

1. [8 p.] Supponete che di voler stimare il modello: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Sapendo che $n=140$;

$$\sum_{i=1}^n (X_i) = 11830; \quad \sum_{i=1}^n (Y_i) = 40487.748; \quad \sum_{i=1}^n (X_i Y_i) = 4108601; \quad \sum_{i=1}^n (X_i^2) = 1228290;$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = 20974.25; \quad \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = 17875.29; \quad \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 3098.96$$

- Calcolate le stime OLS di β_0 e β_1 ;
 - Calcolate R^2 e SER .
 - Per l'osservazione $i=32$: $Y=170$; $X=46$. Calcolate il valore predetto dalla retta di regressione \hat{Y} e il residuo \hat{u} .
2. [6p.] Spiegate in dettaglio quali sono le assunzioni dei minimi quadrati (regressione con un singolo regressore), qual è il ruolo che giocano per gli stimatori OLS, qual è la distribuzione campionaria (forma, media, $SE(\beta_1)$) sotto le assunzioni OLS. Quali ipotesi aggiuntive sono necessarie nella regressione multipla?
3. [8p.] Supponete di aver stimato il seguente modello: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{COMPUTER}_i$ dove Y rappresenta il rendimento degli studenti a scuola (punteggio ai test) e COMPUTER è una dummy uguale a 1 se lo studente i possiede un computer (e zero altrimenti). $\hat{\beta}_1$ risulta positivo e significativamente diverso da zero. Tuttavia, non si hanno informazioni sulla variabile BACKGROUND (che sintetizza le caratteristiche socio-economiche della famiglia). Alla luce del problema della variabile omessa, $\hat{\beta}_1$ è uno stimatore corretto? E' consistente? Mostrate formalmente sotto quali condizioni $\hat{\beta}_1$ è distorto e eventualmente la direzione della distorsione.
4. [10p.] Data la seguente regressione:

Source	SS	df	MS	Number of obs = 3796		
Model	1468.80502	6	244.800836			
Residual	11018.5775	3789	2.90804368			
Total	12487.3825	3795	3.29048287			
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
yrsed						
dist	-.0456809	.0131687				
female	.0558886	.0557061				
black	-.329419	.0714746				
incomehi	.4285605	.0664441				
dadcoll	.8564408	.0808736				
momcoll	.603271	.0898767				
_cons	13.56121	.0557479				

dove $yrsed$ sono gli anni di istruzione, $dist$ è la distanza in km dalla residenza all'università, $female$ è la dummy per donna, $dadcoll=1$ se il padre ha la laurea, $momcoll=1$ se la madre ha la laurea; $incomehi=1$ se la famiglia ha un alto reddito, $black=1$ se l'individuo è afro-americano.

- Date una interpretazione dei coefficienti $dist$, $female$, $incomehi$, $dadcoll$
- Calcolate se $dist$ e $female$ sono statisticamente significativi all'1%.
- Costruite un intervallo di confidenza per $dist$ al livello del 95%.
- Verificate l'ipotesi nulla (al 5%) che il coefficiente di $dadcoll$ sia uguale a 1, contro l'alternativa bilaterale.
- Calcolate R^2 corretto e il SER della regressione.
- Calcolate la statistica F classica per verificare l'ipotesi nulla che tutti i coefficienti (tranne la costante) siano uguali a zero [il valore critico all'1%: $F(6, 3789)=2.8$]