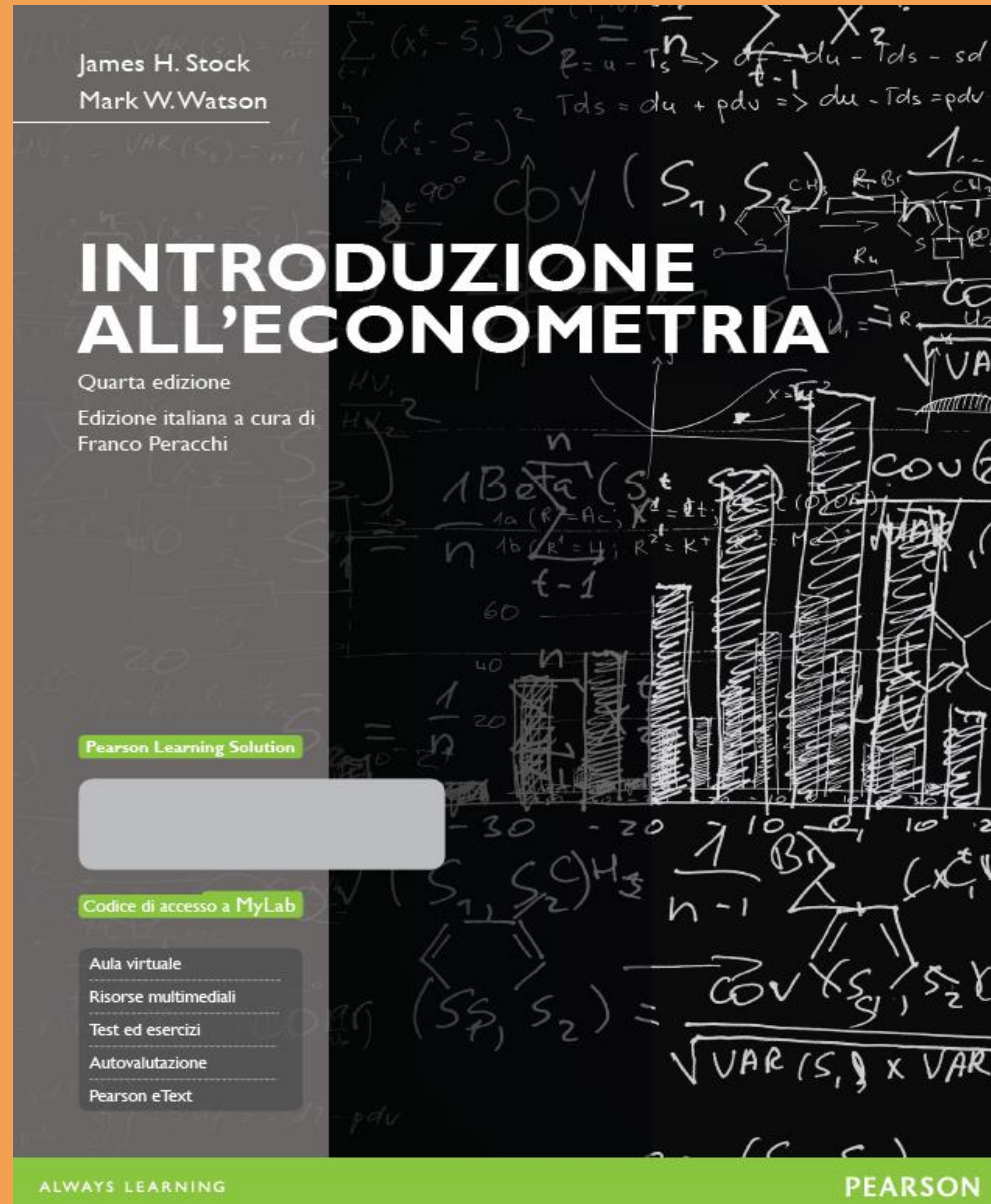
Capitolo 8

Funzioni di regressione non lineari



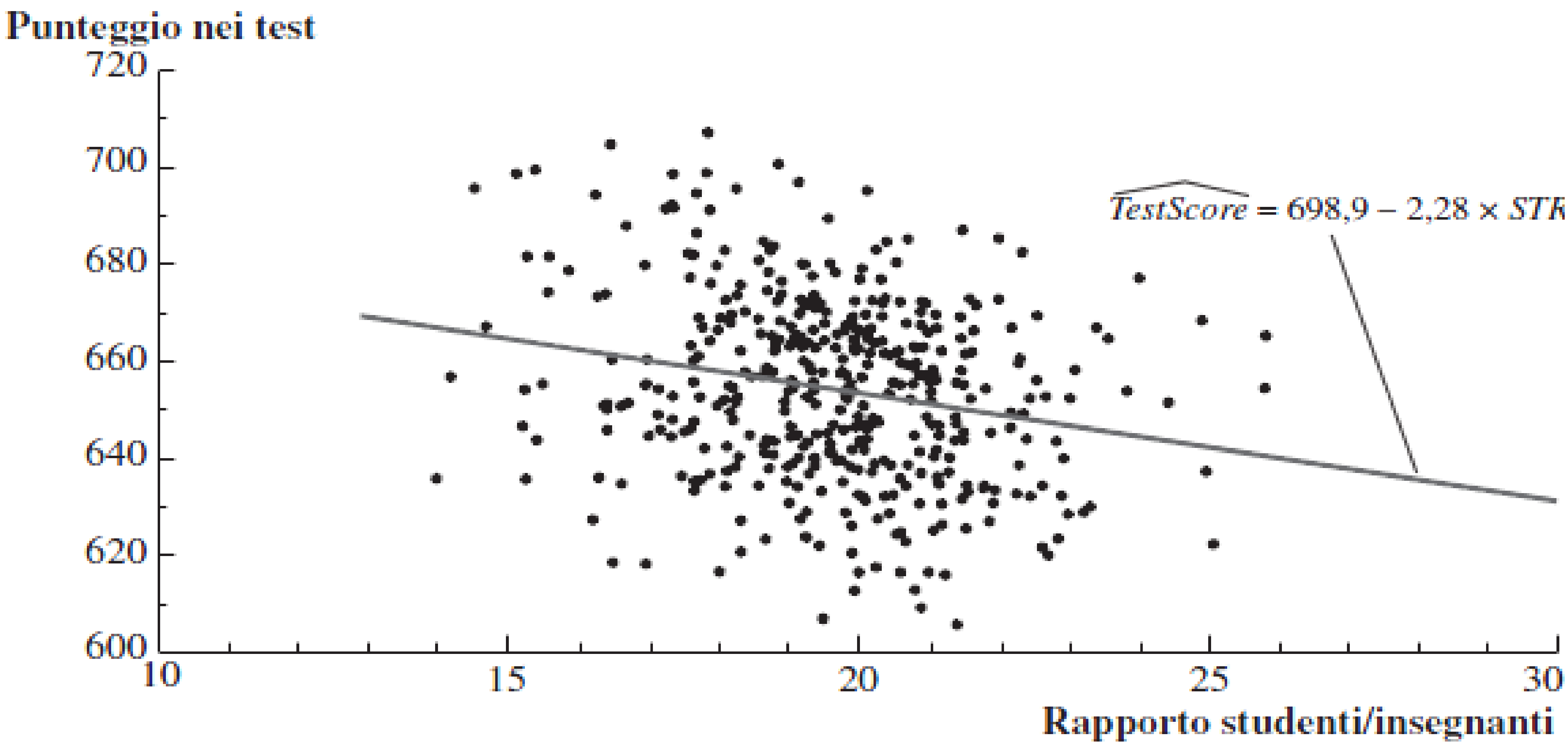
Sommario

1. Funzioni di regressione non lineari – note generali
2. Funzioni non lineari a una variabile
3. Funzioni non lineari a due variabili: interazioni
4. Applicazione al dataset dei punteggi nei test della California

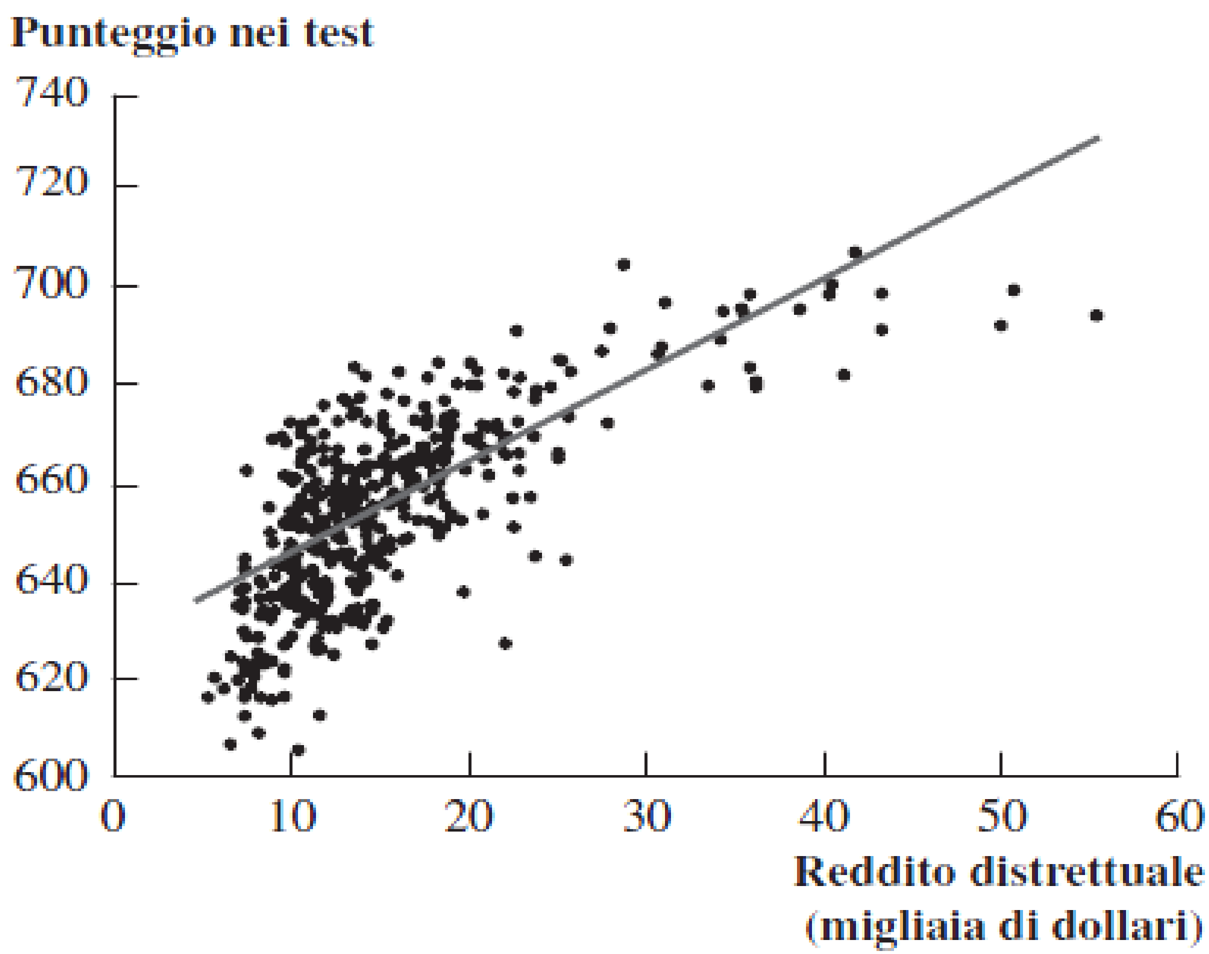
Funzioni di regressione non lineari

- Le funzioni di regressione viste finora erano lineari rispetto alla variabile X
- Ma l'approssimazione lineare non è sempre la migliore
- Il modello di regressione multipla può gestire funzioni di regressione non lineari in una o più X.

La relazione tra punteggio nei test e rapporto studenti/insegnanti sembra lineare (forse)…



Ma la relazione tra punteggio nei test e reddito distrettuale sembra non lineare…



Funzioni di regressione non lineari – concetti generali (Paragrafo 8.1)

Se una relazione tra *Y* e *X* è **non lineare**:

- L'effetto su *Y* di una variazione in *X* dipende dal valore di *X* – ovvero, l'effetto marginale di *X* non è costante
- Una regressione lineare è mal specificata: la forma funzionale è errata
- Lo stimatore dell'effetto su *Y* di *X* è distorto: in generale non è corretto nemmeno sulla media
- La soluzione consiste nell'applicare una funzione di regressione che sia non lineare in *X*

La formula generale per una funzione di regressione non lineare

$$Y_i = f(X_{1i}, X_{2i},\ldots, X_{ki}) + u_i,\ i = 1,\ldots, n$$

Assunzioni

1. $E(u_i| X_{1i}, X_{2i},\ldots, X_{ki}) = 0$ (identica); implica che f è il valore atteso di Y condizionato alle X.
2. $(X_{1i},\ldots, X_{ki}, Y_i)$ sono i.i.d. (identica).
3. Gli outlier sono rari (stessa idea; la condizione matematica precisa dipende dalla f in esame).
4. Assenza di multicollinearità perfetta (stessa idea; la formulazione precisa dipende dalla f in esame).

La variazione in Y associata a una variazione in X_1, mantenendo $X_2,\ldots, X_k$ costanti è:

$$\Delta Y = f(X_1 + \Delta X_1, X_2,\ldots, X_k) - f(X_1, X_2,\ldots, X_k)$$

CONCETTO CHIAVE 8.1

L'effetto atteso su Y di una variazione di X_1, nel modello di regressione non lineare (8.3)

La variazione attesa di Y, ΔY, associata alla variazione di X_1, ΔX_1, tenendo costanti $X_2,\ldots,X_k$, è la differenza tra il valore della funzione di regressione della popolazione prima e dopo la variazione in X_1, tenendo costanti $X_2,\ldots,X_k$. In altri termini, la variazione attesa di Y è la differenza:

$$\Delta Y = f(X_1 + \Delta X_1, X_2,\ldots,X_k) - f(X_1, X_2,\ldots,X_k). \tag{8.4}$$

Lo stimatore di tale differenza ignota è la differenza tra i valori predetti in questi due casi. Sia $\hat{f}(X_1, X_2,\ldots,X_k)$ il valore predetto di Y basato sullo stimatore $\hat{f}$ della funzione di regressione della popolazione. Allora, la variazione predetta di Y è

$$\Delta\hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2,\ldots,X_k) - \hat{f}(X_1, X_2,\ldots,X_k) \tag{8.5}$$

Funzioni non lineari di un'unica variabile indipendente (Paragrafo 8.2)

Vedremo due approcci complementari:

1. Polinomiali in X

   La funzione di regressione della popolazione viene approssimata da una quadratica, una cubica o una polinomiale di grado più alto

2. Trasformazioni logaritmiche

   Le Y e/o le X vengono trasformate prendendone il logaritmo, che ne dà un'approssimazione "percentuale" utile in molte applicazioni

1. Polinomiali in *X*

Approssimiamo la funzione di regressione della popolazione con una polinomiale:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \; X_i^2 + \ldots + \beta_r \; + u_i$$

- È proprio il modello di regressione lineare multipla – salvo che i regressori sono potenze di *X*!
- Per stima, verifica delle ipotesi, ecc. si procede come nel modello di regressione multipla con OLS
- I coefficienti sono difficili da interpretare, ma la funzione risultante è interpretabile

Esempio: la relazione tra punteggio nei test e reddito distrettuale

$Income_i$ = reddito distrettuale medio nel distretto i^{esimo} (migliaia di dollari pro capite)

Approssimazione quadratica:

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 (Income_i)^2 + u_i$$

Approssimazione cubica:

$$TestScore_i = \beta_0 + \beta_1 Income_i + \beta_2 (Income_i)^2 + \beta_3 (Income_i)^3 + u_i$$

Stima dell'approssimazione quadratica in STATA

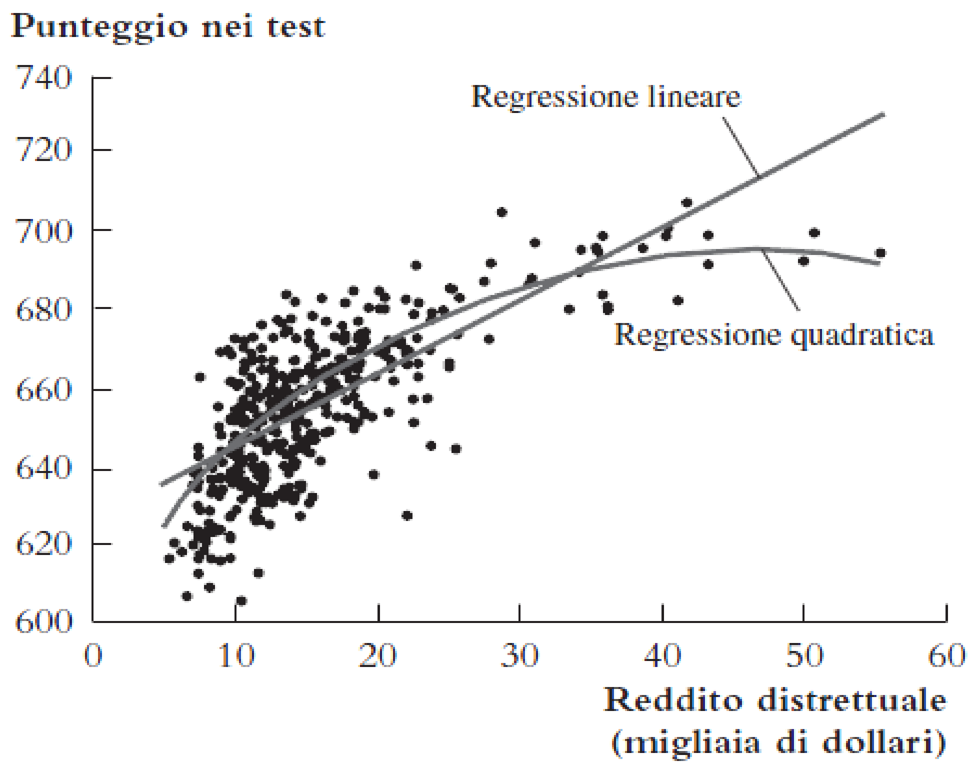
```
generate avginc2 = avginc*avginc;        Crea il regressore cubico
reg testscr avginc avginc2, r;

Regression with robust standard errors                 Number of obs =     420
                                                       F(  2,   417) =  428.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5562
                                                       Root MSE      =  12.724

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      avginc |   3.850995   .2680941    14.36   0.000      3.32401    4.377979
     avginc2 |  -.0423085   .0047803    -8.85   0.000     -.051705   -.0329119
       _cons |   607.3017   2.901754   209.29   0.000     601.5978    613.0056
------------------------------------------------------------------------------
```

Verifica l'ipotesi di linearità confrontandola con l'alternativa che la funzione di regressione sia quadratica….

Interpretazione della funzione di regressione stimata:

(a) Rappresentiamo graficamente i valori della stima

$$\widehat{TestScore} = \underset{(2,9)}{607,3} + \underset{(0,27)}{3,85}Income_i - \underset{(0,0048)}{0,0423}(Income_i)^2$$



Interpretazione della funzione di regressione stimata:

(b) Calcoliamo gli "effetti" per diversi valori di *X*

$$\widehat{TestScore} = 607{,}3 + 3{,}85 Income_i - 0{,}0423(Income_i)^2$$
$$\qquad\qquad (2{,}9) \quad (0{,}27) \qquad\qquad\qquad (0{,}0048)$$

Variazione predetta in *TestScore* per una variazione del reddito da $5.000 pro capite a $6.000 pro capite:

$$\Delta \widehat{TestScore} = 607{,}3 + 3{,}85 \times 6 - 0{,}0423 \times 6^2$$
$$- (607{,}3 + 3{,}85 \times 5 - 0{,}0423 \times 5^2)$$
$$= 3{,}4$$

$$\widehat{TestScore} = \mathbf{607{,}3 + 3{,}85}\,\boldsymbol{Income_i} \mathbf{- 0{,}0423}(\boldsymbol{Income_i})^{\mathbf{2}}$$

"Effetti" attesi in base ai diversi valori di X:

Variazione del reddito ($1000 pro capite)	**Δ** $\widehat{TestScore}$
da 5 a 6	**3,4**
da 25 a 26	**1,7**
da 45 a 46	**0,0**

L'"effetto" di un cambiamento del reddito è maggiore per i redditi più bassi (forse un beneficio marginale decrescente con l'aumento dei budget delle scuole?)

Attenzione! Qual è l'effetto di una variazione da 65 a 66?

Non estrapolate al di fuori dell'intervallo dei dati!

Stima dell'approssimazione cubica in STATA

```
gen avginc3 = avginc*avginc2;                            Crea il regressore cubico
reg testscr avginc avginc2 avginc3, r;

Regression with robust standard errors                  Number of obs =     420
                                                         F(  3,   416) =  270.18
                                                         Prob > F      =  0.0000
                                                         R-squared     =  0.5584
                                                         Root MSE      =  12.707

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      avginc |   5.018677   .7073505     7.10   0.000     3.628251    6.409104
     avginc2 |  -.0958052   .0289537    -3.31   0.001    -.1527191   -.0388913
     avginc3 |   .0006855   .0003471     1.98   0.049     3.27e-06    .0013677
       _cons |    600.079   5.102062   117.61   0.000     590.0499     610.108
------------------------------------------------------------------------------
```

Verifica dell'ipotesi nulla di linearità, contro l'alternativa che la funzione di regressione della popolazione sia quadratica e/o cubica, ovvero sia una polinomiale di grado fino a 3:

H_0: coefficienti di popolazione per $Income^2$ e $Income^3$ = 0
H_1: almeno uno di questi coefficienti è diverso da zero.

test avginc2 avginc3; **Eseguire il comando di test dopo aver eseguito la regressione**

```
 ( 1)   avginc2 = 0.0
 ( 2)   avginc3 = 0.0

       F(  2,   416) =   37.69
            Prob > F =    0.0000
```

L'ipotesi che la funzione di regressione della popolazione sia lineare viene rigettata al livello di significatività dell'1% contro l'alternativa che sia una polinomiale di grado fino a 3.

Riepilogo: funzioni di regressione polinomiali

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \quad + \ldots + \beta_r \quad + u_i$$

- Stima: via OLS dopo aver definito nuovi regressori
- I coefficienti hanno interpretazioni complicate
- Per interpretare la funzione di regressione stimata:
  - rappresentare graficamente i valori predetti come funzione di *x*
  - calcolare gli scarti predetti $\Delta Y/\Delta X$ per i diversi valori di *x*
- Le ipotesi sul grado *r* possono essere verificate tramite test *t* e *F* sugli appropriati blocchi di variabili.
- Scelta del grado *r*
  - rappresentare i dati graficamente, effettuare i test *t* e *F*, verificare la sensibilità e gli effetti stimati, giudicare.
  - *In alternativa usare il criterio di scelta del modello (più avanti)*

2. Funzioni logaritmiche di *Y* e/o *X*

- ln(*X*) = è il logaritmo naturale di *X*
- Le trasformazioni logaritmiche permettono di modellare le relazioni in termini "percentuali" (come l'elasticità) invece che linearmente.

Ecco perché: $\ln(x+\Delta x) - \ln(x) = \ln\left(1+\frac{\Delta x}{x}\right) \cong \frac{\Delta x}{x}$

(calcolo: $\frac{d\ln(x)}{dx} = \frac{1}{x}$)

Numericamente:

ln(1,01) = 0,00995 ≅ 0,01;

ln(1,10) = 0,0953 ≅ 0,10 (circa)

Le tre specificazioni di regressione logaritmica:

Caso	Funzione di regressione della popolazione
I. lineare-log	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$
II. log-lineare	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III. log-log	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$

- L'interpretazione del coefficiente pendenza è diversa in ciascun caso.
- L'interpretazione si trova applicando la regola generale "prima e dopo": predire la variazione in *Y* per una data variazione in *X*."
- Ogni caso ha una diversa interpretazione naturale (per piccole variazioni in *X*)

I. Funzione di regressione della popolazione lineare-logaritmica

Calcolare Y "prima" e "dopo" aver modificato la X:

$Y = \beta_0 + \beta_1 \ln(X)$ ("prima")

Ora cambiamo X: $Y + \Delta Y = \beta_0 + \beta_1 \ln(X + \Delta X)$ ("dopo")

Sottrarre ("dopo") – ("prima"): $\Delta Y = \beta_1[\ln(X + \Delta X) - \ln(X)]$

ora $\ln(X + \Delta X) - \ln(X) \cong \frac{\Delta X}{X}$,

quindi $\Delta Y \cong \beta_1 \frac{\Delta X}{X}$

o $\beta_1 \cong \frac{\Delta Y}{\Delta X / X}$ (per piccole ΔX)

Caso lineare-logaritmico (continua)

$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$

per piccole ΔX,

$$\beta_1 \cong \frac{\Delta Y}{\Delta X / X}$$

Ora $100 \times \dfrac{\Delta X}{X}$ = variazione percentuale in X, quindi ***un incremento dell'1% in X (moltiplicare X per 1,01) è associato a una variazione di 0,01β_1 in Y.***

(1% incremento in X --> 0,01 incremento in ln(X)
--> 0,01β_1 incremento in Y)

Esempio: TestScore su ln(Income)

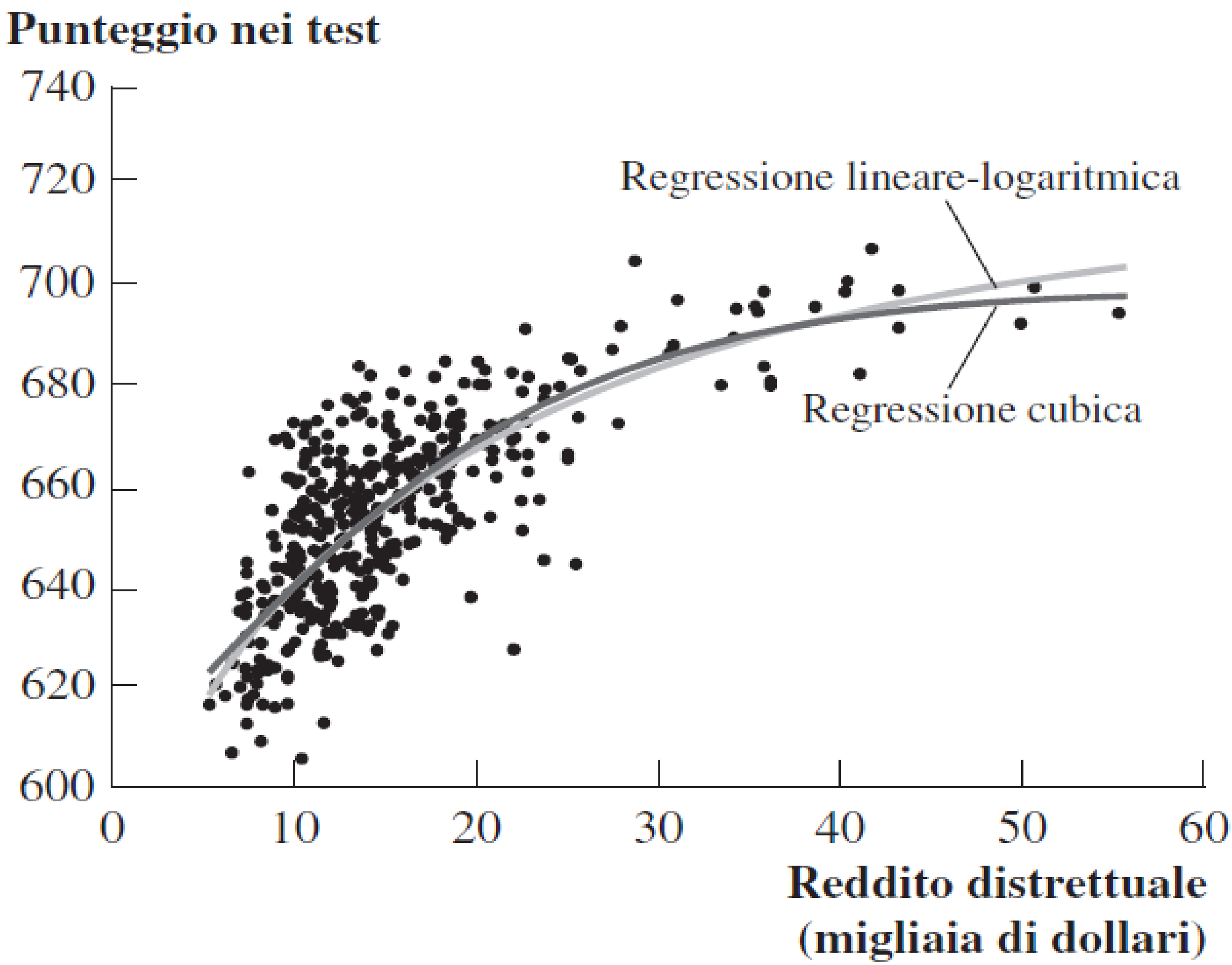
- Definiamo innanzitutto il nuovo regressore, ln(*Income*)
- Il modello è ora lineare su ln(*Income*), quindi possiamo stimare il modello lineare-log tramite OLS:

$$\widehat{TestScore} = \underset{(3,8)}{557{,}8} + \underset{(1{,}40)}{36{,}42} \times \ln(Income_i)$$

quindi un incremento dell'1% in *Income* è associato a un aumento di 0,36 nel punteggio nei test.

- Si applicano tutti i soliti meccanismi di regressione: errori standard, intervalli di confidenza, R^2.
- Come confrontare tutto questo con il modello cubico?

Le funzioni di regressione lineare-logaritmica e cubica



II. Funzione di regressione della popolazione log-lineare

$$\ln(Y) = \beta_0 + \beta_1 X \quad \text{(b)}$$

Variamo X: $$\ln(Y + \Delta Y) = \beta_0 + \beta_1(X + \Delta X) \quad \text{(a)}$$

Sottraiamo (a) – (b): $$\ln(Y + \Delta Y) - \ln(Y) = \beta_1 \Delta X$$

da cui $$\frac{\Delta Y}{Y} \cong \beta_1 \Delta X$$

o $$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X} \quad \text{(per } \Delta X \text{ } \textit{piccole}\text{)}$$

Caso log-lineare (continua)

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

per piccole ΔX, $\beta_1 \cong \dfrac{\Delta Y / Y}{\Delta X}$

- ora $100 \times \dfrac{\Delta Y}{Y}$ = percentuale di variazione in Y, quindi ***una variazione in X di un'unità* (Δ*X* = 1) *si associa a una variazione di* 100β₁% *in Y*.**
- 1 unità di incremento in $X \rightarrow \beta_1$ incremento in $\ln(Y)$
  $\rightarrow 100\beta_1\%$ incremento in Y
- *Nota*: quali sono le unità di u_i e SER?
  - deviazioni frazionali (proporzionali)
  - per esempio *SER* = 0,2 significa…

III. Funzione di regressione della popolazione log-log

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i \qquad \text{(b)}$$

Variamo X: $$\ln(Y + \Delta Y) = \beta_0 + \beta_1 \ln(X + \Delta X) \qquad \text{(a)}$$

Sottraiamo: $$\ln(Y + \Delta Y) - \ln(Y) = \beta_1[\ln(X + \beta X) - \ln(X)]$$

Da cui $$\frac{\Delta Y}{Y} \cong \beta_1 \frac{\Delta X}{X}$$

O $$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X / X}$$ (per piccole ΔX)

Caso log-log (continua)

$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$

per piccole ΔX,

$$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X / X}$$

Ora $100 \times \frac{\Delta Y}{Y}$ = variazione percentuale in Y, e $100 \times \frac{\Delta X}{X}$ = variazione percentuale in X, per cui ***una variazione dell'1% in X produce una variazione del β_1% in Y***.

Nella specifica log-log, β_1 ha l'interpretazione di un coefficiente di elasticità.

Esempio: ln(TestScore) su ln(Income)

- Per prima cosa definiamo una nuova variabile dipendente, ln(*TestScore*) ***e*** il nuovo regressore, ln(*Income*)
- Il modello ora è una regressione lineare di ln(*TestScore*) su ln(*Income*) che può essere stimata mediante OLS:

$$\widehat{\ln(TestScore)} = \underset{(0{,}006)}{6{,}336} + \underset{(0{,}0021)}{0{,}0554} \times \ln(Income_i)$$

A un aumento dell'1% in *Income* si associa un aumento dello 0.0554% in *TestScore* (*Income* aumenta di un fattore 1,01, *TestScore* di un fattore 1,000554)

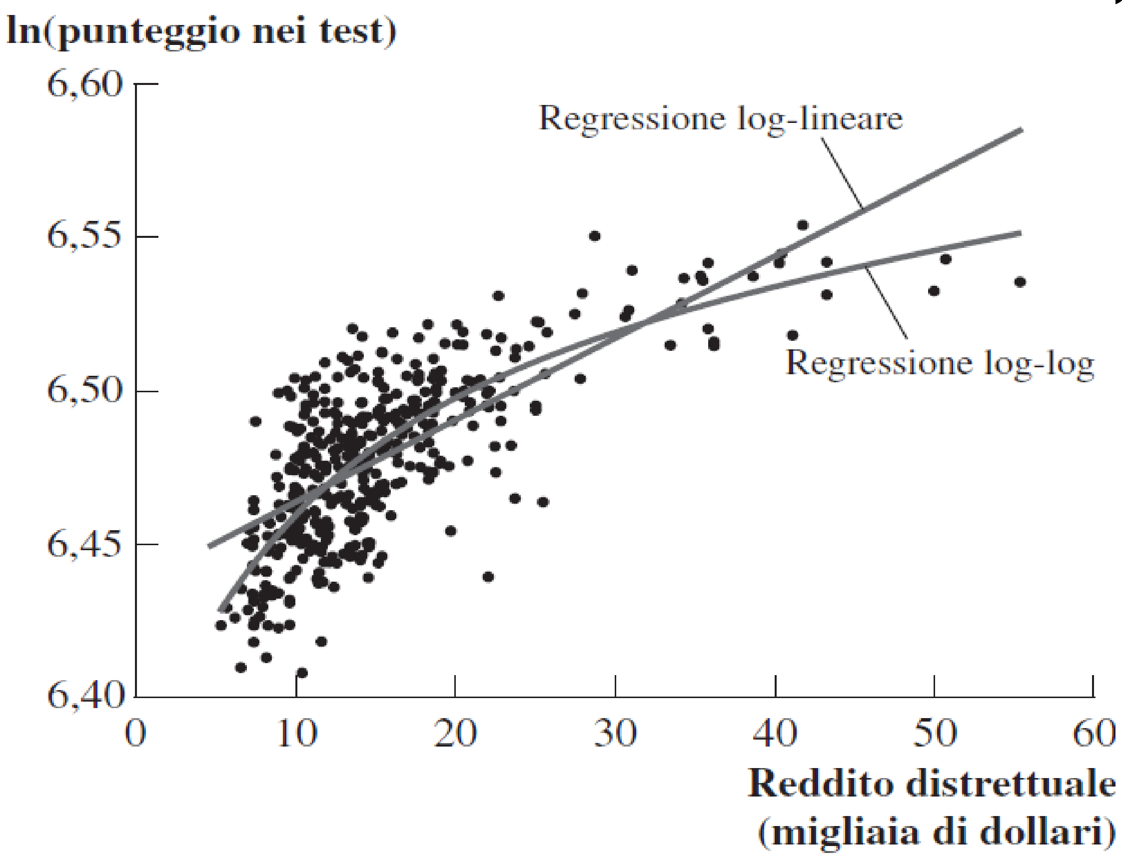
Esempio: ln(TestScore) su ln(Income) (continua)

$$\widehat{\ln(TestScore)} = 6{,}336 + 0{,}0554 \times \ln(Income_i)$$
$$\qquad\qquad (0{,}006) \quad (0{,}0021)$$

- Per esempio, supponiamo che il reddito salga da 10,000$ a 11,000$, o del 10%. Quindi *TestScore* cresce approssimativamente di $0{,}0554 \times 10\%$ = 0,554%. Se *TestScore* = 650, questo corrisponde a un aumento di $0{,}00554 \times 650$ = 3,6 punti.

- Come si confronta rispetto al modello log-lineare?

Le specifiche log-lineare e log-log:



- *Notate l'asse verticale*
- *Niente sembra adattarsi meglio della cubica o lineare-log, almeno in base all'aspetto visivo (il confronto formale è dificile perché le variabili dipendenti differiscono)*

Riepilogo: trasformazioni logaritmiche

- Tre casi, differiscono in base alla o alle variabili *Y* e/o *X* trasformate in logaritmi.
- La regressione diventa lineare sulla(e) nuova(e) variabile(i) ln(*Y*) e/o ln(*X*), mentre i coefficienti possono essere stimati attraverso l'OLS.
- I test di ipotesi e gli intervalli di affidabilità possono essere implementati e interpretati "nel solito modo"
- L'interpretazione di β_1 differisce caso per caso.

La scelta della specificazione (forma funzionale) dev'essere guidata dal ragionamento – quale interpretazione ha più senso nella vostra applicazione? – da test e dall'analisi grafica dei valori predetti

Altre funzioni non lineari (e minimi quadrati non lineari) (Appendice 8.1)

Le funzioni di regressione precedenti hanno delle limitazioni…

- Polinomiali: il punteggio nei test può decrescere all'aumentare del reddito
- Lineare-log: il punteggio aumenta con il reddito, ma senza limite
- Questa è una funzione non lineare in cui la *Y* cresce sempre con *X* e c'è un massimo valore di *Y* (asintoto):

$$Y = \beta_0 - \alpha e^{-\beta_1 X}$$

β_0, β_1 e α sono parametri sconosciuti. Viene chiamata curva di crescita esponenziale negativa. L'asintoto per $X \to \infty$ è β_0.

Crescita esponenziale negativa

Vogliamo stimare i parametri di

$$Y_i = \beta_0 - \alpha e^{-\beta_1 X_i} + u_i$$

o

$$Y_i = \beta_0\left[1 - e^{-\beta_1(X_i - \beta_2)}\right] + u_i \qquad (*)$$

dove $\alpha = \beta_0 e^{\beta_1\beta_2}$ (perché vogliamo farlo?)

Compariamo il modello (*) con quelli lineare-log e cubico:

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_2 X_i^3 + u_i$$

I modelli lineare-log e polinomiale sono *lineari nei parametri* β_0 e β_1 – mentre il modello (*) no.

Minimi quadrati non lineari

- I modelli i cui parametri sono lineari possono essere stimati tramite OLS.
- I modelli non lineari in uno o più parametri possono essere stimati con i minimi quadrati non lineari (NLS) ma non tramite gli OLS.
- Il problema NLS per la specificazione proposta:

$$\min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{n} \left\{ Y_i - \beta_0 \left[1 - e^{-\beta_1 (X_i - \beta_2)} \right] \right\}^2$$

È un problema di minimizzazione non lineare (un problema di "hill-climbing"). Come risolverlo?

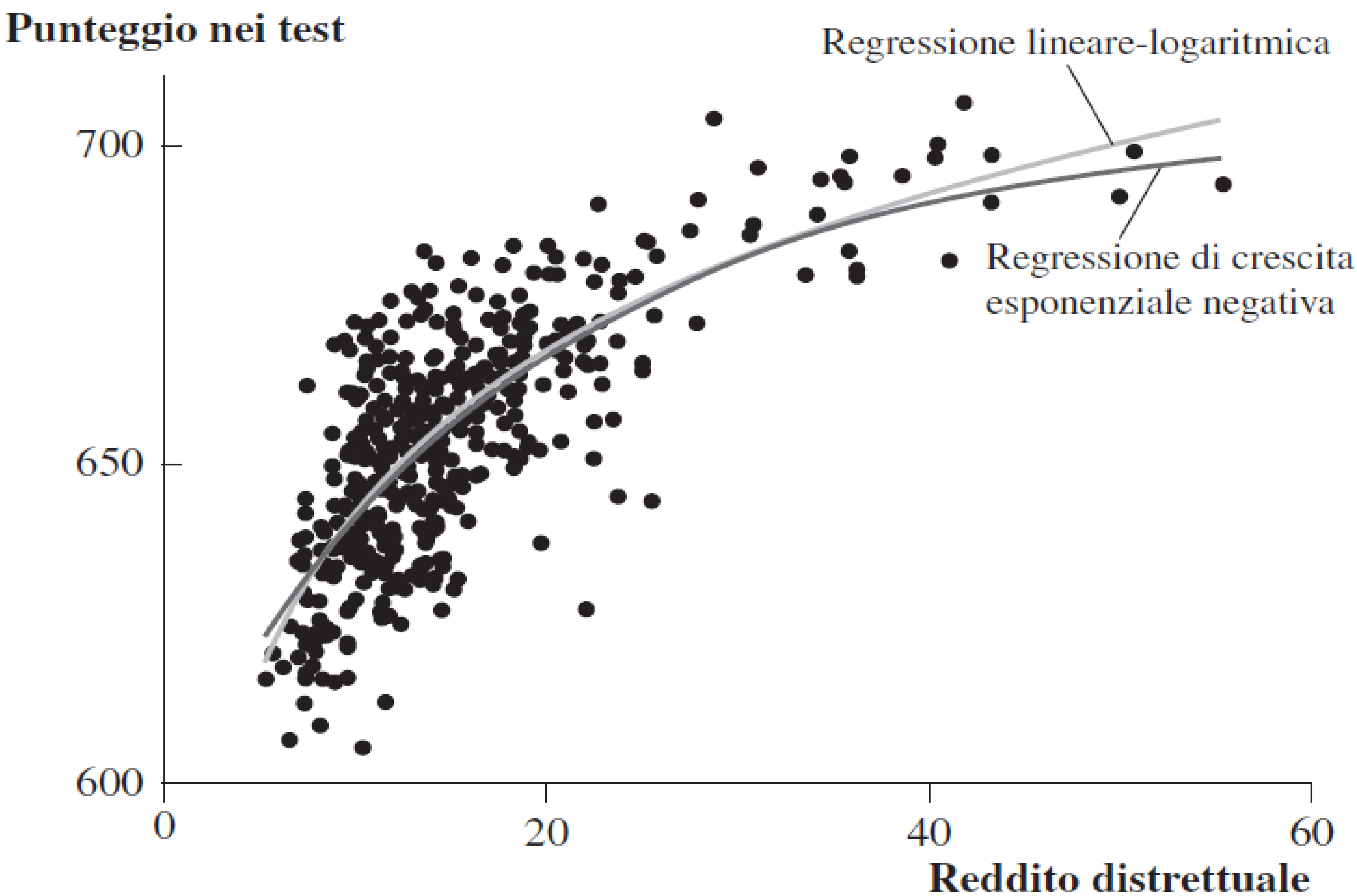
–Tirare a indovinare e verificare

–Ci sono modi migliori…

–Implementazione in STATA…

```
. nl (testscr = {b0=720}*(1 - exp(-1*{b1}*(avginc-{b2})))), r

(obs = 420)
Iteration 0:  residual SS =  1.80e+08                    .
Iteration 1:  residual SS =  3.84e+07                    .
Iteration 2:  residual SS =   4637400                    .
Iteration 3:  residual SS =  300290.9         STATA sta "scalando la collina"
Iteration 4:  residual SS =  70672.13         (minimizzando l'SSR)
Iteration 5:  residual SS =  66990.31                    .
Iteration 6:  residual SS =   66988.4                    .
Iteration 7:  residual SS =   66988.4                    .
Iteration 8:  residual SS =   66988.4

Nonlinear regression with robust standard errors     Number of obs =        420
                                                     F(  3,    417) = 687015.55
                                                     Prob > F      =    0.0000
                                                     R-squared     =    0.9996
                                                     Root MSE      =  12.67453
                                                     Res. dev.     =  3322.157
------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          b0 |   703.2222   4.438003   158.45   0.000     694.4986    711.9459
          b1 |   .0552339   .0068214     8.10   0.000     .0418253    .0686425
          b2 |  -34.00364    4.47778    -7.59   0.000    -42.80547    -25.2018
------------------------------------------------------------------------------
  (SEs, P values, CIs, and correlations are asymptotic approximations)
```

Crescita esponenziale negativa; *RMSE* = 12,675
Linear-log; *RMSE* = 12,618



Interazioni tra variabili indipendenti (Paragrafo 8.3)

- Forse ridurre la dimensione di una classe è più efficace in alcune circostanze che in altre…
- Forse classi più piccole sono migliori se ci sono molti allievi non di madrelingua, che richiedono attenzioni individuali
- Ovvero, $\dfrac{\Delta TestScore}{\Delta STR}$ può dipendere da *PctEL*
- Più in generale, $\dfrac{\Delta Y}{\Delta X_1}$ può dipendere da X_2
- Come modellare queste "interazioni" tra X_1 e X_2?
- Consideriamo prima delle *X binarie*, poi delle *X continue*

(a) Interazioni tra due variabili binarie

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$$

- D_{1i}, D_{2i} sono binarie
- β_1 è l'effetto che si ha cambiando D_1=0 in D_1=1. In questa specificazione, *questo effetto non dipende dal valore di D_2.*
- Per far sì che la modifica di D_1 dipenda da D_2, si inserisce il "termine d'interazione" $D_{1i} \times D_{2i}$ come regressore:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

Interpretazione dei coefficienti

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

Regola generale: confrontare i vari casi

$$E(Y_i|D_{1i}=0, D_{2i}=d_2) = \beta_0 + \beta_2 d_2 \qquad \text{(b)}$$

$$E(Y_i|D_{1i}=1, D_{2i}=d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2 \qquad \text{(a)}$$

sottrarre (a) – (b):

$$E(Y_i|D_{1i}=1, D_{2i}=d_2) - E(Y_i|D_{1i}=0, D_{2i}=d_2) = \beta_1 + \beta_3 d_2$$

- L'effetto di D_1 dipende da d_2 (quel che volevamo)
- β_3 = incremento dell'effetto di D_1, quando $D_2 = 1$

Esempio: TestScore, STR, allievi non di madrelingua

Sia

$HiSTR = \begin{cases} 1 \text{ se } STR \geq 20 \\ 0 \text{ se } STR < 20 \end{cases}$ e $HiEL = \begin{cases} 1 \text{ se } PctEL \geq 10 \\ 0 \text{ se } PctEL < 10 \end{cases}$

$$\widehat{TestScore} = \underset{(1,4)}{664{,}1} - \underset{(2,3)}{18{,}2}HiEL - \underset{(1,9)}{1{,}9}HiSTR - \underset{(3,1)}{3{,}5}(HiSTR \times HiEL)$$

- "Effetto" di *HiSTR* quando *HiEL* = 0 è –1,9
- "Effetto" di *HiSTR* quando *HiEL* = 1 è –1,9 – 3,5 = –5,4
- Sis tima che la riduzione della dimensione della classe abbia un effetto maggiore quando la percentuale degli allievi non di madrelingua è elevata
- Questa interazione non è statisticamente significativa: t = 3.5/3.1

Esempio: *TestScore*, *STR*, allievi non di madrelingu (continua)

Siano

$$HiSTR = \begin{cases} 1 \text{ se } STR \geq 20 \\ 0 \text{ se } STR < 20 \end{cases} \quad \text{e} \quad HiEL = \begin{cases} 1 \text{ se } PctEL \geq 10 \\ 0 \text{ se } PctEL < 10 \end{cases}$$

$$\widehat{TestScore} = \underset{(1,4)}{664{,}1} - \underset{(2,3)}{18{,}2}HiEL - \underset{(1,9)}{1{,}9}HiSTR - \underset{(3,1)}{3{,}5}(HiSTR \times HiEL)$$

- Siete capaci di correlare questi coefficienti con i gruppi ("celle") della tabella seguente e i relativi significati?

	STR basso	***STR elevato***
EL basso	664,1	662,2
EL elevato	645,9	640,5

(b) Interazioni tra variabili continue e binarie

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i$$

- D_i è binaria, X è continua
- Come specificato prima, l'effetto su Y di X (tenendo costante D) = β_2, che non dipende da D
- Per far sì che l'effetto di X *dipenda da D*, includiamo il "termine d'interazione" $D_i \times X_i$ come regressore:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

Interazioni tra variabili continue e binarie: le due rette di regressione

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

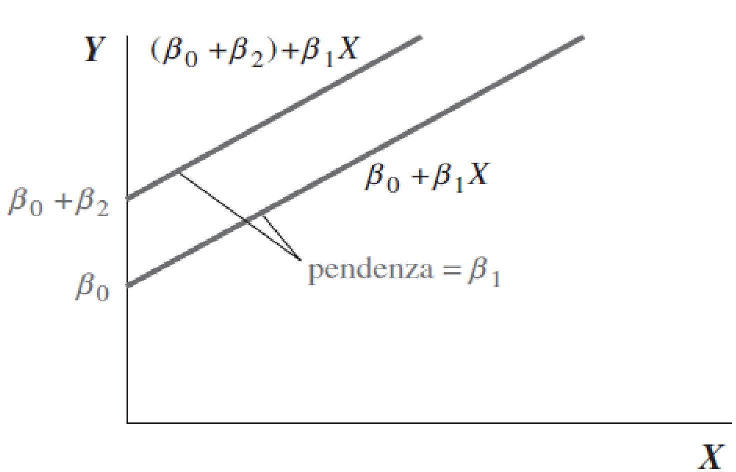
Osservazione con $D_i = 0$ (il gruppo "$D = 0$"):

$$Y_i = \beta_0 + \beta_2 X_i + u_i$$ ***Retta di regressione con D=0***

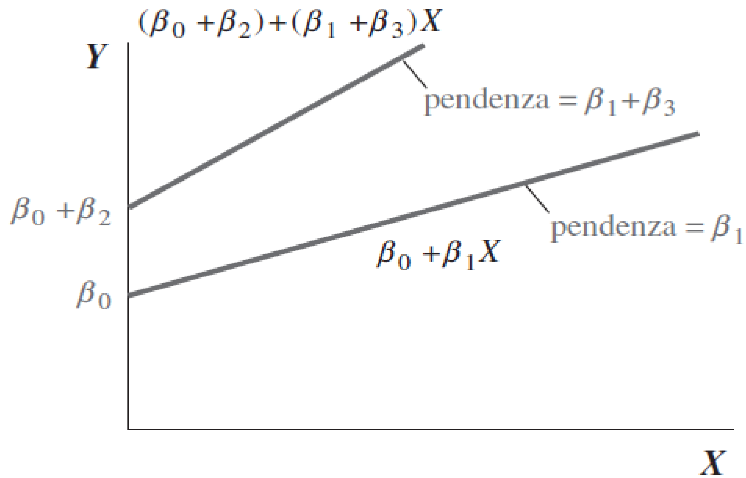
Osservazione con $D_i = 1$ (il gruppo "$D = 1$"):

$$Y_i = \beta_0 + \beta_1 + \beta_2 X_i + \beta_3 X_i + u_i$$
$$= (\beta_0 + \beta_1) + (\beta_2 + \beta_3) X_i + u_i$$ ***Retta di regressione con D=1***

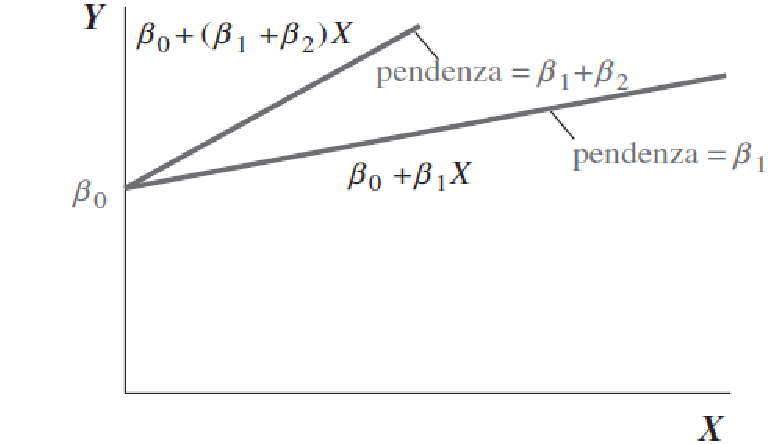
Interazioni tra variabili continue e binarie (continua)



(a) Intercette diverse, pendenze uguali



(b) Intercette diverse, pendenze diverse



(c) Intercette uguali, pendenze diverse

Interpretazione dei coefficienti

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

Regola generale: confrontare i diversi casi

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (D \times X) \qquad \text{(b)}$$

Ora cambiamo X:

$$Y + \Delta Y = \beta_0 + \beta_1 D + \beta_2 (X + \Delta X) + \beta_3 [D \times (X + \Delta X)] \qquad \text{(a)}$$

sottrarre (a) – (b):

$$\Delta Y = \beta_2 \Delta X + \beta_3 D \Delta X \quad \text{o} \quad \frac{\Delta Y}{\Delta X} = \beta_2 + \beta_3 D$$

- L'effetto di X dipende da D (quel che volevamo)
- β_3 = incremento dell'effetto di X, quando $D = 1$

Esempio: *TestScore*, *STR*, *HiEL* (=1 se *PctEL* ≥ 10)

$$\widehat{TestScore} = 682{,}2 - 0{,}97STR + 5{,}6HiEL - 1{,}28(STR \times HiEL)$$
$$\quad (11{,}9)\ (0{,}59) \qquad (19{,}5) \qquad (0{,}97)$$

Quando *HiEL* = 0:

$$\widehat{TestScore} = 682{,}2 - 0{,}97STR$$

- Quando *HiEL* = 1,

$$\widehat{TestScore} = 682{,}2 - 0{,}97STR + 5{,}6 - 1{,}28STR$$

$$\widehat{TestScore} = 687{,}8 - 2{,}25STR$$

- Due rette di regressione: una per ciascun gruppo *HiSTR*.
- Si stima che riduzione della dimensione della classe abbia un effetto maggiore quanto più è ampia la percentuale degli studenti non di madrelingua.

Esempio (continua): verifica delle ipotesi

$$\widehat{TestScore} = 682{,}2 - 0{,}97STR + 5{,}6HiEL - 1{,}28(STR \times HiEL)$$
$$\quad (11{,}9) \quad (0{,}59) \quad (19{,}5) \quad (0{,}97)$$

- Le due rette di regressione hanno la stessa **pendenza** ← → il coefficiente su $STR \times HiEL$ è zero: $t = -1{,}28/0{,}97 = -1{,}32$
- Le due rette di regressione hanno lo stesso **punto di intercetta** ← → il coefficiente di *HiEL* è zero: $t = -5{,}6/19{,}5 = 0{,}29$
- Le due rette di regressione coincidono ← → il coefficiente di *HiEL* = 0 ***e*** quello di $STR \times HiEL = 0$: $F = 89{,}94$
(valore-p < 0,001) **!!**
- Scartiamo le ipotesi congiunte ma non quelle individuali (come può essere?)

(c) Interazioni tra due variabili continue

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

- X_1, X_2 sono continue
- Come specificato, l'effetto di X_1 non dipende da X_2
- Come specificato, l'effetto di X_2 non dipende da X_1
- Per far sì che l'effetto di X_1 dipenda da X_2, includiamo il "termine d'interazione" $X_{1i} \times X_{2i}$ come regressore:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$$

Interpretazione dei coefficienti:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3(X_{1i} \times X_{2i}) + u_i$$

Regola generale: comparazione dei vari casi

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3(X_1 \times X_2) \qquad \text{(b)}$$

Ora cambiamo X_1:

$$Y + \Delta Y = \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2 X_2 + \beta_3[(X_1 + \Delta X_1) \times X_2] \qquad \text{(a)}$$

Sottraiamo (a) – (b):

$$\Delta Y = \beta_1 \Delta X_1 + \beta_3 X_2 \Delta X_1 \quad \text{or} \quad \frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$$

- L'effetto di X_1 dipende da X_2 (quel che volevamo)
- β_3 = incremento dell'effetto di X_1 a seguito dell'aumento di un'unità di X_2

Esempio: *TestScore*, *STR*, *PctEL*

$$\widehat{TestScore} = \underset{(11,8)}{686,3} - \underset{(0,59)}{1,12}STR - \underset{(0,37)}{0,67}PctEL + \underset{(0,019)}{0,0012}(STR \times PctEL),$$

L'effetto stimato della riduzione della dimensione della classe è non lineare, perché la dimensione dell'effetto stesso dipende da *PctEL*:

$$\frac{\Delta TestScore}{\Delta STR} = -1,12 + 0,0012PctEL$$

PctEL	$\frac{\Delta TestScore}{\Delta STR}$
0	–1,12
20%	$-1,12 + 0,0012 \times 20 = -1,10$

Esempio (continua): verifica delle ipotesi

$$\widehat{TestScore} = 686{,}3 - 1{,}12STR - 0{,}67PctEL + 0{,}0012(STR \times PctEL),$$
$$\quad (11{,}8) \quad (0{,}59) \quad (0{,}37) \quad (0{,}019)$$

- Il coefficiente di $STR \times PctEL$ è = 0?

t = 0,0012/0,019 = 0,06 → non si può scartare a livello del 5%

- Il coefficiente di *STR* è = 0?

t = −1,12/0,59 = −1,90 → non si può scartare a livello del 5%

- I coefficienti di **entrambi** *STR* ***e*** $STR \times PctEL$ sono = 0?

F = 3,89 (valore-p = 0,021) → si scarta a livello del 5% (!!) (Perché? Multicollinearità alta ma imperfetta)

Applicazione: effetti non lineari del rapporto studenti/insegnanti sui punteggi nei test (Paragrafo 8.4)

Le specificazioni non lineari ci permettono di esaminare dettagli meno evidenti della relazione tra punteggi nei test e STR, quali:

1. Ci sono effetti non lineari della riduzione della dimensione della classe sui punteggi nei test? (Una riduzione da 35 a 30 ha lo stesso effetto di una riduzione da 20 a 15?)
2. Ci sono interazioni non lineari tra *PctEL* e *STR*? (Le classi piccole sono più efficaci quando ci sono molti studenti non di madrelingua?)

Strategia per la domanda #1 (effetti diversi per STR diversi?)

- Stimare funzioni lineari e non lineari di *STR*, mantenendo costanti le rilevanti variabili demografiche
  - *PctEL*
  - *Income* (si ricordi la relazione non lineare tra punteggio nei test e reddito)
  - *LunchPCT* (pranzo libero /sovvenzionato)
- Verificare se aggiungendo dei termini non lineari si ha una differenza quantitativa “economicamente rilevante” (l’importanza “economica” o “reale” è diversa e quindi statisticamente significativa)
- Verificare se i termini non lineari sono significativi

Strategia per la domanda #2 (interazioni tra *PctEL* e *STR*?)

- Stimare le funzioni lineari e non lineari di *STR*, con l'interazione di *PctEL*.
- Se la specificazione è non lineare (con STR, STR^2, STR^3), allora occorre aggiungere interazioni con tutti i termini, in modo che la risultante forma funzionale possa essere diversa, al variare del livello di *PctEL*.
- Utilizzare una specificazione con interazione binaria-continua aggiungendo $HiEL \times STR$, $HiEL \times STR^2$ e $HiEL \times STR^3$.

Qual è una buona specificazione di "base"?

- La relazione *Punteggio nei test – Reddito*:
- La specificazione logaritmica si comporta meglio verso gli estremi del campione, specialmente per valori di reddito alti.

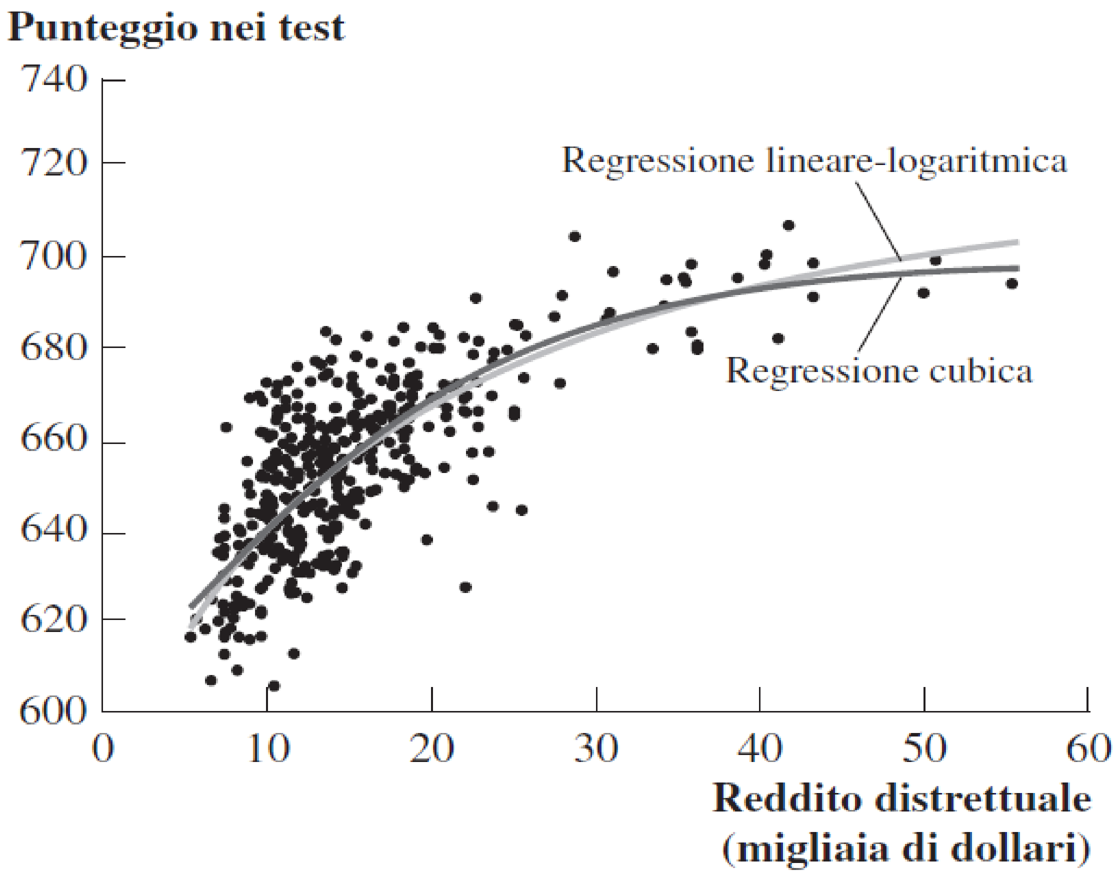


Tabella 8.3 Modelli di regressione non-lineari per il punteggio nei test.

Variabile dipendente: media del punteggio nei test del distretto; 420 osservazioni

Regressori	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
STR	– 1,00** (0,27)	– 0,73** (0,26)	– 0,97 (0,59)	– 0,53 (0,34)	64,33** (24,86)	83,70** (28,50)	65,29** (25,26)
STR^2					– 3,42** (1,25)	– 4,38** (1,44)	– 3,47** (1,27)
STR^3					0,059** (0,021)	0,075** (0,024)	0,060** (0,021)
% studenti non madrelingua	– 0,122** (0,033)	– 0,176** (0,034)					– 0,166** (0,034)
% studenti non madrelingua ≥ 10% (Binario, *HiEL*)			5,64 (19,51)	5,50 (9,80)	– 5,47 (1,03)	816,1* (327,7)	
HiEL × *STR*			– 1,28 (0,97)	– 0,58 (0,50)		– 123,3* (50,2)	
HiEL × STR^2						6,12* (2,54)	
HiEL × STR^3						– 0,101* (0,043)	
% aventi diritto al sussidio mensa	– 0,547** (0,024)	– 0,398** (0,033)		– 0,411** (0,029)	– 0,420** (0,029)	– 0,418** (0,029)	– 0,402** (0,033)
Reddito medio nel distretto (logaritmo)		11,57** (1,819		12,12** (1,80	11,75** (1,78	11,80** (1,78	11,51** (1,81
Intercetta	700,2** (5,6)	658,6** (8,6)	682,2** (11,9)	653,6** (9,9)	252,0 (163,6)	122,3 (185,5)	244,8 (165,7)

Verifica di ipotesi congiunte:

Statistiche F e valori-p per le ipotesi congiunte							
(a) tutte le variabili STR e le interazioni = 0			5,64 (0,004)	5,92 (0,003)	6,31 (< 0,001)	4,96 (< 0,001)	5,91 (< 0,001)
(b) STR^2 e $STR^3 = 0$					6,17 (< 0,001)	5,81 (0,003)	5,96 (0,003)
(c) $HiEL \times STR$, $HiEL \times STR^2$, $HiEL \times STR^2 = 0$						2,96 (0,046)	
SER	9,08	8,64	15,88	8,63	8,56	8,55	8,57
$\bar{R}^2$	0,773	0,794	0,305	0,795	0,798	0,799	0,798

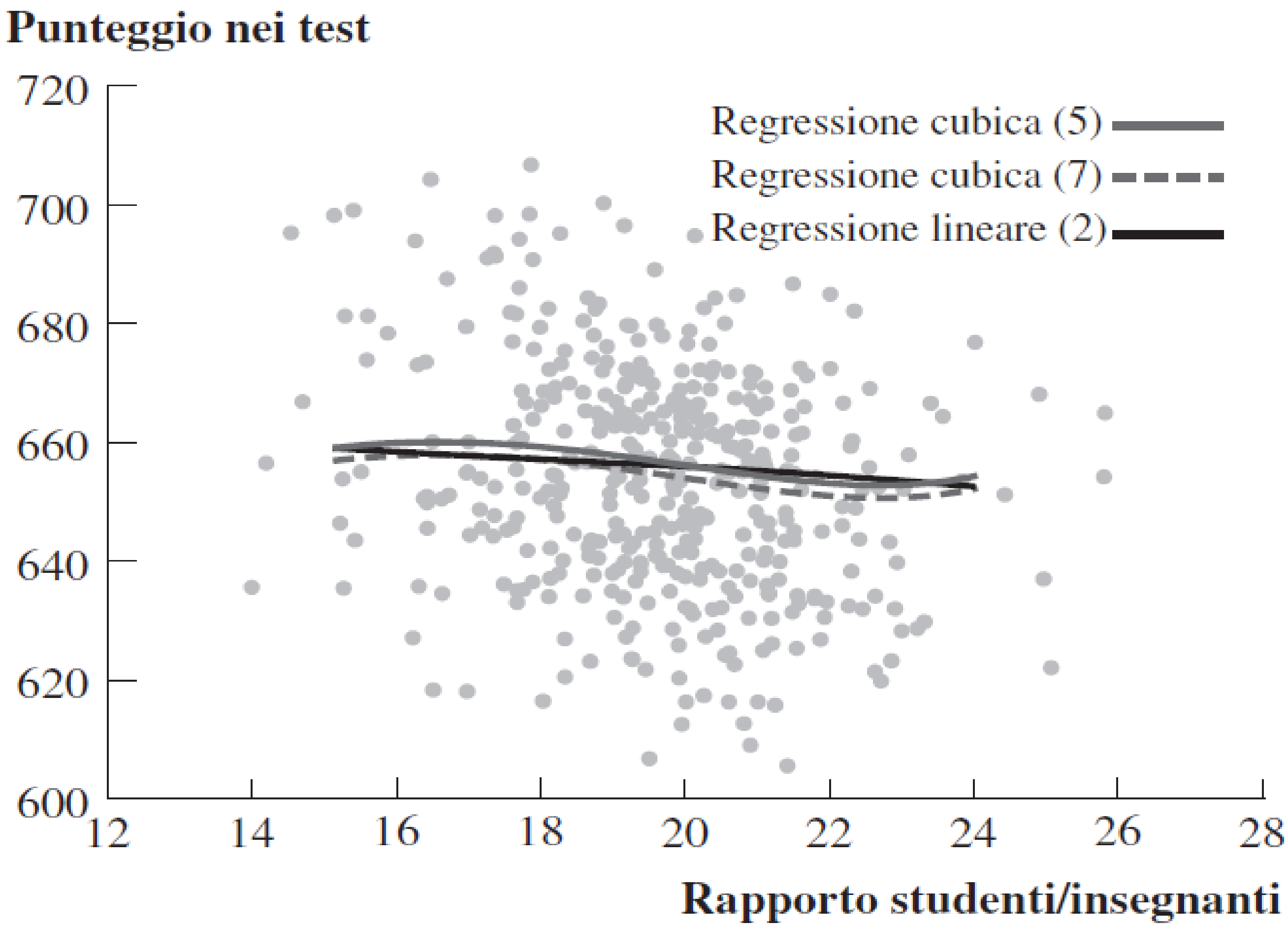
Queste regressioni sono state stimate utilizzando i dati sui distretti scolastici K-8 della California, descritti nell'Appendice 4.1. Gli errori standard sono riportati in parentesi sotto i coefficienti e i valori-p sono riportati in parentesi sotto le statistiche F. I coefficienti sono singolarmente significativi al livello del *5% o dell'**1%.

Che cosa potete concludere sulla domanda #1?

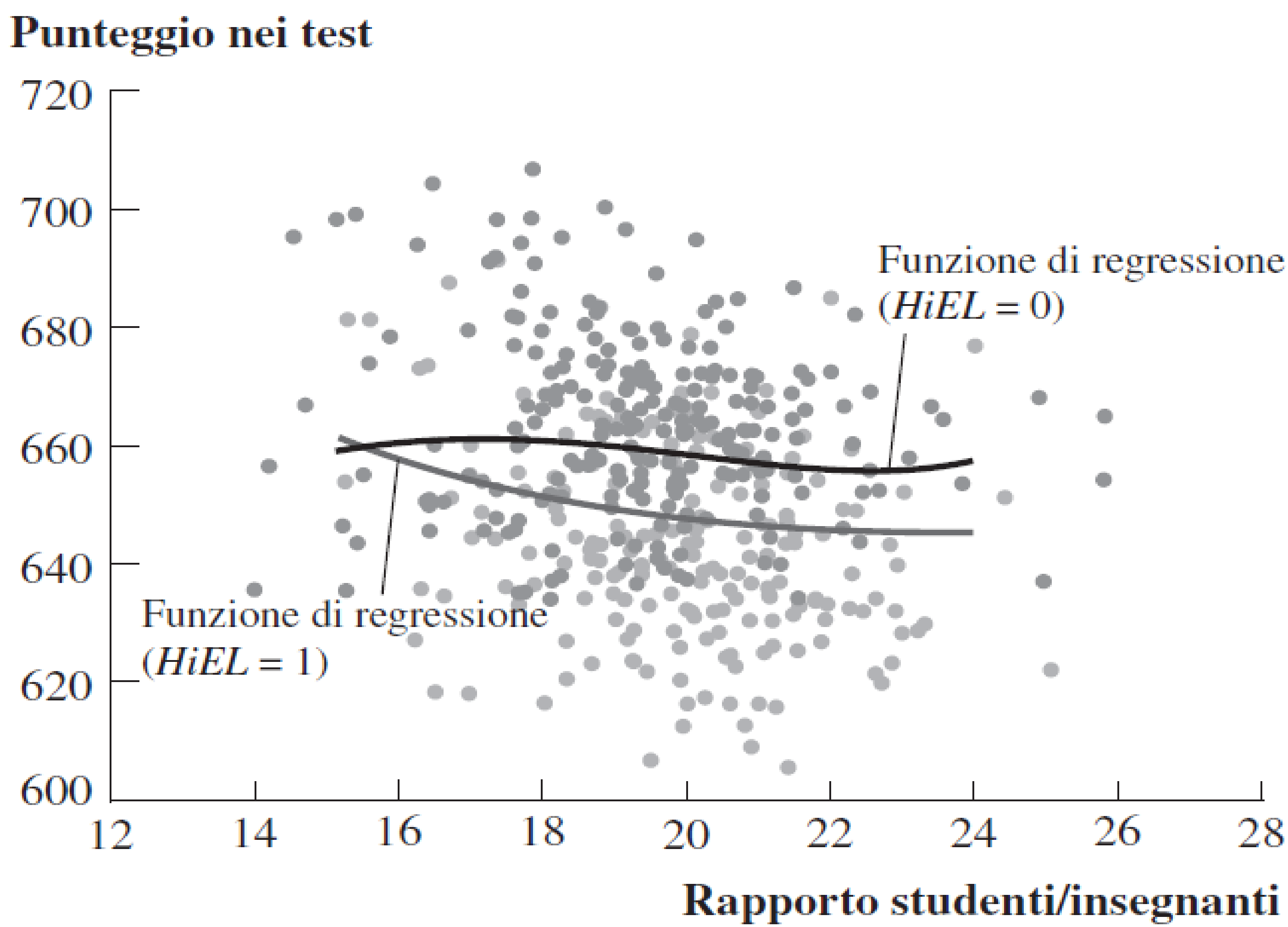
E sulla domanda #2?

Interpretazione delle funzioni di regressione per via grafica:

Per prima cosa, confrontate le specificazioni lineari e non lineari:



Quindi confrontate le regressioni con le interazioni:



Riepilogo: funzioni di regressione non lineari

- Utilizzando funzioni di variabili indipendenti come $\ln(X)$ o $X_1 \times X_2$, possiamo riformulare una vasta famiglia di funzioni di regressione lineare come regressioni multiple.
- La stima e l'inferenza procedono in modo analogo al modello di regressione lineare multiplo.
- L'interpretazione dei coefficienti è specifica del modello utilizzato, ma la regola generale consiste nel calcolare gli effetti confrontando i casi diversi (i diversi valori delle X originali)
- Sono possibili molte specificazioni non lineari, per cui è necessario riflettere:
  - Quali effetti non lineari si vogliono analizzare?
  - Quale ha senso nella particolare applicazione considerata?