15-07-2019

2022-10-27

ES 3)

Studente	X	Y	Studente	X	Y
1	29	22	8	23	25
2	29	28	9	21	24
3	27	30	10	21	24
4	24	25	11	19	25
5	24	27	12	19	24
6	23	23	13	18	25
7	23	22	14	18	23

X = VOTO ESAME SCRITTO

Y = VOTO ESAME ORALE

```
(Tabella = data.frame(
  X = c(29,29,27,24,24,23,23,23,21,21,19,19,18,18),
  Y = c(22,28,30,25,27,23,22,25,24,24,25,24,25,23)
))
```

```
##     X  Y
## 1  29 22
## 2  29 28
## 3  27 30
## 4  24 25
## 5  24 27
## 6  23 23
## 7  23 22
## 8  23 25
## 9  21 24
## 10 21 24
## 11 19 25
## 12 19 24
## 13 18 25
## 14 18 23
```

Y=ESAME ORALE

```
(TabellaY = Tabella %>% group_by(Y) %>% summarise(fi = n()))
```

```
## # A tibble: 7 x 2
##       Y    fi
##   <dbl> <int>
## 1    22     2
## 2    23     2
## 3    24     3
## 4    25     4
## 5    27     1
## 6    28     1
## 7    30     1
```

```
(TabellaY = cbind(TabellaY,Fi=cumsum(TabellaY$fi)))
```

```
##    Y fi Fi
## 1 22  2  2
## 2 23  2  4
## 3 24  3  7
## 4 25  4 11
## 5 27  1 12
## 6 28  1 13
## 7 30  1 14
```

```
(TabellaY = TabellaY %>% mutate(pi=fi/sum(fi)))
```

```
##    Y fi Fi         pi
## 1 22  2  2 0.14285714
## 2 23  2  4 0.14285714
## 3 24  3  7 0.21428571
## 4 25  4 11 0.28571429
## 5 27  1 12 0.07142857
## 6 28  1 13 0.07142857
## 7 30  1 14 0.07142857
```

```
(TabellaY = cbind(TabellaY,Pi=cumsum(TabellaY$pi)))
```

```
##    Y fi Fi         pi        Pi
## 1 22  2  2 0.14285714 0.1428571
## 2 23  2  4 0.14285714 0.2857143
## 3 24  3  7 0.21428571 0.5000000
## 4 25  4 11 0.28571429 0.7857143
## 5 27  1 12 0.07142857 0.8571429
## 6 28  1 13 0.07142857 0.9285714
## 7 30  1 14 0.07142857 1.0000000
```

MEDIA

$$E(Y) = \frac{1}{n} * \sum_{i=1}^{n} y_i = \sum_{j=1}^{J} y_j * p_j$$

```
(Y_Media = mean(Tabella$Y))
```

```
## [1] 24.78571
```

```
sum(TabellaY$Y*TabellaY$pi)
```

```
## [1] 24.78571
```

VARIANZA

$$V(Y) = E[(Y - E(Y))^2] = E(Y^2) - (E(Y))^2 = \sum_{j=1}^{J}(y_j - E(Y))^2 * p_j$$

```
# valori grezzi
(Y_Varianza = mean((Tabella$Y - Y_Media)**2))
```

```
## [1] 4.739796
```

```
(Y_sigma = Y_Varianza**0.5)
```

```
## [1] 2.177107
```

```
# valori con frequenze relative
sum(((TabellaY$Y-Y_Media)**2)*TabellaY$pi)
```

```
## [1] 4.739796
```

X = ESAME SCRITTO

```
(TabellaX = Tabella %>% group_by(X) %>% summarise(fi = n()))
```

```
## # A tibble: 7 x 2
##       X    fi
##   <dbl> <int>
## 1    18     2
## 2    19     2
## 3    21     2
## 4    23     3
## 5    24     2
## 6    27     1
## 7    29     2
```

```
(TabellaX = cbind(TabellaX,Fi=cumsum(TabellaX$fi)))
```

```
##    X fi Fi
## 1 18  2  2
## 2 19  2  4
## 3 21  2  6
## 4 23  3  9
## 5 24  2 11
## 6 27  1 12
## 7 29  2 14
```

```
(TabellaX = TabellaX %>% mutate(pi=fi/sum(fi)))
```

```
##    X fi Fi         pi
## 1 18  2  2 0.14285714
## 2 19  2  4 0.14285714
## 3 21  2  6 0.14285714
## 4 23  3  9 0.21428571
## 5 24  2 11 0.14285714
```

```
## 6 27  1 12 0.07142857
## 7 29  2 14 0.14285714
```

```
(TabellaX = cbind(TabellaX,Pi=cumsum(TabellaX$pi)))
```

```
##    X fi Fi         pi        Pi
## 1 18  2  2 0.14285714 0.1428571
## 2 19  2  4 0.14285714 0.2857143
## 3 21  2  6 0.14285714 0.4285714
## 4 23  3  9 0.21428571 0.6428571
## 5 24  2 11 0.14285714 0.7857143
## 6 27  1 12 0.07142857 0.8571429
## 7 29  2 14 0.14285714 1.0000000
```

MEDIA

$$E(Y) = \frac{1}{n} * \sum_{i=1}^{n} y_i = \sum_{j=1}^{J} y_j * p_j$$

```
(X_Media = mean(Tabella$X))
```

```
## [1] 22.71429
```

```
sum(TabellaX$X*TabellaX$pi)
```

```
## [1] 22.71429
```

VARIANZA

$$V(Y) = E[(Y - E(Y))^2] = E(Y^2) - (E(Y))^2 = \sum_{j=1}^{J} (y_j - E(Y))^2 * p_j$$

```
# valori grezzi
(X_Varianza = mean((Tabella$X - X_Media)**2))
```

```
## [1] 12.77551
```

```
(X_sigma = X_Varianza**0.5)
```

```
## [1] 3.574285
```

```
# valori con frequenze relative
sum(((TabellaX$X-X_Media)**2)*TabellaX$pi)
```

```
## [1] 12.77551
```

COVARIANZA

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

```
(XY_Covarianza = mean(Tabella$X*Tabella$Y) - Y_Media*X_Media)
```

```
## [1] 2.795918
```

CORRELAZIONE LINEARE

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_Y * \sigma_X}$$

```
(rhoXY = XY_Covarianza/( Y_sigma*X_sigma))
```

```
## [1] 0.3592986
```

$$\beta = \frac{Cov(X,Y)}{V(X)} = \rho_{XY} * \frac{\sigma_Y}{\sigma_X}$$

$$\alpha = E(Y) - \beta * E(X)$$

```
(beta = XY_Covarianza / X_Varianza)
```

```
## [1] 0.2188498
```

```
(beta = rhoXY*Y_sigma/X_sigma)
```

```
## [1] 0.2188498
```

```
(alfa = Y_Media - beta*X_Media)
```

```
## [1] 19.8147
```

$$Ys = \beta X + \alpha$$

```
(Ystimata = beta*Tabella$X + alfa)
```

```
##  [1] 26.16134 26.16134 25.72364 25.06709 25.06709 24.84824 24.84824 24.84824
##  [9] 24.41054 24.41054 23.97284 23.97284 23.75399 23.75399
```

```
(Ys_Varianza = mean(Ystimata**2) - mean(Ystimata)**2)
```

```
## [1] 0.6118863
```

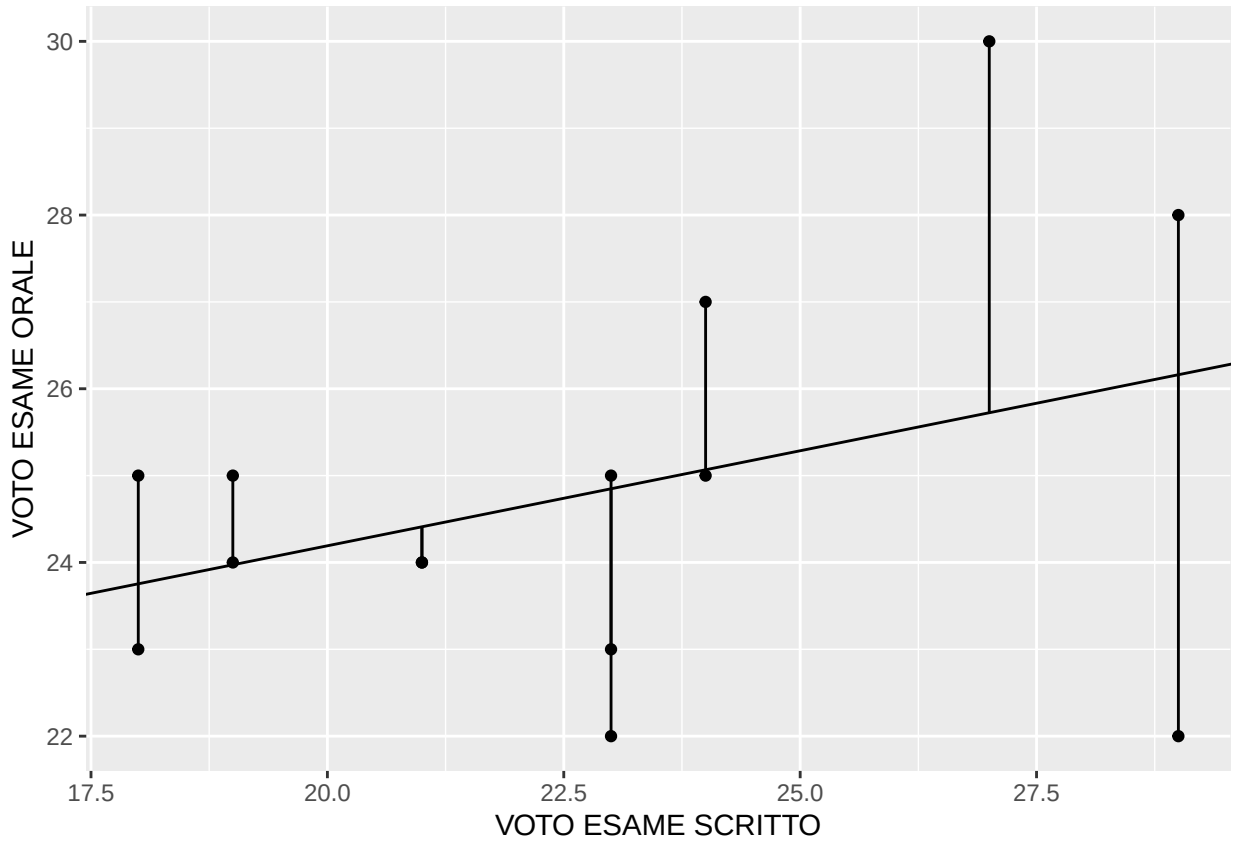
```
(ErroriStimati = Tabella$Y - Ystimata)
```

```
##  [1] -4.16134185  1.83865815  4.27635783 -0.06709265  1.93290735 -1.84824281
##  [7] -2.84824281  0.15175719 -0.41054313 -0.41054313  1.02715655  0.02715655
## [13]  1.24600639 -0.75399361
```

```
(Es_Varianza = mean(ErroriStimati**2) - mean(ErroriStimati)**2)
```

```
## [1] 4.12791
```

```
ggplot(Tabella,aes(x=X,y=Y)) + geom_point()  + geom_abline(aes(intercept=alfa,slope=beta) ) +geom_segmen
```



$$R^2 = \frac{V(Ys)}{V(Y)} = 1 - \frac{V(Es)}{V(Y)}$$

```
(Rquadro = Ys_Varianza / Y_Varianza)
```

```
## [1] 0.1290955
```

```
1-(Es_Varianza/Y_Varianza)
```

```
## [1] 0.1290955
```