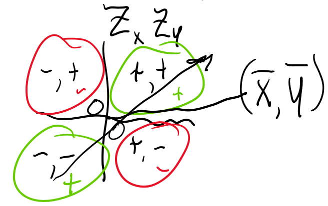
Coefficient of Determination (R^2) r^2

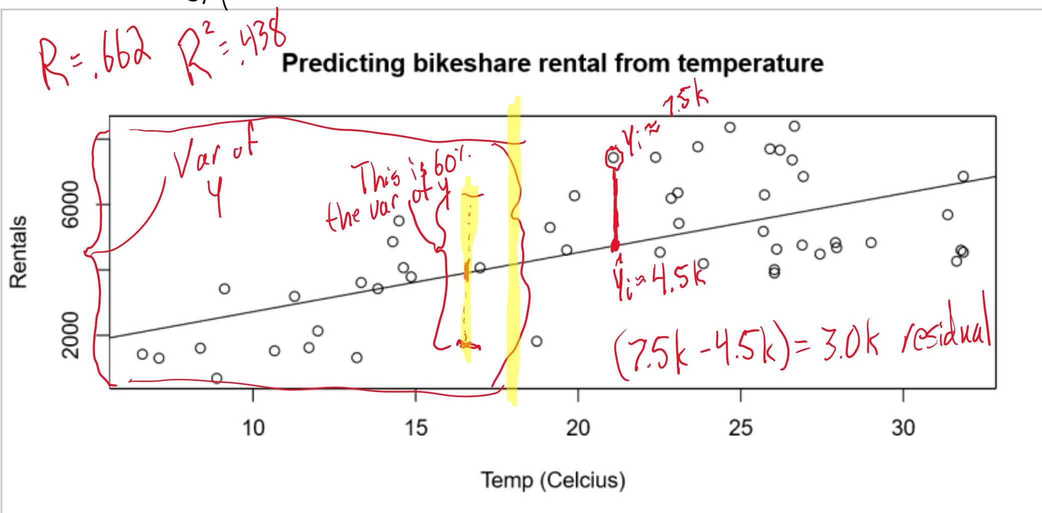
X, Y we want to predict Y from X.

$$R = \text{Correlation Coefficient} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{S_x}\right) \times \left(\frac{Y_i - \bar{Y}}{S_y}\right)$$

$$= \sum_{i=1}^{n} Z_{x_i} \times Z_{y_i}$$

43.8% of var of y is explained by X.





Residual is the difference between the observed value Y_i and $\hat{Y}_i = \hat{b}_0 + \hat{b}_1 \cdot X_i$

$(Y_i - \hat{Y}_i)$

The line of best fit $\left(\hat{b}_1 = \frac{S_y}{S_x} \cdot R,\ \hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}\right)$ minimizes the sum of squared residuals $\left(\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2\right) = \sum_{i=1}^{n} (Y_i - (\hat{b}_0 + \hat{b}_1 X_i))$

This is our best guess informed by X_i.

$R^2 = R^2$ Square of the correlation coefficient

$$= 1 - \frac{RSS}{TSS}$$

RSS: Residual sum of squares

TSS: Total sum of squares (for y) $\sum_{i=1}^{n} (Y_i - \bar{Y})^2$

This is our best guess of Y_i for X_i unknown

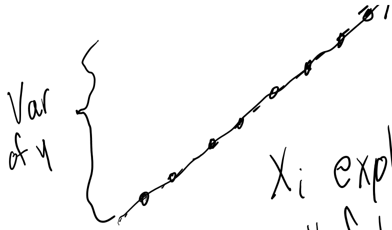
i.e. not using regression

This tells us what percentage of variability is still unexplained.

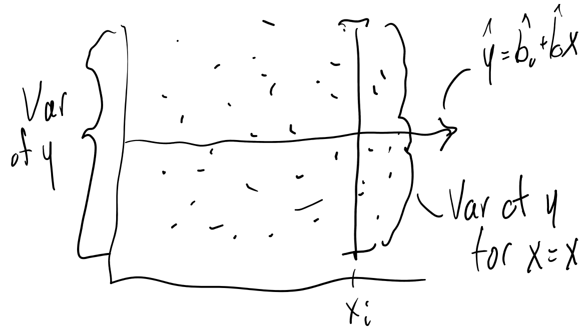
If $RSS = 0$ $R^2 = 100\%$

If $RSS = TSS$ $R^2 = 0\%$

$\hat{y} = \hat{b}_0 + \hat{b}x$



X_i explain 100% of the variability of y_i.



R^2 tells us what percentage of the variability in y is explained by X.