

Econometria

Prof. Valerio Potì

Esercizi Empirici I

DOMANDE

ESERCIZIO EMPIRICO 8.1.

- a) Si effettui una regressione della retribuzione oraria media (AHE) sull'età(AGE), sul genere (FEMALE) e sull'istruzione(BACHELOR). Se Age aumenta da 25 a 26 come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34 come ci si aspetta che cambino le retribuzioni medie?
- b) Si effettui una regressione del logaritmo della retribuzione oraria media (AHE), $\ln(AHE)$, su Age, Female e Bachelor. Se Age aumenta da 25 a 26, come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34, come ci si aspetta che cambino le retribuzioni orarie?
- c) Si effettui una regressione del logaritmo della retribuzione oraria media (AHE), $\ln(AHE)$, su $\ln(Age)$, Female e Bachelor. Se Age aumenta da 25 a 26, come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34, come ci si aspetta che cambino le retribuzioni orarie?
- d) Si effettui una regressione del logaritmo della retribuzione oraria media (AHE), $\ln(AHE)$, su Age, Age², Female e Bachelor. Se Age aumenta da 25 a 26, come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34, come ci si aspetta che cambino le retribuzioni orarie?
- e) Si preferisce la regressione della (c) a quella della (b)? Si argomenta la risposta
- f) Si preferisce la regressione della (d) a quella della (b)? Si argomenta la risposta
- g) Si preferisce la regressione della (d) a quella della (c)? Si argomenta la risposta
- h) Si raffigurino le relazioni di regressione tra Age e $\ln(AHE)$ corrispondenti a (b), (c) e (d) per gli uomini con un diploma di scuola superiore. Si descrivano le somiglianze e le differenze tra le funzioni di regressione stimate. La risposta sarebbe diversa se si raffigurasse la funzione di regressione delle donne laureate?
- i) Si effettui una regressione di $\ln(AHE)$ su Age, Age², Female e Bachelor e il termine d'interazione Female x Bachelor. Che cosa misura il coefficiente del termine d'interazione? Alexis è una donna di 30 anni con una laurea. Che cosa predice la regressione per il suo valore di $\ln(AHE)$? Jane è una donna di 30 anni con un diploma. Che cosa predice la regressione per il suo valore di $\ln(AHE)$? Qual è la differenza predetta tra la retribuzione oraria di Alexis e quella di Jane? Bob è un uomo di 30 anni con una laurea. Che cosa predice la regressione per il suo valore di $\ln(AHE)$? Jim è un uomo di 30 anni con un diploma. Che cosa predice la regressione per il suo valore di $\ln(AHE)$? Qual è la differenza predetta tra la retribuzione oraria di Bob e quella di Jim?
- j) L'effetto di Age sulle retribuzioni è differente per gli uomini e per le donne? Si specifichi e si stimi una regressione che si possa usare per rispondere a questa domanda.
- k) L'effetto di Age sulle retribuzioni orarie è differente per i diplomati e per i laureati? Si specifichi e si stimi una regressione utilizzabile per rispondere a questa domanda.
- l) Dopo aver effettuato tutte queste regressioni (e tutte quelle che si vogliono effettuare), si riassume l'effetto dell'età sulle retribuzioni orarie dei giovani lavoratori.

ESERCIZIO EMPIRICO 9.1.

- a) Si discuta la validità interna delle regressioni usate per rispondere all'esercizio empirico 8.1 (1). Si includa una discussione sulla possibile distorsione da variabili omesse, sulla incorretta specificazione della forma funzionale della regressione, sugli errori nelle variabili, sulla selezione campionaria, sulla causalità simultanea e sull'inconsistenza degli errori standard OLS.
- b) L'insieme di dati CPS92_08 descritto nell'Esercizio empirico 3.1 include dati del 2008 e del 1992. Si usino tali dati per esaminare la validità esterna (temporale) delle conclusioni raggiunte nell'esercizio empirico 8.1(1).

DOMANDE E RISPOSTE

ESERCIZIO EMPIRICO 8.1.

a) Si effettui una regressione della retribuzione oraria media (AHE) sull'età (AGE), sul genere (FEMALE) e sull'istruzione (BACHELOR). Se Age aumenta da 25 a 26 come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34 come ci si aspetta che cambino le retribuzioni medie?

Come mostrato nella colonna (1) della *Tabella 8.a*, per una variazione unitaria dell'Age, sia nel caso che vari da 25 a 26 o da 33 a 34, la variazione sulla variabile dipendente sarà pari al valore del coefficiente della variabile Age: la retribuzione oraria media aumenterà di 0,48\$ per ogni anno di età in più.

b) Si effettui una regressione del logaritmo della retribuzione oraria media (AHE), $\ln(AHE)$, su Age, Female e Bachelor. Se Age aumenta da 25 a 26, come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34, come ci si aspetta che cambino le retribuzioni orarie?

Come mostrato nella colonna (2) della *Tabella 8.a*, poiché la variabile dipendente AHE è espressa in termini logaritmici, ma la variabile indipendente Age non lo è, una variazione unitaria di Age da 25 a 26 comporta una variazione di 0,03 del logaritmo di AHE, ovvero

$$\ln(AHE_i) = \ln(AHE_0) + 0,03 \Rightarrow \ln(AHE_i) - \ln(AHE_0) = 0,03 \Rightarrow \ln\left(\frac{AHE_i}{AHE_0}\right) = 0,03$$

Quindi, se Age varia di un'unità provoca un aumento del 3% di AHE.

Allo stesso modo una variazione di Age da 33 a 34 comporta una variazione unitaria dello 0,03 di AHE.

c) Si effettui una regressione del logaritmo della retribuzione oraria media (AHE), $\ln(AHE)$, su $\ln(Age)$, Female e Bachelor. Se Age aumenta da 25 a 26, come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34, come ci si aspetta che cambino le retribuzioni orarie?

Come mostrato nella colonna (3) della *Tabella 8.a*, un aumento della variabile indipendente Age da 25 a 26, ovvero un aumento del $\ln(26/25) = \ln 26 - \ln 25 = 3,9\%$, comporta un aumento di AHE del: $0,76 \times (\ln 26 - \ln 25) = 0,0298 = 3\%$ circa.

Quindi variando l'età del 3,9% si ha un incremento della retribuzione media oraria del 2,98%.

Allo stesso modo un incremento dell'età da 33 a 34 provoca un incremento della retribuzione di $0,76 \times (\ln 34 - \ln 33) = 0,023 = 2,3\%$.

d) Si effettui una regressione del logaritmo della retribuzione oraria media (AHE), $\ln(AHE)$, su Age, Age^2 , Female e Bachelor. Se Age aumenta da 25 a 26, come ci si aspetta che cambino le retribuzioni orarie? Se Age aumenta da 33 a 34, come ci si aspetta che cambino le retribuzioni orarie?

Come mostrato nella colonna (4) della *Tabella 8.a*, una variazione di Age da 25 a 26 porta ad una variazione del logaritmo della retribuzione oraria media AHE pari a:

$$(0,17 \times 26 - 0,002 \times 26^2) - (0,17 \times 25 - 0,002 \times 25^2) = 0,068 = 6,8\%$$

Allo stesso modo un incremento di Age da 33 a 34 comporta un incremento di AHE pari al coefficiente di Age 0,17:

$$(-0,002 \times 34^2) - (-0,002 \times 33^2) = -0,134 \text{ per } Age^2$$

Quindi sommando i due effetti si ha un incremento del 3,6% ($0,17 - 0,134 = 0,036$) della retribuzione oraria media.

e) Si preferisce la regressione della (c) a quella della (b)? Si argomenta la risposta

La regressione (c) è leggermente preferita alla regressione (b); osservando la *Tabella 8.a*, dal confronto dei \bar{R}^2 delle due regressioni risulta che \bar{R}^2 è leggermente maggiore nella regressione che considera come variabile dipendente $\ln(\text{AHE})$ e variabili indipendenti $\ln(\text{Age})$, Female e Bachelor; infatti è pari a 0,1850, mentre \bar{R}^2 della regressione calcolata al punto b) è pari a 0,1854.

f) Si preferisce la regressione della (d) a quella della (b)? Si argomenta la risposta

In questo caso, come mostra la *Tabella 8.a*, il confronto dei due \bar{R}^2 , ci porta a preferire la regressione svolta al punto d. Infatti \bar{R}^2 è pari a 0,1850, mentre al punto b) è pari a 0,1861. Si giunge alla medesima conclusione osservando la statistica t associata al coefficiente di age^2 (-0,25) poichè non è statisticamente significativa; inoltre il coefficiente stesso è molto vicino a zero, per questo si preferisce la regressione (d).

g) Si preferisce la regressione della (d) a quella della (c)? Si argomenta la risposta

La regressione differisce dalla (c) per l'utilizzo del coefficiente Age - in (c) si utilizza $\ln(\text{Age})$ mentre in (d) Age e Age^2 . Confrontando \bar{R}^2 delle due regressioni, si preferisce la (d) - 0,1854 in (c) e 0,1861 in (d).

h) Si raffigurino le relazioni di regressione tra Age e $\ln(\text{AHE})$ corrispondenti a (b), (c) e (d) per gli uomini con un diploma di scuola superiore. Si descrivano le somiglianze e le differenze tra le funzioni di regressione stimate. La risposta sarebbe diversa se si raffigurasse la funzione di regressione delle donne laureate?

Ricordiamo innanzitutto che le variabili Female e Bachelor sono variabili dummy, che di conseguenza possono assumere valori pari a 0 ed a 1. Nel nostro caso, siccome consideriamo uomini con il solo diploma di scuola superiore, assumono entrambe valori pari a zero.

Come evidenzia la *Figura 1*, i tre grafici sono molto simili, gli ultimi due sono relazioni perfettamente lineari mentre il primo, avendo una componente di secondo grado, ha una forma curvilinea; le pendenze delle tre rette sono quasi identiche, come ci si aspettava già osservando i coefficienti.

Se analizzassimo la stessa distribuzione per le donne laureate, numericamente cambierebbe soltanto l'intercetta (che varierebbe della somma dei coefficienti delle due variabili Female e Bachelor), mentre graficamente le rette sarebbero semplicemente traslate verso l'alto.

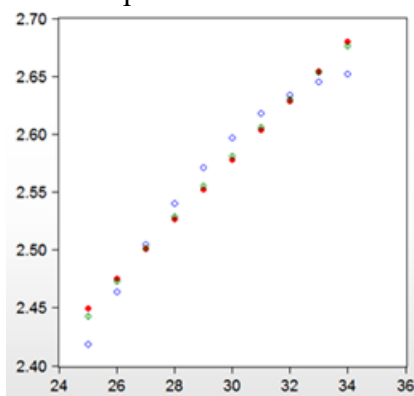


Figura 1. Si riportano le relazioni stimate tra Age, $\ln(\text{AHE})$ per gli uomini con diploma di scuola superiore. La retta rappresentata dai punti blu riporta la regressione di $\ln(\text{AHE})$ su Age, Age^2 , Female e Bachelor; la retta rappresentata dai punti rossi riporta la regressione di $\ln(\text{AHE})$ su Age, Female e Bachelor; la retta rappresentata dai punti verdi riporta la regressione di $\ln(\text{AHE})$ su $\ln(\text{Age})$, Female e Bachelor.

i) Si effettui una regressione di $\ln(\text{AHE})$ su Age, Age^2 , Female e Bachelor e il termine d'interazione Female x Bachelor. Che cosa misura il coefficiente del termine d'interazione? Alexis è una donna di 30 anni con una laurea. Che cosa predice la regressione per il suo valore

di ln(AHE)? Jane è una donna di 30 anni con un diploma. Che cosa predice la regressione per il suo valore di ln(AHE)? Qual è la differenza predetta tra la retribuzione oraria di Alexis e quella di Jane? Bob è un uomo di 30 anni con una laurea. Che cosa predice la regressione per il suo valore di ln(AHE)? Jim è un uomo di 30 anni con un diploma. Che cosa predice la regressione per il suo valore di ln(AHE)? Qual è la differenza predetta tra la retribuzione oraria di Bob e quella di Jim?

Il coefficiente di interazione misura quanto l'effetto di variazione di una variabile sia spiegabile dalla variazione di un'altra variabile. Come mostra la colonna (5) della **Tabella 8.a**, il coefficiente di interazione evita che si sovrastimi l'effetto di essere maschio o femmina sui salari indipendentemente dall'aver una laurea o meno (l'aver una laurea per una donna potrebbe essere più significativo che per un uomo).

Per i differenti soggetti i valori previsti sono:

$$\text{Alexis: } -0,28 + 0,17*30 + (-0,002)*30^2 + (-0,20*1) + 1*0,37 + 0,07*1 = 3,26$$

$$\text{Jane: } -0,28 + 0,17*30 + (-0,002)*30^2 + (-0,20*1) + 0*0,37 + 0,07*0 = 2,82$$

$$\text{Bob: } -0,28 + 0,17*30 + (-0,002)*30^2 + (-0,20*0) + 1*0,37 + 0,07*0 = 3,39$$

$$\text{Jim: } -0,28 + 0,17*30 + (-0,002)*30^2 + (-0,20*0) + 0*0,37 + 0,07*0 = 3,02$$

$$\text{Differenza su ln(AHE): Alexis - Jane} = 3,26 - 2,82 = 0,44$$

$$\text{Differenza su ln(AHE): Bob - Jim} = 3,39 - 3,02 = 0,37$$

Si noti che la differenza tra le differenze negli effetti previsti è $0,44 - 0,37 = 0,07$, il valore del coefficiente del termine di interazione.

j) L'effetto di Age sulle retribuzioni è differente per gli uomini e per le donne? Si specifichi e si stimi una regressione che si possa usare per rispondere a questa domanda.

Per rispondere a questo quesito creiamo un termine di interazione tra il sesso e l'età. La nuova regressione, riportata in **Tabella 8.a** colonna (6), comprenderà quindi due variabili aggiuntive: Female*Age e Female *Age².

La statistica F (test di Wald) che verifica il vincolo che i coefficienti di questi termini di interazione siano uguali a zero è $F = 2,88$, con valore-p 0,06. Ciò implica che Age potrebbe avere un effetto differente su ln(AHE) per gli uomini rispetto alle donne; tuttavia non possiamo affermarlo con certezza poiché i singoli coefficienti t dei due nuovi regressori sembrano suggerirne una dubbia significatività.

k) L'effetto di Age sulle retribuzioni orarie è differente per i diplomati e per i laureati? Si specifichi e si stimi una regressione utilizzabile per rispondere a questa domanda.

Per rispondere a questo quesito creiamo un termine di interazione tra il sesso e l'età. La nuova regressione, riportata in **Tabella 8.a** colonna (7), comprenderà quindi due variabili aggiuntive: Bachelor*Age e Bachelor *Age².

La statistica F (test di Wald) che verifica il vincolo che i coefficienti di questi termini di interazione siano uguali a zero è $F = 3,76$, con valore-p 0,02. Ciò implica che Age potrebbe avere un effetto differente su ln(AHE) per i diplomati rispetto ai laureati; tuttavia non possiamo affermarlo con certezza poiché i singoli coefficienti t dei due nuovi regressori sembrano suggerirne una dubbia significatività.

l) Dopo aver effettuato tutte queste regressioni (e tutte quelle che si vogliono effettuare), si riassume l'effetto dell'età sulle retribuzioni orarie dei giovani lavoratori.

Come mostrato nella colonna (8) della **Tabella 8.a**, l'effetto dell'età ha, come ci si aspettava, un segno positivo.

In particolare, sembra favorire i maschi laureati, subito dopo le femmine laureate.

TABELLE ESERCIZIO 8.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variabile dipendente								
	AHE	ln(AHE)	ln(AHE)	ln(AHE)	ln(AHE)	ln(AHE)	ln(AHE)	ln(AHE)
Age	0,48 (0,00)	0,03 (0,00)		0,17 (0,002)	0,17 (0,002)	0,22 (0,002)	0,18 (0,02)	0,22 (0,01)
ln(Age)			0,76 (0,00)					
Age ²				-0,002 (0,008)	-0,002 (0,008)	-0,003 (0,007)	-0,003 (0,04)	-0,003 (0,02)
Female	-2,90 (0,00)	-0,17 (0,00)	-0,17 (0,00)	-0,16 (0,00)	-0,20 (0,00)	1,77 (0,27)	-0,20 (0,00)	2,10 (0,20)
Female x Age						-0,12 (0,26)		-0,14 (0,20)
Female x Age ²						0,002 (0,30)		0,002 (0,24)
Bachelor	6,72 (0,00)	0,40 (0,00)	0,40 (0,00)	0,40 (0,00)	0,37 (0,00)	0,37 (0,00)	0,50 (0,75)	-0,10 (0,95)
Bachelor x Age							-0,02 (0,84)	0,017 (0,88)
Bachelor x Age ²							0,00 (0,75)	-0,00 (0,98)
Female x Bachelor					0,07 (0,007)	0,07 (0,01)	0,08 (0,01)	0,07 (0,01)
Intercetta	0,48	0,81	0,002	-0,31	-0,28	-1,14	-0,28	-0,99
F-statistic	338,29 (0,00)	341,44 (0,00)	342,37 (0,00)	258,20 (0,00)	208,31 (0,00)	149,74 (0,00)	150,05 (0,00)	117,77 (0,00)
SER	7,72	0,45	0,45	0,45	0,45	0,45	0,45	0,45
R ²	0,184	0,1855	0,1859	0,1868	0,188	0,189	0,189	0,191
Adjusted R squared	0,183	0,1850	0,1854	0,1861	0,187	0,187	0,188	0,189
n	4500	4500	4500	4500	4500	4500	4500	4500

Tabella 8.a. La tabella riporta otto diverse regressioni: la colonna (1) riporta la regressione della retribuzione oraria media (AHE) sull'età (Age), sul sesso (Female) e sull'istruzione (Bachelor); la colonna (2) indica la regressione di ln(AHE) su Age, Female e Bachelor; la colonna (3) evidenzia la regressione di ln(AHE) su ln(Age), Female e Bachelor; la colonna (4) mostra la regressione di ln(AHE) su Age, Age², Female e Bachelor; la colonna (5) riporta la regressione di ln(AHE) su Age, Age², Female, Bachelor e Female x Bachelor; dalla colonna (6) si vede la regressione di ln(AHE) su Age, Age², Female, Bachelor, Female x Bachelor, Female x Age e Female x Age²; dalla colonna (7) si osserva la regressione di ln(AHE) su Age, Age², Female, Bachelor, Female x Bachelor, Bachelor x Age e Bachelor x Age²; infine la colonna (8) mostra la regressione su ln(AHE) di tutte le precedenti variabili. Per ciascun valore dei regressori viene riportato in parentesi il p-value associato. Il campione oggetto di studio contiene n = 4500 osservazioni. Si riportano inoltre: la statistica F per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e due misure di bontà di adattamento del modello ai dati (R² e adjusted R squared). Le stime sono ottenute con un campione che comprende le prime 4500 osservazioni.

ESERCIZIO EMPIRICO 9.1.

a) Si discuta la validità interna delle regressioni usate per rispondere all'esercizio empirico 8.1 (1). Si includa una discussione sulla possibile distorsione da variabili omesse, sulla incorretta specificazione della forma funzionale della regressione, sugli errori nelle variabili, sulla selezione campionaria, sulla causalità simultanea e sull'inconsistenza degli errori standard OLS.

Variabili omesse: si ha una potenziale distorsione da omissione di variabili quando dalla regressione viene esclusa una variabile che (i) influisce su $\ln(\text{AHE})$ e (ii) è correlata con una variabile inclusa nella regressione. Un esempio può essere l'esperienza lavorativa: i lavoratori con maggiore esperienza, e quindi anzianità di servizio, riceveranno un compenso maggiore. Escludere questa variabile dalla regressione può condurre a una distorsione da variabili omesse (in quanto correlata con Age e determinante per il livello di retribuzione) per l'effetto stimato dell'istruzione sulla retribuzione.

La errata forma funzionale dipende dal tipo di equazione che utilizziamo per descrivere il modello. Per esempio, potremmo utilizzare una forma lineare quando invece è più consigliabile una forma quadratica. Questo aspetto è stato analizzato nell'esercizio empirico 8.1 (vedi sopra) con particolare riferimento alla variabile Age: da questa analisi è emerso che l'effetto di Age sulla retribuzione è non lineare, ma è adeguatamente colto da una regressione con termini di interazione.

Per quanto riguarda gli errori nelle variabili invece si apre una finestra di possibilità piuttosto ampia: infatti le possibilità di commettere errori nelle variabili può partire dalla rilevazione, fino all'interpretazione, fino addirittura a errori di battitura stessi. Per esempio, gli intervistati potrebbero aver riportato un livello di reddito non corrispondente a verità; oppure il campione da noi scelto è stato selezionato da un'area dove la componente sessuale o quella del titolo di studio abbiano un'importanza particolarmente bassa.

Selezione del campione: i dati riguardano solamente lavoratori a tempo pieno, quindi esiste la possibilità di una distorsione da selezione del campione.

Si parla di causalità simultanea quando non è solo la variabile indipendente ad essere causa della dipendente ma anche viceversa; questo comporta OLS distorti e inconsistenti. In questo caso la variabile dipendente "reagisce" alle variazioni delle variabili indipendenti modificando quindi gli errori della regressione che saranno infine correlati con essa. Per quanto riguarda la regressione in esame, è improbabile che costituisca un problema, in quanto è improbabile che AHE influisca su Age o sul genere.

La prima fonte di inconsistenza degli standard error degli OLS è l'eteroschedasticità degli errori: per ovviare a questo problema è sufficiente utilizzare standard error robusti all'eteroschedasticità. La seconda fonte di inconsistenza è la correlazione tra errori e osservazioni; in questo caso i dati sono raccolti, almeno approssimativamente, utilizzando campionamenti i.i.d., quindi è improbabile che vi sia un problema di questo tipo.

b) L'insieme di dati CPS92_08 descritto nell'Esercizio empirico 3.1 include dati del 2008 e del 1992. Si usino tali dati per esaminare la validità esterna (temporale) delle conclusioni raggiunte nell'esercizio empirico 8.1(1).

Un modello è considerato esternamente valido se le sue inferenze e conclusioni possono essere generalizzate dalla popolazione e dal contesto studiati ad altre popolazioni o contesti. La validità esterna può essere quindi influenzata da differenze nelle popolazioni, oltretutto da differenze di contesto.

Le regressioni fatte sui dati del campione riferito all'anno 1992 e quello 2008 permettono di analizzare un numero ristretto di valori, il cui modello generale è valido anche esternamente. Utilizziamo questi dati per effettuare regressioni sui dati riferiti al solo anno 2008 e ad ambi gli anni 1992-2008, per meglio comprendere la validità esterna del modello di riferimento.

Nella **Tabella 9.a** colonna (1) abbiamo effettuato la regressione sulle variabili Age, Age², Female; Bachelor e Female*Bachelor considerando i dati riferiti all'anno 2008.

Successivamente (**Tabella 9.a** colonna (2)) abbiamo effettuato la medesima regressione su un campione di osservazioni “allargato” in quanto considera sia i dati del 2008, sia quelli del 1992.

Vediamo ora le principali analogie e differenze tra le due regressioni. Innanzitutto l'R-squared è simile in entrambe le regressioni, attestandosi a circa 0,2 nella regressione per i dati del 2008 e 0,18 per i dati riferiti agli anni 1992/2008. I coefficienti stimati delle variabili indipendenti non variano sostanzialmente dalla regressione dell'anno 2008 a quella degli anni 1992-2008. Inoltre è possibile notare come il coefficiente del termine d'interazione (bachelor*female) è piuttosto basso in entrambe le regressioni; in particolare è rispettivamente 0,07 nella prima e 0,11 nella seconda.

Ciò che è differente è il valore della t-statistic per Age (vedere **Tabella 9.a** in nota) che, mentre nella prima regressione risulta attorno a 1,86 circa portandoci a non rifiutare l'ipotesi nulla di non significatività del regressore, nel secondo caso è superiore a due (2,62) e quindi ci fa rifiutare l'ipotesi nulla. Anche il valore della t-statistic associata a Age squared è piuttosto basso: -1,23 nella prima regressione, -1,85 nella seconda.

Nel secondo caso è stato preso in considerazione un campione molto più ampio rispetto al primo: per la regressione del 2008 abbiamo analizzato 7711 osservazioni, contro le 15316 di entrambi gli anni 1992/2008. Aumentando la numerosità del campione il valore della F statistic aumenta considerevolmente da 389,70 a 696,81.

Si considera anche la variabile Year per aumentare la validità esterna del modello; come si nota dalla **Tabella 9.a** colonna (3), inserendo tale regressore è variato il valore dell'intercetta (da 0,80 a -54,34), sono aumentati l'R-squared corretto (da 0,1851 a 0,3362) e anche il valore della F statistic (da 696,81 a 1293,83).

TABELLE ESERCIZIO 9.1.

	(1)	(2)	(3)
	ln(AHE)	ln(AHE)	ln(AHE)
Age	0,08 (0,06)	0,09 (0,009)	0,12 (0,00)
Age ²	-0,001 (0,22)	-0,001 (0,06)	-0,002 (0,002)
Female	-0,22 (0,00)	-0,23 (0,00)	-0,21 (0,00)
Bachelor	0,40 (0,00)	0,40 (0,00)	0,37 (0,00)
Female x Bachelor	0,07 (0,001)	0,11 (0,00)	0,07 (0,00)
Year			0,03 (0,00)
Intercetta	1,10	0,80	-54,34
F-statistic	389,70 (0,00)	696,81 (0,00)	1293,83 (0,00)
SER	0,47	0,51	0,46
R ²	0,2018	0,1853	0,3364
Adjusted R squared	0,2013	0,1851	0,3362
n	7711	15316	15316

Tabella 9.a. La tabella riporta tre diverse regressioni effettuate su tre campioni differenti: il primo campione nella colonna (1) considera 7711 osservazioni relative all'anno 2008 e riporta la regressione della retribuzione oraria media (AHE) sull'età (Age), su Age², sul sesso (Female), sull'istruzione (Bachelor) e Female x Bachelor; il secondo campione nella colonna (2) considera 15316 osservazioni relative all'anno 2008 e 1992 e riporta la regressione della retribuzione oraria media (AHE) sull'età (Age), su Age², sul sesso (Female), sull'istruzione (Bachelor) e Female x Bachelor; l'ultimo campione nella colonna (3) considera 15316 osservazioni relative all'anno 2008 e 1992 e riporta la regressione della retribuzione oraria media (AHE) sull'età (Age), su Age², sul sesso (Female), sull'istruzione (Bachelor), Female x Bachelor e Year. Come richiesto dall'esercizio i dati relativi ad AHE sono stati corretti per l'inflazione dei prezzi tra il 1992 e il 2008. Per ciascun valore dei regressori viene riportato in parentesi il p-value associato. Si riportano inoltre: la statistica F per la significatività congiunta dei regressori, la deviazione standard campionaria dei residui e due misure di bontà di adattamento del modello ai dati (R² e adjusted R squared). Di seguito si specificano le statistiche t riferite alle regressioni delle colonne (1) e (2) in quanto trattate nelle risposte ai quesiti. Partendo dalla regressione in colonna (1) i valori sono: Age 1,86, Age² 1,23, Female -14,30, Bachelor 27,72, Female*Bachelor 3,16; in colonna (2) i valori sono: Age 2,62, Age² -1,85, Female -20,22, Bachelor 35,99, Female*Bachelor 6,49.