R Squared

Pritam Prakash Shete
Computer Division, BARC
Centre for Excellence in Basic Sciences

Topics

- Interpretation
- Equations
- Examples

R^2 – Interpretation

- Coefficient of determination
- Coefficient of multiple determination
- Strength of relationship
- Value between 0.0 – 1.0
- Percentage value
- Independent variable explains p percent of variation in dependent variable
- Independent variable reduces p percent of variation in dependent variable

R^2 – Equation

$$\mathrm{Var}(v) = \frac{1}{n_x} \times \mathrm{SS}(v) \qquad R^2 = \frac{\mathrm{Var}(\mathrm{mean}) - \mathrm{Var}(\mathrm{fit})}{\mathrm{Var}(\mathrm{mean})}$$

- SS – Sum of Squared differences
- Var – Variation around variable
- n_x – Number of samples
- Var(mean) – Variation around mean
- Var(fit) – Variation around fit

R^2 – Equation

$$R^2 = \frac{SS(\text{mean}) - SS(\text{fit})}{SS(\text{mean})}$$

- SS – Sum of squared differences
- SS(mean) – Sum of squared differences around mean
- SS(fit) – Sum of squared differences around fit

R^2 – Example

x	y
75	167
35	95
0	32
-35	-31
-75	-103

- Celcius to Farenheit
- n_x – 5 samples
- Mean value – 32
- Var(mean) – 8877.6

$$\frac{1}{5}\left(167+95+32-31-103\right)=32$$

$$\frac{1}{5}\begin{pmatrix}(167-32)^2+(95-32)^2\\+(32-32)^2+(-31-32)^2\\+(-103-32)^2\end{pmatrix}=8877.6$$

R^2 – Example

x	y	ŷ
75	167	167
35	95	95
0	32	32
-35	-31	-31
-75	-103	-103

$$\frac{1}{5}\left(\begin{array}{l}(167-167)^2+(95-95)^2\\+(32-32)^2+(-31+31)^2\\+(-103+103)^2\end{array}\right)=0$$

$$\frac{8877.6-0}{8877.6}=1$$

- Celcius to Farenheit
- W_0 – 32
- W_1 – 1.8
- Var(mean) – 8877.6
- Var(fit) – 0
- R^2 – 100%

R^2 – Example

x	y	ŷ
75	167	142.5
35	95	82.5
0	32	30
-35	-31	-22.5
-75	-103	-82.5

$$\frac{1}{5}\left(\begin{array}{l}(167-142.5)^2+(95-82.5)^2\\+(32-30)^2+(-31+22.5)^2\\+(-103+82.5)^2\end{array}\right)=250.6$$

$$\frac{8877.6-250.6}{8877.6}=0.9717$$

- Celcius to Farenheit
- W_0 – 30
- W_1 – 1.5
- Var(mean) – 8877.6
- Var(fit) – 250.6
- R^2 – 97.17%

Adjusted R^2

- R^2
  - Increase independent variables – Increase R^2
  - Increase independent variables – Constant R^2
- Adjusted R^2
  - Increase independent variables (then)
  - Increase model accuracy (then only)
  - Increase adjusted R2

Adjusted R^2

$$\text{Adjusted } R^2 = 1 - \frac{\left(1 - R^2\right)\left(N - 1\right)}{N - p - 1}$$

- N – Number of samples
- p – Number of independent variables
- Degree of freedom for model

Questions?

Thank you