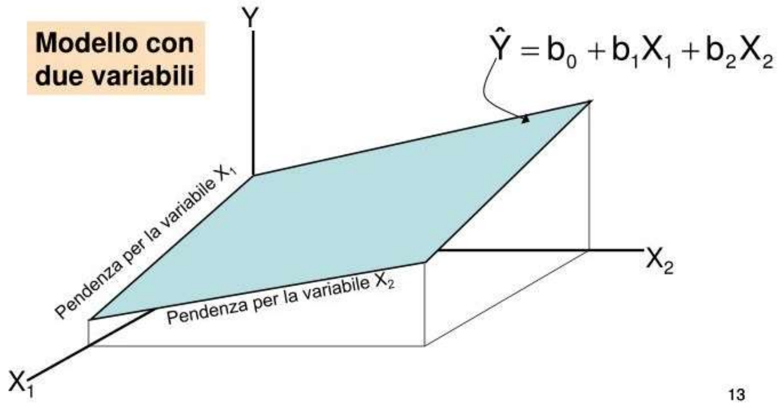
Regressione lineare multipla

Regressione lineare multipla

Il modello di regressione lineare semplice può essere esteso per includere due o più variabili indipendenti X. Il termine «regressione multivariata» è spesso (erroneamente) usato per riferirsi alla regressione multipla.

Ad esempio, la relazione tra due variabili indipendenti X_1, X_2 e la risposta Y può essere scritta ha:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$



Regressione lineare multipla

Il metodo dei minimi quadrati può ancora essere utilizzato per stimare i coefficienti b_0, b_1 e b_2 che definiscono il piano di regressione (o «iperpiano» con 3 o più variabili X). L'obiettivo è ridurre al minimo la quantità:

$$SS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i})]^2$$

(esistono soluzioni algebriche ma il calcolo diventa più complesso e richiede la spiegazione della notazione matriciale)

Anche qui la bontà dell'adattamento può essere valutata dall'indice R^2 :

$$R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2} = \frac{SS_{regressione}}{SS_{tot}} =$$ Proporzione di varianza di Y spiegata da X_1 e X_2

Esempio Regressione lineare multipla

$$\text{Pressione sistolica} = \beta_0 + \beta_1 WAIST + \beta_2 AGE + \varepsilon$$

$$\text{Pressione sistolica} = 72.91 + 0.42\ WAIST + 0.13\ AGE$$

SBP	coefficient	SE	95%CI	T	P-value
WAIST	0.422	0.127	[0.164;0.680]	3.31	0.002
AGE	0.134	0.113	[-0.096;0.363]	1.18	0.246

$r^2=0.4244$

Regressione lineare semplice:

SBP	coefficient	SE	95%CI	T	P-value	r^2
WAIST	0.51	0.101	[0.312-0.708]	5.06	<0.001	0.4028

Regressione lineare multipla

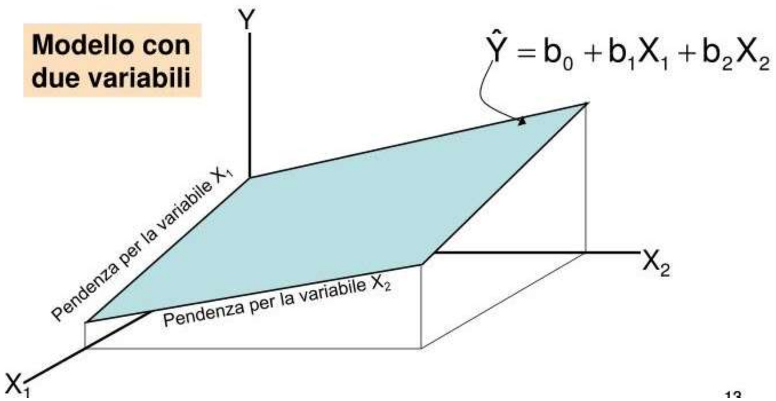
Come si interpretano i coefficienti?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

β_0 = valore atteso di Y quando X_1=0 e X_2=0

β_1 = variazione attesa di Y quando X_1 aumenta di una unità **e X_2 rimane costante**

β_2 = variazione attesa di Y quando X_2 aumenta di una unità **e X_1 rimane costante**



Regressione lineare multipla -

MODELLO 1 $\hat{y}$= 69.6 + 0.51 WAIST

MODELLO 2 $\hat{y}$= 72.91 + 0.42 WAIST + 0.13 AGE

Basandosi sul modello 1, come varia la pressione sistolica attesa tra un uomo (A) con circonferenza vita di 101 cm rispetto a un uomo (B) con circonferenza vita di 100 cm?

$\widehat{y_A}$= 69.6 + 0.51 101 =121.11

$\widehat{y_B}$= 69.6 + 0.51 100 =120.6

$\widehat{y_A}$-$\widehat{y_B}$ = 121.11-120.6=0.51=b_2

Regressione lineare multipla -

MODELLO 1 $\hat{y} = 69.6 + 0.51$ WAIST

MODELLO 2 $\hat{y} = 72.91 + 0.42$ WAIST $+ 0.13$ AGE

Basandosi sul modello 2, come varia la pressione sistolica attesa tra un uomo (A) di 50 anni con circonferenza vita di 101 cm rispetto a un uomo (B) di 50 anni con circonferenza vita di 100 cm?

$\widehat{y_A} = 72.91 + 0.42\ 101 + 0.13\ 50 = 121.83$

$\widehat{y_B} = 72.91 + 0.42\ 100 + 0.13\ 50 = 121.41$

$$\widehat{y_A} - \widehat{y_B} = 121.83 - 121.41 = 0.42 = b_2$$

… se A e B avessero 70 anni?
… se la circonferenza vita di A fosse 96 cm e quella di B 95 cm?

b_2: variazione attesa nella pressione sistolica per l'incremento di 1 cm della circonferenza vita quando **l'età rimane costante**

Confondimento

$\bar{y}_M = 8\, feet$

$\hat{y} = 7.5 + 0.5\, man$

$\bar{y}_F = 7.5\, feet$

Can we conclude that men have higher jumping abilities?

$$\begin{tikzcd} \text{Sesso} \arrow[rr] & & \text{Altezza salto} \\ & \text{Altezza} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Come possiamo valutare se gli uomini hanno maggiori abilità nel salto?

1. Bilanciamo il campione di uomini e donne sull'altezza (disegno dello studio)
2. Usiamo un modello di regressione per valutare come varia l'altezza del salto tra uomini e donne a parità di altezza

Confondimento

Can we conclude that men have higher jumping abilities?

Esempio Regressione lineare multipla

Supponiamo che X sia una variabile dicotomica (es. BMI>25). Come può essere inclusa questa variabile in un modello di regressione?

$$\text{Pressione sistolica} = \beta_0 + \beta_1 X + \varepsilon$$

X: SOVRAPPESO

X=1 se l'indice di massa corporea è maggiore di 25 (BMI>25)

X=0 altrimenti (BMI$\leq$25)

$$\hat{y} = 112.9 + 10.7 \text{ (se SOVRAPPESO)}$$

Che pressione sistolica ci attendiamo per un uomo non sovrappeso? E per un uomo sovrappeso?

$\hat{y} = 112.9$ mmHg **se BMI$\leq$25**

$\hat{y} = 112.9 + 10.7 = 123.6$ mmHg **se BMI$>$25**

Come interpetiamo b_2=10.7?

Esempio Regressione lineare multipla

Modelli sulla pressione sistolica

MODEL1	coefficient	SE	95%CI	T	P-value	r^2
WAIST	0.51	0.101	[0.312-0.708]	5.06	<0.001	0.4028

MODEL2	coefficient	SE	95%CI	T	P-value
WAIST	0.422	0.127	[0.164;0.680]	3.31	0.002
AGE	0.134	0.113	[-0.096;0.363]	1.18	0.246

r^2=0.4244

MODEL3	coefficient	SE	95%CI	T	P-value
WAIST	0.55	0.18	[0.18;0.93]	2.99	0.005
AGE	0.12	0.12	[-0.116;0.349]	1.02	0.317
BMI>25	-5.2	5.35	[-16.1;5.65]	-0.97	0.337

r^2=0.4392

Dummy variables

Supponiamo che X sia una variabile categoriale con K livelli. Come può essere inclusa questa variabile in un modello di regressione? Dobbiamo creare variabili binarie K-1 assumendo il valore 0 o 1 per descrivere i contrasti tra i livelli.

Ad esempio, se la variabile X2 ha K=3 livelli (basso, medio, alto) possiamo creare il K-1=2 seguenti variabili dummy Dm e Dh:

X	$\mathbf{D_m}$	$\mathbf{D_h}$
Basso	0	0
Medio	1	0
Alto	0	1

In questo caso si sceglie come livello di riferimento «basso» e:
- D_m definisce il contrasto di «medio» vs «basso»
- D_h definisce il contrasto di «alto» vs «basso»

Dummy variables

Il modello sarebbe il seguente (X_1 è una variabile continua):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 D_m + \beta_3 D_h + \varepsilon$$

β_0 = valore atteso di Y quando $X_1=0$ e $D_m=0$, $D_h=0$ (i.e. X_2=basso)

β_1 = variazione attesa di Y quando X_1 aumenta di una unità
e X_2 rimane costante

β_2 = variazione attesa di Y quando D_m aumenta di una unità (i.e. X_2 passa da «basso» a «medio») and X_1 rimane costante

β_3 = variazione attesa di Y quando D_h aumenta di una unità (i.e. X_2 passa da «basso» a «alto») and X_1 rimane costante

Esempio Dummy variables

Studio osservazionale «body data» su campione di 40 uomini

$$\text{Pressione sistolica} = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \varepsilon$$

$D_1 = 1$ se $25 < BMI < 30$
$D_1 = 0$ altrimenti
$D_2 = 1$ se $BMI > 30$
$D_2 = 0$ altrimenti

MODEL	coefficient	SE	95%CI	T	P-value
25<BMI<30	**4.16**	**4.65**	[-5.3;13.58]	**0.89**	**0.377**
BMI>30	**15.92**	**4.35**	[7.10;24.75]	**3.66**	**0.001**

R-squared = 0.2759

La pressione sistolica per un uomo con BMI>30 è attesa esser di circa 16 mmHg superiore agli uomini con BMI≤25.

Usefulness of waist circumference for the identification of childhood hypertension

Simonetta Genovesi[a,b], Laura Antolini[c], Marco Giussani[a,d], Federico Pieruzzi[a,b], Sara Galbiati[a], Maria Grazia Valsecchi[c], Paolo Brambilla[d,e] and Andrea Stella[a,b]

Objective To investigate the ability of BMI and waist circumference, single and combined, in identifying children who are at risk of hypertension and in influencing absolute blood pressure values.

Methods The body weight, height, waist circumference and blood pressure of 4177 5–11-year-old school children [2005 (48%) girls] were collected. Elevated blood pressure was defined if either systolic or diastolic blood pressure values or both were more than the 95th percentile according to sex, age and height (US normative blood pressure tables). Overweight and obese children were defined according to International Obesity Task Force BMI cut-offs.

Results The prevalence of hypertension was 4.1% and increased together with weight class: 1.4% ($n = 42/3076$) in normal weight, 7.1% ($n = 59/827$) in overweight and 25% ($n = 69/274$) in obese ($P < 0.001$). Only BMI and waist circumference showed a remarkable ability to discriminate hypertensive children (areas under receiver operating characteristic curves, 0.84 and 0.76, respectively). The multivariate analysis showed that *z*-scores for both BMI and waist circumference were significantly related to the risk of hypertension with odds ratios of 3.59 (95% confidence interval, 2.55, 5.06) and 1.20 (95% confidence interval, 1.04, 1.39), respectively, after adjusting for sex and age. When the weight class was included in the multivariate analysis, waist circumference retained its ability to identify hypertensive children only in the obese class (odds ratio, 1.44; 95% confidence interval, 1.21, 1.72; $P < 0.01$). When considering blood pressure as a continuous variable, both weight class and waist circumference showed a significant effect on systolic and diastolic blood pressure absolute values ($P < 0.01$). The effect of waist circumference on blood pressure values was maintained even when corrected for BMI.

Conclusion High blood pressure is strongly associated with excess weight. Waist circumference improves the ability of BMI to identify hypertension in obese children. Waist circumference is related to absolute blood pressure values in all weight classes. *J Hypertens* 26:1563–1570

Journal of Hypertension 2008, 26:1563–1570

Keywords: blood pressure, central adiposity, children, obesity, overweight, waist circumference

Abbreviations: AUC, area under the curve; BMI, body mass index; CDC, central of diseases control; DBP, diastolic blood pressure; IOTF, international obesity task force; ROC, receiver operating characteristic; SBP, systolic blood pressure; WtHr, waist circumference to height ratio

[a]Department of Clinical and Preventive Medicine, [b]Division of Nephrology, San Gerardo Hospital, [c]Center of Biostatistics for Clinical Epidemiology, Department of Clinical and Preventive Medicine, University of Milano-Bicocca, Monza, [d]Italian Federation of Pediatricians (FIMP) and [e]International Center for the Assessment of Nutritional Status (ICANS), University of Milan, Milan, Italy

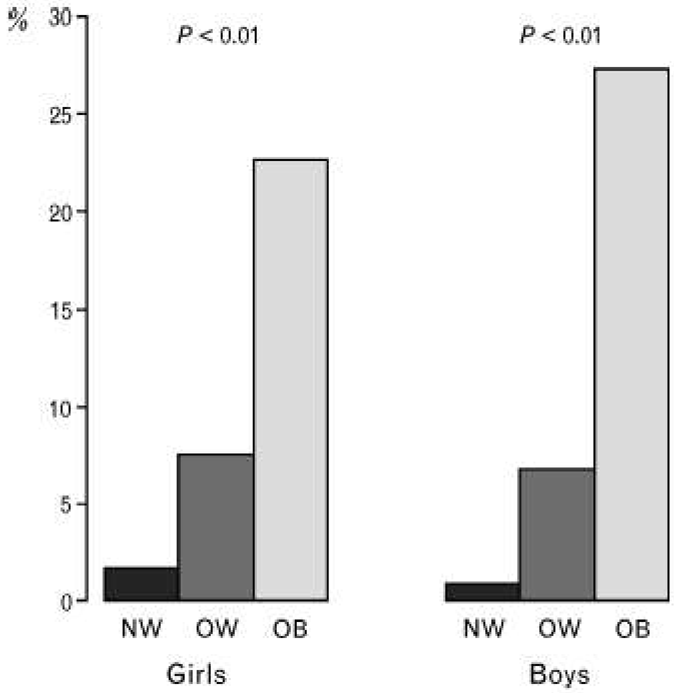
Correspondence to Dr Simonetta Genovesi, Dipartimento di Medicina Clinica e Prevenzione (DIMEP), Università degli Studi di Milano-Bicocca, Via Cadore 48, 20052 Monza (MI), Italy
Tel: +39 0392332375; fax: +39 0392332376;
e-mail: simonetta.genovesi@unimib.it

Received 13 July 2007 **Revised** 29 February 2008
Accepted 27 March 2008

Table 1 Descriptive characteristics according to sex

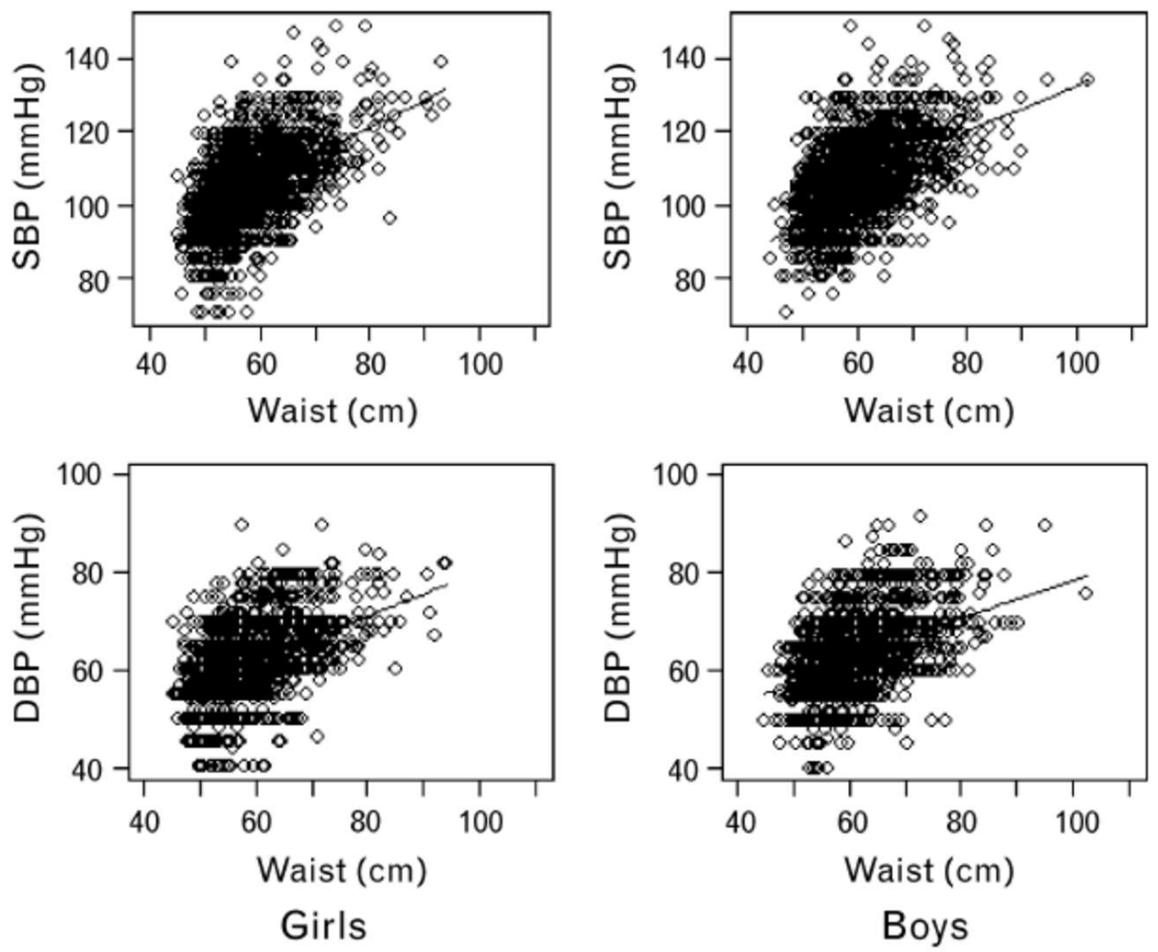
Variable	Girls ($n = 2005$)	Boys ($n = 2172$)	P
Age (years)	8.1 (1.5)	8.1 (1.5)	0.51
Weight (kg)	30.0 (8.4)	30.7 (8.2)	<0.01
Height (cm)	130.7 (10.4)	131.8 (9.8)	<0.01
BMI (kg/m^2)	17.3 (2.8)	17.4 (2.8)	0.06
Waist (cm)	57.9 (7.0)	59.7 (7.0)	<0.01
Systolic BP (mmHg)	104.1 (11.0)	105.7 (10.8)	<0.01
Diastolic BP (mmHg)	61.4 (7.1)	62.9 (7.3)	<0.01

Values are expressed as mean (standard deviation). BMI, body mass index; BP, blood pressure.



Prevalence of hypertension (%) by sex according to weight class. NW, normal weight; OB, obese; OW, overweight. *P* values refer to comparison between weight classes.

Fig. 5



Scatter plot of systolic and diastolic blood pressure levels against waist circumference measurements, with smoother. Girls $n = 2005$, and boys $n = 2172$. DBP, diastolic blood pressure; SBP, systolic blood pressure.

Table 4 **Coefficients *b* for the effect of BMI and waist circumference *z*-scores and BMI and waist-to-height ratio *z*-scores on systolic and diastolic blood pressure, according to a linear regression model adjusted for sex, age and height**

	Systolic BP		Diastolic BP	
Variable	*b* (95% CI)	*P*	*b* (95% CI)	*P*
(a) BMI and waist circumference *z*-scores				
BMI *z*-score	2.07 (1.70, 2.44)	<0.001	1.31 (1.05, 1.57)	<0.001
Waist circumference *z*-score	1.61 (1.31, 1.91)	<0.001	0.88 (0.67, 1.09)	<0.001
Height	0.20 (0.15, 0.25)	<0.001	0.13 (0.10, 0.17)	<0.001
Sex (girl vs. boy)	−0.99 (−1.63, −0.35)	0.002	0.01 (−0.44, 0.46)	0.067
Age (years)	1.34 (1.02, 1.66)	<0.001	0.54 (0.31, 0.77)	<0.001
(b) BMI and waist-to-height ratio *z*-scores				
BMI *z*-score	1.72 (1.33, 2.12)	<0.001	1.22 (0.94, 1.50)	<0.001
WtHr *z*-score	1.85 (1.53, 2.16)	<0.001	0.90 (0.68, 1.12)	<0.001
Height	0.36 (0.31, 0.41)	<0.001	0.21 (0.18, 0.25)	<0.001
Sex (girl vs. boy)	1.59 (1.04, 2.13)	<0.001	1.38 (0.99, 1.77)	<0.001
Age (years)	1.01 (0.68, 1.33)	<0.001	0.37 (0.14, 0.60)	0.001

b, coefficient *b*; BP, blood pressure; CI, confidence intervals; WtHr, waist-to-height ratio.

Alcuni problemi generali da considerare quando si costruisce un modello di regressione

Selezione delle variabili

La selezione dei predittori da includere in un modello di regressione è uno dei problemi più impegnativi per chi analizza di dati.

Possono essere scelte strategie diverse a seconda che il modello sia costruito per valutare l'associazione tra uno o più fattori e il risultato che si aggiusta per altre variabili o se il modello è sviluppato a fini di previsione.
Sono disponibili diverse procedure automatiche per la selezione delle variabili completamente guidate dai dati osservati (es. Non ci sono regole pratiche da consigliare valide per ogni situazione).
Un suggerimento generale quando si costruisce il modello è di tenere in considerazione:

- Conoscenze scientifiche precedenti
- Analisi descrittiva dei dati osservati
- La dimensione del campione disponibile
- L'interpretabilità dei risultati

Multicollinearità

Il problema della multicollinearità si riferisce alla possibile presenza in un modello di regressione di due o più variabili esplicative che sono altamente linearmente correlate. In pratica, ciò significa che nel modello sono inclusi più fattori che portano informazioni molto simili sul risultato.

Ciò provoca un'inflazione delle stime dell'errore standard che porta a una non valida inferenza sui parametri di regressione. Quando il grado di multicollinearità è estremamente elevato, può accadere che gli algoritmi numerici per la stima di massima verosimiglianza non convergano.

La multicollinearità per una variabile X può essere misurata utilizzando il Variance Inflation Factor (VIF) definito come il rapporto tra la varianza del modello di regressione multipla e la varianza del modello dove X è l'unico predittore.
Come regola generale, VIF ≈ 1 denota l'assenza di multicollinearità, VIF>4 è sospetto e VIF>10 identifica un grosso problema di multicollinearità.

Esempio

Studio osservazionale «body data» su campione di 40 uomini

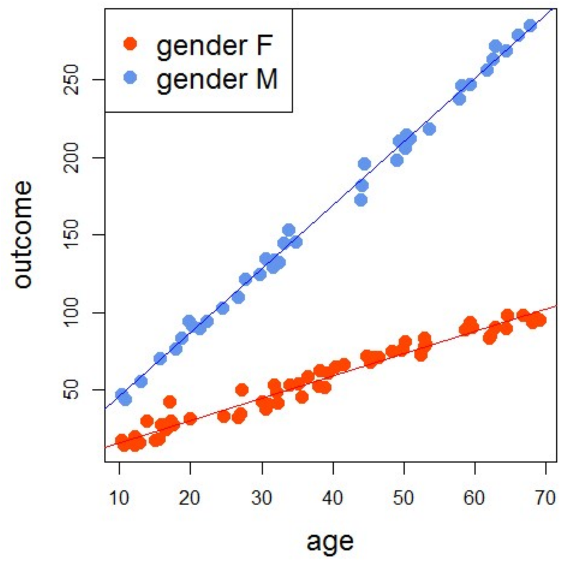
```
------------------------------------------------------------------------------
    SYSTOLIC |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       WAIST |   .8174599   .2796309     2.92   0.006     .2508738    1.384046
         BMI |  -1.028161   .8823345    -1.17   0.251    -2.81594     .7596187
       _cons |     68.045   10.06694     6.76   0.000     47.64743    88.44257
------------------------------------------------------------------------------
```

R-squared = 0.4239

MODEL1	coefficient	SE	95%CI	T	P-value	r^2
WAIST	0.51	0.101	[0.312-0.708]	5.06	<0.001	0.4028

Interazioni

Quando il valore di X_1 (es. genere) ha un impatto sull'effetto (es. variazione provocata all'esito) di X_2 (es. età) si ha interazione tra le due variabili.



Oltre a scegliere correttamente le variabili esplicative da includere in un modello di regressione, si dovrebbe anche decidere se consentire al modello di catturare alcune interazioni tra predittori.

In pratica, questo viene fatto testando i coefficienti dei termini di interazione. Tuttavia, è spesso impossibile testare tutte le possibili interazioni in un modello (sono possibili anche interazioni tra tre o più variabili X!). Pertanto, è importante testare solo le interazioni più plausibili basate sulla conoscenza precedente.

Esempio

BMI<=25

SYSTOLIC	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
WAIST	.1460829	.3564868	0.41	0.689	-.62406	.9162257
_cons	101.0548	29.11252	3.47	0.004	38.16108	163.9486

25<BMI<=30

SYSTOLIC	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
WAIST	.9088671	.4065356	2.24	0.052	-.0107803	1.828514
_cons	28.56726	39.7657	0.72	0.491	-61.38901	118.5235

BMI>30

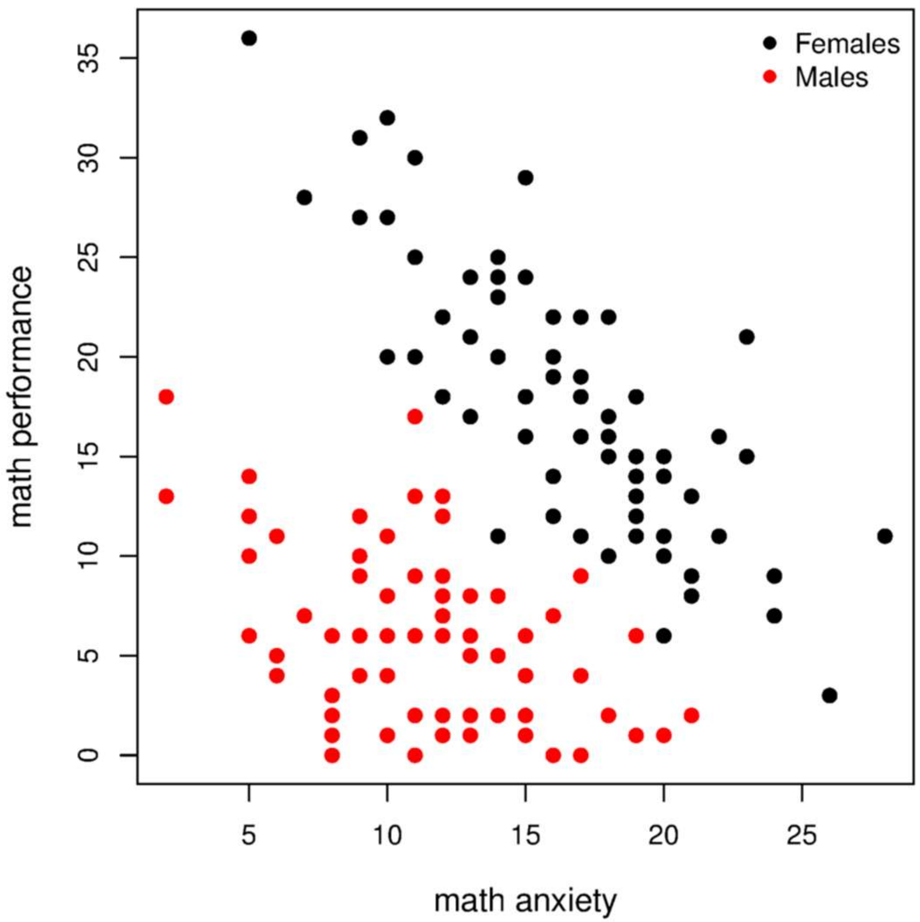
SYSTOLIC	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
WAIST	.8308569	.3124903	2.66	0.021	.149999	1.511715
_cons	33.48664	35.9515	0.93	0.370	-44.84495	111.8182

Esempio

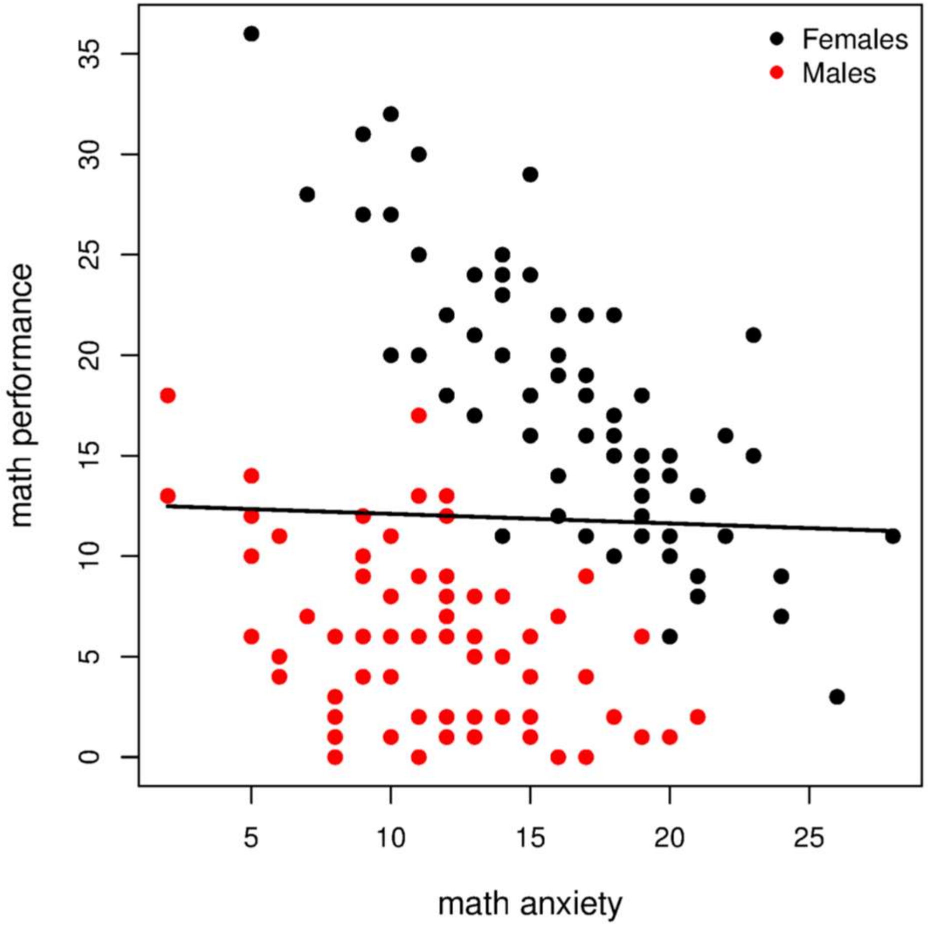
- In a study, data on math anxiety (math.anx), math performances (math.perf) and gender were collected in a sample of Sociology students.
- The aim of the study is to evaluate whether there is a relationship between math anxiety and math performance and in particular whether the effect of anxiety on performance is the same for males and females.
- Since the gender is a categorical variable, it was encoded as a dummy variable (genderM) assuming 0 for females and 1 for males.

$$math.perf = \beta_0 + \beta_1 math.anx + \beta_2 Males + \varepsilon$$

Math performances – raw data



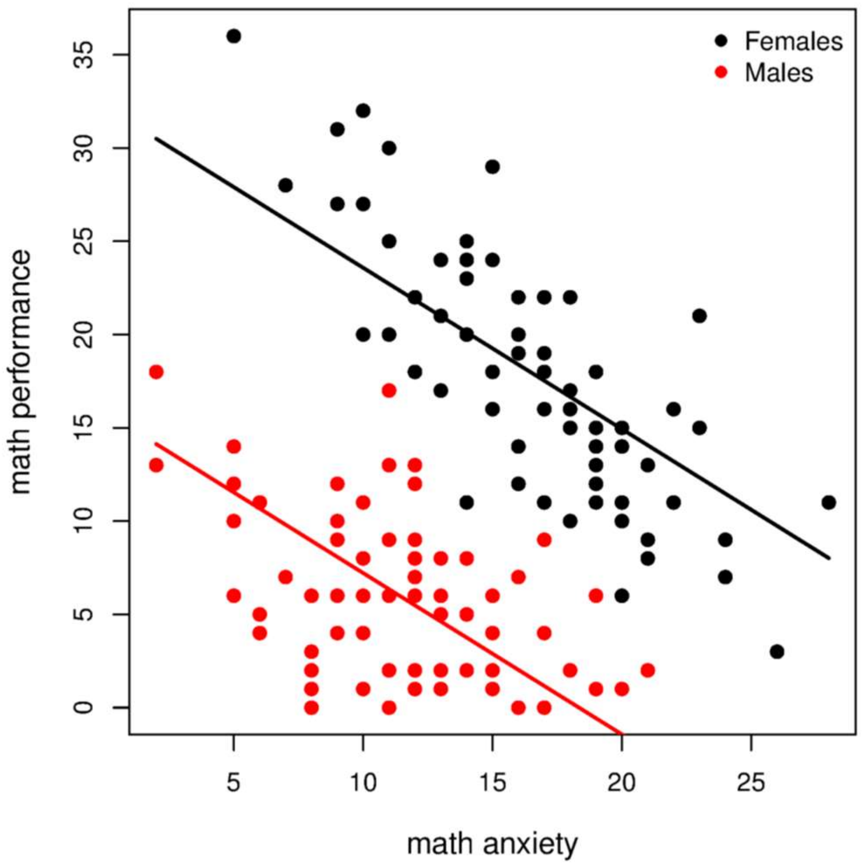
Math performances – model 1



r^2=0.008

MODEL 1	coefficient	SE	T	P-value
Intercept/constant	12.58	2.06	6.10	<0.001
Math anxiety	-0.047	0.138	-0.342	0.733

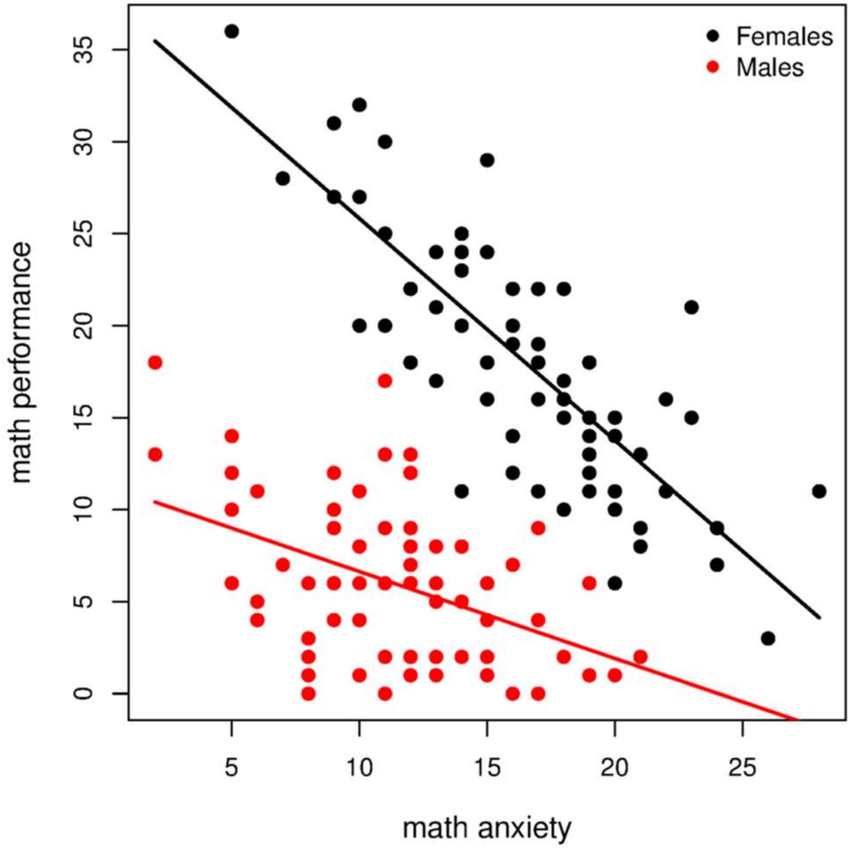
Math performances – model 2



$r^2 = 74\%$

MODEL 2	coefficient	SE	T	P-value
Intercept/constant	32.23	1.46	22.08	<0.001
Math anxiety	-0.86	0.08	-10.5	<0.001
Males vs Females	-16.37	0.83	-19.61	<0.001

Math performances – model 3



$r^2 = 78\%$

MODEL 3	coefficient	SE	T	P-value
Intercept/constant	37.88	1.80	21.04	<0.001
Math anxiety in Females	-1.2	0.11	-11.5	<0.001
Math anxiety in Males	-0.47	0.18	-2.6	0.01

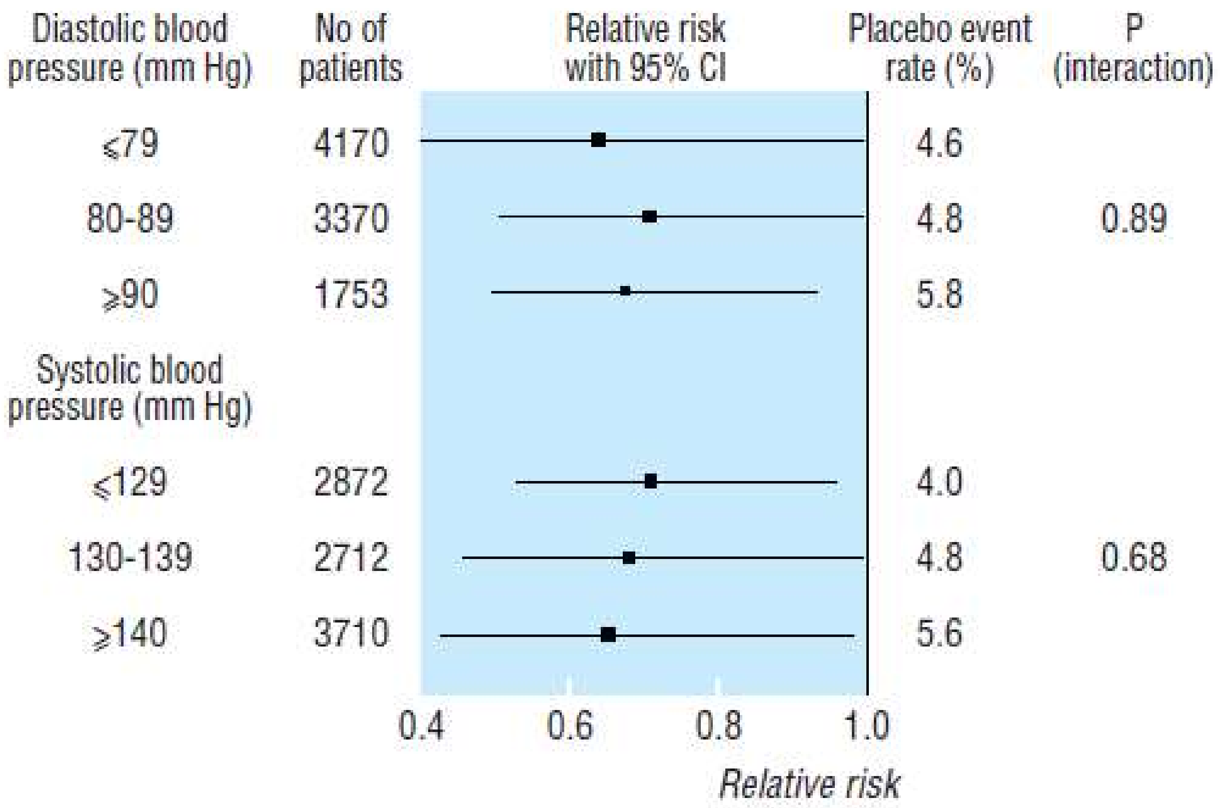


Fig 2 Impact of ramipril on stroke based on baseline blood pressure

Confondimento – parte 2

- Coffee drinkers had higher rate of cancer

Overall

	Lung Cancer Yes	Lung Cancer No	
Heavy Drinker	1057	1949	3006
Light Drinker	1896	6220	8116
	2953	8169	11122

OR = 1.78

It appears that heavy drinking may be associated with lung cancer
The OR above is called the «crude odds ratio»

Smoker

Lung Cancer Yes	Lung Cancer No	
786	665	1451
702	591	1293
1488	1256	2744

OR = 1.0

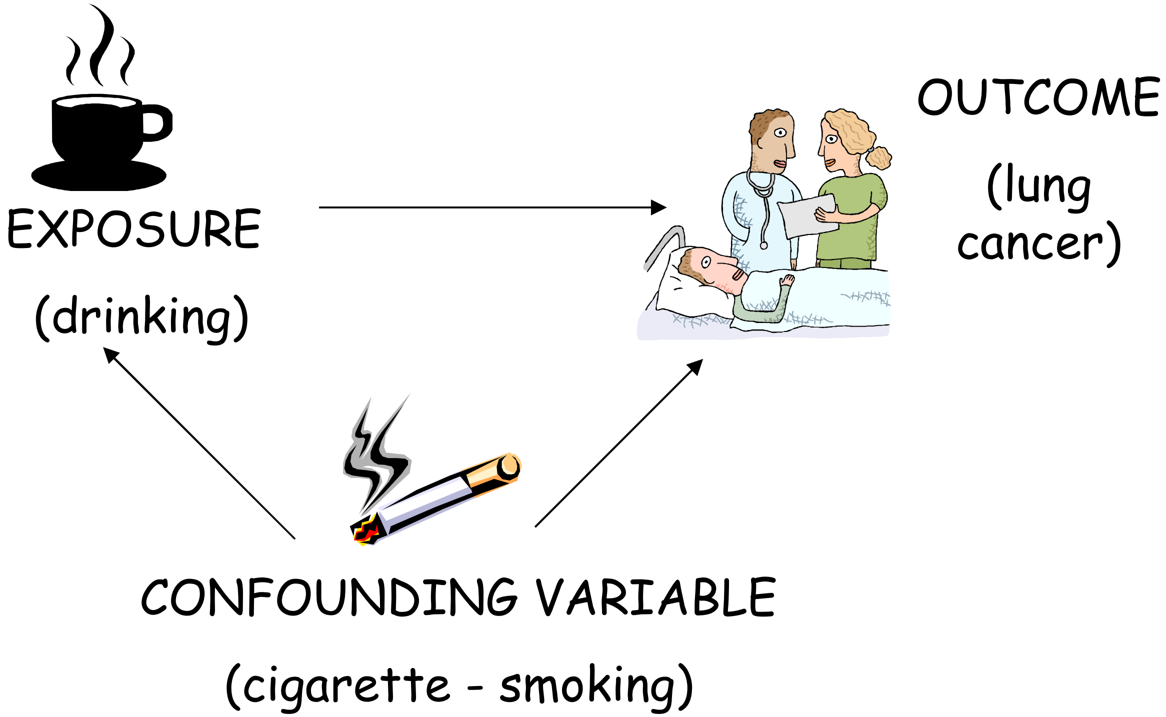
Non-smoker

Lung Cancer Yes	Lung Cancer No	
271	1248	1555
1194	5629	6823
1463	6913	8378

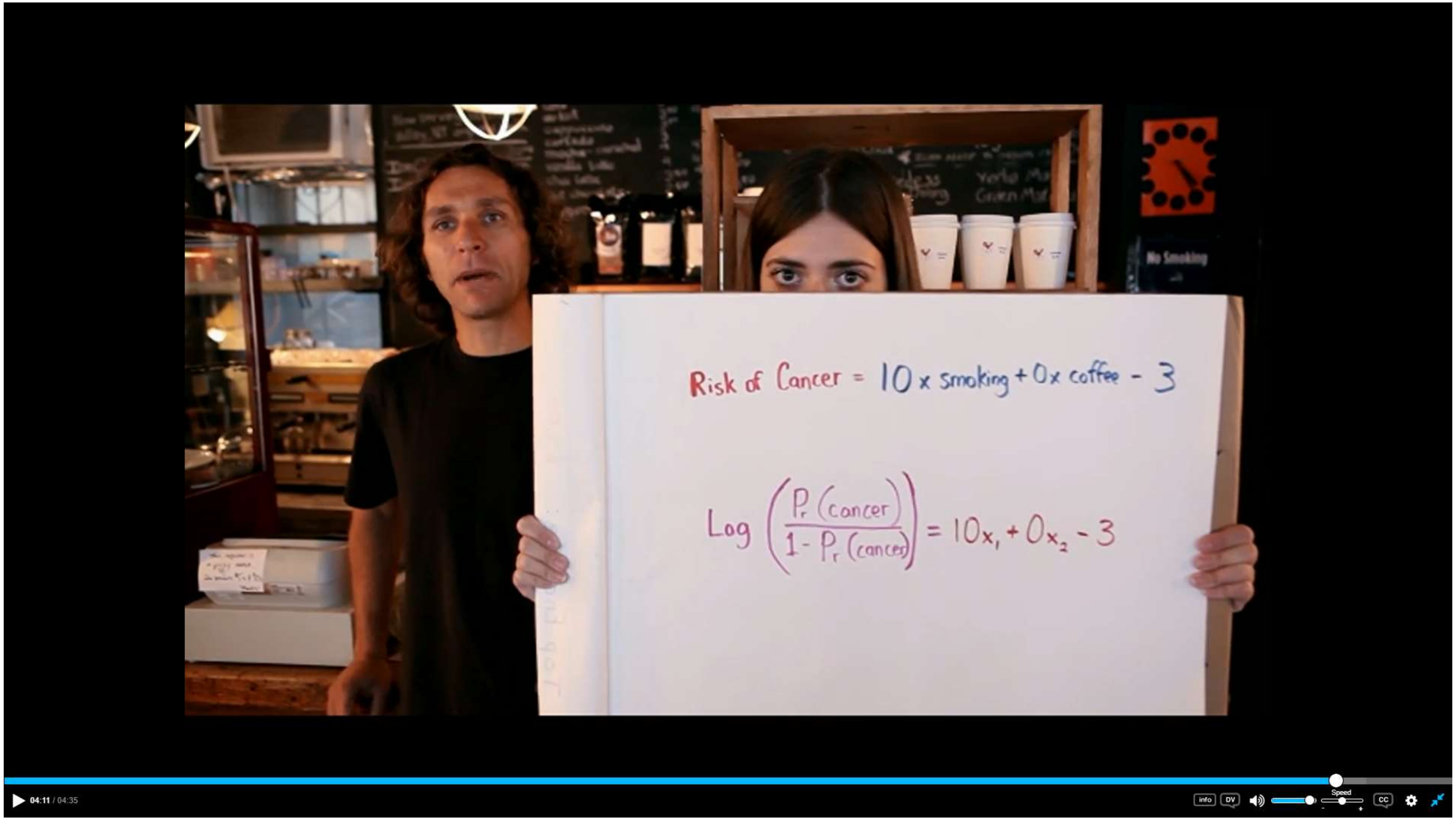
OR = 1.0

Illustration of confounding

Smoking is a potential "confounder":



Confondimento – parte 2



Forma funzionale

I modelli lineari generalizzati (GLM) sono una classe di modelli di regressione in cui il risultato o una funzione del risultato (ad esempio la funzione "logit" nella regressione logistica) è espresso come una funzione lineare delle variabili esplicative.

Il termine “lineare” si riferisce alla linearità nei parametri (b0, b1, b2, …). Tuttavia, è consentita una trasformazione dei predittori.
Esempi tipici:

- funzione polinomiale

$$g(Y)= \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_j X_1^j + \dots + \varepsilon$$

- Logaritmo

$$g(Y)= \beta_0 + \beta_1 \log(X_1) + \dots + \varepsilon$$

Una scelta corretta della forma funzionale delle covariate potrebbe migliorare l'adattamento del modello ma in generale riduce l'interpretabilità dei coefficienti.

Table 3 Odds ratios for hypertension with multivariable logistic regression adjusted for sex and age

Variable	OR (95% CI)	*P*
(a) Weight class and waist circumference *z*-score		
Weight class (OW vs. NW)	4.03 (2.32, 6.98)	<0.01
(OB vs. NW)	6.94 (3.22, 14.96)	<0.01
Waist circumference *z*-score (NW)	1.01 (0.74, 1.39)	0.47
(OW)	1.17 (0.98, 1.42)	0.05
(OB)	1.44 (1.21, 1.72)	<0.01
Sex (girl vs. boy)	0.53 (0.33, 0.85)	<0.01
Age (years)	1.32 (1.16, 1.49)	<0.01
(b) Weight class and waist-to-height ratio *z*-score		
Weight class (OW vs. NW)	4.32 (2.65, 7.05)	<0.01
(OB vs. NW)	4.20 (1.85, 9.52)	<0.01
WtHr *z*-score (NW)	1.59 (1.12, 2.25)	<0.01
(OW)	0.99 (0.74, 1.32)	0.53
(OB)	1.83 (1.43, 2.35)	<0.01
Sex (girl vs. boy)	0.77 (0.54, 1.08)	<0.01
Age (years)	1.42 (1.26, 1.61)	0.13

CI, confidence intervals; NW, normal weight; OB, obese; OR, odds ratios; OW, overweight; WtHr, waist-to-height ratio.

Very famous remark

"All models are wrong but some are useful"

[…] Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen ***parsimonious models often do provide remarkably useful approximations****. For example, the law PV = RT relating pressure P, volume V and temperature T of an "ideal" gas via a constant R is not exactly true for any real gas, but it frequently provides a useful approximation and furthermore its structure is informative since it springs from a physical view of the behavior of gas molecules.*
For such a model there is no need to ask the question ***"Is the model true?".*** *If "truth" is to be the "whole truth"* ***the answer must be "No".*** *The only question of interest is "Is the model illuminating and useful?". […]*

George E. P. Box

Esercizio

- Nello studio sulla valutazione dell'utilità della circonferenza vita nell'identificazione della ipertensione infantile si sono trovati i seguenti risultati.

Table 5 Coefficients *b* for the effect of weight class and waist circumference z-scores and weight class and waist-to-height ratio *z*-scores on systolic and diastolic blood pressure, according to a linear regression model adjusted for sex, age and height

	Systolic BP		Diastolic BP	
Variable	*b* (95% CI)	*P*	*b* (95% CI)	*P*
(a) Weight class and waist circumference *z*-scores				
Weight class (OW vs. NW)	3.79 (2.96, 4.61)	<0.01	2.57 (1.99, 3.16)	<0.01
(OB vs. NW)	7.43 (5.60, 8.86)	<0.01	4.96 (3.94, 5.97)	<0.01
Waist circumference *z*-score	1.61 (1.31, 1.90)	<0.01	0.83 (0.62, 1.04)	<0.01
Height	0.21 (0.16, 0.26)	<0.01	0.14 (0.11, 0.18)	<0.01
Sex (girl vs. boy)	−0.72 (−1.38, −0.07)	0.02	0.25 (−0.22, 0.71)	0.15
Age (years)	1.26 (0.94, 1.57)	<0.01	0.50 (0.27, 0.72)	<0.01
(b) Weight class and waist-to-height ratio *z*-scores				
Weight class (OW vs. NW)	3.29 (2.45, 4.13)	<0.01	2.47 (1.87, 3.07)	<0.01
(OB vs. NW)	6.31 (4.83, 7.80)	<0.01	4.70 (3.64, 5.75)	<0.01
WtHr *z*-score	1.84 (1.53, 2.14)	<0.01	0.85 (0.64, 1.07)	<0.01
Height	0.37 (0.32, 0.41)	<0.01	0.22 (0.18, 0.25)	<0.01
Sex (girl vs. boy)	1.81 (1.27, 2.35)	<0.01	1.53 (1.15, 1.91)	<0.01
Age (years)	0.95 (0.63, 1.27)	<0.01	0.35 (0.12, 0.57)	<0.01

b, coefficient *b*; BP, blood pressure; CI, confidence intervals; NW, normal weight; OB, obese; OW, overweight; WtHr, waist-to-height ratio.

- Quali variabili sono associate alla pressione sistolica (modello a)?
- Quale è la differenza media di sistolica tra femmine e maschi (a parità delle altre variabili)?
- La circonferenza vita è associata alla pressione sistolica?