ECONOMETRIA: Laboratorio I

Luca De Angelis

CLASS - Università di Bologna

Programma Laboratorio I

- Valori attesi e varianze
- Test di ipotesi
- Stima di un modello lineare attraverso OLS

Valore atteso

- Data una variabile aleatoria X con pdf $f(x)$, il valore atteso di X è:

$$E(X) = \mu_x = \sum xf(x) \quad \text{v.c. discreta}$$
$$E(X) = \mu_x = \int xf(x) \quad \text{v.c. continua}$$

- Proprietà:
  - $E(a) = a$
  - $E(aX + b) = aE(X) + b$
  - $E(aX_1 + ... + nX_n) = aE(X_1) + ... + nE(X_n)$
  - se X e Y sono indipendenti allora: $E(XY) = E(X)E(Y)$

Varianza

- Data una variabile aleatoria X, la varianza di X, è definita:

$$\begin{aligned} Var(X) &= \sigma_x^2 = E[(X - E(X))^2] = E(X^2) - E(X)^2 \\ &= E\left[(X - \mu_x)^2\right] \end{aligned}$$

- Proprietà:
  - $Var(aX + b) = a^2 Var(X)$
  - $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
  - $Var(aX_1 + ... + nX_n + d) = a^2 Var(X_1) + ... + n^2 Var(X_n)$

Esercizio 1

- X é la variabile aleatoria "somma di due dadi"
- Scrivere la distribuzione di probabilità
- Calcolare $E(X)$ e $Var(X)$

Covarianze e correlazione

- Date due variabili aleatorie X e Y

$$
\begin{aligned}
Cov\,(X, Y) = \gamma(k) &= E\left[(X - \mu_X)\,(Y - \mu_Y)\right] \\
&= E\,(XY) - E\,(X)\,E\,(Y) \\
Corr\,(X, Y) = \rho(k) &= \frac{Cov\,(X, Y)}{\sqrt{Var\,(X)\,Var\,(Y)}}
\end{aligned}
$$

- Nel caso di Time-Series, se il processo è stazionario

$$Corr\,(X, Y) = \rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Valore atteso condizionato

- Date due variabili aleatorie X e Y il valore atteso condizionato:

$$
\begin{aligned}
E(X|Y=y) &= \sum_{x \subset X} xP\left(X=x|Y=y\right) \\
&= \sum_{x \subset X} x \frac{P(X=x, Y=y)}{P(Y=y)}
\end{aligned}
$$

- Nota:

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$condizionata = \frac{congiunta}{marginale}$$

Esercizio 2

- Data la seguente distribuzione di probabilità:

	$X = 0$	$X = 1$	*Totale*
$Y = 0$	0.15	0.07	0.22
$Y = 1$	0.15	0.63	0.78
Totale	0.30	0.70	1

- Calcolare:
  - $E(X), E(Y), Var(X), Var(Y), Corr(X, Y)$
- Data la variabile V=20-7Y calcolare:
  - $E(V), Var(V)$
- Calcolare:
  - $E(Y|X = 1)$

Test d'ipotesi

- Test d'ipotesi: valuta un congettura su un parametro $\theta \in \Theta$ di interesse della popolazione, avendo osservato un determinato campione.
- H_0:ipotesi nulla **vs** H_1:ipotesi alternativa.
- Per valutare H_0 si utilizza una *statistica test*, ossia una funzione di H_0 e dei dati osservati.
- Spesso la distribuzione di probabilità della statistica test sotto l'ipotesi nulla è nota (es: Normale, t di Student, Chi quadrato).

Test d'ipotesi

Come stabiliamo quando un campione fornisce una chiara indicazione contro l'ipotesi nulla?

- Un test statistico è una regola per discriminare i campioni che, se osservati, portano al rifiuto di H_0
- La regione di rifiuto di un test statistico è formata dai campioni che contengono "abbastanza" evidenza contro H_0.
- La regione di accettazione di un test statistico è formata dai campioni che <u>non</u> contengono "abbastanza" evidenza contro H_0.
- La regione di rifiuto e quella di accettazione bipartiscono lo spazio campionario, consentendoci di rifiutare o meno l'ipotesi nulla sulla base del campione osservato

Test d'ipotesi

Errori del test

- La regione di rifiuto consisterà di valori estremi sotto l'ipotesi nulla.
- La verifica di H_0 darà un risultato errato in due casi
  - H_0 è vera e noi la rifiutiamo, errore di prima specie
  - H_0 è falsa e noi la accettiamo, errore di seconda specie
- $\text{P}(\text{Rifiutare } H_0 | H_0 \text{ vera}) = \alpha$
- $\text{P}(\text{Accettare } H_0 | H_0 \text{ falsa}) = \beta$
- $\text{P}(\text{Rifiutare } H_0 | H_0 \text{ falsa}) = 1 - \beta = \gamma$

- Supponiamo di dover testare se la media di una popolazione μ sia uguale ad un determinato valore μ_0
- Avremo quindi l'ipotesi nulla $H_0 : \mu = \mu_0$ contro l'ipotesi alternativa $H_1 : \mu \neq \mu_0$ (test bidirezionale).
- Dopo aver estratto un campione di numerosità n, possiamo calcolare la media campionaria $\bar{Y}$ e la deviazione standard campionaria corretta $\hat{\sigma}$.
- A questo punto, possiamo testare H_0 considerando la statistica test:

$$T = \frac{\bar{Y} - \mu_0}{\hat{\sigma}/\sqrt{n}} \sim t_{(n-1)}$$

dove $\hat{\sigma}$ è la deviazione standard corretta e t è una v.c t di Student con $n-1$ gradi di libertá.

Test d'ipotesi

Valore critico

Come valutiamo se la statistica test ottenuta ci porta a rifiutare o a non rifiutare l'ipotesi nulla H_0?

- Definiamo la regione di rifiuto come l'insieme dei campioni per i quali
$$|T| > t_\alpha \quad (T < -t_\alpha, T > t_\alpha)$$
- Il valore critico t_α è scelto in modo che
$$P(T > t_\alpha | H_0) = \alpha$$
- Per un opportuno α e sulla base della distributione sotto H_0 (in questo caso abbiamo una t di Student con n-1 gradi di libertà) costruiremo il valore critico t_α.

Test d'ipotesi

p-value

- Le conclusioni del che otteniamo dal nostro test dipendono dal livello di significatività α che scegliamo.
- Nella pratica quindi si calcola il p-value.

$$P\left(|T| > t|H_0\right) = p$$

- Rappresenta la proporzione di campioni in cui, sotto l'ipotesi nulla, la statistca test T assume valore più estremo del valore osservato t.
- In pratica, valori piccoli di p indicano evidenza contro H_0.

Gretl e Test d'ipotesi

- Dal menù di Gretl: *Strumenti - Tavole statistiche*
- Dal menù di Gretl: *Strumenti - Calcolo p-value*
- Dal menù di Gretl: *Strumenti - Calcola Test*

Esercizio 3: test per la media

Si dispone di tre campioni di $n = 8$ elementi

I Campione	II Campione	III Campione
15	17	31
16	18	18
13	16	27
15	14	22
17	22	15
17	20	26
13	21	26
17	22	30

Table: Esercizio 3

- Ipotesi nulla da testare: $H_0 : \mu = 15$
- Utilizzando Gretl calcolare la statistica test t per i tre diversi campioni osservati
- Calcolare il p-value associato ai tre test.

Esercizio 3: test per la media

Indicazioni per lo svoglimento

- Creare un nuovo dataset direttamente su Gretl
  - Dal menù di Gretl: *File - Nuovo Dataset* e inserire il numero delle osservazioni e la tipologia di dati che vogliamo analizzare (8 osservazioni e dati cross-section). Barrare la casella 'Inizia a inserire i valori dati' per inserire i dati relativi ai tre campioni.
- Definire nuove variabili per la media (mean), la deviazione standard (sd) e la statistica t per ogni campione.
- Utilizzare il tool di Gretl per calcolare i p-value.
- In alternativa: *Strumenti - Calcola Test - Media*.

Test d'ipotesi
Confronto tra medie

- Un caso di possibile interessere è quello in cui vogliamo confrontare le medie di delle popolazioni da cui abbiamo estratto due campioni differenti.
- Denotiamo μ_X e μ_Y le medie delle due popolazioni di interesse, per cui la nostra ipotesi nulla sarà data da

$$H_1 : \mu_X = \mu_Y$$

  contro l'ipotesi alternativa

$$H_1 : \mu_X \neq \mu_Y$$

- Esercizio: Costruire un test per il confronto tra medie con i dati dell'esercizio 3. (Suggerimento: Utilizzare il menù di Gretl)

Capital Asset Pricing Model (CAPM)

Teoria economica

- Il CAPM rappresenta il primo modello di equilibrio dei prezzi delle attività finanziarie in una situazione di incertezza e consente di fornire una soluzione ad uno dei principali problemi della finanza, la relazione tra rischio e rendimento.

Capital Asset Pricing Model (CAPM)

Formulazione

- Il CAPM si può rappresentare come una relazione lineare tra il rischio non diversificabile e il rendimento atteso

$$E\left(r_i\right) = r_f + \left(E\left(r_m\right) - r_f\right)\beta_i$$

- Un'importante classe di test sul CAPM è stata sviluppata nel contesto della regressione lineare con osservazioni temporali

$$r_{i,t} - r_{f,t} = \alpha + \beta(r_{m,t} - r_{f,t}) + \varepsilon_{i,t} \quad t = 1, \ldots T$$

Capital Asset Pricing Model (CAPM)

Return - Log-return

- Dato P_t il prezzo di un'attività finanziaria il rendimento può essere calcolato

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}}$$

oppure in termini di log-return

$$\begin{aligned} r_t &= \log\left(\frac{P_t}{P_{t-1}}\right) = \log\left(\frac{P_t - P_{t-1} + P_{t-1}}{P_{t-1}}\right) \\ &= \log\left(1 + \frac{\Delta P_t}{P_{t-1}}\right) = \log(1 + R_t) \end{aligned}$$

- Se il valore di R_t non è troppo alto o troppo basso il log-return è una buona approssimazione del Rendimento poichè

$$\log(1 + R_t) \approx R_t$$

Stimatore OLS

In caso di omoschedasticità

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \left(X'X\right)^{-1}\left(X'y\right) \\
E\left(\hat{\beta}_{OLS}\right) &= \beta \\
Var\left(\hat{\beta}_{OLS} \mid X\right) &= Var\left(\beta + (X'X)^{-1}X'u \mid X\right) \\
&= Var\left((X'X)^{-1}X'u \mid X\right) \\
&= (X'X)^{-1}X'\left(\sigma_u^2 I_n\right)X(X'X)^{-1} \\
&= \sigma_u^2 (X'X)^{-1}
\end{aligned}
$$

Capital Asset Pricing Model (CAPM)

Su Gretl

- Importare il file excel ASSETS.
- Generare le seguenti variabili:
  - Differenze prime del prezzo $(X_t - X_{t-1})$
  - Logaritmo del prezzo
  - Log-return
- Stima del modello di mercato:

$$r_{i,t} - r_{f,t} = \alpha + \beta(r_{m,t} - r_{f,t}) + \varepsilon_{i,t} \quad t = 1, \ldots T$$

Output del modello di Regressione

Su Gretl

- Coefficienti stimati $\hat{\beta}$
- Standard Error delle stime $\sqrt{\hat{V}(\hat{\beta})}$
- Rapporto t sotto l'ipotesi nulla $H_0 : \beta = 0$
- *p-value* relativo al rapporto t
- F e p-value(F): test di significatività globale dei coefficienti
- Coefficiente di determinazione lineare R^2

Esercizio 4: Capital Asset Pricing Model (CAPM)

- Stimare lo stesso modello modificando il titolo i e commentare l'output.

Stimatore vincolato

Teoria

- Stimatore Restricted Least Square:

$$\hat{\beta}_{RLS} = \hat{\beta}_{OLS} - \left(X'X\right)^{-1} R \left[\left(R' \left(X'X\right)^{-1} R\right)\right]^{-1} \left(R'\hat{\beta}_{OLS} - r\right)$$

- Varianza dello stimatore β_{RLS}:

$$= V\left(\hat{\beta}_{OLS}\right) - \sigma^2 \left(X'X\right)^{-1} R \left[\left(R' \left(X'X\right)^{-1} R\right)\right]^{-1} R' \left(X'X\right)^{-1}$$

Capital Asset Pricing Model (CAPM)

Test d'ipotesi sui parametri

1. $H_0 : \beta = 0$
2. $H_0 : \beta = 1$
3. $H_0 : \alpha = 0$

- Stimare il modello di mercato sotto il vincolo 1.
- Stimare il modello di mercato sotto il vincolo 2.
- Stimare il modello di mercato sotto i vincoli 2 e 3.