

ESEMPIO INTERPRETAZIONE MODELLO LINEARE

```
altezza = c(175, 168, 170, 171, 169, 165, 165, 160, 180, 186)
```

```
peso = c(80, 68, 72, 75, 70, 65, 62, 60, 85, 90)
```

```
fumatore = c("si", "no", "si", "si", "no", "no", "si", "no", "si", "no")
```

```
dati = data.frame(altezza, peso, fumatore)
```

```
# Modello lineare semplice
```

```
modello = lm(formula = altezza ~ peso, data=dati)
```

```
summary(modello)
```

VARIABILE DIPENDENTE / di INTERESSE
e le VARIABILI che VOGLIAMO SPIEGARE

VARIABILE INDIPENDENTE /
COVARIATA / REGRESSORE serve per SPIEGARE Y

```
Call:
lm(formula = altezza ~ peso, data = dati)
```

← FORMULA del MODELLO che stiamo IMPLEMENTANDO

Residuals: → DISTANZA VERI VALORI di Y e QUELLI STIMATI con IL MODELLO

Min	1Q	Median	3Q	Max
-1.6622	-0.9683	-0.1622	0.5679	2.2979

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Coefficients: → INTERCETTA e COEFF. ASSOCIATI alle COVARIATE

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	115.20021	3.48450	33.06	7.64e-10 ***
peso	0.76616	0.04754	16.12	2.21e-07 ***

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

Residual standard error: 1.405 on 8 degrees of freedom
Multiple R-squared: 0.9701, Adjusted R-squared: 0.9664
F-statistic: 259.7 on 1 and 8 DF, p-value: 2.206e-07

INDICE di
BONTÀ del
MODELLO
NB: ADJUSTED
PENALIZZA R^2
PER IL N° di
COVARIATE

→ H_0 : tutti i $\beta = 0$ con 1 SOLA COVARIATA ci ASPETTIAMO
COERENZA con p-value
Sopra

ESTIMATE: $\hat{\beta}$

STD.ERROR: stimato dei coefficienti

t VALUE: e il t-statistic del TEST d'IPOTESI:

$H_0: \beta = 0$ vs $H_1: \beta \neq 0$

$Pr(>|t|)$: e il P-VALUE più *** troviamo

vuol dire che il coeff. è significativo a quel livello di
significatività

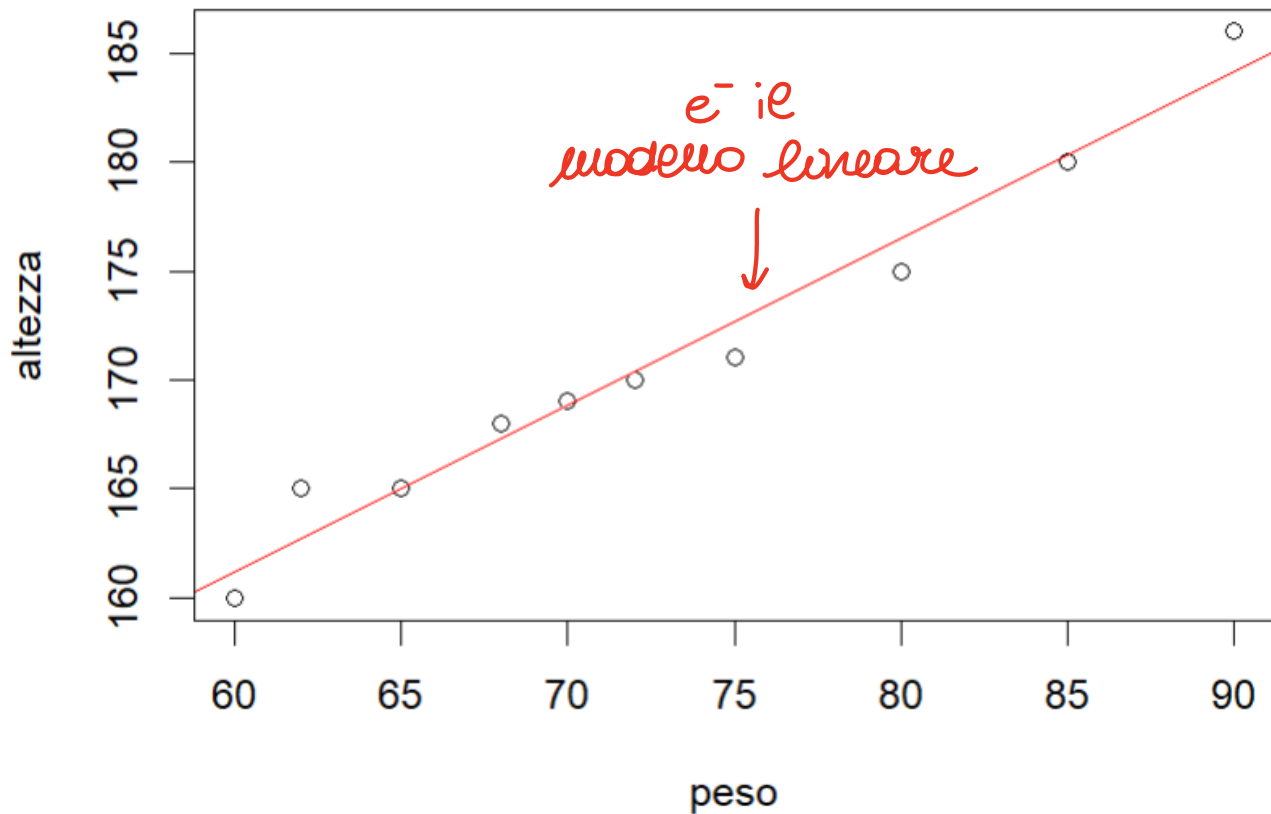
NB:
 $p\text{-value} < \alpha$
RIFIUTO
 H_0

β_1 = coeff. associato alle variabile peso

$$\hat{\beta}_1 = 0.466 \approx 0.47$$

INTERPRETAZIONE: AU' AUMENTARE UNITARIO del PESO,
L'ALTEZZA MEDIA AUMENTA di 0.47
(a PARTITA delle ALTRE VARIABILI - quando presenti)

con una variabile dipendente possiamo visualizzare il plot del modello



`plot(altezza ~ peso)`

`abline(modello, col = "red")`

Modello lineare multiplo: tutte le covariate

```
modello1 = lm(formula = altezza ~ ., data=dati)
```

```
summary(modello1)
```

↳ tutte le covariate

equivale a `lm(formula = altezza ~ peso + fumatore, data=dati)`

```
Call:
lm(formula = altezza ~ ., data = dati)

Residuals:
    Min       1Q   Median       3Q      Max
-1.3959 -0.9849 -0.1151  0.3010  2.7069

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 114.95730    3.62185   31.74 7.96e-09 ***
peso         0.77398    0.05047   15.34 1.21e-06 ***
fumatoresi  -0.65070    0.94367   -0.69  0.513
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.454 on 7 degrees of freedom
Multiple R-squared:  0.972,    Adjusted R-squared:  0.964
F-statistic: 121.6 on 2 and 7 DF, p-value: 3.663e-06
```

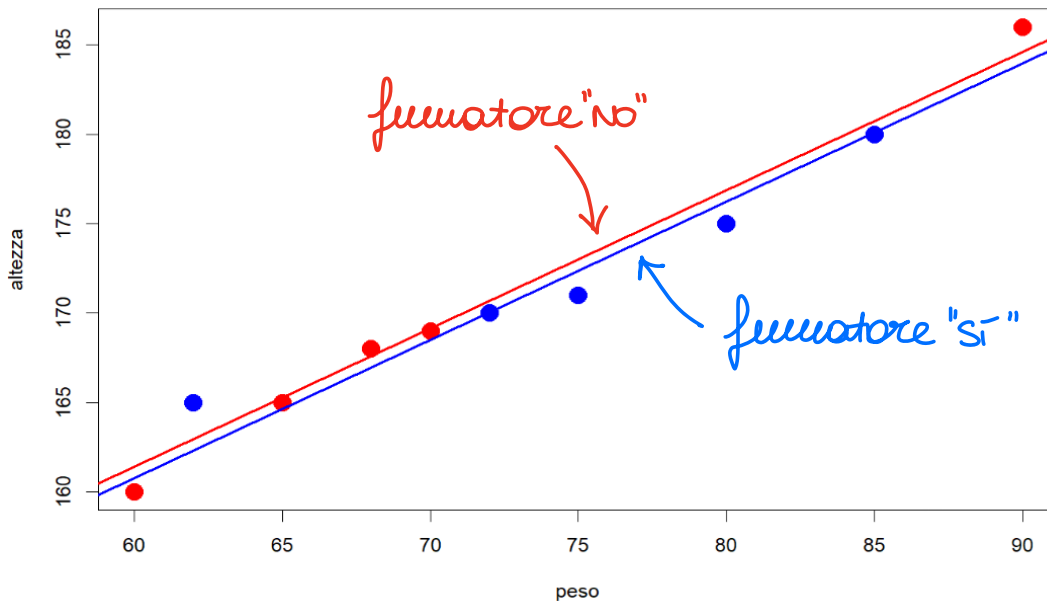
FUMATORE è una DUMMY e prende quindi 2 valori "NO" e la BASELINE

INTERPRETAZIONE: Il coefficiente associato alla variabile fumatore non è significativo, quindi non ci fidiamo che influisca sull'altezza, ma se dovessimo darle comunque un'interpretazione:

"Passando da un non fumatore ad una persona fumatore, a parità del peso, l'altezza media diminuisce di 0.65."

```
plot(altezza ~ peso, pch = 20, cex=3, col = c('red', 'blue')[as.factor(fumatore)])
b = coef(modello1)
abline(b[1], b[2], col = "red", lwd = 2)
abline(b[1]+b[3], b[2], col = "blue", lwd = 2)
```

coeff. associato a fumatore = "sì"



IN PARALLELO SI SPOSTA la RETTA di
REGRESSIONE perché CAMBIA
L'INTERCETTA