Econometria: prima parte

Note per preparare l'esame con i miei appunti:

Gli appunti che seguono comprendono tutte le nozioni per poter passare bene l'esame di Econometria del professor Deana.

Oltre agli appunti presi in classe, sono presenti, all'inizio del documento, alcune nozioni fondamentali di statistica che servono per ripassare e rinfrescare temi fondamentali per l'approccio all'Econometria. Non provare nemmeno a preparare l'esame senza aver ben chiaro questi argomenti perché sarà un vero e proprio casino! 😉

Alla fine degli appunti, invece, c'è una parte di domande che sono ricorrenti nei testi di esame. Si tratta di 32 domande con risposta nelle ultime tredici pagine di questi appunti

Infine, la materia è molto ostica. Purtroppo, da parte del professore, non c'è sempre stata la chiarezza necessaria per poter capire profondamente tutto. Noterai che ci sono parti dei miei appunti più o meno chiare e questo è proprio dovuto al fatto di cui sopra. Me ne dispiaccio e ti auguro in bocca al lupo!

Per info scrivimi liberamente a nicolo.badino@gmail.com

Richiami di Probabilità e Statistica

Spazio campionario: è l'insieme di tutti i risultati possibili, un evento è un sottoinsieme dello spazio campionario.

Distribuzione di probabilità: è l'elenco di tutti i possibili valori di una variabile e della probabilità con cui ciascuno di essi si verifica. La somma delle probabilità di tutti questi possibili valori è uguale a 1.

Probability Density Function: Poiché una variabile continua può assumere un numero infinito di valori la pdf rappresenta, attraverso la sua area sottesa, la probabilità che vengano assunti un determinato range di valori.

Valore Atteso: il valore atteso di una variabile casuale Y, indicato con E(Y), è il valore medio di tale variabile casuale calcolato sulla base di un numero elevato di prove ripetute. Il **valore atteso** di y è detto anche media di y.

Varianza: misura, come la deviazione standard, la dispersione di probabilità. La varianza di una variabile Y è il valore atteso del quadrato della deviazione di Y dalla sua media. Cioè → $var(Y) = E[(Y - \mu Y)^2]$

Deviazione Standard: è la radice quadrata della varianza ed è indicata con δ_Y ed ha la stessa unità di misura di Y.

Momenti: La media di Y è detta anche momento primo di Y e il valore atteso del quadrato di Y è anche detto momento secondo. In generale il valore atteso di Y è detto momento r-esimo della variabile casuale di Y

Indipendenza: Due variabili casuali X e Y sono indipendenti se conoscere il valore di una di esse non fornisce alcuna informazione dell'altra. In particolare, X e Y sono indipendenti se la distribuzione condizionata di Y data X è uguale alla distribuzione marginale di Y. In altri termini se la probabilità di Y è 23%, conoscendo X tale probabilità rimane uguale.

Covarianza: Una misura dell'intensità con la quale due variabili casuali si muovono assieme è la loro covarianza. La covarianza fra X e Y è il valore atteso $E\left[\left(X_j - \mu x\right)(Y_i - \mu y)\right]$ ovvero quanto si distacca il valore dell'osservazione j-esima della x dalla sua media e quanto si distacca l'osservazione i-esima della ipsilon dalla sua media. Ipotizziamo che X sia minore della media e che Y lo sia anche allora avremo il prodotto di due valori negativi che restituiranno una covarianza positiva (covariano assieme). Se invece X è maggior della sua media e Y minore (o viceversa) avremo una correlazione negativa. Se invece Y e X sono indipendenti allora la loro covarianza sarà pari a zero.

Correlazione: Dal momento che la covarianza è il prodotto fra X e Y espresse in deviazione dalle rispettive medie, la sua unità di misura è espressa nell'unità di misura di X moltiplicata per l'unità di misura di Y. La Correlazione risolve questo problema. La correlazione fra X e Y è la covarianza fra esse divisa per il prodotto delle loro deviazioni standard.

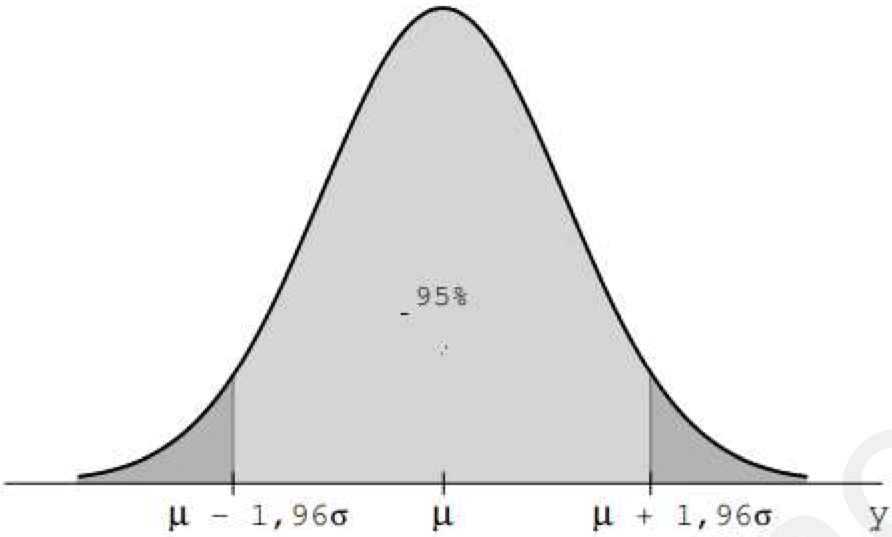
$$corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$

La correlazione è sempre compresa tra -1 e 1.

Proprietà interessanti:

1. **Media di somma:** $E(X + Y) = E(X) + E(Y) = \mu x + \mu y$
2. **Var v. dipendenti:** $var(X + Y) = var(X) + var(Y) + 2cov(X, Y) = \delta_X^2 + \delta_Y^2 + 2\delta_{X\,Y}$
3. **Var. v. indipendenti** $var(X + Y) = var(X) + var(Y) + 2cov(X, Y) = \delta_X^2 + \delta_Y^2$

Distribuzione normale: è una distribuzione di probabilità che è simmetrica attorno alla sua media e concentra il 95% della sua probabilità tra $\mu - 1{,}96\delta \; e \; \mu + 1{,}96\delta$.



La distribuzione normale si indica con $N(\mu, \delta^2)$ in cui la media è il centro della campana mentre la varianza è la larghezza.

Si chiama **distribuzione normale standard** quella particolare distribuzione con media uguale a 0 e varianza uguale a 1.

Le variabili casuali aventi distribuzione N(0,1) sono spesso indicate con Z.

Distribuzione chi-quadrato: è la distribuzione della **somma dei quadrati di m variabili casuali indipendenti** ognuna con una distribuzione normale standard. Questa distribuzione dipende da m che è chiamato numero di gradi di libertà della distribuzione chi-quadrato.

Ipotizziamo di avere tre variabili casuali che si distribuiscono attraverso una normale Z_1, Z_2, Z_3.

Allora per costruire la chi-quadro si fa la somma dei quadrati $Z_1^2 + Z_2^2 + Z_3^2$ che ha tre gradi di libertà. Si chiama chi-quadro perché è rappresentata simbolicamente così: χ_m^2

Distribuzione F: è una distribuzione strettamente legata alla chi-quadro. Consiste in una distribuzione chi-quadro divisa per i gradi di libertà.

Prendendo l'esempio di prima:

$$\frac{Z_1^2 + Z_2^2 + Z_3^2}{3}$$

Distribuzione t di Student: è la distribuzione del rapporto di due variabili casuali indipendenti, la prima delle quali è normale standard e l'altra è la radice quadrata di una variabile casuale chi-quadro con m gradi di libertà divisa per m. Matematicamente:

Sia Z una variabile casuale normale standard, sia W una variabile casuale con distribuzione chi-quadro con m gradi di libertà e siano Z e W indipendentemente distribuite, allora:

$$\frac{Z}{\sqrt{\frac{W}{m}}} \; si\ distribuisce\ con\ una\ t\ di\ Student$$

La distribuzione t di Student dipende dai gradi di libertà m. Ha una forma campanulare simile a quella della distribuzione normale ma quando m è piccolo (<20) si dice che ha più “massa” nelle code. Quando m è grande (>30) allora la t di Student è approssimata bene dalla normale standard.

Stimatori e rispettive proprietà:

Uno stimatore è una funzione di un campione di dati estratti casualmente di una popolazione. La **stima** è il valore numerico dello stimatore

Non distorsione: Supponiamo di calcolare uno stimatore diverse volte. Se lo stimatore funziona bene è ragionevole pensare che in media, con tante osservazioni, si ottenga il valore giusto. Una caratteristica molto ricercata in uno stimatore è che in media dia i risultati corretti. Si dice dunque che uno stimatore generico Pippo è corretto o non distorto quando E(Pippo)=Pippo

Consistenza: Un’altra proprietà desiderabile per uno stimatore è che, quando il campione è molto grande ci sia poca incertezza relativa al mio stimatore derivante da deviazioni casuali. In altre parole, che, in presenza di grandi campioni, Pippo stimato si ritrovi in un intorno molto piccolo rispetto al vero valore di Pippo.

Verifica delle Ipotesi

Molti quesiti economici o sociali non possono avere ipotesi in modo da avere come risposta un semplice sì o no. La retribuzione mensile dei lavoratori italiani è uguale a 1000 euro? Boh

Per rispondere a questo tipo di domande è necessario l’utilizzo della “**verifica delle ipotesi**”.

Ipotesi nulla e ipotesi alternativa: Il punto di partenza dell’analisi di ipotesi statistiche è la specificazione dell’ipotesi nulla che viene confrontata con l’ipotesi alternativa.

L’ipotesi nulla precede che, nel caso dell’esempio sopra, la media dei redditi (Y) nella popolazione, E(Y), assuma il valore specifico 1000 indicato $\mu_{Y,0}$

Allora si può scrive in termini matematici come segue:

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

Nell’ipotesi alternativa si nega quello affermato nell’ipotesi nulla

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

Ovvero si suppone che i lavoratori in Italia non guadagnino, in media, 1000 euro.

Dove C è il jjesimo elemento di $(x'x)^{-1}$

Proprietà di δ^2

1) $E(\delta^2) = \delta^2$
2) $\frac{\hat{\delta}^2(n-k)}{\delta^2} \sim X^2(n-k)$ Dove X sta per chi
3) $\widehat{\delta^2}$ è consistente sotto A4'
4) $\hat{\beta}\ e\ \delta^2\ sono\ incorrelati$

La Scomposizione della varianza e R^2

Ora vogliamo costruire una misura che mi indichi quanto è buona la mia stima e, cioè, quanto si avvicina alla realtà.

Vogliamo capire in soldoni quanto della variabilità complessiva riesce a spiegare il mio modello.

$$Y = \hat{Y} + e$$

$$Var(Y) = Var(\hat{Y}) + Var(e) + 2Cov(\hat{Y}, e)$$

$$TSS = ESS + RSS + 0$$

L'identità superiore ci dice che la Varianza Totale (TSS) è uguale alla Varianza Spiegata (ESS)+ Varianza residuo + una parte che vale zero perché è la covarianza fra Y stimato e c che contiene tutto ciò che non sono riuscito a spiegare e quindi non può covariare con Y.

La varianza totale è costituita da due componenti: quello che siamo riusciti a spiegare (o a predire) + quello che rimane. La varianza spiegata ci dà un'idea di quanto il modello sia stato preciso a stimare qualcosa. Tanto più ESS è più grande rispetto a RSS tanto più il nostro modello è buono.

Riscrivendo tutto in maniera più tangibile:

$$\sum (Y_i - \bar{Y})^2 = \sum \left(\hat{Y} - \bar{\hat{Y}}\right)^2 + \sum \left(e_i - \underset{residuo\ medio=0}{\bar{e}}\right)^2 + 0$$

$$TSS = ESS + RSS + 0$$

Se noi dividiamo la varianza spiegata per la varianza totale otteniamo un indice che ci visualizza immediatamente quanto della varianza complessiva sia stata spiegata dal nostro modello.

Ricaviamo il nostro indicatore R^2

Riassumendo:

$$\frac{\widehat{\beta_J} - \beta_j}{Se(\widehat{\beta_J})} \sim t(n-k)$$

È la nostra distribuzione t di Student, mentre

$$-t_{c\frac{a}{2}} * SE(\widehat{\beta_J}) \qquad \widehat{\beta_J} \qquad +t_{c\frac{a}{2}} * SE(\widehat{\beta_J})$$

$$\widehat{\beta_J} \pm t_{c\frac{a}{2}} * SE(\widehat{\beta_J})$$

È l'intervallo di confidenza.

Dunque, se vogliamo eseguire un test hp (ipotesi) su un singolo parametro dobbiamo fare il cosiddetto TEST T e quello che mi serve è:

1) Ipotesi (hp) nulla (h0)
2) Ipotesi alternativa (h1)
3) Statistica test (test t)
4) Regione di rifiuto

Questo tipo di test mi permette di capire se il valore ottenuto si trova nella regione di accettazione o in quella di rifiuto. Essendo un test su un singolo parametro allora lo chiamiamo TEST T (altrimenti, come vedremo dopo si sarebbe chiamato TEST F)

Riprendendo dalla statistica noi sappiamo che l'ipotesi nulla è quell'ipotesi già validata da precedenti ricerche o che noi, comunque, vogliamo sfatare.

L'ipotesi alternativa è l'ipotesi che noi riteniamo essere vera o che comunque vogliamo verificare.

In un ipotetico esercizio quando ci si chiede: "Verificare che $\beta_j > x$ …" allora $\beta_j > x$ diventa H_1 mentre H_0 è, o la sua negazione, oppure quello che si ritiene essere vero oppure quello che si vuole contro-provare.

Facciamo dei generici esempi

1) Ipotesi nulla:
   H_0: $\beta_j = c$
   o
   H_0: $\beta_j = 1$
   o
   H_0: $\beta_j = 0$

2) Ipotesi alternativa:
   H_1: $\beta_j > c$
   o

Stimatori GLS

Questa classe di stimatori ha proprietà migliori degli OLS in presenza di errori non-sferici. Se per qualche ragione:

$$V(\varepsilon) = \delta^2\Omega \neq \delta^2 I,$$

allora gli OLS smetteranno di essere BLUE. In particolare, gli stimatori dei minimi quadrati rimarranno corretti e consistenti, visto che per dimostrare queste proprietà sono sufficienti le assunzioni da A1 ad A4 fatte introducendo il modello di regressione lineare. Ma il teorema di Gauss-Markov (che richiede anche A5) non sarà in genere valido. E la varianza degli stimatori OLS diventerà:

$$\delta^2(\mathrm{x}'\mathrm{x})^{-1}\mathrm{x}'\Omega\,\mathrm{x}(\mathrm{x}'\mathrm{x})^{-1} \neq \delta^2(x'x)^{-1}$$

Proviamo ad aggiustare il mio modello:

Assumiamo di avere una matrice di trasformazione T di dimensioni NxN tale che $T\Omega\mathrm{T}' = I$ e di applicarla al nostro modello di regressione lineare:

$$Ty = Tx\beta + T\varepsilon$$

Gli errori di tale modello trasformato sarebbero sferici e i minimi quadrati di nuovo BLUE. Se conosciamo Ω è facile individuare tale matrice T con queste caratteristiche.

Dato che Ω è simmetrica e positiva, esiste una matrice P non singolare per la quale vale la relazione $\Omega = \mathrm{PP}'$. Da cui otteniamo $P^{-1}\Omega\mathrm{P}'^{-1} = I$.

Ponendo $T = P^{-1}$ possiamo applicare i minimi quadrati al modello trasformato e ottenere uno stimatore che è BLUE.

Dal momento che $T = P^{-1}$ possiamo scrive:

$$P^{-1}y = P^{-1}x\beta + P^{-1}\varepsilon$$

E semplificarne la struttura assumendo che:

$P^{-1}y = \tilde{y}$

$P^{-1}x = \tilde{x}$

$P^{-1}\varepsilon = \tilde{\varepsilon}$

Con $\tilde{y} = \beta\tilde{x} + \tilde{\varepsilon}$

La varianza sarà:

$$V(\tilde{\varepsilon}) = E(\tilde{\varepsilon}\tilde{\varepsilon}') = E(P^{-1}\tilde{\varepsilon}\tilde{\varepsilon}'P'^{-1}) = P^{-1}E(\tilde{\varepsilon}\tilde{\varepsilon}')P'^{-1} = P^{-1}\delta^2\Omega(\mathrm{P}^{-1})'$$

$$= \delta^2 P^{-1}\Omega(\mathrm{P}^{-1})' = \delta^2\mathrm{I_n}$$

Dove $P^{-1}\Omega(\mathrm{P}^{-1})'$ è I

La soluzione all'eteroschedasticità consiste nell'utilizzo dei minimi quadrati ponderati. Si stima prima il modello con OLS ottenendo prima una stima consistente di β, $\hat{\beta}_{OLS}$; Quindi si stima nuovamente con OLS il modello trasformato dividendo per $\sqrt{x_i'\hat{\beta}_{OLS}\,(1 - x_i'\hat{\beta}_{OLS}\,)}$

22) $x_i'\,\hat{\boldsymbol{\beta}}_{OLS}$ assume solo valori compresi tra 0 e 1?

$x_i'\,\hat{\beta}_{OLS}$ può essere esterno all'intervallo [0,1]

Nel modello di probabilità lineare $E(y_i|x_i) = \Pr\,(y_i = 1|x_i)$, dunque deve valere:

$$0 \le E(y_i|x_i) \le 1$$

Tuttavia (dannata econometria) può accadere che la stima di $\Pr\,(y_i = 1|x_i)$ risulti maggiore o minore di zero.

Questo è il serio problema che ci porta ad arrivare al modello della variabile latente.

23) Supponiamo di avere dati su un campione di famiglie e di essere interessati a determinare le variabili rilevanti nella scelta di acquisto di una barca per le vacanze estive. Scrivere il modello e sviluppare.

Il modello in questione deve restituirci valori pari a 1 o a 0 nel caso in cui la famiglia j-esima decidesse o meno di comprare una barca.

In tal caso, l'acquisto della barca dipende dalla disponibilità finanziaria della famiglia. Utilizzeremo il modello con variabile latente in quanto assumiamo che esista una soglia di costo della barca che fa sì che se il costo della barca va al di là la famiglia non compra mentre se sta al di sotto la famiglia compra la barca. La variabile latente misura proprio la soglia di soddisfazione nell'acquistare la barca. Se è uguale a zero o inferiore allora la barca non viene comprata.

Il modello teorico è il seguente:

$$y_i^* = x_i'\beta + \varepsilon_i^*$$

Dove y_i^* è la variabile latente.

Il modello stimabile è dato così:

$$y_i = 1 \; se \; y_i^* > 0$$
$$y = 0 \; se \; y_i^* \le 0$$

Possiamo scrivere il modello così:

$$y_i = E(y_i|x_i) + \varepsilon_i$$

Sappiamo che:

$$E(y_i|x_i) = \Pr\,(y_i = 1|x_i)$$

Che possiamo riscriverlo com:

$$E(y_i|x_i) = \Pr(y_i = 1|x_i) = \Pr(y_i^* > 0|x_i) =$$
$$= \Pr(x_i'\beta + \varepsilon_i^* > 0|x_i) =$$
$$= \Pr\,(\varepsilon_i^* > -x_i\beta|x_i)$$

In quanto normale è simmetrica, quindi:

$$= \Pr(\varepsilon_i^* < x_i\beta|x_i) =$$
$$= F(x_i\beta)$$