Modelli Lineari Generalizzati (Cap. 11)

Marcello Gallucci

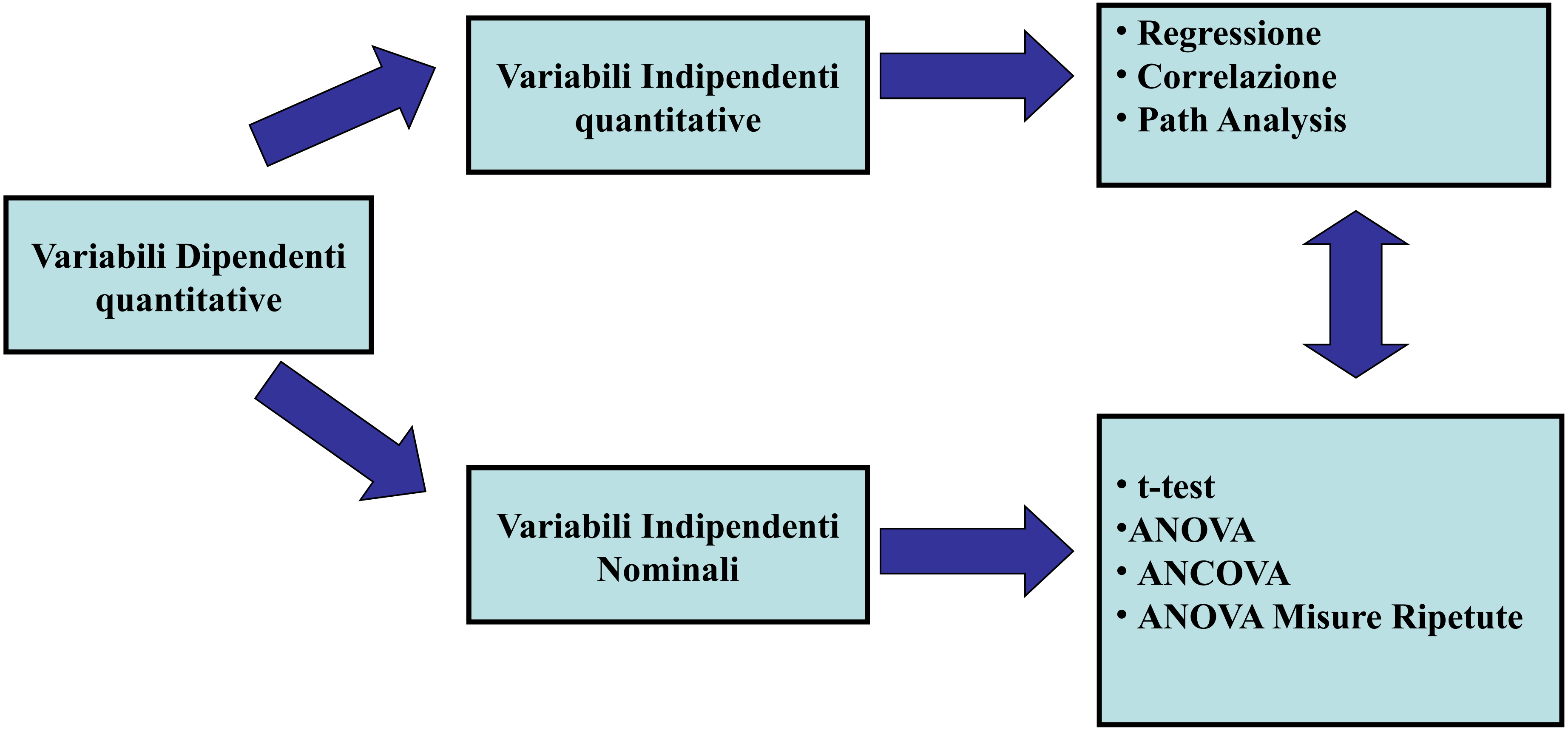
AMD

Lezione 12

Preludio

- La maggior parte delle comuni tecniche statistiche volte ad individuare le relazioni fra variabili, quali Correlazioni, Regressione, ANOVA, ANCOVA, sono riconducibili al Modello Lineare Generale (GLM)
- Il GLM ci permette di studia gli effetti di variabili indipendenti di vario tipo su variabili quantitative (variabili dipendenti continue)
- La ricerca empirica è disseminata di variabili dipendenti qualitative (scelte dicotomiche, scelte multiple, frequenze di eventi, classificazioni, etc)
- I Modelli Lineari Generalizzati ci consentono di studiare gli effetti di variabili indipendenti su variabili dipendenti qualitative

Modello Lineare Generale



Modello Lineare Generalizzato

Modello Lineare Generale

- **Variabili dipendenti Dicotomiche** → **Chi-quadro; Regressione logistica; ANOVA logistica**
- **Variabili dipendenti Multinomiali** → **Regressione logistica multinomiale; ANOVA logistica multinomiale; Analisi di tabelle multidimensionali**
- **Variabili dipendenti basate su frequenze** → **Regressione Poisson; ANOVA Poisson**
- **Variabili dipendenti ripetute** → **Equazioni di stima generalizzate**

Tecniche volte a studiare e quantificare gli effetti di una o più variabili indipendenti *continue o nominali* su una variabile dipendente *nominale (qualitativa)*

Tecniche statistiche:

- La regressione logistica: Variabile dipendente dicotomica
- La regressione ordinale: Variabile dipendente ordinabile
- La regressione multinomiale: Variabile dipendente politomica
- La regressione di Poisson: Variabile dipendente basata su frequenze
- Tutto ciò, anche a misure ripetute

Violazioni assunzioni

- Quando le assunzioni non sono soddisfatte, i risultati sono da considerarsi dubbi

- Se la violazione delle assunzioni è grave, la regressione/ANOVA non può essere applicata

- Nella pratica, con variabili dipendenti continue normalmente distribuite, queste assunzioni sono abbastanza semplici da soddisfare

- Ma cosa succede se volessimo usare una variabile dipendente dicotomica?

La regressione logistica

- La regressione logistica si propone di studiare e quantificare le relazioni tra una o più variabili indipendenti quantitative (es. età, salario, atteggiamenti, personalità) e una variabile dipendente dicotomica (es. stato civile, voto al referendum, appartenenza ad un gruppo, etc.)

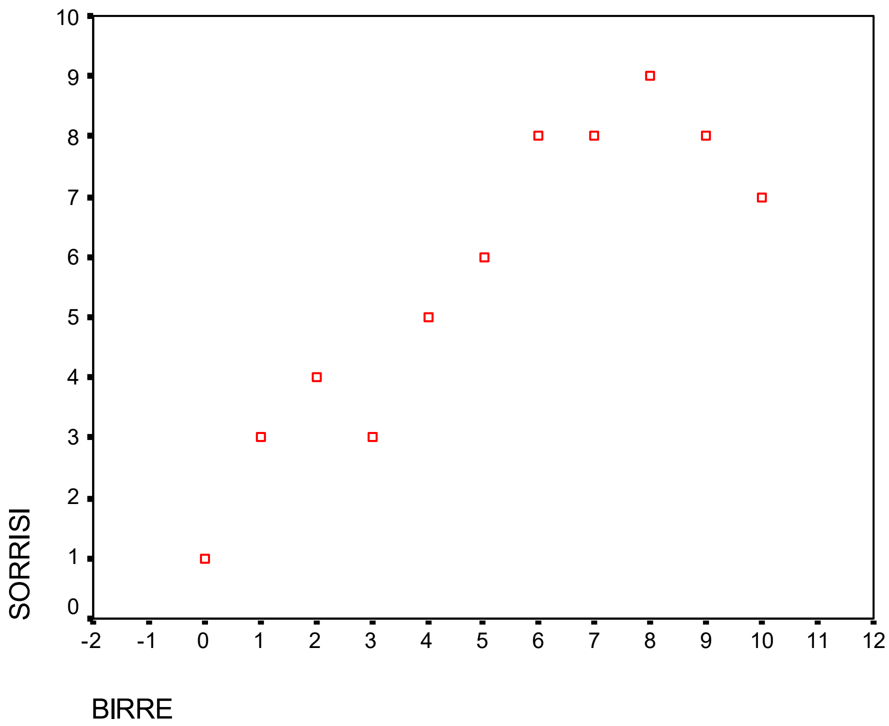
Assunzioni e Dicotomiche

- Le assunzione della regressione lineare non possono essere soddisfatte nel caso di VD dicotomiche

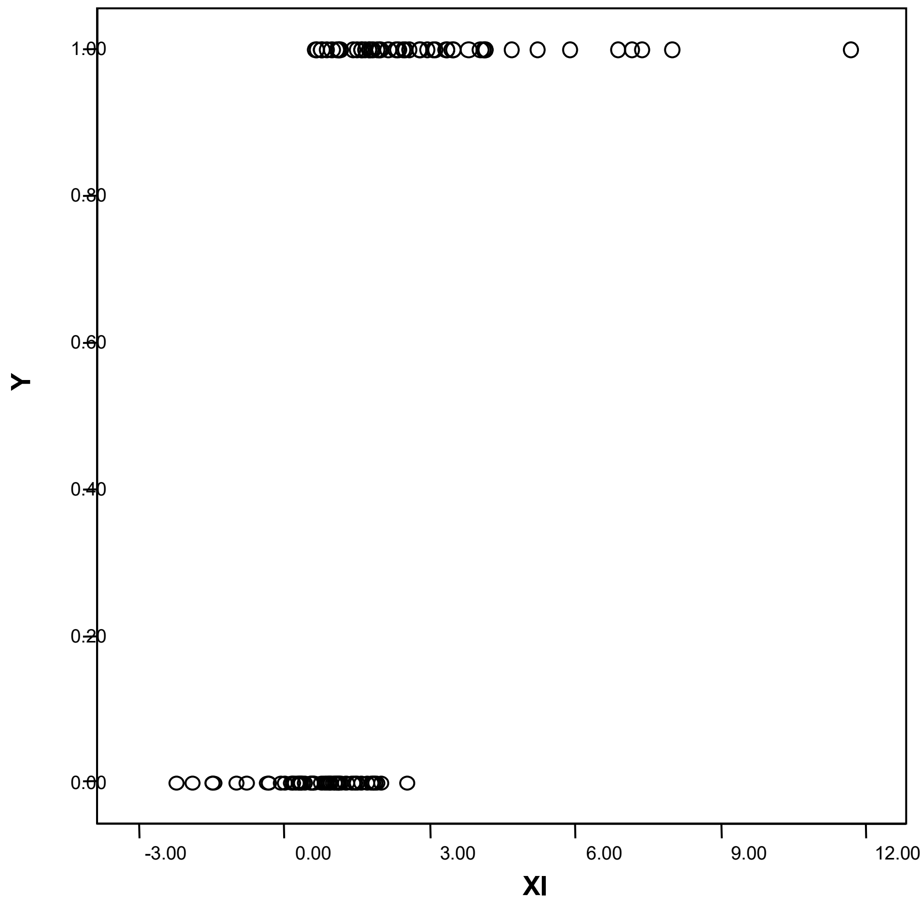
Assunzione	Se VD dicotomica
Linearità	Non può essere lineare
Omoschedasticità	Sicuramente la varianza dipende dal valore predetto
Normalità errori	Gli errori saranno sempre distribuiti con due gobbe

Assunzioni regressione lineare

- Come si può intuire dal fatto che la regressione interpola una retta tra i punti, la relazione che stima è una relazione lineare, omoschedastica e normale

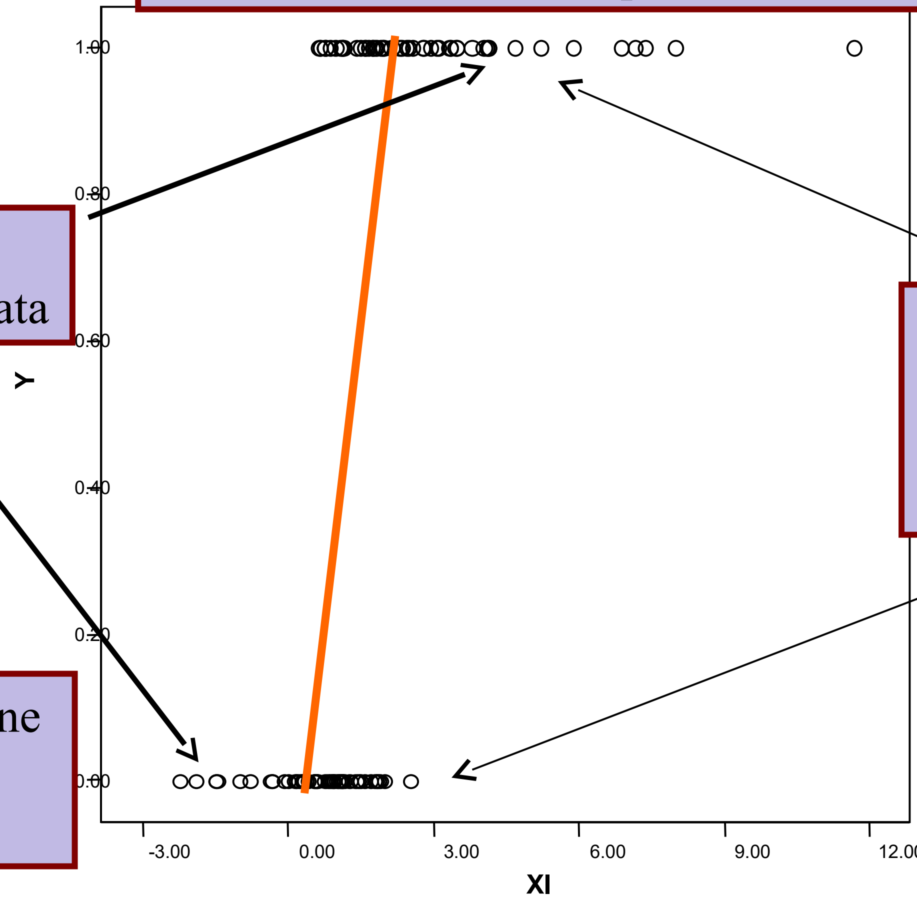


VD dicotomica



VD dicotomica

La relazione non protrà mai essere lineare



Gran parte della varianza non è spiegata

I punteggi saranno necessariamente raggruppati in due rette piatte

La retta di regressione avrà moltissimo margine di errore

Assunzione 1: Linearità

- Come visto precedentemente la relazione che riusciamo a catturare con la regressione è una relazione lineare

- **Se la regressione è condotta con una variabile dipendente dicotomica, l'assunzione di linearità non può essere soddisfatta, creando problemi sia nella bontà della predizione, che nella sua interpretazione**

- **Dunque l'assunzione di linearità è <u>sicuramente</u> violata**

VD Categorica

- Quando abbiamo una variabile dipendente dicotomca, ogni soggetto ha o 1 o 0 come valore della variabile dipendente

VD=sesso (Maschi=0, Femmine=1), VD=acquisto (Si=1, No=0), voto al referendum (Si=1, No=0)

- La media della variabile dipendente è la probabilità di ottenere il valore 1

$$\bar{Y} = \frac{n_1}{n_{tot}}$$

- Ciò che prediciamo è la probabilità *p* di appartenere al gruppo con valore 1 (e 1-p sarà la probabilità di appartenere al gruppo 0).

Soluzione

- Necessitiamo dunque di un tipo di regressione che:
- Risolva il problema della omoschedasticità, linearità e normalità degli errori
- Ammetta valori non assurdi
- Ci esprima le relazioni sulla base di probabilità o qualcosa di comparabile
- **Cioè dobbiamo usare un modello che trasformi la variabile dipendente tale da linearizzare la relazione, rendere la variabile dipendente continua, e farla variare su tutto l'asse (valori positivi e negativi)**

Soluzione: parte 1

- Intanto decidiamo di **non cercare di predire la probabilità**, ma il rapporto tra probabilità

- Tale rapporto è detto **odd (rapporto di probabilita')**

$$P_i = a + b_{yx} x_i \quad \Rightarrow \quad \frac{P_i}{1 - P_i} = a + b_{yx} x_i$$

Odd

- L'odd è il rapporto tra la probabilità di un evento (appartenere ad un gruppo) rispetto alla probabilità del non evento (appartenere all'altro gruppo)

$$Odd = \frac{P_i}{1 - P_i}$$

- **Esempi: se la probabilità di avere una figlia femmina è .50**

$$Odd = \frac{.5}{1 - .5} = 1$$

- **se la probabilità di votare Si ad un referendum è .70**

$$Odd = \frac{.7}{1 - .7} = 2.33$$

Odd

- L'odd indica quanto più probabile è un evento rispetto al suo complemento

$$\frac{P_i}{1-P_i}$$

- **Esempi: Una figlia femmina è tanto probabile quanto un maschio**

$$Odd = \frac{.5}{1-.5} = 1$$

- **Il voto Si è 2.33 volte più probabile del No**

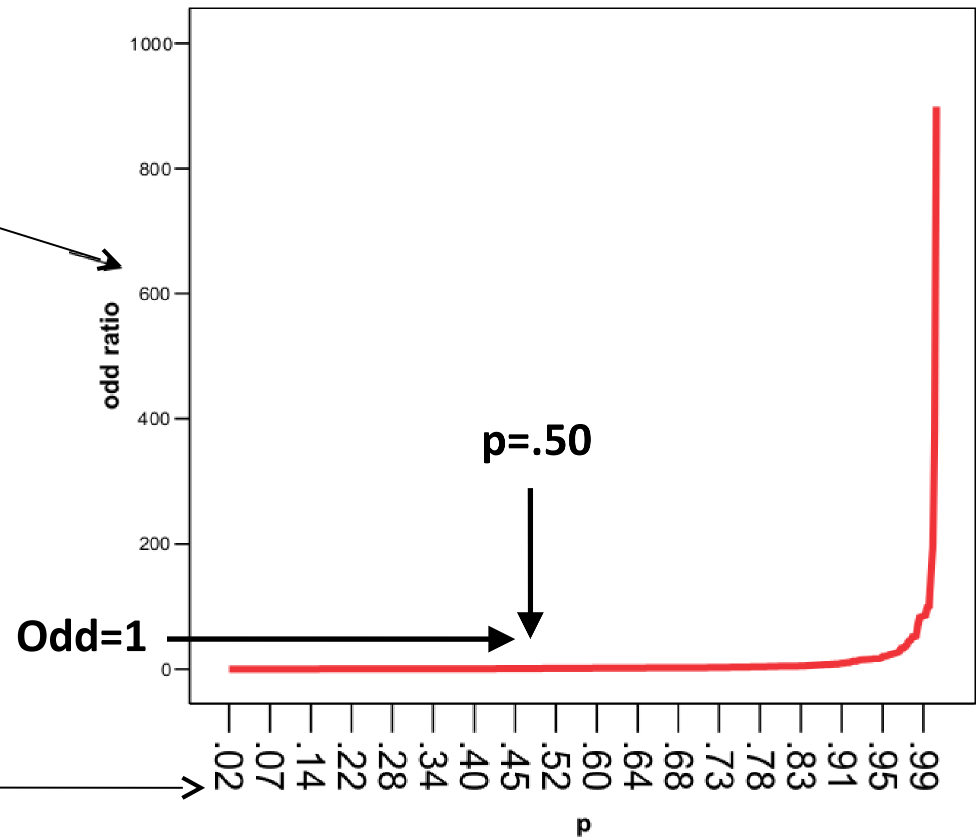
$$Odd = \frac{.7}{1-.7} = 2.33$$

Odd: Interpretazione

- L'odd consente di esprimere la probabilità mediante valori che variano da 0 ad infinito

Odd da 0 ad infinito

$$Odd = \frac{P_i}{1 - P_i}$$



Probabilità da 0 a 1

Odd: Interpretazione

- L'odd varia da 0 ad infinito

Se gli eventi sono equiprobabili

$$p = .5 \rightarrow or = \frac{.5}{1-.5} = 1$$

É maggiore di 1 se l'evento è più probabile del contrario

$$p = .7 \rightarrow or = \frac{.7}{1-.7} = 2.33 > 1$$

É minore di 1 se l'evento è meno probabile del contrario

$$p = .2 \rightarrow or = \frac{.2}{1-.2} = .25 < 1$$

Problema con odd

- Se usassimo gli odd come variabile dipendente, potremmo ottenere predizioni impossibili, come predizioni di valori negativi

$$\frac{P_i}{1-P_i} = a + b_{yx} x_i$$

Se a=1, b=3 e x=-2

$$\frac{P_i}{1-P_i} = 1 + 3 * (-2) = -5$$

Soluzione: parte 2

- Decidiamo di non cercare di predire l'odd, ma il logaritmo dell'odd

$$\frac{P_i}{1-P_i} = a + b_{yx}x_i \Rightarrow \boxed{\ln\left(\frac{P_i}{1-P_i}\right) = a + b_{yx}x_i}$$

La trasformazione con il logaritmo si chiama **logit**

$$\log it = \ln\left(\frac{P_i}{1-P_i}\right)$$

La regressione che cerca di predire il **logit** si chiama regressione **logistica**

Logaritmo

- Il logaritmo in base A di B è quel numero a cui dobbiamo elevare A per ottenere B.

$$Log_{10} 100 = 2 \quad \Rightarrow \quad 10^{2} = 100$$

- Spesso si usa il logaritmo naturale, cioè il logaritmo con base **e** o numero neperiano (da Napier – Giovanni Nepero - che lo scoprì) o di Eulerio (che lo formulò nei termini che lo conosciamo)

$$e = 2.71828182845904523536028747135266249775 7\ldots$$

$$Ln(100) = 4.605 \qquad e^{4.605} = 100$$

Vantaggi del Logaritmo

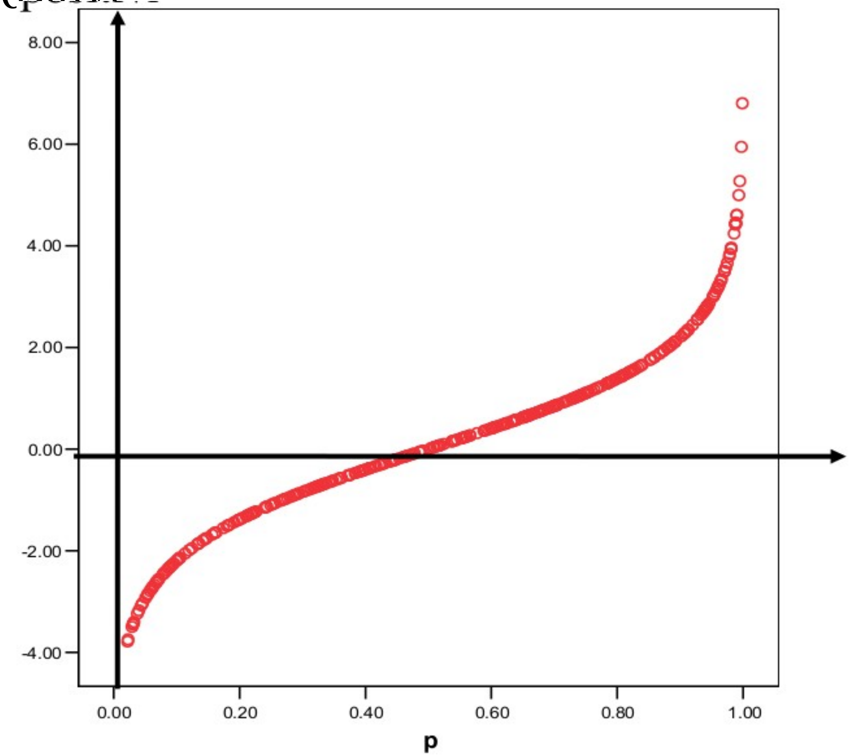
- La logistica usa il logaritmo in quanto:
  - Trasforma una variabile positiva (odd ratio) in negativi e positivi
  - É positivo se l'argomento è maggiore di 1
    (es., Ln (**5**) = ***1.6*** $2.718^{1.6} = 5$)
  - É negativo se l'argomento è minore di 1
    (es., Ln (**0.2**) = ***-1.6*** $2.718^{-1.6} = 0.2$)
  - É zero se l'argomento è uguale ad 1
    (Ln (**1**) = ***0*** $2.718^{0} = 1$)

Perché il logaritmo?

- Il logaritmo di una variabile che varia da 0 ad infinito (come gli odd), varia per tutti i valori possibili, da negativi a positivi

Il logaritmo dell'Odd permette di esprimere la probabilità mediante valori sia positivi che negativi

$$\log it = \ln\left(\frac{p}{1-p}\right)$$

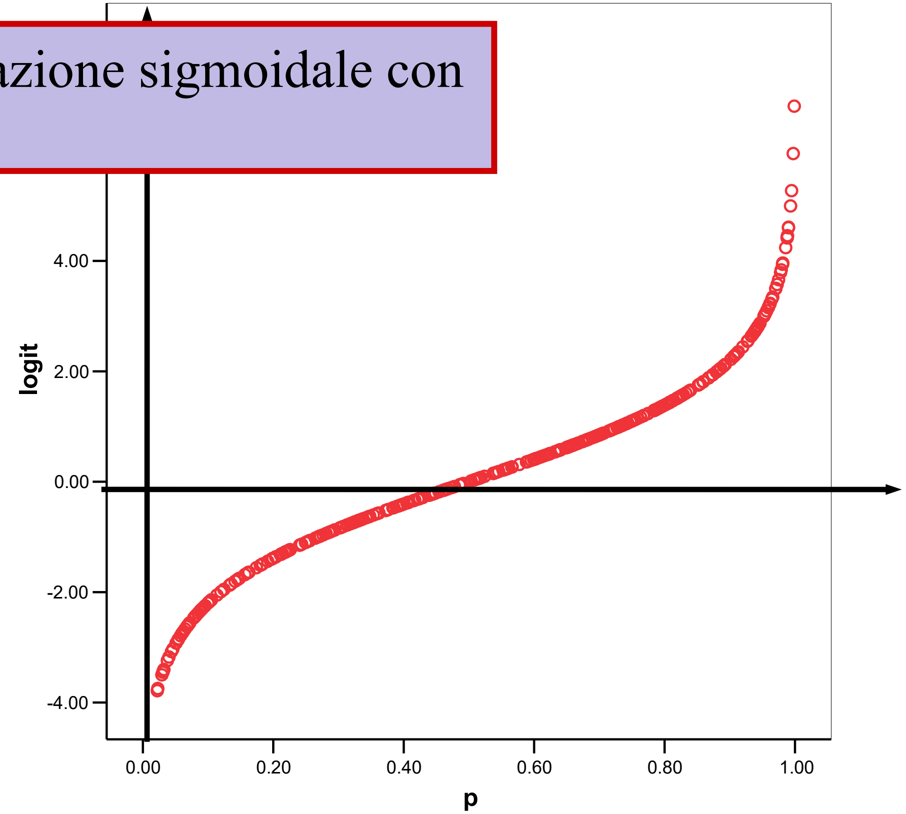


Perché il logaritmo?

- Il logaritmo di una variabile che varia da 0 ad infinito (come gli odd ratio), varia per tutti i valori possibili, da negativi a positivi

Il logaritmo dell'OR sta in relazione sigmoidale con la probabilità

$$\log it = \ln\left(\frac{p}{1-p}\right)$$

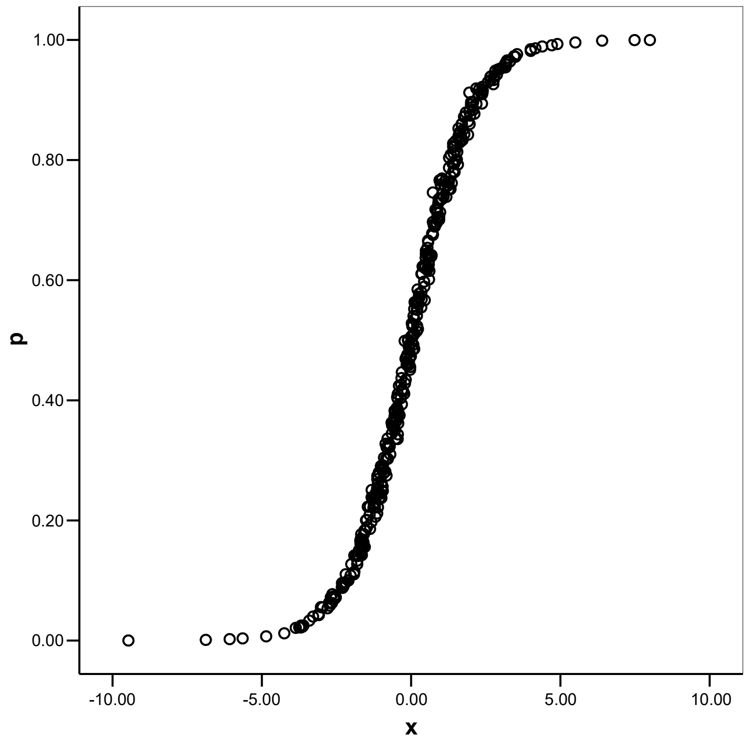


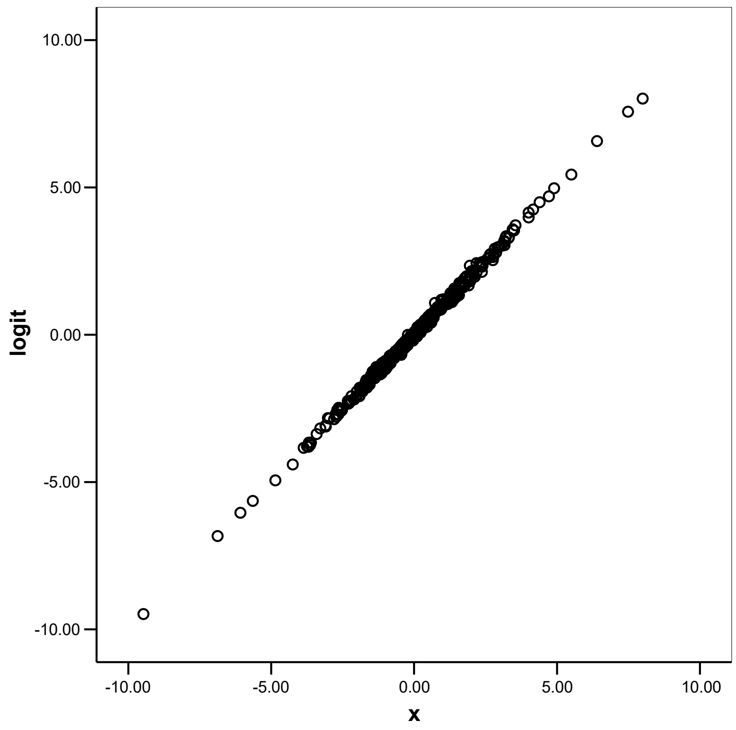
Linearizzazione della relazione

- Grazie al fatto che il **logit** sta in rapporto sigmoidale con la probabilità, il logit sarà in rapporto lineare con le variabili dipendenti

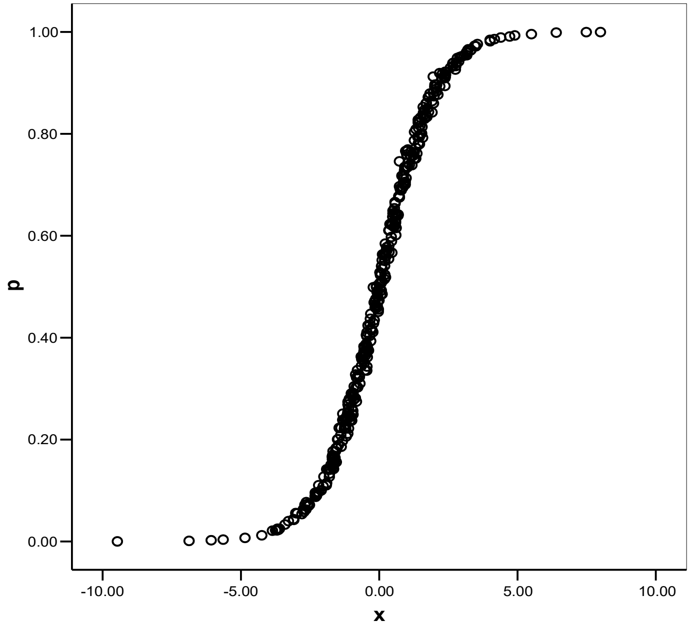
Se X predice P grazie ad una sigmoidale

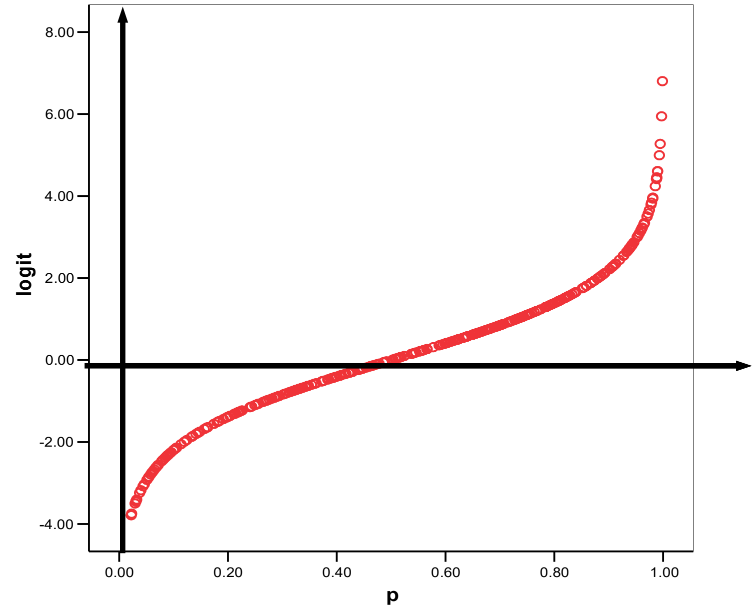
X predirrà LOGIT grazie ad una retta

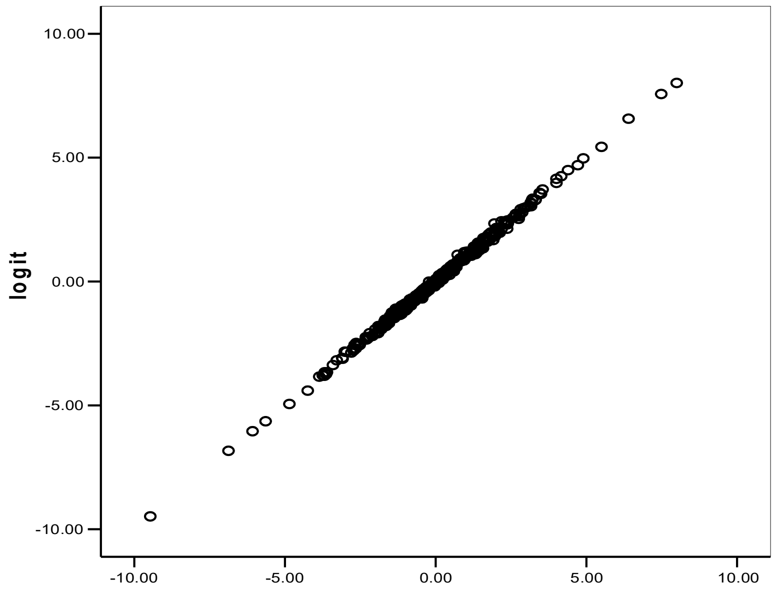




Linearizzazione della relazione II







Centramento della relazione

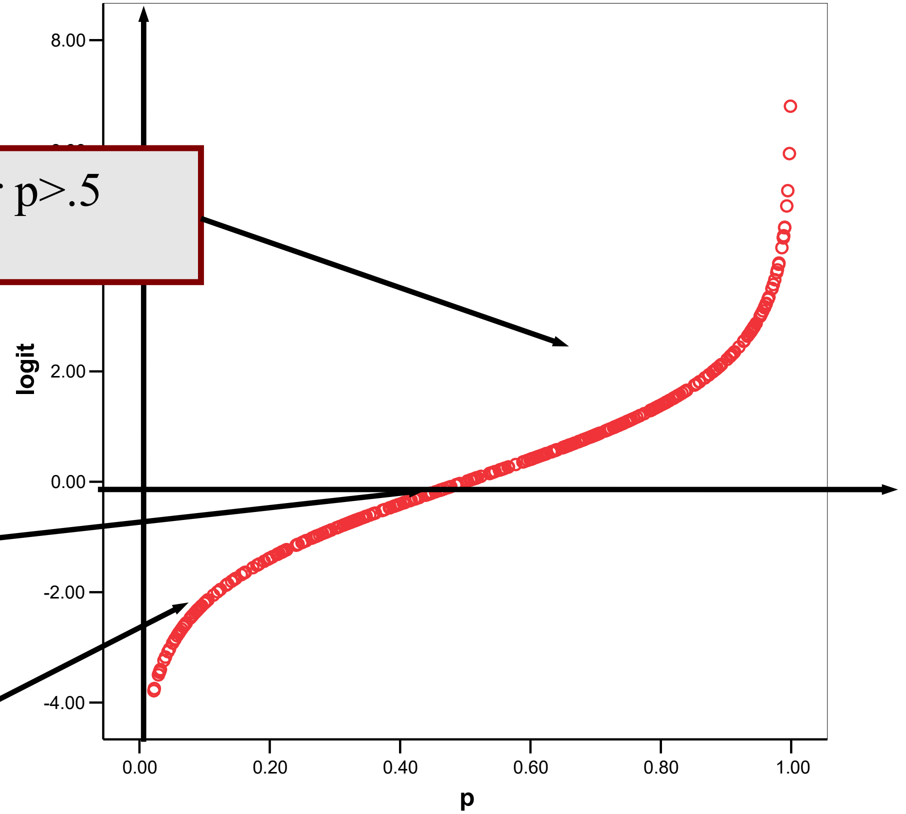
- Il Logit e' centrato rispetto alle probabilita'

$$\log it = \ln\left(\frac{p}{1-p}\right)$$

Positivo per p>.5

Centrato a zero quando p=.5

Negativo per p<.5



Trasformazione Logit

- Per ovviare a ciò, la regressione logistica non predice la variabile dipendente così come è, ma la trasforma

Variabile Originale	Odd	Logit
Probabilità	Rapporto di probabilità	Logaritmo del Odd
$P_i(Y=1)$ $1-P_i(Y=0)$	$Odd = \frac{P_i}{1-P_i}$	$\ln(\frac{P_i}{1-P_i})$
Appartenenza ad un gruppo	Quanto è più (o meno) probabile un gruppo rispetto all'altro	Variabile continua definita su tutto l'asse delle Y

Regressione logistica

La regressione logistica è una regressione in cui la variabile dipendente è dicotomica, e dunque si predice mediante una regressione lineare il logaritmo del rapporto tra la probabilità di essere in un gruppo piuttosto che l'altro

$$\ln\left(\frac{P_i}{1-P_i}\right) = a + b_{yx} x_i$$

Regressione Logistica

- Dato che la variabile è stata trasformata, la regressione ora è possibile
- Rispetto alla regressione che già conosciamo, cambierà:
    - **<u>Come interpretare i coefficienti</u>**
    - Il test di significatività (si usa il Wald test, ma la logica e' come con il t-test)
    - **<u>Come interpretare l'R^2</u>**

Esempio

- Vogliamo vedere se una variabile continua estroversione influenza la preferenza della birra rispetto al vino

$$\ln\left(\frac{Beer}{Wine}\right) = a + b_{yx}\, EXTRO_i$$

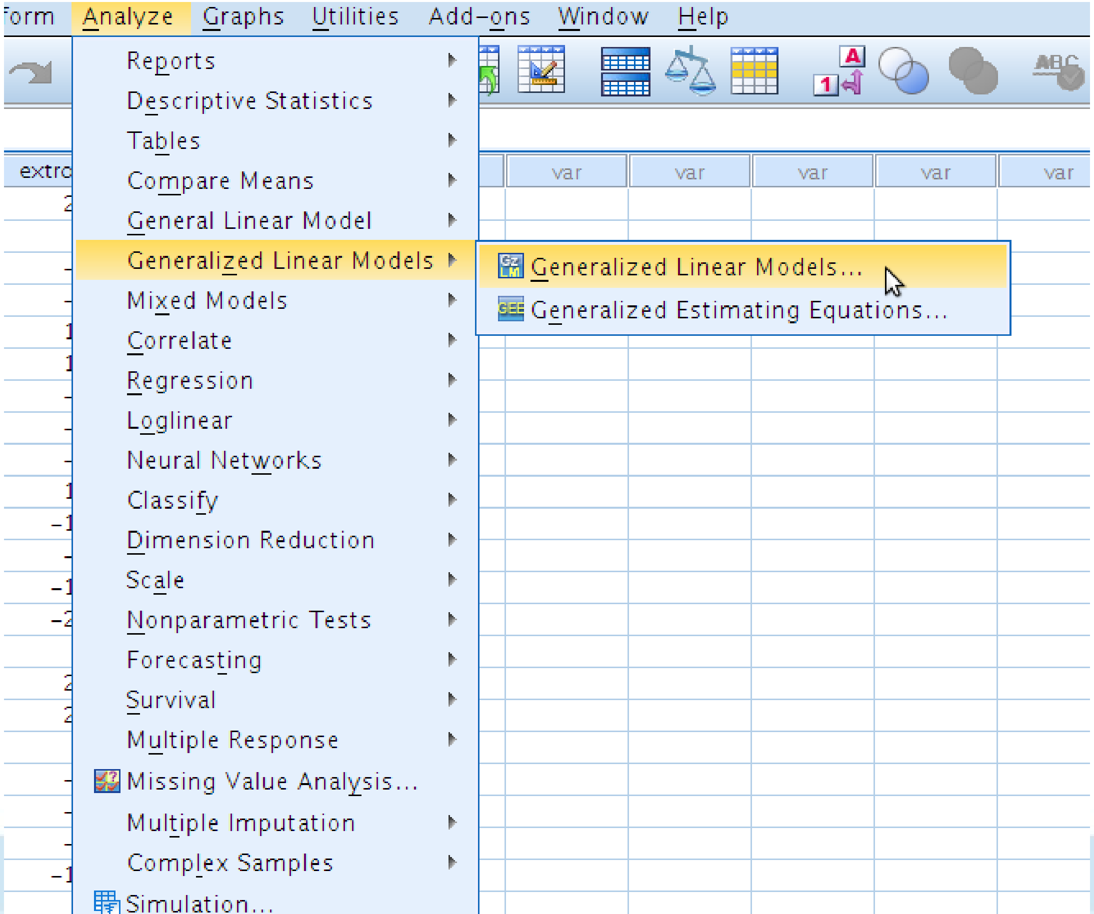
Dati SPSS

	case	sseek	extro	beer	nation	var
1	259	-1.71	2.49	1.00	1	
2	264	-.02	.55	.00	2	
3	265	1.26	-.20	1.00	1	
4	271	-.24	-.14	.00	3	
5	272	1.03	1.25	1.00	1	
6	292	-1.89	1.57	.00	2	
7	299	.09	-.14	.00	3	
8	301	-.07	-.55	1.00	1	
9	305	.32	-.06	.00	2	
10	306	-.57	1.76	1.00	1	
11	311	.00	-1.63	1.00	1	
12	317	-.49	-.56	1.00	1	
13	319	-.31	-1.38	.00	1	
14	320	-.32	-2.40	.00	1	
15	321	.00	.24	1.00	2	
16	[illegible]	-.71	2.12	1.00	1	
17	324	1.61	2.02	1.00	3	
18	[illegible]	[illegible]	.61	1.00	3	
19	331	1.89	-.52	1.00	2	

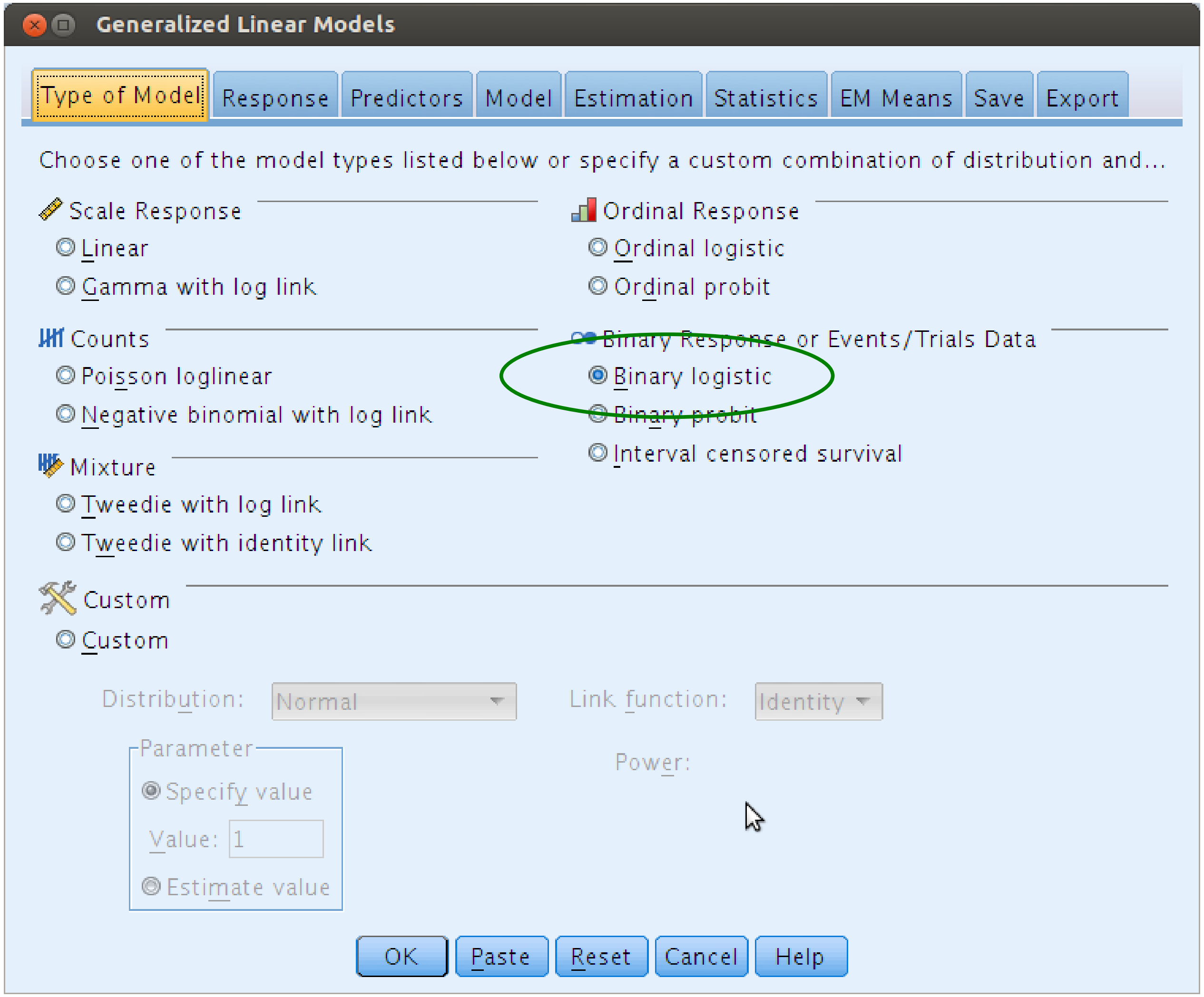
Variabile dicotomica

SPSS GzLM

- Usiamo “generalized linear models” (modelli lineari generalizzati)

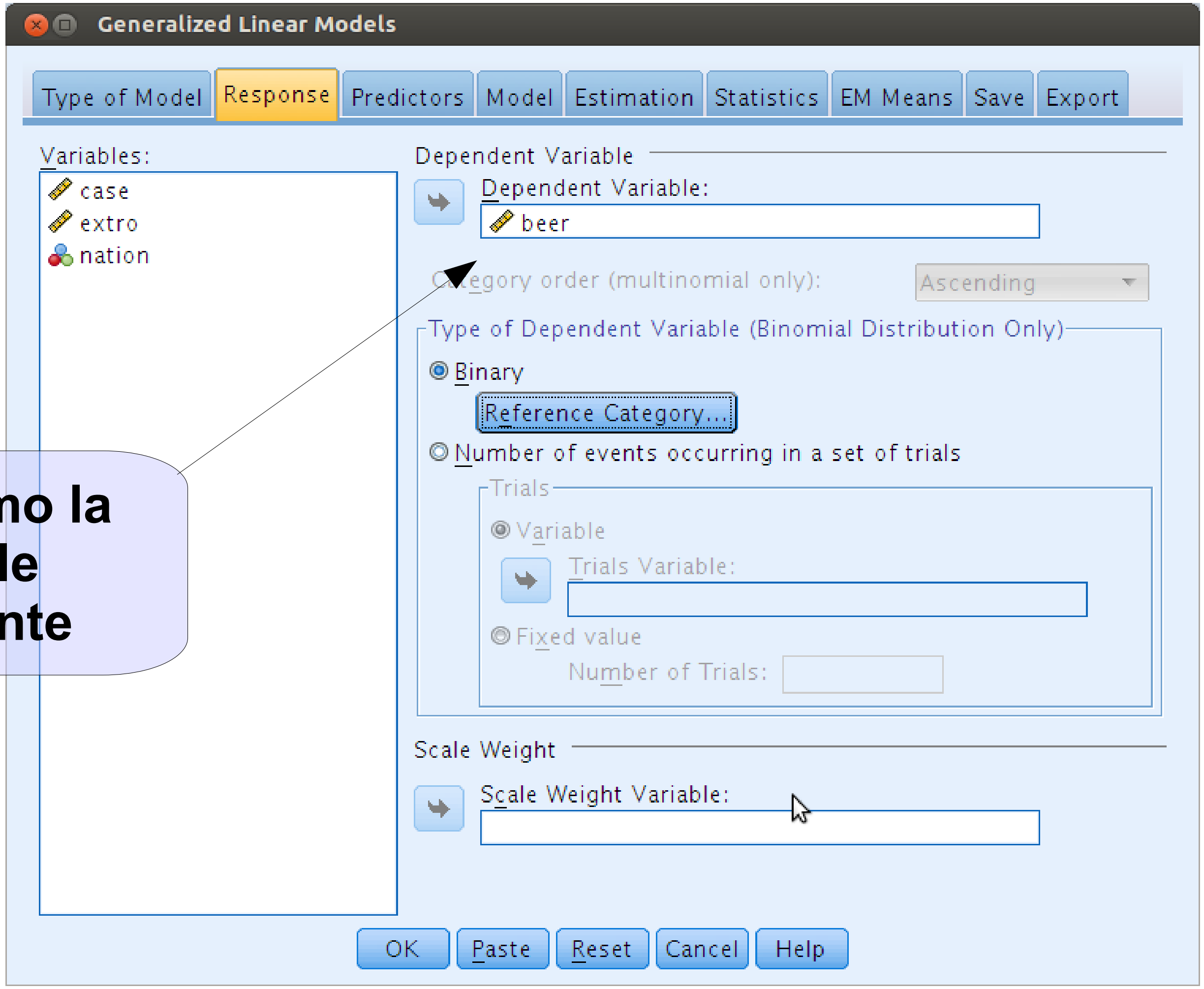


SPSS


Generalized Linear Models
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Choose one of the model types listed below or specify a custom combination of distribution and...
Scale Response
Linear
Gamma with log link
Counts
Poisson loglinear
Negative binomial with log link
Mixture
Tweedie with log link
Tweedie with identity link
Ordinal Response
Ordinal logistic
Ordinal probit
Binary Response or Events/Trials Data
Binary logistic
Binary probit
Interval censored survival
Custom
Custom
Distribution: Normal
Link function: Identity
Parameter
Specify value
Value: 1
Estimate value
Power:
OK
Paste
Reset
Cancel
Help

...

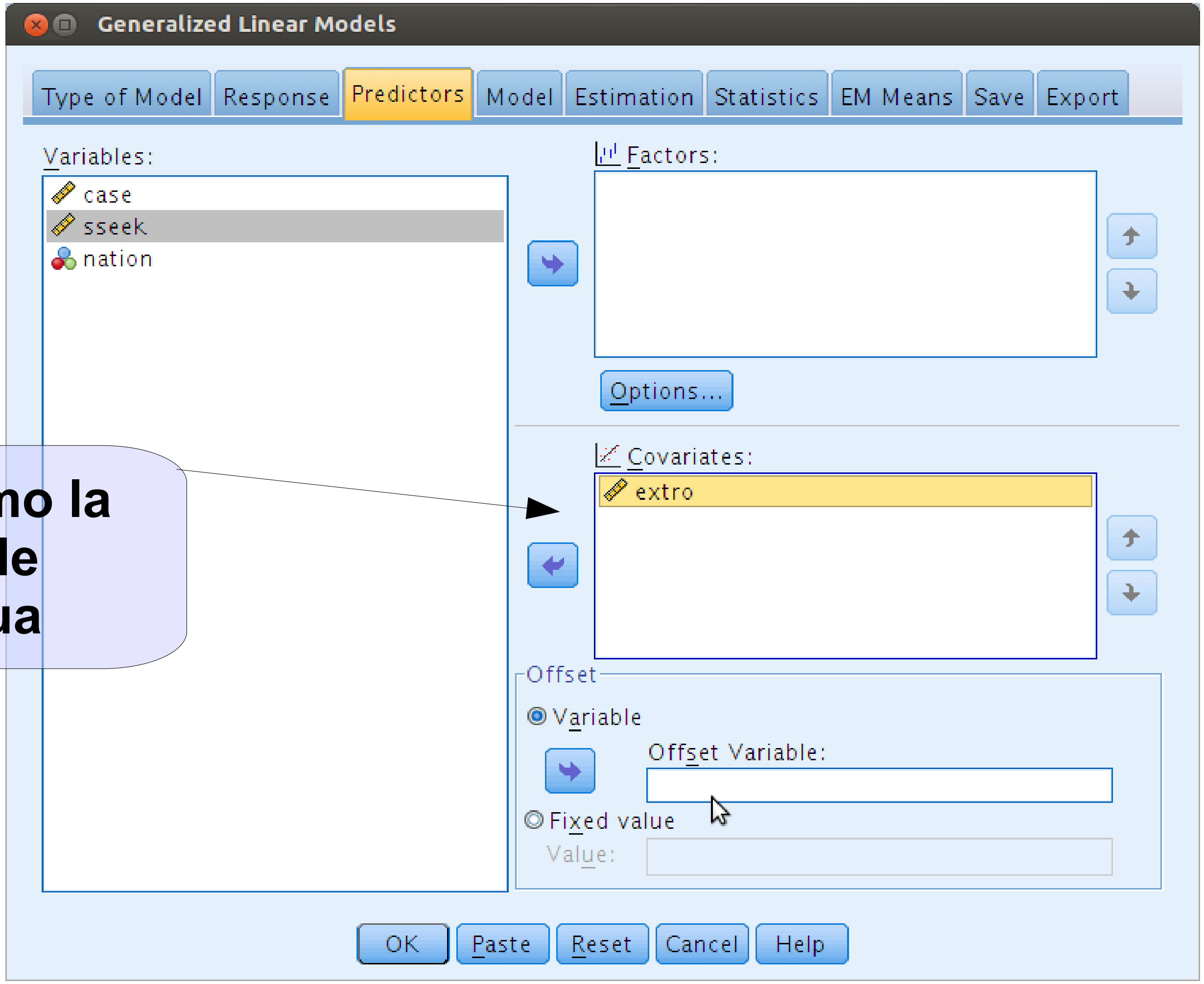
SPSS



Dichiariamo la variabile dipendente

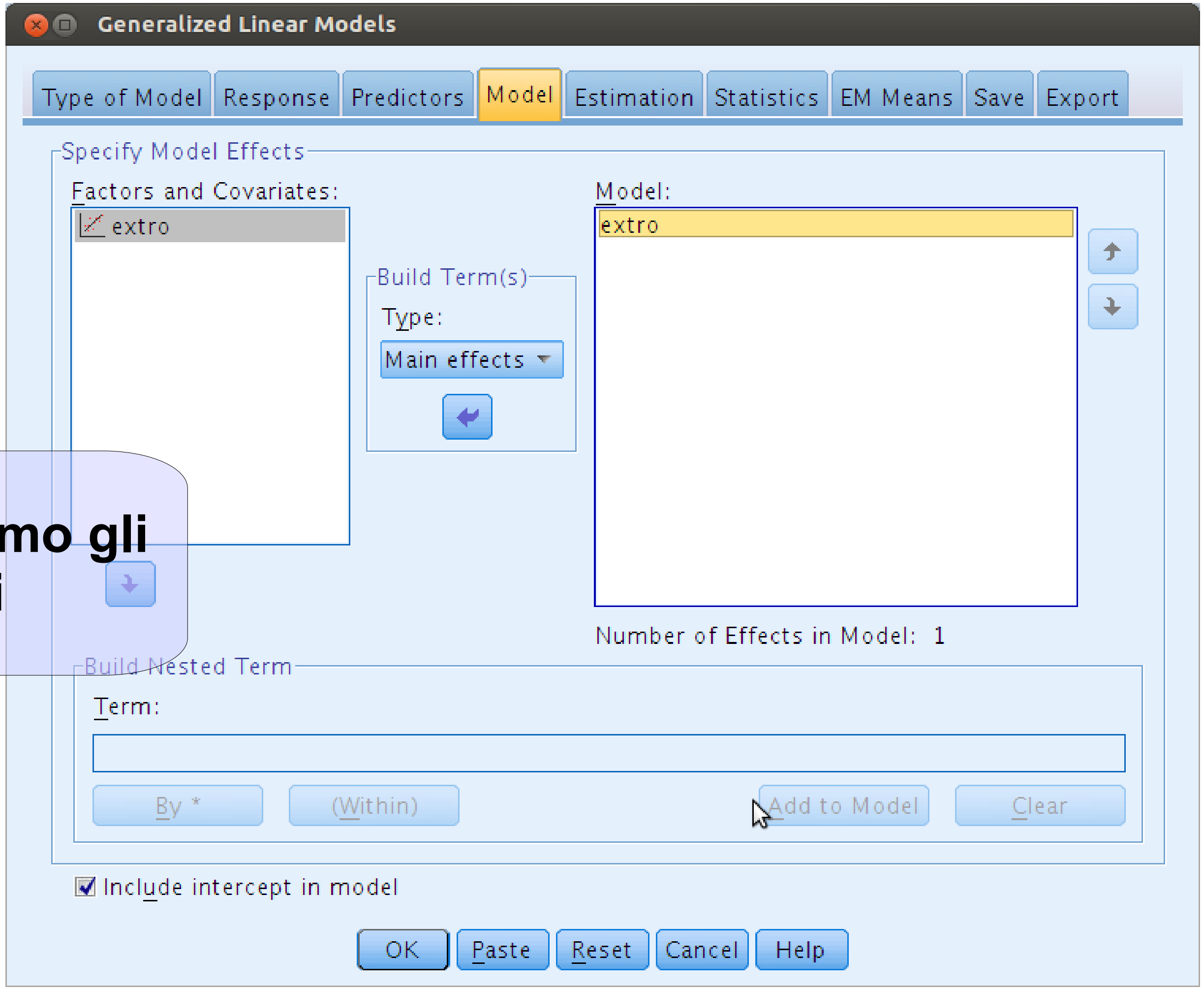
...

SPSS



Dichiariamo la variabile continua

SPSS

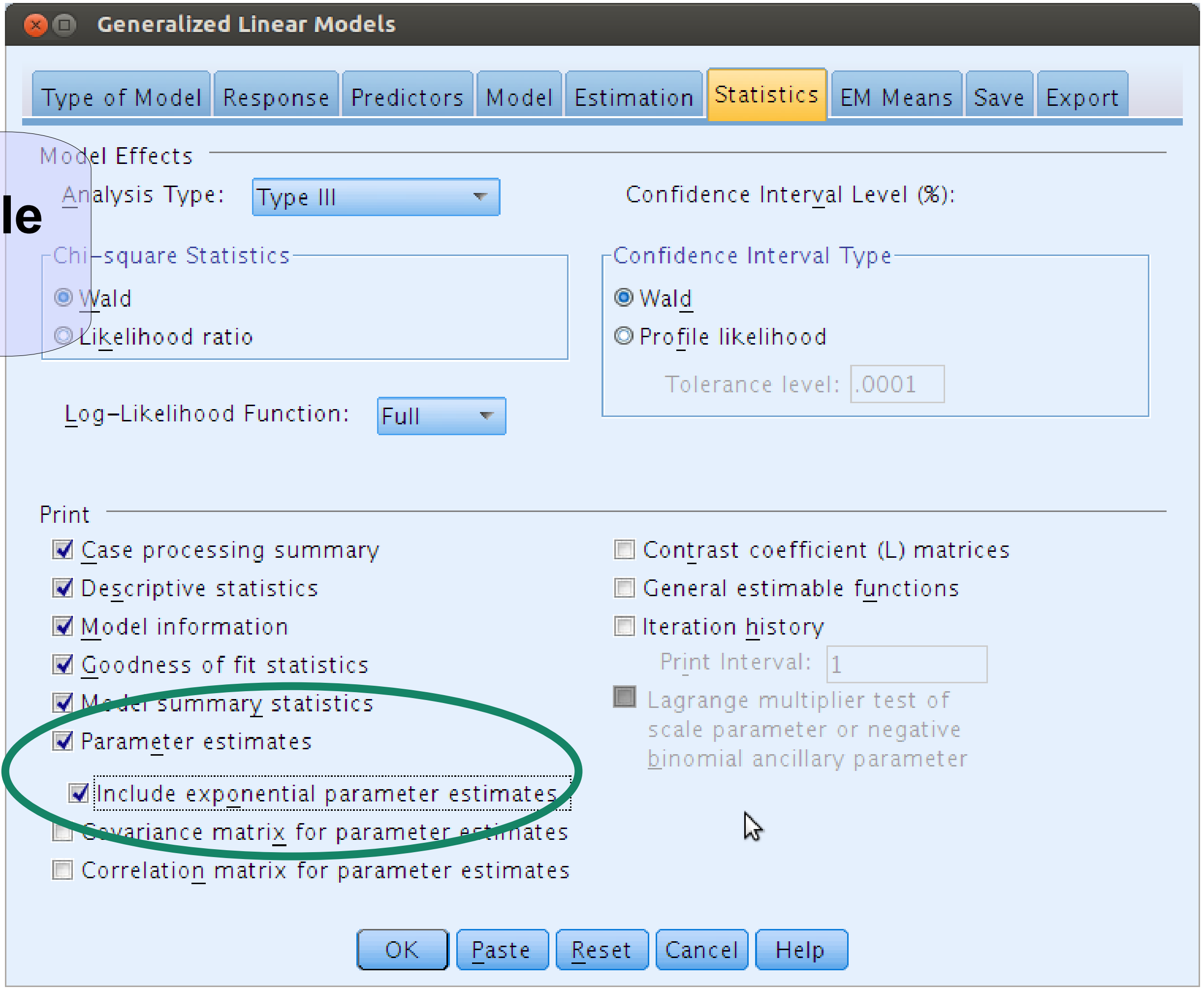


Specifichiamo gli effetti

...

SPSS

Specifichiamo le opzioni



Results

- Info about the model

Modello

Model Information

Dependent Variable	beer[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models beer as the response, treating wine as the reference category.

Missing

Case Processing Summary

	N	Percent
Included	133	100.0%
Excluded	0	0.0%
Total	133	100.0%

Dipendente

Categorical Variable Information

			N	Percent
Dependent Variable	beer	wine	60	45.1%
		beer	73	54.9%
		Total	133	100.0%

Output: Risultati

- Essendo una regressione, ci aspettiamo di trovare un coefficiente B, un coefficiente costante, e un test di significatività del coefficiente

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Output: Risultati

- Essendo una regressione, ci aspettiamo di trovare un coefficiente B, un coefficiente costante, e un test di significatività del coefficiente

Effetti

Test di significatività

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Significatività (valore-p)

Non c'è il coefficiente standardizzato, ma c'è exp(B)

Interpretazione

- Essendo una regressione, possiamo interpretare i coefficiente come al solito

Valore atteso nel logaritmo di Odd, quando la VI è zero

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Cambiamento atteso nel logaritmo di Odd, per uno spostamento nella di una unità nella VI

Significatività (valore-p): se minore di 0.05, rifiutiamo l'ipotesi nulla di B=0

Interpretazione: Problema

- Il problema sta nel fatto che tutte le informazioni (come in ogni regressione) sono espresse nell'unità di misura della VD

- Nel caso del logaritmo di Odd, questa unità è non intuitiva e poco informativa

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Per ogni unità in più di *estro*, ci aspettiamo un aumento del logaritmo di Odd di .907! Ma è tanto o poco?

Svantaggi del Logaritmo

- Il problema del logaritmo è che la sua unità di misura non è intuitivamente interpretabile
  - Una differenza di .903 nella scala logaritmica è tanto o poco in termini di probabilità?

- Per ovviare a ciò, le quantità espresse su scala logaritmiche possono essere riportate all'unità originale mediante la funzione esponenziale

$$\exp(4.605) = 100$$

$$Ln(100) = 4.605 \qquad e^{4.605} = 100$$

n=e

Unità più comprensibili

- Dato che nella logistica le informazioni sono ottenute sulla base di una VD logaritmica, la funzione esponenziale le riporta all'unità precedente (*funzione inversa*)

- L'unità precedente è l'odd ratio

Logit

$$Ln\left(\frac{P_i}{1-P_i}\right)$$

base

argomento

$\Longleftrightarrow$

Odd ratio

logaritmo

$$\exp\left(Ln\left(\frac{P_i}{1-P_i}\right)\right) = \frac{P_i}{1-P_i}$$

funzione esponenziale

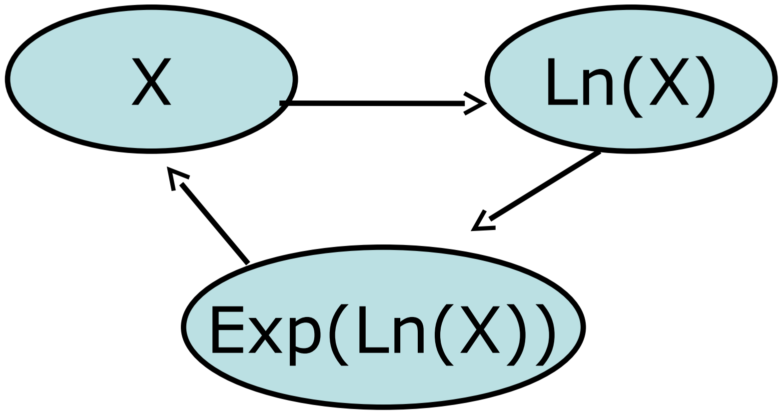
argomento

In generale

$$\exp(Ln(q)) = q$$

L’esponenziale

- La funzione esponenziale di un logaritmo ci da l’argomento orginale del logaritmo



Relazioni tra unità di misure

- Al fine interpretativo è importante ricordare che:
- La somma tra due logaritmi, equivale al prodotto tra gli argomenti

Logit

Odd ratio

$$q = Ln(a) + Ln(b) \quad \Longleftrightarrow \quad \exp(q) = a * b$$

$$Ln(2) + Ln(3) = 1.79$$

$$\exp(1.79) = 3 \times 2 = 6$$

$$B = Ln(Odd_1) + Ln(Odd_2)$$

$$\exp(B) = Odd_1 * Odd_2$$

B espresso come OR

- Per facilitare l'interpretazione, il legame tra VD e VI si esprime mediante l'esponenziale di B

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Exp(B) trasforma il B espresso in scala logaritmica in un B expresso in termini di odd ratio

Interpretazione di exp(B)

- Exp(B) è il tasso con cui aumenta Odd per ogni unità della VI

- Cioè il tasso con cui cambia il rapporto di probabilità al variare della VI

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Per ogni unità in più di *extra*, il rapporto di probabilità di preferire birra rispetto al vino aumenta di 2.47 volte

Interpretazione di exp(Costante)

- Exp(Costante) è l'Odd atteso quanto la VI è zero

Se *extra* è uguale a zero, ci aspettiamo un odd ratio di 1.141, cioè preferire la birra è 1.141 volte più probabile che preferire il vino

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	.132	.1885	-.238	.501	.487	1	.485	1.141	.788	1.650
extro	.907	.2239	.468	1.346	16.399	1	.000	2.476	1.597	3.841
(Scale)	1[a]									

Dependent Variable: beer
Model: (Intercept), extro

a. Fixed at the displayed value.

Se extra è uguale a zero , preferire la birra o il vino sono quasi equiprobabili

Regressione logistica multipla

- Tutto cio' che sappiamo sulla regressione lineare (interazione, effetti parziali, mediazione, path analysis) rimane concettualmente equivalente per la logistica

- Cambia cosa si predice ed il calcolo dei coefficient

- Cambia anche come valutare la bontà complessiva dell'equazione di regressione

Regressione logistica multipla

- Tutto cio' che sappiamo sulla regressione lineare (interazione, effetti parziali, mediazione, path analysis) rimane concettualmente equivalente per la logistica

- Cambia cosa si predice ed il calcolo dei coefficienti

- Esempio di RL multipla: Vogliamo predire se le persone preferiscono il vino (1) o la birra (0) in funzione della loro nazionalita' (Italiani, 1, vs. Inglesi, 0) tenendo sotto controllo la quantita' (quanti bicchieri bevono ogni giorno)

Coefficienti

- I coefficienti sono espressi nella scala logaritmica [**B**] e nella scala dei rapporti di probabilita' o odd ratio [**exp(B)**]

Ln B

Test di significativita'

Odd Ratio B

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	NATION	1.935	.467	17.164	1	.000	6.921	2.771	17.283
	GLASSESC	.694	.229	9.157	1	.002	2.002	1.277	3.138
	Constant	-1.663	.420	15.694	1	.000	.189		

a. Variable(s) entered on step 1: NATION, GLASSESC.

Interpretazione

- B e' il cambiamento predetto nella logit per un'unita di cambiamento nella VI, **<u>mantenendo costante</u>** (al netto, parzializzando) l'altra VI

- exp(B) e' di quanto il OR (odd ratio) di appartenere al gruppo 1 si moltiplica quando muoviamo la VI di 1 unita'

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	NATION	1.935	.467	17.164	1	.000	6.921	2.771	17.283
	GLASSESC	.694	.229	9.157	1	.002	2.002	1.277	3.138
	Constant	-1.663	.420	15.694	1	.000	.189		

a. Variable(s) entered on step 1: NATION, GLASSESC.

Interpretazione

OR per Inglesi (Nation=0) sulla preferenza di vino rispetto alla birra e' di 0.189

Se ci spostiamo da Inglesi ad Italiani, OR diviene .189*6.921=1.3

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	NATION	1.935	.467	17.164	1	.000	6.921	2.771	17.283
	GLASSESC	.694	.229	9.157	1	.002	2.002	1.277	3.138
	Constant	-1.663	.420	15.694	1	.000	.189		

a. Variable(s) entered on step 1: NATION, GLASSESC.

OR per preferire il vino per gli Italiani e' 6.92 volte maggiore che per gli Inglesi

Interpretazione

- exp(B) della VI e' di quanto il OR di appartenere al gruppo 1 si moltiplica quano muoviamo la VI di 1 unita'

- exp(B) della costante e' OR quando la VI e' eguale a 0 (Nation=Inglesi)

OR per Inglesi=.189 → Inglesi preferiscono birra

OR per vino per Italiani e' 6.92 piu' grande che per gli Inglesi → Italiani preferiscono vino 7 volte di piu' degli Inglesi

OR per Italiani, 189*6.921=1.3 → Italiani preferiscono vino

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	NATION	1.935	.467	17.164	1	.000	6.921	2.771	17.283
	GLASSESC	.694	.229	9.157	1	.002	2.002	1.277	3.138
	Constant	-1.663	.420	15.694	1	.000	.189		

a. Variable(s) entered on step 1: NATION, GLASSESC.

Riassumendo

- La **Regressione Logistica** e' una regressione con una VD binaria
- Si focalizza sulla probabilita' di appartenza al gruppo
- I coefficienti sono espressi in scala logaritmica (B) come Odd Ratio exp(B)
- Il exp(B) e' la quantita' per la quale OR viene moltiplicato quando muoviamo la VI di 1 unita'
- La logica di fondo e' come per la Regressione Lineare Multipla
- E' comprensibile come un caso particolare di Modello Lineare Generalizzato