

# METODI STATISTICI PER LA RICERCA SOCIALE

## REGRESSIONE MULTIPLA IN

Alessandra Mattei

Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)  
Università degli Studi di Firenze  
[mattei@disia.unifi.it](mailto:mattei@disia.unifi.it)

LM-88 SOCIOLOGIA E RICERCA SOCIALE

# Abbandono degli studi

- Abbandoni nelle scuole superiori professionali
- Campione casuale di  $n = 50$  classi
- Variabili
  - $Y$  = Percentuali abbandoni al termine del primo anno
  - $X_1$  = Percentuale ripetenti
  - $X_2$  = Percentuale studenti la cui famiglia ha un reddito al di sotto della soglia di povertà
  - $X_3$  = Percentuale di studenti i cui genitori hanno titolo di studio pari alla scuola dell'obbligo
  - $X_4$  = Percentuale di studenti maschi

# Abbandono degli studi: Dati

```
> rm(list=ls())
> abbandoni <- read.table("Dati/Abbandoni.txt", header=T)
> head(abbandoni)
  y x1 x2 x3 x4
1 19  9 25 16 59
2  9  4  4 23 57
3 17 15  1 20 72
4 17 10  7 22 73
5 14  7 10 19 74
6 19 15  9 33 58

> apply(abbandoni, 2, summary)
      y      x1      x2      x3      x4
Min.   4.00   1.0   0.00   8.00  57.00
1st Qu. 10.25   4.0   5.00  15.25  62.00
Median  14.00   7.0   9.00  21.00  65.00
Mean    14.14   7.5  10.58  20.88  65.96
3rd Qu.  17.00  10.0  15.00  25.75  70.75
Max.    23.00  18.0  27.00  44.00  77.00

> attach(abbandoni)
```

# Abbandono degli studi: Dati

```
> me <- lm(y ~ x1 + x2 + x3 + x4, data=abbandoni)
> me
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4, data = abbandoni)
```

Coefficients:

(Intercept)	x1	x2	x3	x4
3.18044	0.58526	0.20896	0.11460	0.02981

# Abbandono degli studi: Modello di regressione

```
> summary(me)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4, data = abbandoni)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6.2355	-1.8816	0.0923	1.8683	7.0119

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.18044	5.89108	0.540	0.59195
x1	0.58526	0.12050	4.857	1.48e-05 ***
x2	0.20896	0.06674	3.131	0.00306 **
x3	0.11460	0.06821	1.680	0.09984 .
x4	0.02981	0.08418	0.354	0.72490

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.267 on 45 degrees of freedom

Multiple R-squared: 0.4677, Adjusted R-squared: 0.4203

F-statistic: 9.883 on 4 and 45 DF, p-value: 7.959e-06

# Abbandono degli studi: Modello di regressione

```
> ##Valori teorici e residui
> yhat <- me$fit
> ehat <- me$res
>
> ##Coefficiente di correlazione multipla e R2
> cor(y, yhat)
[1] 0.683855
> cor(y, yhat)^2
[1] 0.4676576
>
> summary(me)$r.squared
[1] 0.4676576
```

# Abbandono degli studi: Modello di regressione

```
> ##Intervalli di confidenza per i coefficienti  
> confint(me)
```

	2.5 %	97.5 %
(Intercept)	-8.68480236	15.0456912
x1	0.34256761	0.8279602
x2	0.07454642	0.3433694
x3	-0.02277186	0.2519769
x4	-0.13974452	0.1993686

```
>
```

```
> confint(me, level=0.90)
```

	5 %	95 %
(Intercept)	-6.713199e+00	13.0740877
x1	3.828955e-01	0.7876322
x2	9.688108e-02	0.3210348
x3	5.512414e-05	0.2291499
x4	-1.115699e-01	0.1711940

# Abbandono degli studi: Confronto tra modelli

```
> ##Modello nullo  
> m0<- lm(y ~ 1, data=abbandoni)  
> summary(m0)
```

```
Call:  
lm(formula = y ~ 1, data = abbandoni)
```

Residuals:

Min	1Q	Median	3Q	Max
-10.14	-3.89	-0.14	2.86	8.86

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	14.1400	0.6068	23.3	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.291 on 49 degrees of freedom



# Abbandono degli studi: Confronto tra modelli

```
> anova(m0, me)
Analysis of Variance Table

Model 1: y ~ 1
Model 2: y ~ x1 + x2 + x3 + x4
  Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1      49 902.02
2      45 480.18  4    421.84 9.883 7.959e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Abbandono degli studi: Confronto tra modelli

```
> ##Modello esteso versus modello ridotto  $Y \sim b_0 + b_1 x_1 + e$   
> m.x1 <-lm(y ~ x1, data=abbandoni)  
> summary(m.x1)
```

Call:  
lm(formula = y ~ x1, data = abbandoni)

Residuals:

Min	1Q	Median	3Q	Max
-7.575	-2.715	0.365	2.967	6.855

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.8650	1.0627	9.283	2.73e-12 ***
x1	0.5700	0.1242	4.589	3.22e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.614 on 48 degrees of freedom

Multiple R-squared: 0.3049, Adjusted R-squared: 0.2904

F-statistic: 21.05 on 1 and 48 DF, p-value: 3.222e-05

# Abbandono degli studi: Confronto tra modelli

```
> anova(m.x1, me)
```

```
Analysis of Variance Table
```

```
Model 1: y ~ x1
```

```
Model 2: y ~ x1 + x2 + x3 + x4
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	48	627.00				
2	45	480.18	3	146.81	4.5862	0.006938 **

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Abbandono degli studi: Confronto tra modelli

```
> ##Modello esteso versus modello ridotto  $Y \sim b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + e$   
> m12 <-lm(y ~ x1+x2, data=abbandoni)  
> summary(m12)
```

Call:

```
lm(formula = y ~ x1 + x2, data = abandoni)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6.8808	-2.5603	0.3101	1.9458	7.5626

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	6.98033	1.31197	5.320	2.82e-06	***
x1	0.64794	0.11583	5.594	1.10e-06	***
x2	0.21740	0.06661	3.264	0.00205	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.298 on 47 degrees of freedom

Multiple R-squared: 0.4333, Adjusted R-squared: 0.4092

F-statistic: 17.97 on 2 and 47 DF, p-value: 1.596e-06

# Abbandono degli studi: Confronto tra modelli

```
> anova(m12, me)
```

Analysis of Variance Table

Model 1:  $y \sim x1 + x2$

Model 2:  $y \sim x1 + x2 + x3 + x4$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	47	511.14				
2	45	480.18	2	30.953	1.4504	0.2452

# Abbandono degli studi: Modello con interazione

```
> ##Modello con interazione  
> m.i <- lm(y ~ x1 + x2 + x1:x2, data=abbandoni)  
> m.i
```

Call:

```
lm(formula = y ~ x1 + x2 + x1:x2, data = abbandoni)
```

Coefficients:

(Intercept)	x1	x2	x1:x2
6.68010	0.68868	0.24433	-0.00396

# Abbandono degli studi: Modello con interazione

```
> summary(m.i)
```

Call:

```
lm(formula = y ~ x1 + x2 + x1:x2, data = abbandoni)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6.8382	-2.5887	0.2645	1.9242	7.6520

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	6.68010	1.83769	3.635	0.000699	***
x1	0.68868	0.20863	3.301	0.001868	**
x2	0.24433	0.13253	1.844	0.071688	.
x1:x2	-0.00396	0.01679	-0.236	0.814601	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.331 on 46 degrees of freedom

Multiple R-squared: 0.434, Adjusted R-squared: 0.3971

F-statistic: 11.76 on 3 and 46 DF, p-value: 7.669e-06

# Abbandono degli studi: Modello con interazione

```
> cor(x1, x1*x2)
[1] 0.504533
> cor(x2, x1*x2)
[1] 0.6238434
>
> x1.c <- x1-mean(x1)
> x2.c <- x2-mean(x2)
> x12.c <- x1.c*x2.c
>
> m.i.c<- lm(y ~ x1.c + x2.c + x1.c:x2.c)
> m.i.c
```

Call:

```
lm(formula = y ~ x1.c + x2.c + x1.c:x2.c)
```

Coefficients:

(Intercept)	x1.c	x2.c	x1.c:x2.c
14.11596	0.64678	0.21463	-0.00396



# Abbandono degli studi: Modello con interazione

```
> summary(m.i.c)
```

Call:

```
lm(formula = y ~ x1.c + x2.c + x1.c:x2.c)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6.8382	-2.5887	0.2645	1.9242	7.6520

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	14.11596	0.48203	29.284	< 2e-16	***
x1.c	0.64678	0.11712	5.522	1.5e-06	***
x2.c	0.21463	0.06831	3.142	0.00293	**
x1.c:x2.c	-0.00396	0.01679	-0.236	0.81460	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.331 on 46 degrees of freedom

Multiple R-squared: 0.434, Adjusted R-squared: 0.3971

F-statistic: 11.76 on 3 and 46 DF, p-value: 7.669e-06

## Abbandono degli studi: Intervalli di confidenza per la risposta media e di previsione

```
> ## Intervalli di confidenza e di previsione:  
> ## Modello  $Y=b_0+b_1*x_1+b_2*x_2+e$   
> m12
```

Call:

```
lm(formula = y ~ x1 + x2, data = abbandoni)
```

Coefficients:

(Intercept)	x1	x2
6.9803	0.6479	0.2174

## Abbandono degli studi: Intervalli di confidenza per la risposta media e di previsione

```
> new<-data.frame(x1=c(5,10,15), x2=c(5,10,20))
> predict.lm(m12,new, se.fit=T)
$fit
      1      2      3
11.30704 15.63376 21.04749

$se.fit
      1      2      3
0.6956125 0.5461178 1.2612171

$df
[1] 47

$residual.scale
[1] 3.297764
```

## Abbandono degli studi: Intervalli di confidenza per la risposta media e di previsione

```
> predict.lm(m12,new, interval="c")
      fit      lwr      upr
1 11.30704  9.90765 12.70643
2 15.63376 14.53511 16.73241
3 21.04749 18.51025 23.58473
>
> predict.lm(m12,new, interval="p")
      fit      lwr      upr
1 11.30704  4.526812 18.08727
2 15.63376  8.909158 22.35836
3 21.04749 13.944614 28.15036
```