

REGRESSIONE LINEARE con VARIABILI CATEGORICA

Come nel caso della regressione semplice, consideriamo due variabili numeriche X, Y .

Consideriamo inoltre una variabile categorica D su due livelli. A ciascun livello associamo i valori $\{0, 1\}$.

Estendiamo il modello di regressione semplice nel seguente modo:

$$Y = \alpha + \beta X + \gamma D + \delta XD + \varepsilon \quad \text{dove} \quad \varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$$

includendo D , per tenere in considerazione la divisione in gruppi.

Come nel caso della regressione lineare semplice, è possibile stimare i parametri $\alpha, \beta, \gamma, \delta$ e ottenere il modello:

$$E[Y] = \hat{Y} = \hat{\alpha} + \hat{\beta} X + \hat{\gamma} D + \hat{\delta} XD$$

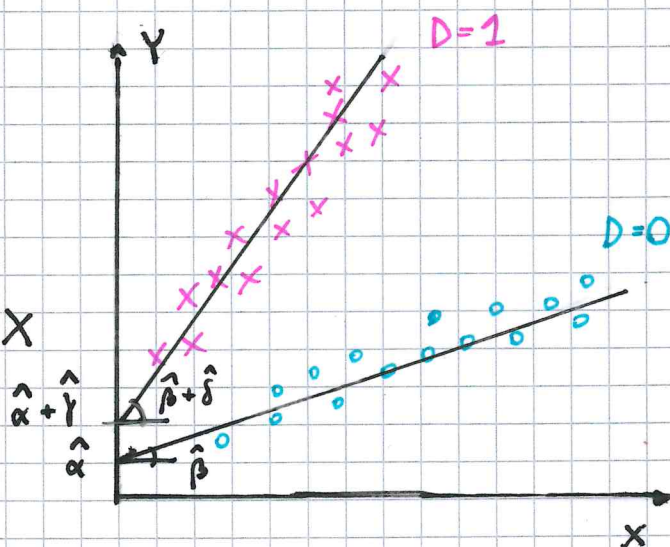
Sostituendo i due possibili valori di D , otteniamo le equazioni delle rette nei due gruppi:

- GRUPPO $D=0$:

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X$$

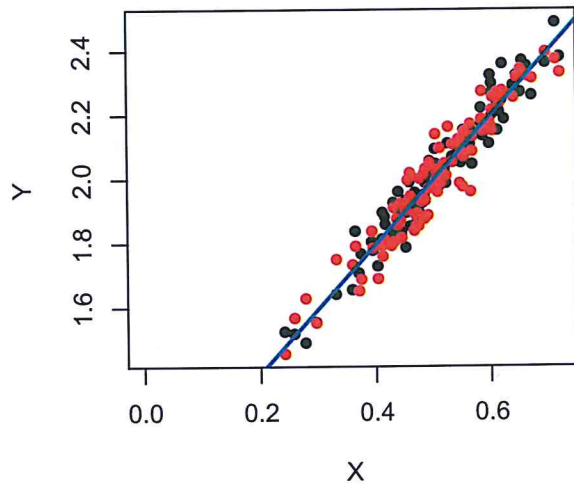
- GRUPPO $D=1$:

$$\hat{Y} = (\hat{\alpha} + \hat{\gamma}) + (\hat{\beta} + \hat{\delta}) X$$



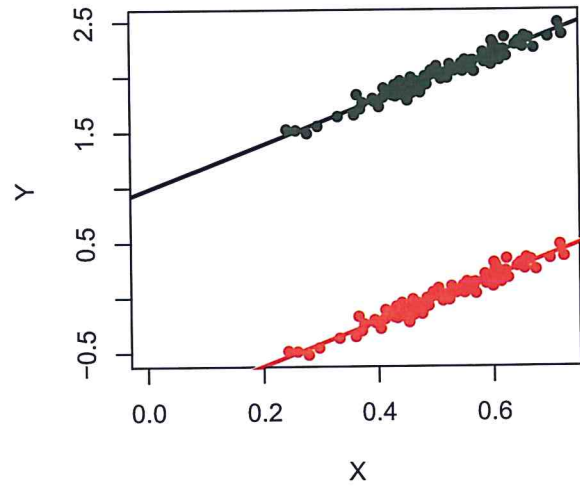
$$\gamma = \delta = 0$$

Gruppi indistinguibili



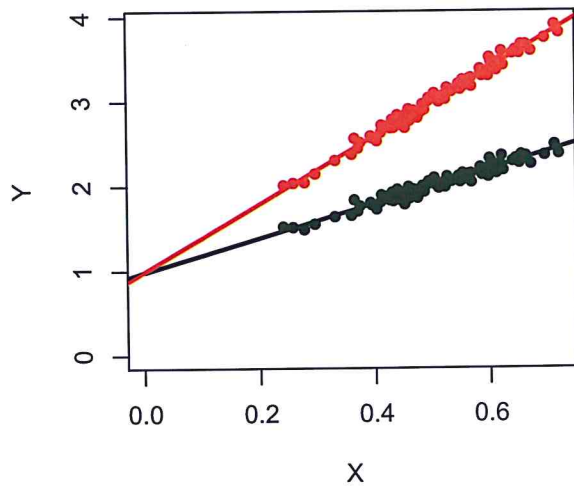
$$\gamma \neq 0 \wedge \delta = 0$$

Intercette diverse e pendenze uguali



$$\gamma = 0 \wedge \delta \neq 0$$

Intercette uguali e pendenze diverse



$$\gamma \neq 0 \wedge \delta \neq 0$$

Intercette e pendenze diverse

