The Simple Linear Regression Model

GR 5205 / GU 4205 Columbia University
Section 2/ Section 3 Xiaofei Shi

Regression Analysis

- Regression

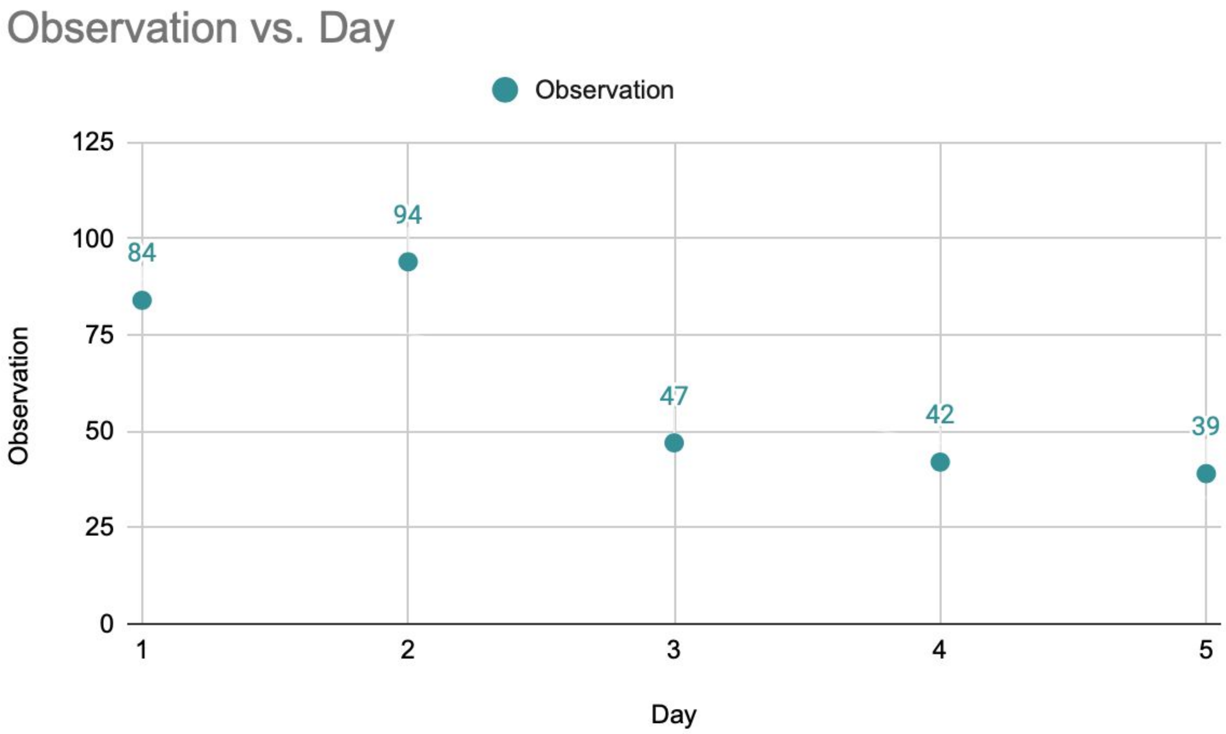
  Statistical method to study dependencies between variables in the presence of noise.

- Linear Regression

  Statistical method to study linear dependencies between variables in the presence of noise.

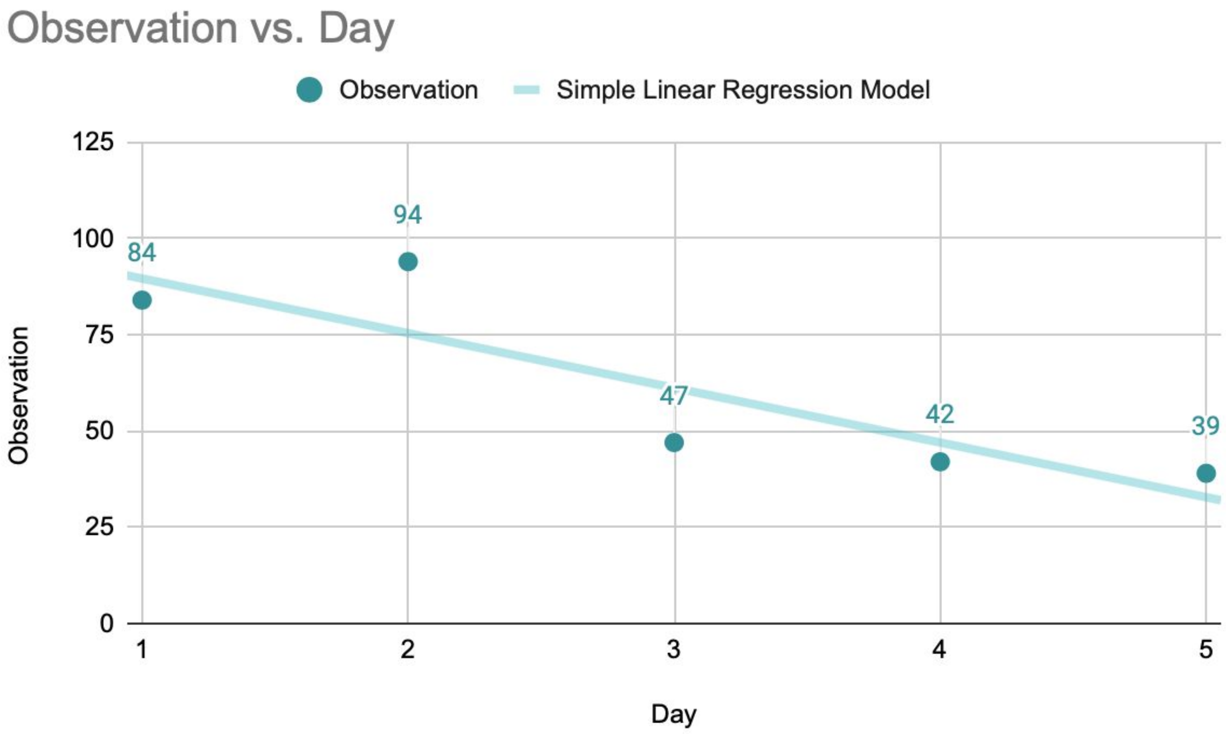
Example

Day	Observation
1	84
2	94
3	47
4	42
5	39



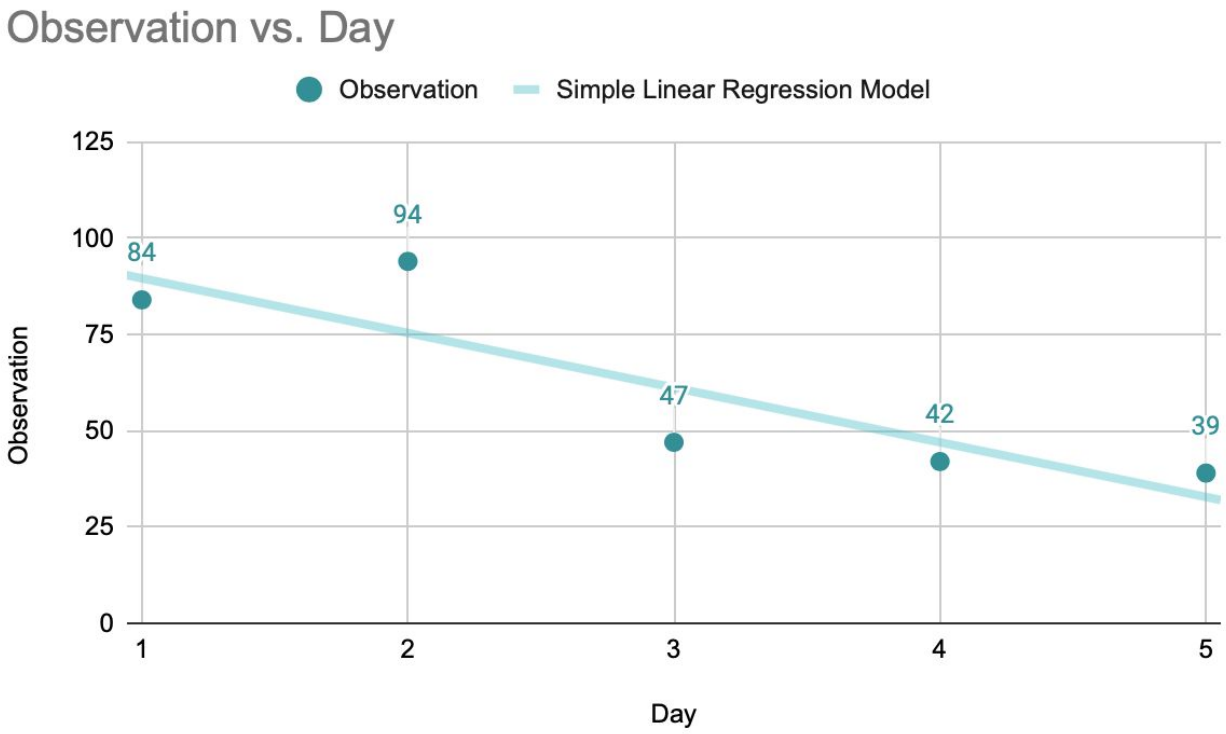
Example

Day	Observation
1	84
2	94
3	47
4	42
5	39

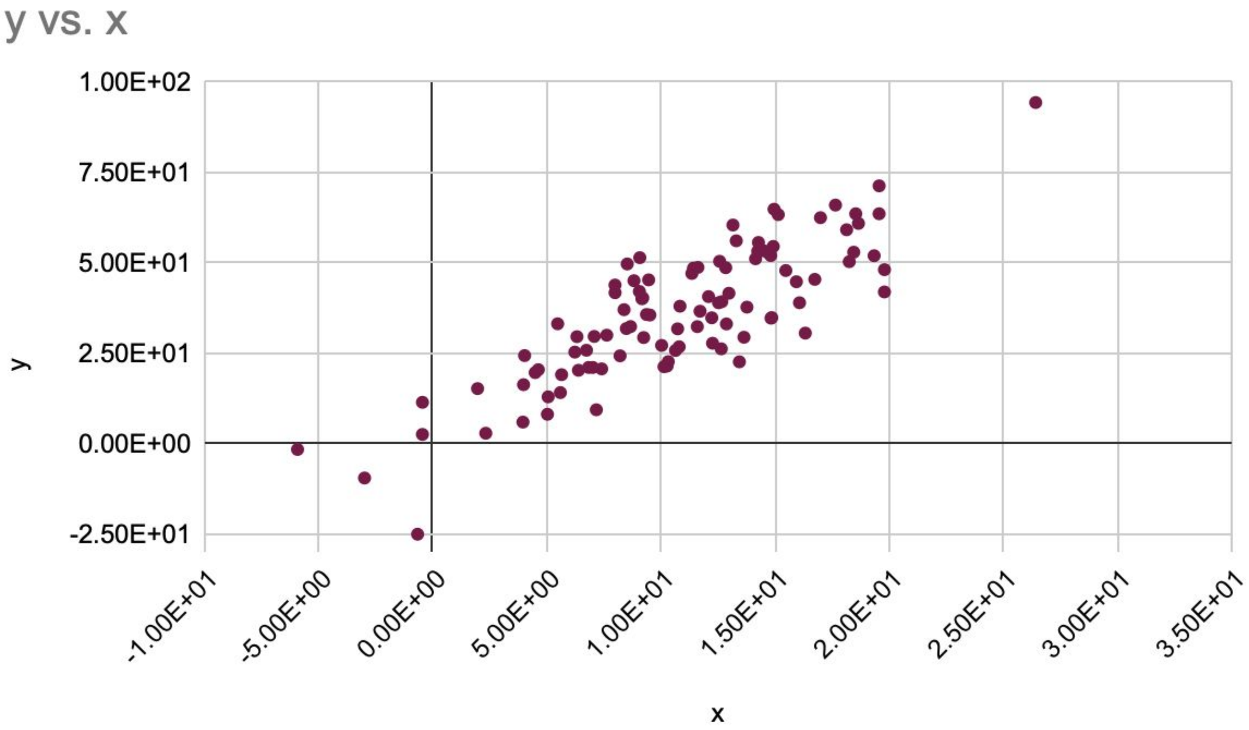


Example

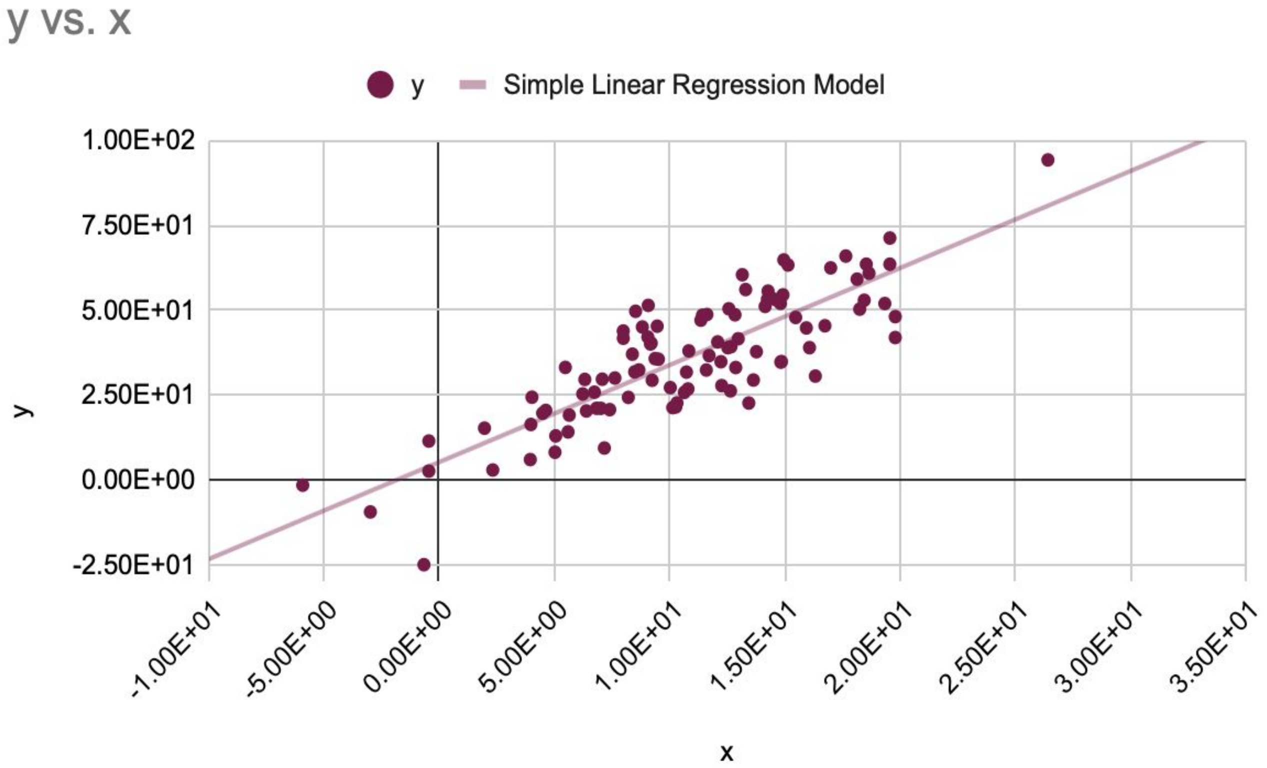
Day	Observation
1	84
2	94
3	47
4	42
5	39



Another example...



Another example…



(Simple) linear regression procedures:

X - predictor (random) variable **Y - response random variable**

- Build your model:

  1) relationship: $Y = \beta_0 + X\beta_1 + \epsilon \Rightarrow Y_i = \beta_0 + X_i\beta_1 + \epsilon_i$

  2) preference: choose $\beta_0, \ \beta_1$ to minimize $\mathbb{E}\left[(Y - \beta_0 - X\beta_1)^2\right]$

- Estimate your model parameters: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

  1) using observed data to express your preference: $\min_{\beta_0, \beta_1} Q := \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$

  2) get parameters estimation for your model: Today's mission!

- Understand your model:

  1) properties of estimations: HWK 1!

  2) predictions

(Simple) linear regression procedures:

X - predictor (random) variable **Y - response random variable**

- Build your model:

  1) relationship:

  2) preference:

- Estimate your model parameters: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

  1) using observed data to express your preference:

  2) get parameters estimation for your model:

- Understand your model:

  1) properties of estimations:

  2) predictions

A theoretical example: bivariate normal distribution

$$(X, Y) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

A theoretical example: bivariate normal distribution

$$(X,Y) \sim \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2\end{bmatrix}\right)$$

- Useful Trick: normalization!

A theoretical example: bivariate normal distribution

$$(X,Y) \sim \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2\end{bmatrix}\right)$$

- Conditional expectation $\mathbb{E}[Y|X] = \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)+\mu_Y$
- Conditional variance $\mathrm{Var}[Y|X] = (1-\rho^2)\sigma_Y^2$
- Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ be i.i.d samples from (X,Y)
- SLR model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

$$\Rightarrow \quad \beta_0 = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X \qquad \beta_1 = \rho\frac{\sigma_Y}{\sigma_X}$$

$$\epsilon_i \sim \mathcal{N}(0,\sigma^2) \qquad \sigma^2 = (1-\rho^2)\sigma_Y^2$$

UMVUL for Bivariate Normal

$$\beta_1 = \rho \frac{\sigma_Y}{\sigma_X} \qquad \Rightarrow \widehat{\beta}_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}[X]}$$

$$\beta_0 = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X \quad \Rightarrow \widehat{\beta}_0 = Y - X\widehat{\beta}_1$$

- Uniformly: convergens property as sample size going to infinity
- Unbiased:
- Linear: with respect to response variable
- Minimum Variance:
- Rigorous proof will be shown later in this course

Adding data

$$\beta_1 = \rho \frac{\sigma_Y}{\sigma_X} \quad \Rightarrow \widehat{\beta}_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}\left[X\right]}$$

$$\beta_0 = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X \quad \Rightarrow \widehat{\beta}_0 = Y - X\widehat{\beta}_1$$

- Recall the estimation for μ_X, μ_Y

  using data: $\hat{\mu}_X = \bar{x}, \hat{\mu}_Y = \bar{y}$

- And the estimation for σ_X, σ_Y

  using data: $\sigma_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \; \sigma_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

  in matrix form: $\widehat{\sigma}_X^2 = \frac{1}{n-1} \|x - \bar{x}\mathbf{1}_n\|^2, \; \widehat{\sigma}_Y^2 = \frac{1}{n-1} \|y - \bar{y}\mathbf{1}_n\|^2$

- As well as for $\mathrm{Cov}(X,Y) \Rightarrow \widehat{\mathrm{Cov}}(X,Y) = \frac{1}{n-1} (x - \bar{x}\mathbf{1}_n)^\top (y - \bar{y}\mathbf{1}_n)$

$$\rho = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \Rightarrow \hat{\rho} = \frac{\widehat{\mathrm{Cov}}(X,Y)}{\widehat{\sigma}_X \widehat{\sigma}_Y}$$

In other words...

$$\beta_1 = \rho \frac{\sigma_Y}{\sigma_X} \Rightarrow \widehat{\beta}_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}[X]} \Rightarrow b_1 = \frac{(x - \bar{x}1_n)^\top (y - \bar{y}1_n)}{\|x - \bar{x}1_n\|^2}$$

$$\beta_0 = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X \Rightarrow \widehat{\beta}_0 = Y - X\widehat{\beta}_1 \Rightarrow b_0 = \bar{y} - \bar{x} b_1$$

- Recall the estimation for μ_X, μ_Y

  using data: $\hat{\mu}_X = \bar{x}, \hat{\mu}_Y = \bar{y}$

- And the estimation for σ_X, σ_Y

  using data: $\sigma_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \ \sigma_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$

  in matrix form: $\widehat{\sigma}_X^2 = \frac{1}{n-1} \|x - \bar{x}1_n\|^2, \ \widehat{\sigma}_Y^2 = \frac{1}{n-1} \|y - \bar{y}1_n\|^2$

- As well as for $\mathrm{Cov}(X,Y) \Rightarrow \widehat{\mathrm{Cov}}(X,Y) = \frac{1}{n-1} (x - \bar{x}1_n)^\top (y - \bar{y}1_n)$

$$\rho = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \Rightarrow \hat{\rho} = \frac{\widehat{\mathrm{Cov}}(X,Y)}{\widehat{\sigma}_X \widehat{\sigma}_Y}$$

The Simple Linear Regression Model

More general case...

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be samples from the same model
- If the SLR model holds, we write $Y_i = \beta_0 + X_i\beta_1 + \epsilon_i\,,$
- Here, ϵ_i satisfies $\mathbb{E}[\epsilon_i] = 0$ and $\mathbb{E}[\epsilon_i\epsilon_j] = \sigma^2\delta_{ij}$
- Observations: predictor : $x_1, x_2, \ldots, x_n$ response :$y_1, y_2, \ldots, y_n$
- Preference: $Q = \sum_{i=1}^{n}(y_i - \beta_0 - x_i\beta_1)^2$
- Model parameters: $\beta_0, \beta_1(, \sigma^2)$

General Methodology

- Preference + data $\Longrightarrow$ Q = Q(model parameters; data)
- Estimation of model parameters $\Longleftrightarrow$ Minimizing Q wrt model parameters

$\Longrightarrow$ Taking partial derivatives of Q wrt model parameters and set them to 0!

Least Square Estimator

$$b_1 = \frac{(x - \bar{x}1_n)^\top (y - \bar{y}1_n)}{\|x - \bar{x}1_n\|^2}$$

$$b_0 = \bar{y} - \bar{x}b_1$$

Prediction and residual

$$b_1 = \frac{(x - \bar{x}1_n)^\top (y - \bar{y}1_n)}{\|x - \bar{x}1_n\|^2} \qquad b_0 = \bar{y} - \bar{x} b_1$$

- Prediction: $\hat{y}_i = b_0 + x_i b_1$
- Residual: $e_i = y_i - \hat{y}_i = y_i - b_0 - x_i b_1$
- Residual can be viewed as the estimation of unobservable error terms

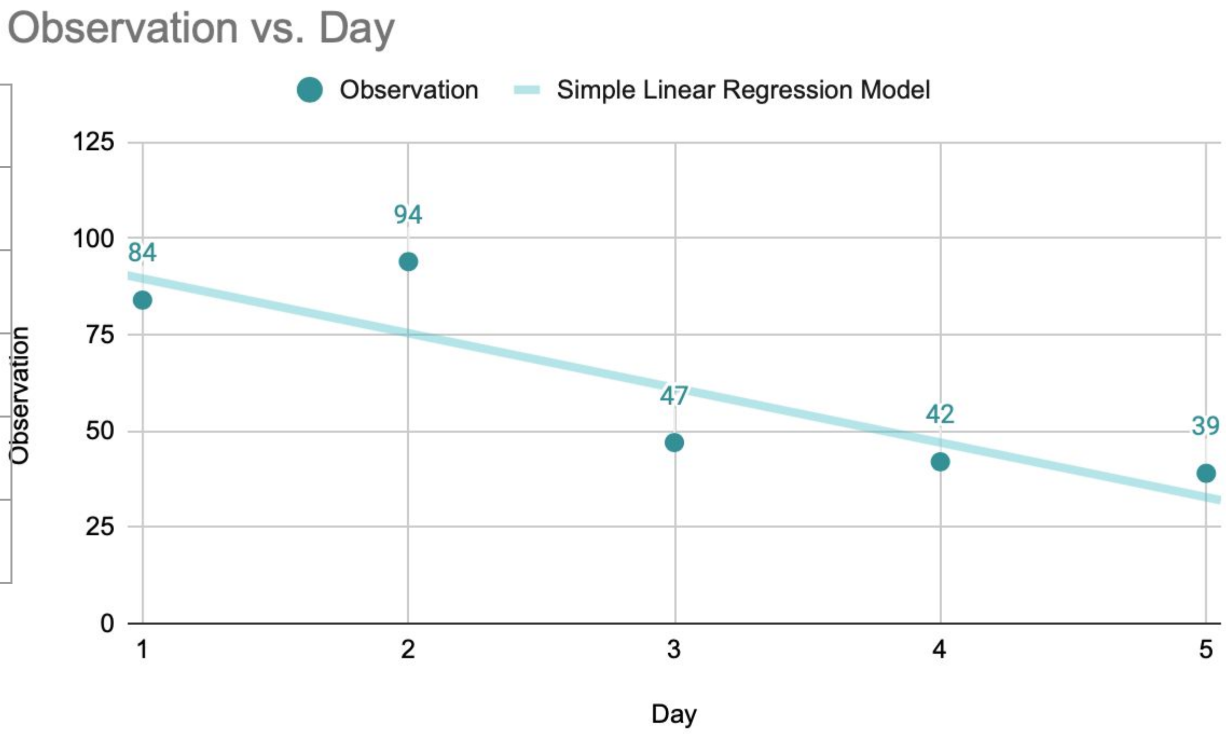
$$\hat{\epsilon}_i = e_i = y_i - \hat{y}_i = y_i - b_0 - x_i b_1$$

- Estimation of $\widehat{\sigma}^2 = \mathrm{MSE} = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} = \frac{\|y - \hat{y}\|^2}{n-2}$

n-2?

More about residuals...

Day	Observation	Residual
1	84	-5.6
2	94	18.6
3	47	-14.2
4	42	-5
5	39	6.2



Properties of the line: $y = b_0 + xb_1$

$$b_1 = \frac{(x-\bar{x}1_n)^\top (y-\bar{y}1_n)}{\|x-\bar{x}}$$

$$b_0 = \bar{y} - \bar{x}b_1$$