Example: linear regression on dataset (gambler win/loss)

Pretend you tracked a gambler's net win/loss across 8 sessions. Shocking result: sometimes they win, sometimes they cry.

Data:

session (x)	net \$ win/loss (y)
1	-5
2	-2
3	0
4	3
5	1
6	4
7	6
8	5

Step 1 — sample means (calmly)

Formula if your wondering:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$$

or

$$\mu = \frac{\sum x_i}{n}$$

Sum of x is $1 + 2 + \cdots + 8 = 36$, so

$$\bar{x} = \frac{36}{8} = 4.5.$$

Sum of y is $-5 - 2 + 0 + 3 + 1 + 4 + 6 + 5 = 12$, so

$$\bar{y} = \frac{12}{8} = 1.5.$$

Step 2 — slope formula (the point where math does the heavy lifting)

Slope β_1 is

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Compute numerator and denominator explicitly (yes, we actually compute):

- Numerator: $\sum(x_i - \bar{x})(y_i - \bar{y}) = 60$.
- Denominator: $\sum(x_i - \bar{x})^2 = 42$.

Plugging:

$$\beta_1 = \frac{(1-4.5)(-5-1.5)+(2-4.5)(-2-1.5)+(3-4.5)(0-1.5)+(4-4.5)(3-1.5)+(5-4.5)(1-1.5)+}{(1-4.5)^2+(2-4.5)^2+(3-4.5)^2+(4-4.5)^2+(5-4.5)^2+(6-4.5)^2}$$

So

$$\beta_1 = \frac{60}{42} = \frac{10}{7} \approx 1.428571\ldots$$

Step 3 — intercept

Intercept β_0 is

$$\beta_0 = \bar{y} - \beta_1\bar{x}$$
$$\beta_0 = 1.5 - \frac{10}{7}(4.5).$$

If u prefer decimals:

$$\beta_0 = 1.5 - 1.428571(4.5).$$

Numerically,

$$\beta_0 = 1.5 - \frac{45}{7} = \frac{21}{14} - \frac{90}{14} = -\frac{69}{14} \approx -4.9285695\ldots$$

Final fitted line (the glorious lie)

$$\hat{y} = \beta_0 + \beta_1 x = -\frac{69}{14} + \frac{10}{7}x.$$

If decimals are more comforting:

$$\hat{y} \approx -4.9285695 + 1.428571x.$$

Predictions (so you can point at numbers and feel smart)

For each session x the predicted $\hat{y}$ is:

x	$\hat{y}$ (approx)
1	-3.500000
2	-2.071429
3	-0.642857
4	0.785714

x	$\hat{y}$ (approx)
5	2.214286
6	3.642857
7	5.071429
8	6.500000

How wrong are we? (Residual Sum of Squares)

Residuals $r_i = y_i - \hat{y}_i$. The Residual Sum of Squares is

$$\text{RSS} = \sum_{i=1}^{n} r_i^2 \approx 12.285714.$$

Total Sum of Squares (variation in y):

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 98.0.$$

Coefficient of determination (aka R^2, the smugness metric):

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} \approx 1 - \frac{12.285714}{98.0} \approx 0.8746.$$

Interpretation: about 87.5% of the variance in the gambler's net win/loss is explained by the session index in this toy dataset. Which is to say: session number kind of predicts performance here — or the gambler's behavior followed a trend for these eight sessions. Or it was a fluke. Your call.

R snippet to reproduce (because code is honest)

```r
x <- c(1,2,3,4,5,6,7,8) y <- c(-5,-2,0,3,1,4,6,5) model <- lm(y ~ x) summary(model) # coefficients will match the numbers shown above
```