 Una misura di adattamento del modello: $R^2$

Per la valutazione della bontà di adattamento è possibile calcolare il coefficiente di determinazione  $R^2$  che misura la variazione in  $y_i$  che è spiegata dal modello.  $y_i$  è diviso nella componente spiegata dal modello e quella non spiegata dal modello utilizzando il fatto che come mostrato nella figura la retta di regressione stimata mediante i minimi quadrati passa attraverso il baricentro della nostra nuvola di punti.



## Slide 27 Coefficiente di determinazione $R^2$

Il Coefficiente di determinazione può essere calcolato come rapporto tra la somma dei quadrati dovuta al modello di regressione che misura la parte della variazione totale in  $\$y$  attorno alla media campionaria spiegata dal modello (SQM) e la somma dei quadrati totale che misura la variazione totale in  $\$y$  attorno alla media campionaria (SQT). Quanto  $R^2$  è più vicino a 1, tanto più i valori del campione  $y_i$  sono vicini all'equazione di regressione adattata. Se  $R^2 = 1$ , tutti i punti campionari sono esattamente sulla linea dei minimi quadrati, quindi  $SQR = 0$ , e il modello si adatta ai dati "perfettamente". Se i dati per  $y$  e  $x$  non sono correlati e non mostrano alcuna associazione lineare, allora la stima della retta di regressione è orizzontale e identica a  $y$  barrato, quindi  $SQM = 0$  e  $R^2 = 0$ .

## Slide 28 Coefficiente di determinazione $R^2$ in R

Dall'esempio per la relazione tra spesa alimentare e reddito abbiamo che l' $R$ -quadro è pari a 0.385 concludendo che il 38,5% della variazione nella spesa alimentare (rispetto alla sua media campionaria) è spiegato dal nostro modello di regressione, che utilizza solo il reddito come variabile esplicativa. Tuttavia, altri criteri da tener presente al fine di valutare il modello sono il segno e la grandezza delle stime, la loro significatività statistica ed economica, la loro precisione. Inoltre, anche la capacità del modello stimato di prevedere valori della variabile dipendente che non appartengono al campione di stima.

## Slide 29 Analisi grafica dei residui

Per concludere il tema della verifica della bontà di adattamento del modello è importante considerare i residui sui cui si possono effettuare test. Tuttavia, un'ispezione grafica già fornisce indicazioni importanti per individuare non corrette specificazioni della dipendenza dalle variabili esplicative, come ad esempio:

- dipendenze non lineari (mediante il grafico variabili esplicative verso i residui)
- se le ipotesi di omoschedasticità, media nulla e non correlazione dei residui sono verificate (grafico valori stimati dal modello verso i residui)
- se i residui hanno una distribuzione normale (normal probability plot)

## Slide 30 Corretta specificazione

In R è facile produrre i grafici per l'analisi dei Residui mediante la funzione plot. In questi grafici i risultati mostrano che il modello è stato specificato bene. I Residui hanno media nulla, non sono correlati e hanno distribuzione normale.

## Slide 31 Un Modello di regressione lineare multipla

Concludiamo questa sezione sulla stima di modelli per dati cross-section introducendo il modello di Regressione multipla, riprendendo l'esempio iniziale della vendita di auto Honda.

## Slide 32 Esempio: due variabili esplicative

Semplifichiamo considerando un modello economico in cui i ricavi dipendono da una o più variabili esplicative, ovvero che i ricavi siano linearmente correlati al prezzo e alla spesa pubblicitaria. I risultati individuati per la Regressione semplice rimangono validi. In questo caso cambia solo l'interpretazione dei parametri.

In particolare,  $\beta_1$  è la variazione dei ricavi (in migliaia di dollari) quando il prezzo (Price) è aumentato di un'unità (in dollari) e la spesa pubblicitaria (Advert) è tenuta costante e  $\beta_2$  è la variazione dei ricavi (in migliaia di dollari) quando la spesa pubblicitaria (Advert) è aumentata di un'unità (in migliaia di dollari) e il prezzo (Price) è tenuto costante

## Slide 33 Stima e test di ipotesi: parametri



La stima in R nel caso di Regressione multipla viene effettuata sempre utilizzando la funzione `lm()`, l'output è lo stesso e anche l'interpretazione del p-value rimane la stessa. In questo esempio il valore del p-value molto basso ci porta a concludere che tutti i parametri sono diversi da 0.

### Slide 34 Test d'ipotesi congiunta

Rispetto alla Regressione semplice, nel caso di quella multipla si considerano anche test di ipotesi congiunte sui parametri mediante la statistica test F, calcolata usando una semplice formula basata sulla somma dei quadrati dei residui di due Regressioni:

- 1) una prima Regressione, chiamata Regressione vincolata (V), l'ipotesi nulla assume che i coefficienti siano zero escludendo, quindi, i regressori rilevanti dalla Regressione
- 2) nella seconda Regressione, chiamata Regressione non vincolata (NV), è valida invece l'ipotesi alternativa. R calcola nel suo output la statistica F riportata in questa formula

### Slide 35 Verifica complessiva del modello

Nel nostro modello introduciamo una Regressione quadratica, introducendo come variabile esplicativa anche il quadrato della variabile `Advert` e verifichiamo l'ipotesi nulla che tutti i parametri siano uguali a 0, ovvero che il modello non abbia buone capacità esplicative.

### Slide 36 Statistica F in R `lm()`

Utilizzando la funzione `summary` per l'oggetto in cui abbiamo salvato la stima della Regressione otteniamo il seguente output. Anche per la statistica F è sufficiente considerare il p-value- Nel nostro esempio il p-value è molto basso e quindi possiamo rifiutare l'ipotesi nulla.

### Slide 37 Scelta del modello

Un aspetto molto importante è la costruzione di criteri per la scelta di un modello tra diverse alternative. Gli strumenti principali per costruire un modello dovrebbero essere la teoria economica, un valido ragionamento basato su principi economici e fare in modo che il modello soddisfi le ipotesi sugli errori. Si potrebbero poi verificare le situazioni di distorsioni da variabili omesse e dell'esclusione di variabili irrilevanti che possono aumentare la variabilità delle stime.

Dopo aver considerato tutti questi aspetti e stabilito un modello, ci sono alcune grandezze che aiutano a confrontare diversi modelli, quali il criterio di informazione di **Akaike** (AIC), e il criterio di **Schwarz** (o Bayesian information) (SC o BIC) in base ai quali il miglior modello è quello che minimizza i criteri.

### Slide 38 Multicollinearità

Nel caso della regressione multipla è necessario verificare se uno dei regressori è una esatta combinazione lineare degli altri regressori. A seconda dei software, il pacchetto gestisce la perfetta multicollinearità in due modi: ometterà una delle variabili o rifiuterà di calcolare le stime OLS e darà un messaggio di errore (Perfetta multicollinearità). Inoltre, è necessario verificare i casi di multicollinearità imperfetta che si verifica quando uno dei regressori è altamente correlato, anche se non perfettamente con gli altri regressori. A differenza della perfetta multicollinearità, la multicollinearità imperfetta non impedisce la stima della regressione. Tuttavia, implica che uno o più coefficienti della regressione possa essere stimato in modo impreciso.

### Conclusioni Slide 39

Bene, siamo giunti alla fine di questa video lezione.

Ti ricordo che abbiamo introdotto che cos'è l'Econometria.



In particolare abbiamo visto:

- la Regressione semplice e multivariata
- come questi aspetti sono implementati in R

Grazie per l'attenzione!