

Ana Karen Priscila Rodrigo Flores

$$\textcircled{1} \hat{\beta} = \arg \min \|y - X'\beta\|^2 = \langle y - X'\beta, y - X'\beta \rangle$$

$$= \sum_{i=1}^n (y_i - x_i' \beta)^2$$

$$X = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$\underbrace{\hspace{10em}}_{\bar{x}_1 (1 \times K)}$

$$\text{FOC } \beta : \sum_{i=1}^n (y_i - y_i \hat{\beta}) x_i' = 0$$

$$\sum_{i=1}^n y_i x_i' = \sum_{i=1}^n x_i' x_i \hat{\beta}$$

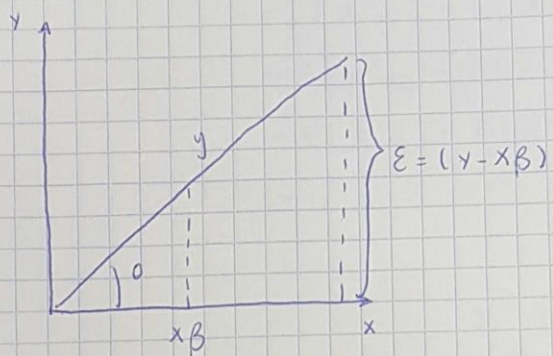
$$\hat{\beta} = \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \sum_{i=1}^n y_i x_i' = (X'X)^{-1} (X'Y) //$$

Ajuste lineal: X y y asume implícitamente relación lineal.
 $E(y|x) = x\hat{\beta} //$

Si pudiéramos ajustarlo para $y = x^2$ considero $z = x^2$

$$\hat{\beta} = \arg \min \|y - z\beta\|^2 //$$

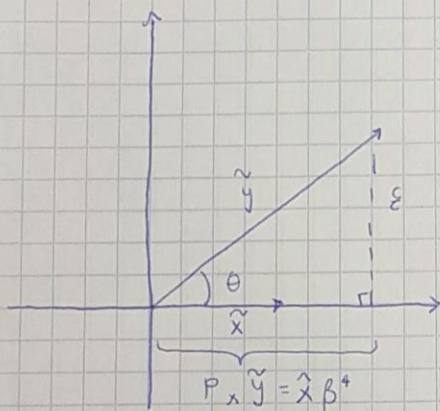
② Relación con proyección en espacio vectorial



por pitagoras, esta $\hat{\beta}$ es la menor distancia
entre y a $x\beta$

Prueba $E[\tilde{x}'\tilde{\varepsilon}] = E[\tilde{x}'(\tilde{y} - \tilde{x}\beta^*)]$

$$\begin{aligned}
 &= E[\tilde{x}'\tilde{y}] - E[\tilde{x}'\tilde{x}]\beta^* \\
 &= E[\tilde{x}'\tilde{y}] - E[\tilde{x}'\tilde{x}](E[\tilde{x}'\tilde{x}])^{-1}E[\tilde{x}'\tilde{y}] \\
 &= E[\tilde{x}'\tilde{y}] - E[\tilde{x}'\tilde{y}] \\
 &= 0
 \end{aligned}$$



Si \exists finita $E(\tilde{x}'\tilde{y})$ & $E(\tilde{x}'\tilde{x})$
(no singular) i.e. invertibles

$$\Rightarrow \tilde{y} = \tilde{x}\beta^* + \tilde{\varepsilon}$$

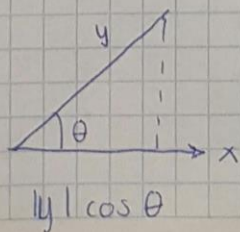
$|x| \quad |x| \cdot |x|^{-1} \quad |x|$

donde $\beta^* = \operatorname{argmin} E[(\tilde{y} - \tilde{x}\beta)^2] =$
 $= E(\tilde{x}'\tilde{x})^{-1} E(\tilde{x}'\tilde{y})$

\Downarrow
 $E(\tilde{x}'\tilde{\varepsilon}) = 0$ (Prueba 1)

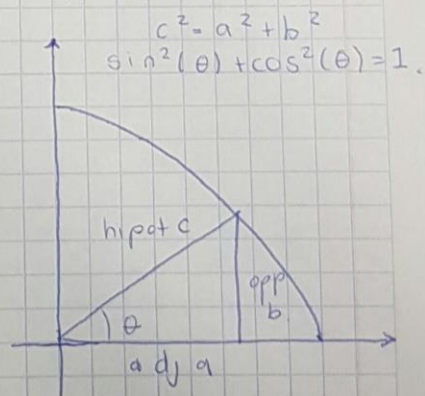
ahoras $\|\tilde{y}\|^2 = \|\tilde{x}\beta^*\|^2 + \|\tilde{\varepsilon}\|^2$

$$R^2 = \frac{\|\tilde{x}\beta^*\|^2}{\|\tilde{y}\|^2} = \frac{1}{\cos(\theta)} = \sec(\theta)$$



$$x' \tilde{y} = |x| |y| \cdot \cos(\theta)$$

$$\cos(\theta) = \frac{x' \tilde{y}}{\|x\| \|y\|}$$



$$\sin(\theta) = \frac{\text{opp}}{\text{hip}} = \frac{b}{c}$$

$$\cos(\theta) = \frac{\text{adj}}{\text{hip}} = \frac{a}{c}$$

$$\tan(\theta) = \frac{\sin \theta}{\cos \theta} = \frac{a}{b}$$

$$f(y_i | x_i) = \frac{1}{\sqrt{2\pi} \sigma^2} \exp \left\{ -\frac{(y_i - x_i \beta)^2}{2\sigma^2} \right\}$$

$$l_i(\beta, \sigma^2) = \left[-\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y_i - x_i \beta)^2 \right]$$

$$\max_{\beta} \sum_{i=1}^n l_i(\beta, \sigma^2) = \sum_{i=1}^n \downarrow (*)$$

$$= \sum_{i=1}^n \left[-\frac{1}{2\sigma^2} (y_i - x_i \beta)^2 \right]$$

(FOC β :)

$$\Rightarrow \hat{\beta}_{OLS} = (\sum x_i x_i')^{-1} (\sum x_i y_i)$$

```

> library(ggplot2)

> data(diamonds)

>

> ggplot(diamonds,aes(x=carat,y=price))+
+   geom_point(color='blue',fill='blue')+
+   xlim(0,quantile(diamonds$carat,0.99))+
+   ylim(0,quantile(diamonds$price,0.99))+
+   ggtitle('Diamond price vs. carat')

```

Warning message:

Removed 926 rows containing missing values (geom_point).

```

>

> linmodel <- lm(price ~ cut + color + clarity + (x*y*z),
+               data = diamonds)

> summary(linmodel) # to show the results

```

Call:

```
lm(formula = price ~ cut + color + clarity + (x * y * z), data = diamonds)
```

Residuals:

Min	1Q	Median	3Q	Max
-16260.7	-561.7	-161.7	371.2	11891.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18191.875	341.204	53.317	< 2e-16 ***
cut.L	271.273	19.987	13.572	< 2e-16 ***
cut.Q	-111.783	17.566	-6.364	1.99e-10 ***
cut.C	138.708	15.268	9.085	< 2e-16 ***
cut^4	30.242	12.100	2.499	0.0124 *
color.L	-1934.049	17.115	-113.006	< 2e-16 ***
color.Q	-659.710	15.596	-42.300	< 2e-16 ***
color.C	-160.126	14.549	-11.006	< 2e-16 ***
color^4	53.593	13.364	4.010	6.07e-05 ***
color^5	-97.719	12.624	-7.741	1.01e-14 ***
color^6	-53.904	11.475	-4.697	2.64e-06 ***
clarity.L	3898.696	29.865	130.545	< 2e-16 ***
clarity.Q	-1826.088	27.896	-65.461	< 2e-16 ***
clarity.C	915.093	23.861	38.350	< 2e-16 ***
clarity^4	-308.229	19.059	-16.172	< 2e-16 ***
clarity^5	192.269	15.559	12.357	< 2e-16 ***
clarity^6	18.024	13.548	1.330	0.1834
clarity^7	83.751	11.959	7.003	2.53e-12 ***
x	-6437.213	116.440	-55.283	< 2e-16 ***
y	-1913.808	120.064	-15.940	< 2e-16 ***
z	-2541.423	173.117	-14.680	< 2e-16 ***
x:y	893.877	16.756	53.347	< 2e-16 ***
x:z	1119.344	19.762	56.640	< 2e-16 ***
y:z	-238.581	27.662	-8.625	< 2e-16 ***

x:y:z -53.058 2.516 -21.092 < 2e-16 ***

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

Residual standard error: 1117 on 53915 degrees of freedom

Multiple R-squared: 0.9217, Adjusted R-squared: 0.9217

F-statistic: 2.645e+04 on 24 and 53915 DF, p-value: < 2.2e-16

>

