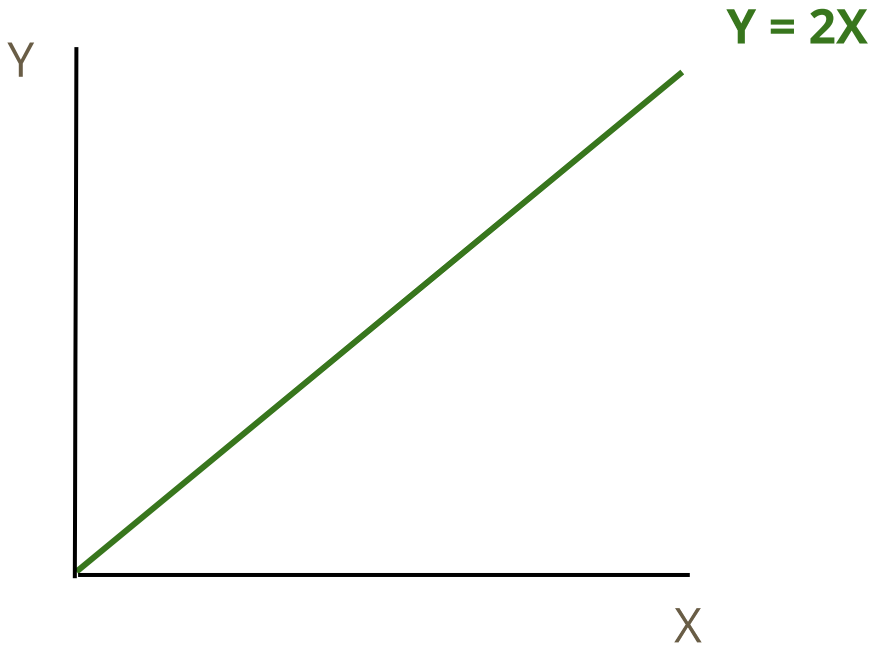
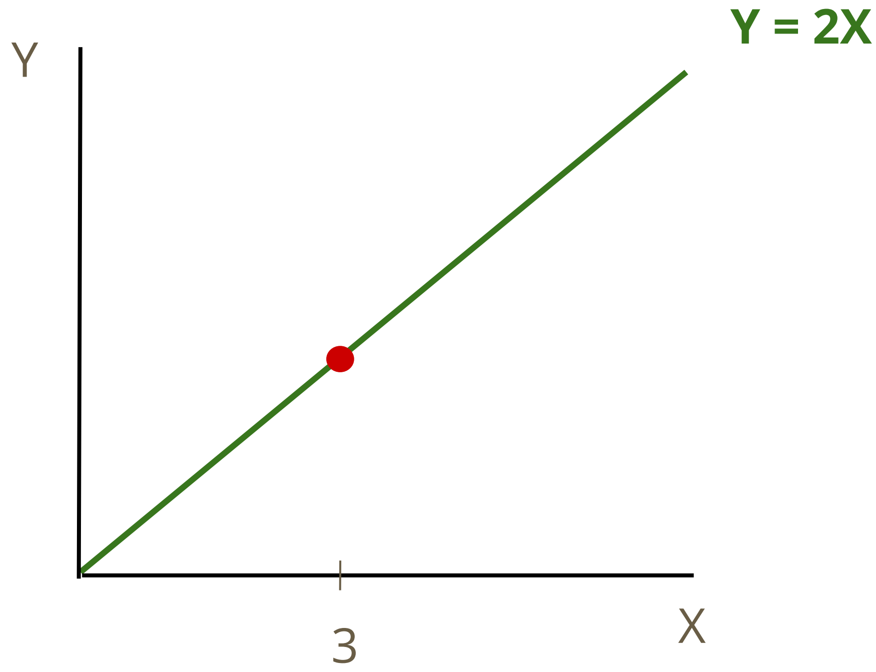
Linear Model Evaluation

Boston University CS 506 - Lance Galletti

Linear Function

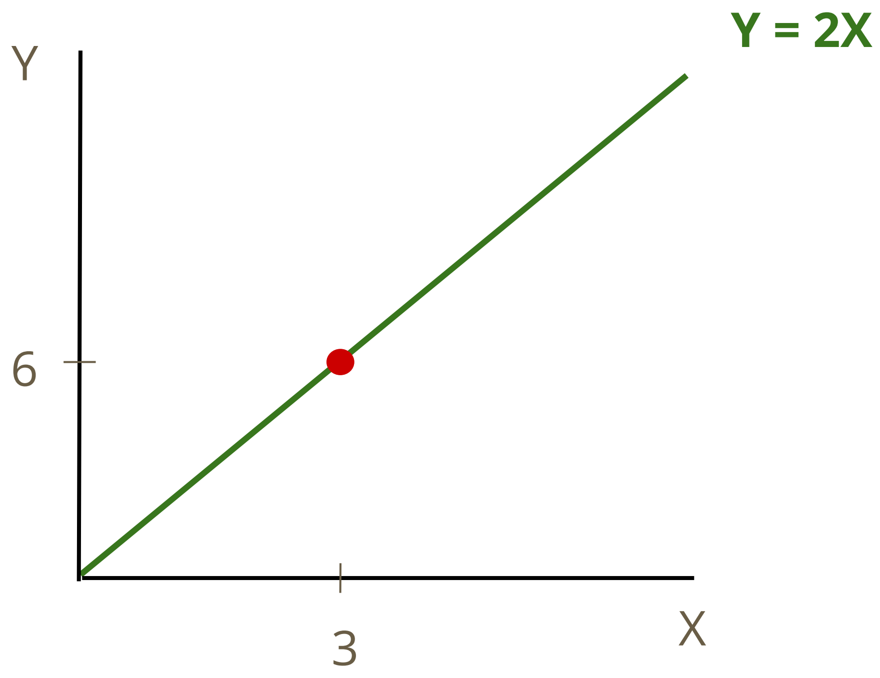


Linear Function

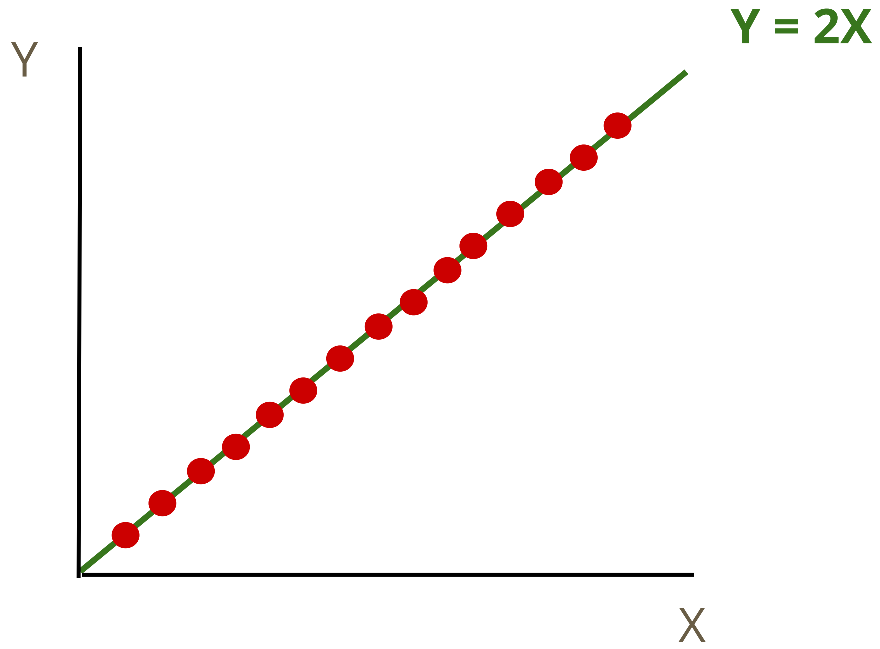


Linear Function

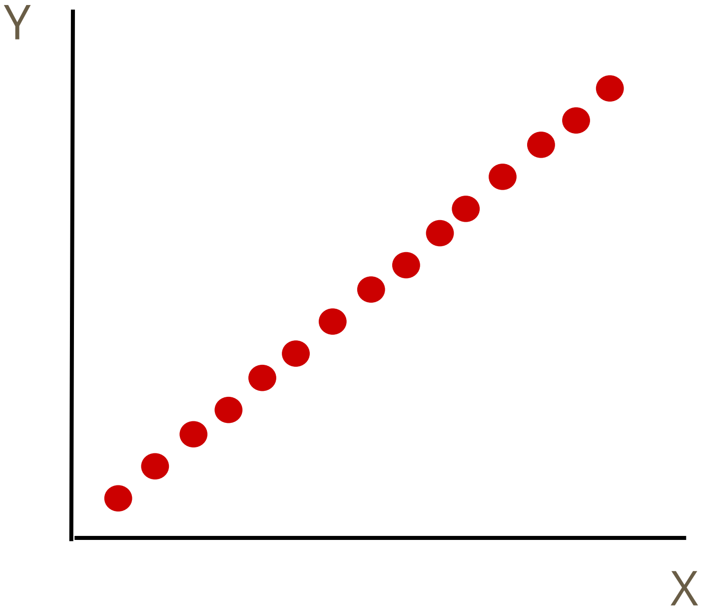
I know for sure what value of Y I'm gonna get



Linear Function



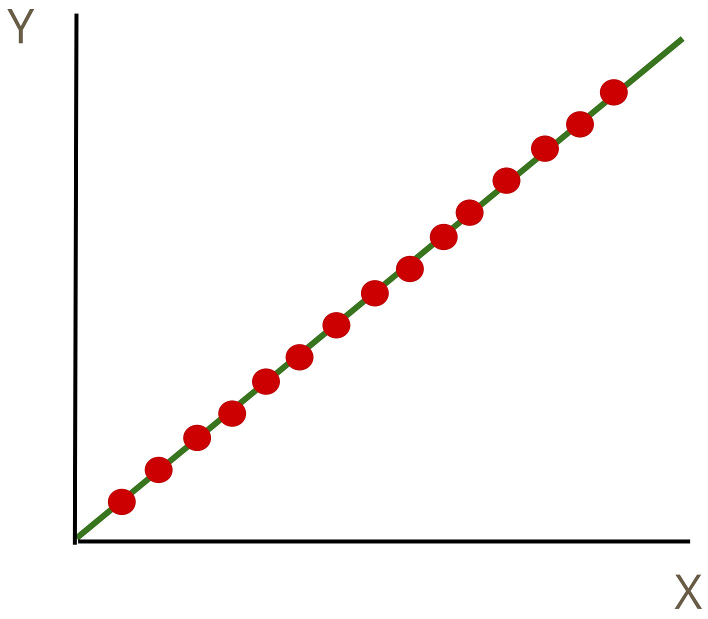
Ideally



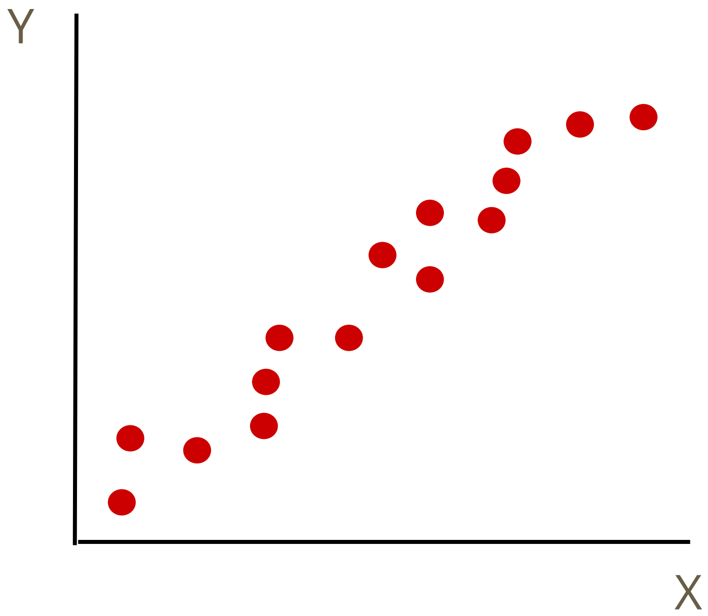
Ideally

Guess the relationship
(**a** and **b**) from the data

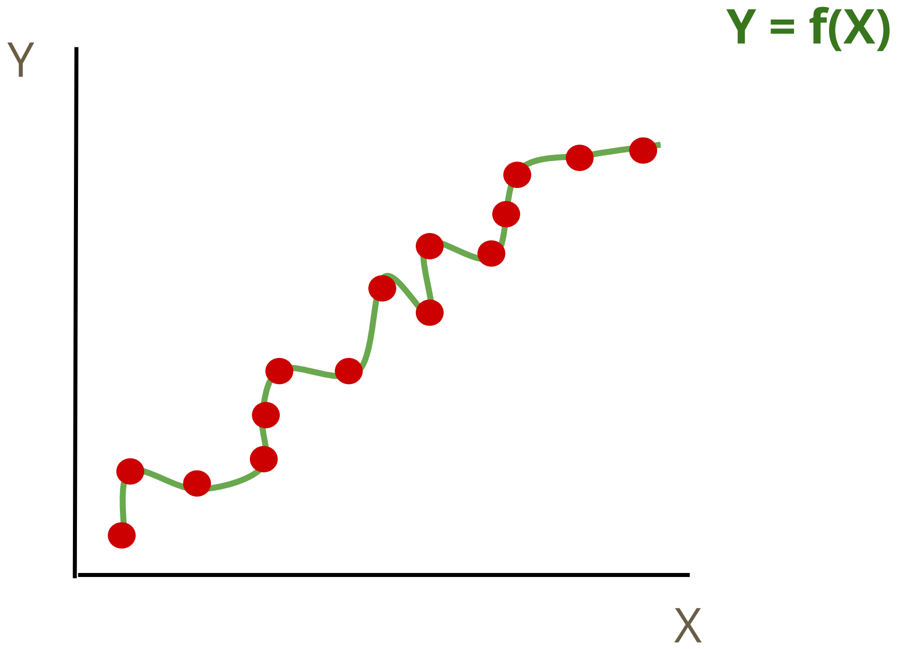
$Y = aX + b$



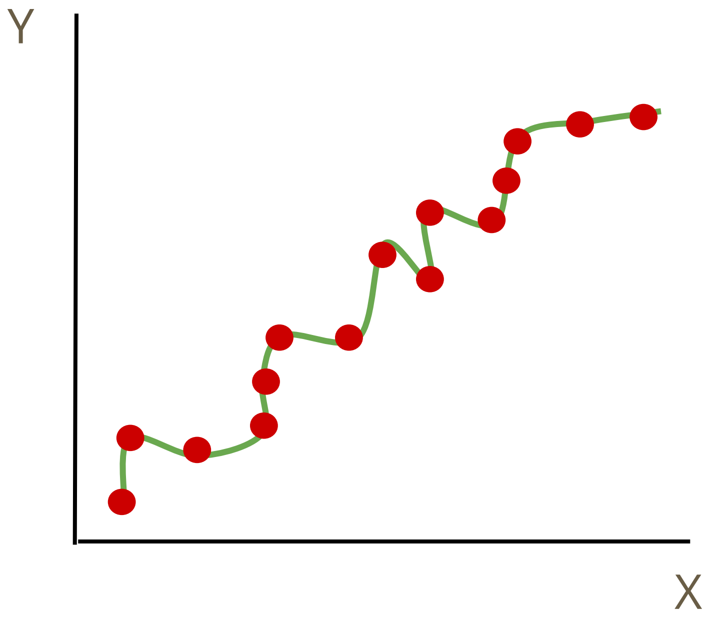
Practically



Practically



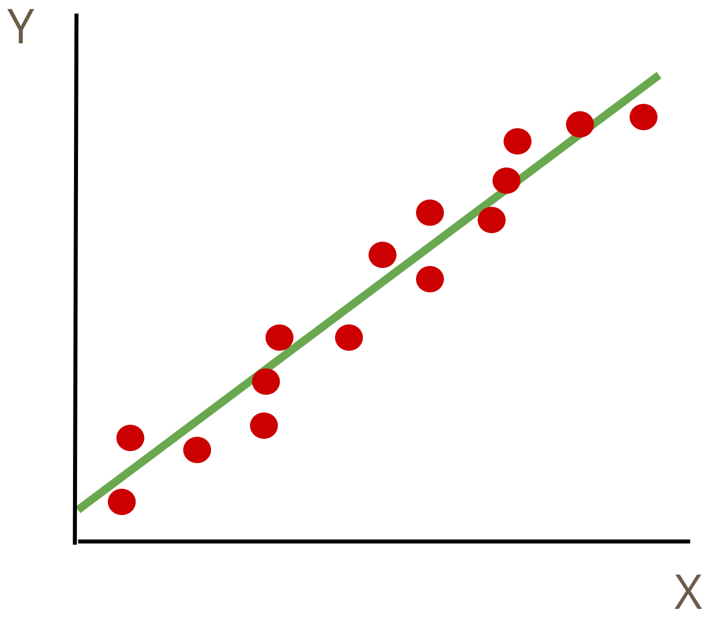
Practically



Y = f(X)

f can be anything: too complex a problem to solve…

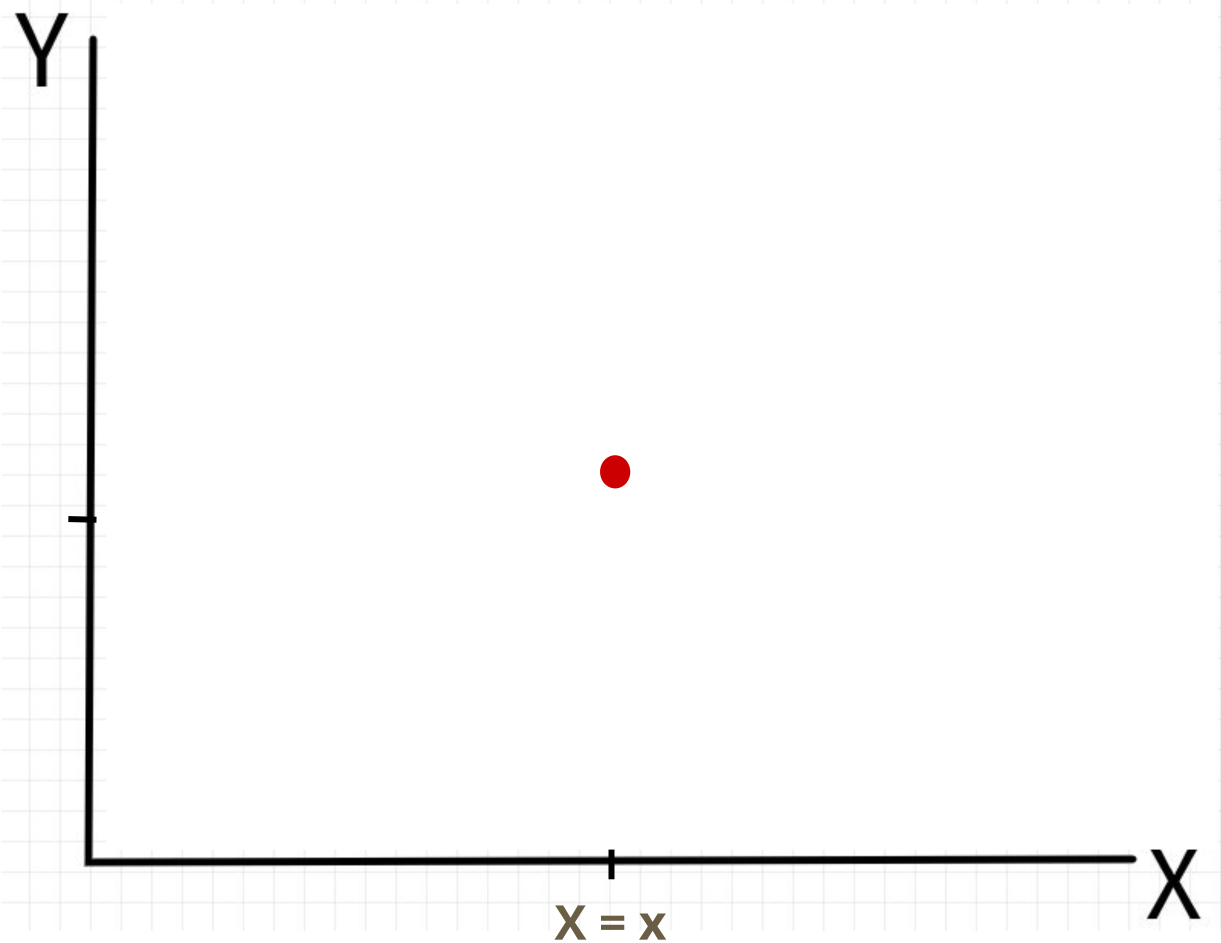
Practically



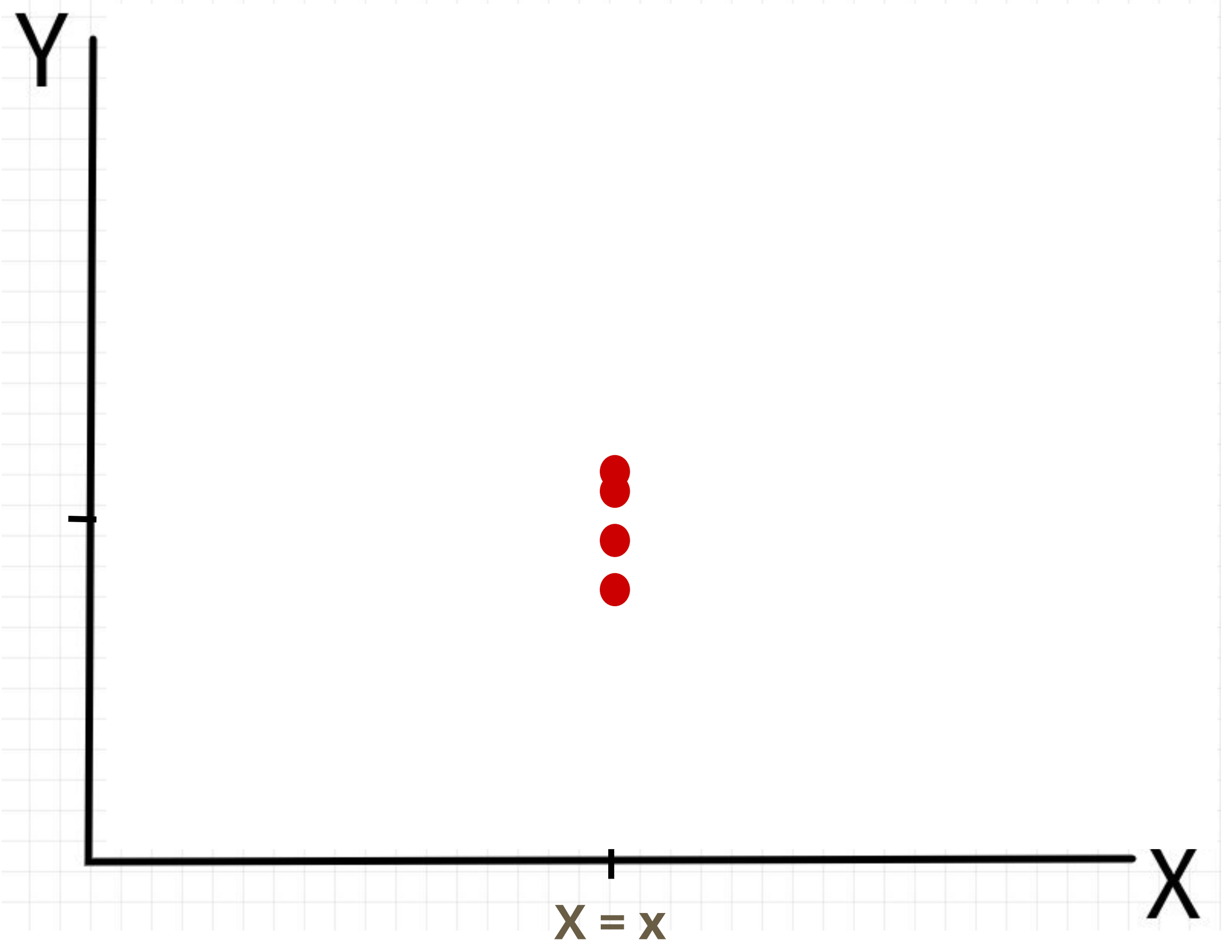
Y = f(X) + noise

Assume f is linear and the variation we're seeing is noise

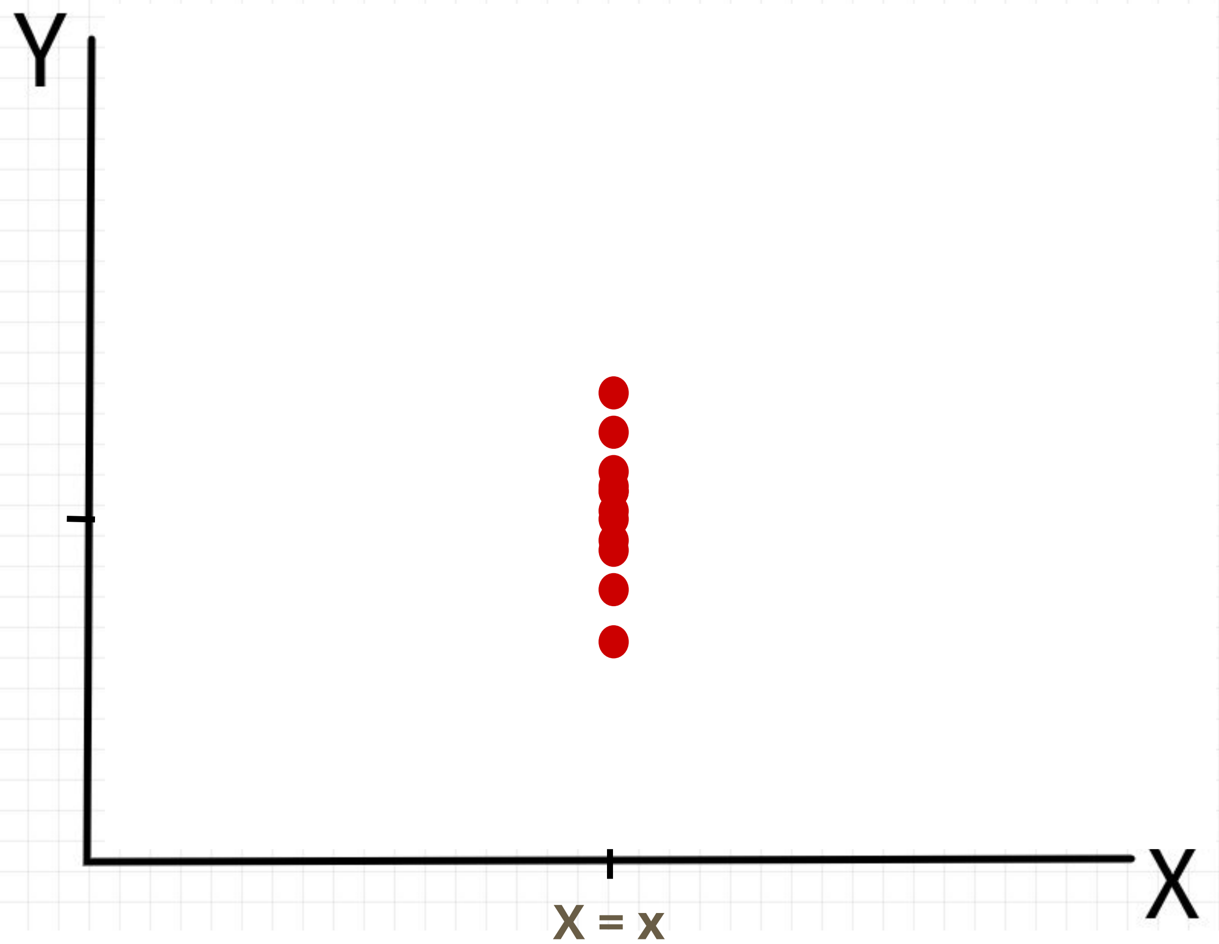
Assumptions



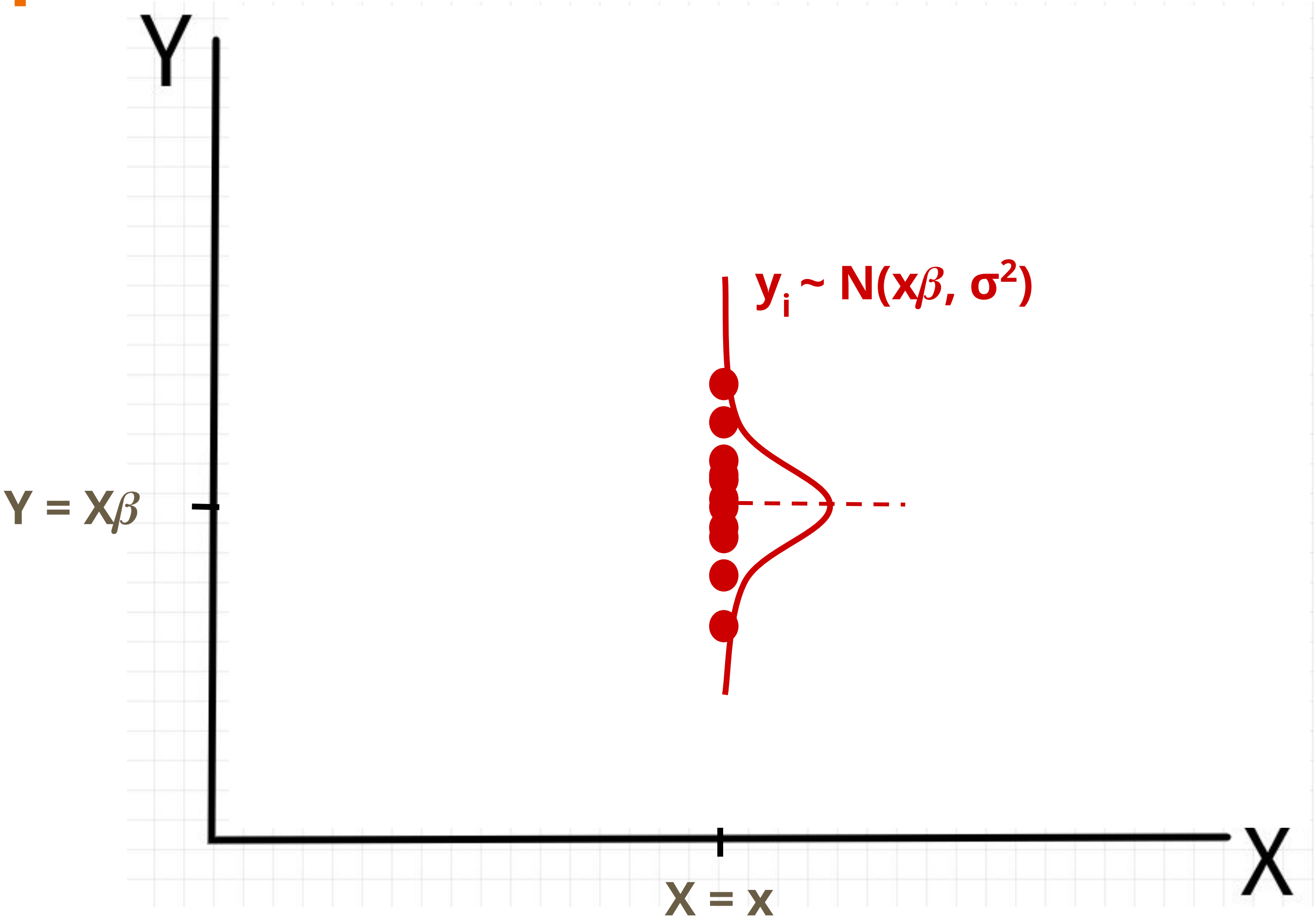
Assumptions



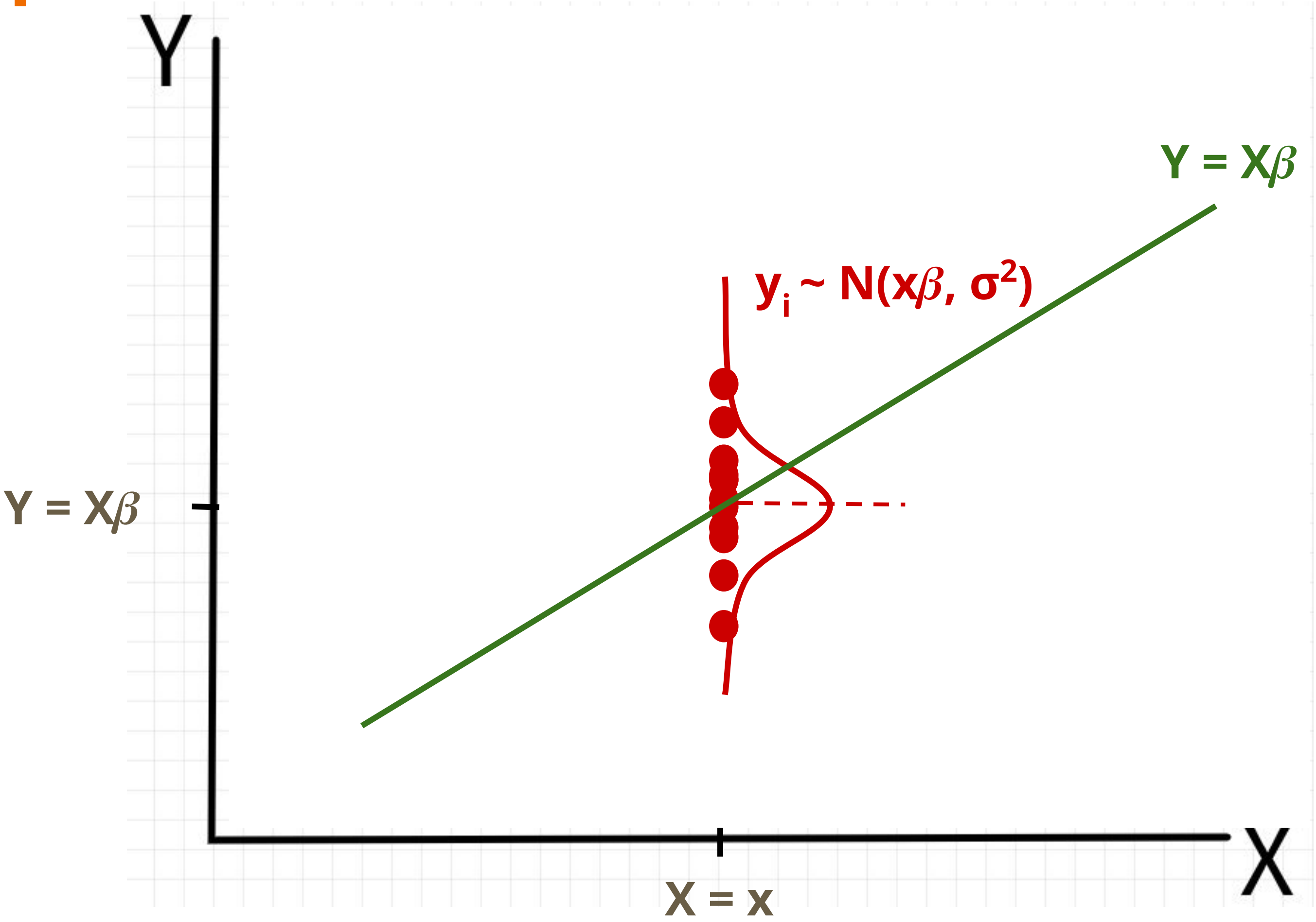
Assumptions



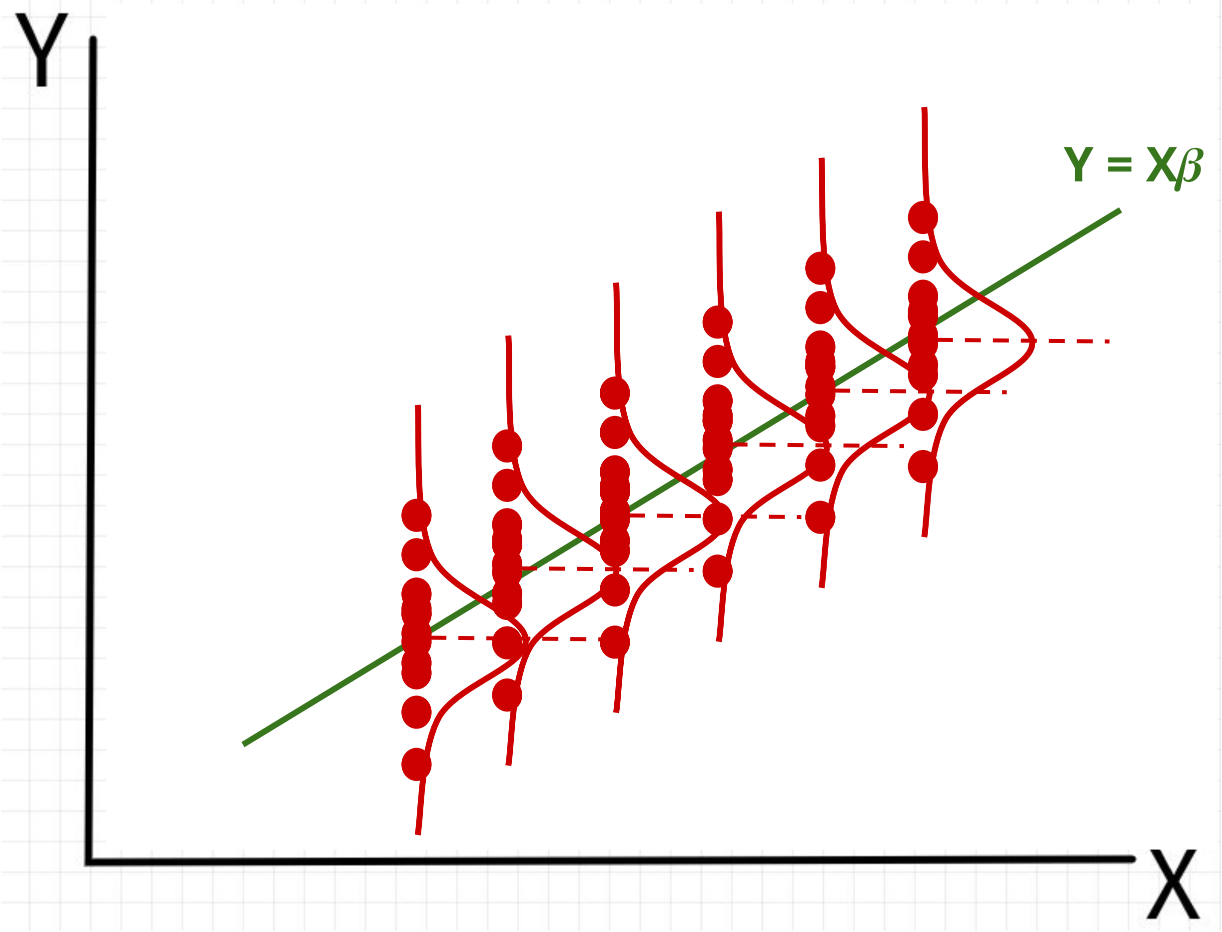
Assumptions



Assumptions

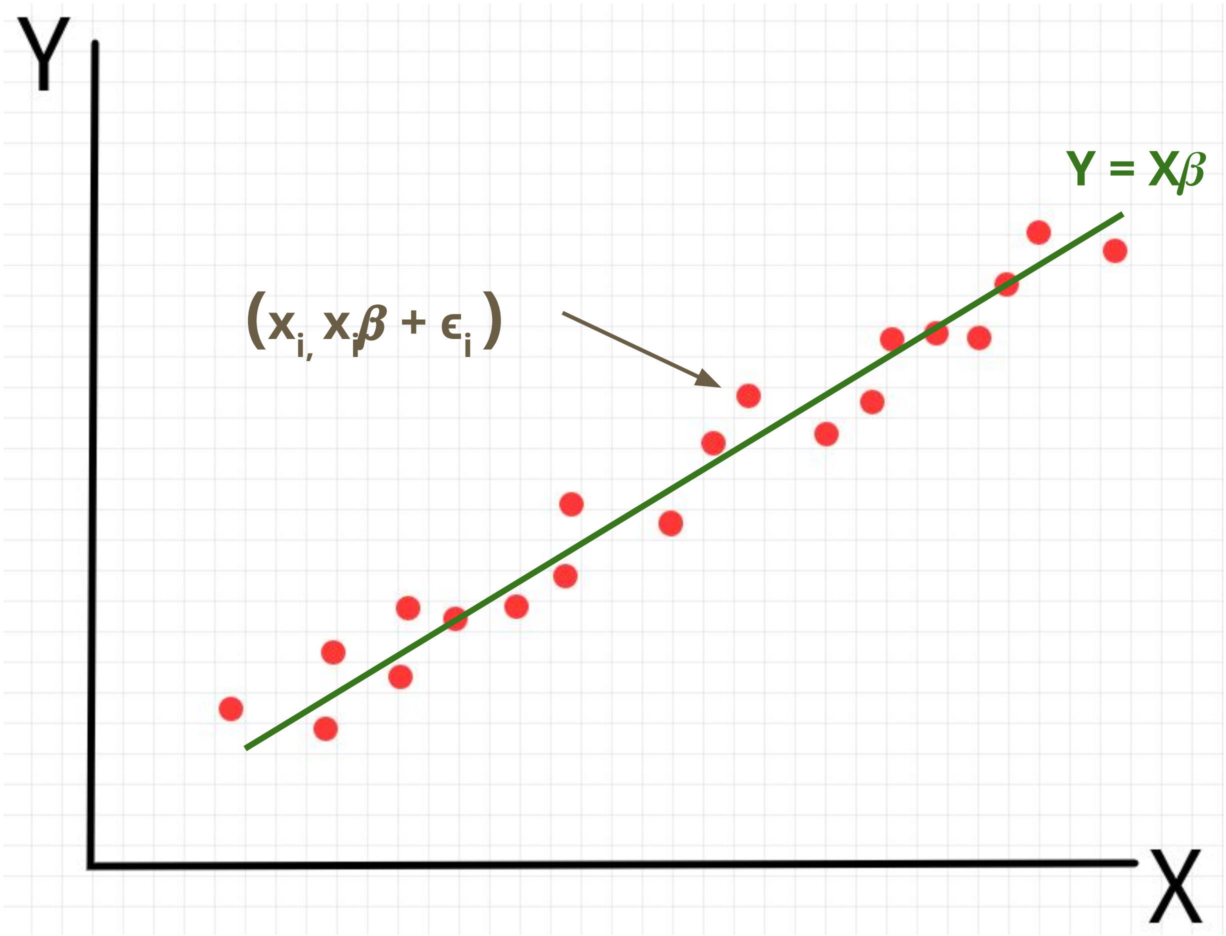


Assumptions

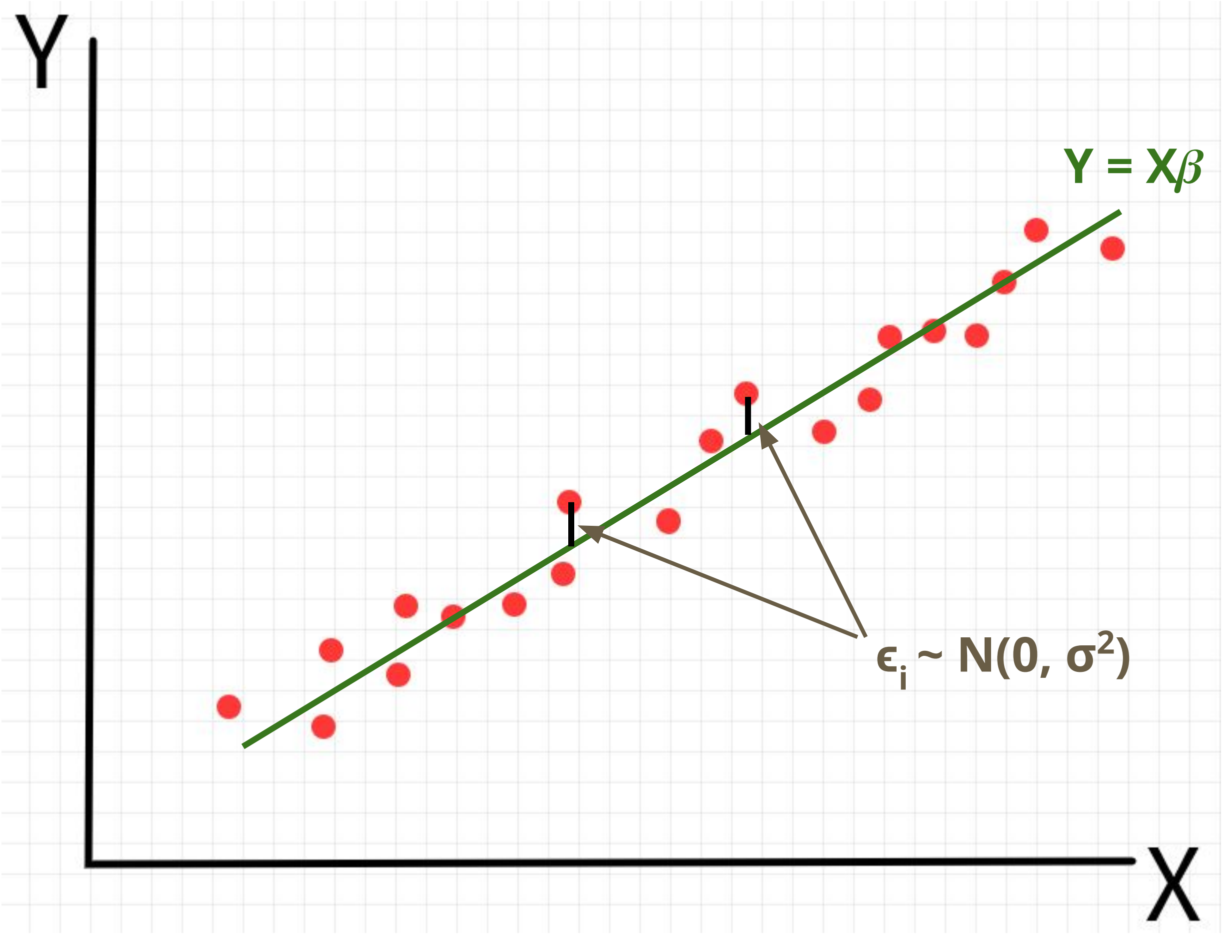


Assumptions

Assumptions



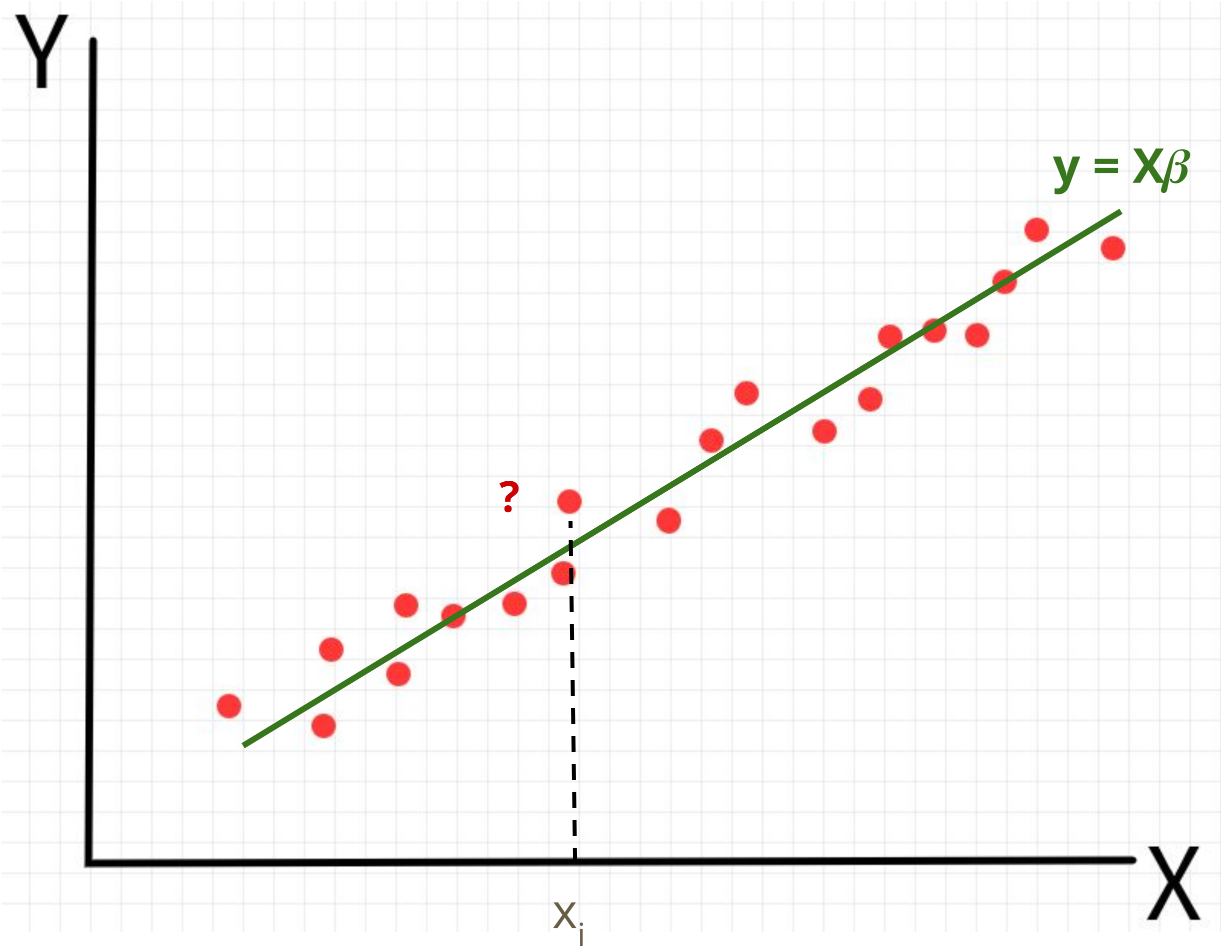
Assumptions

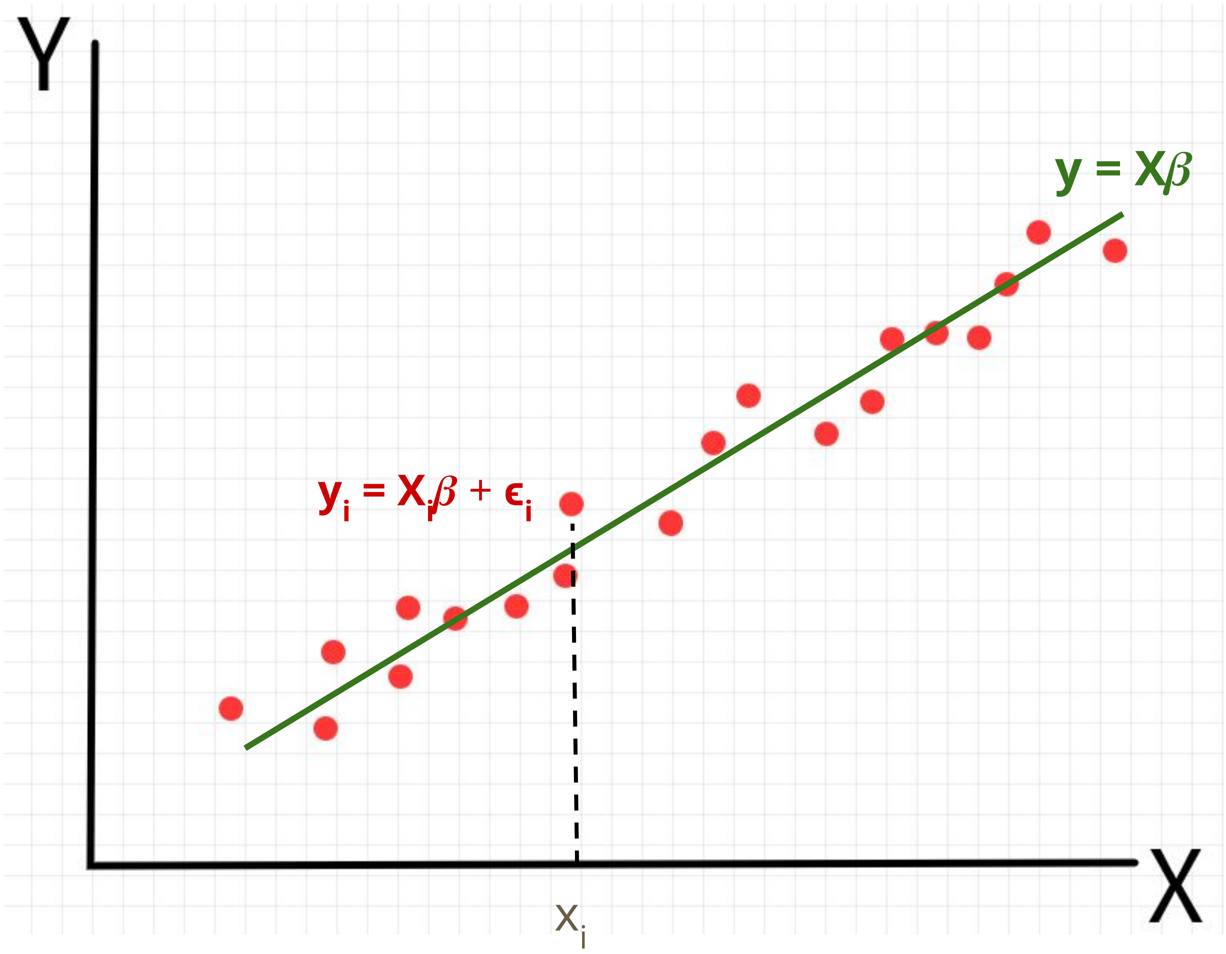


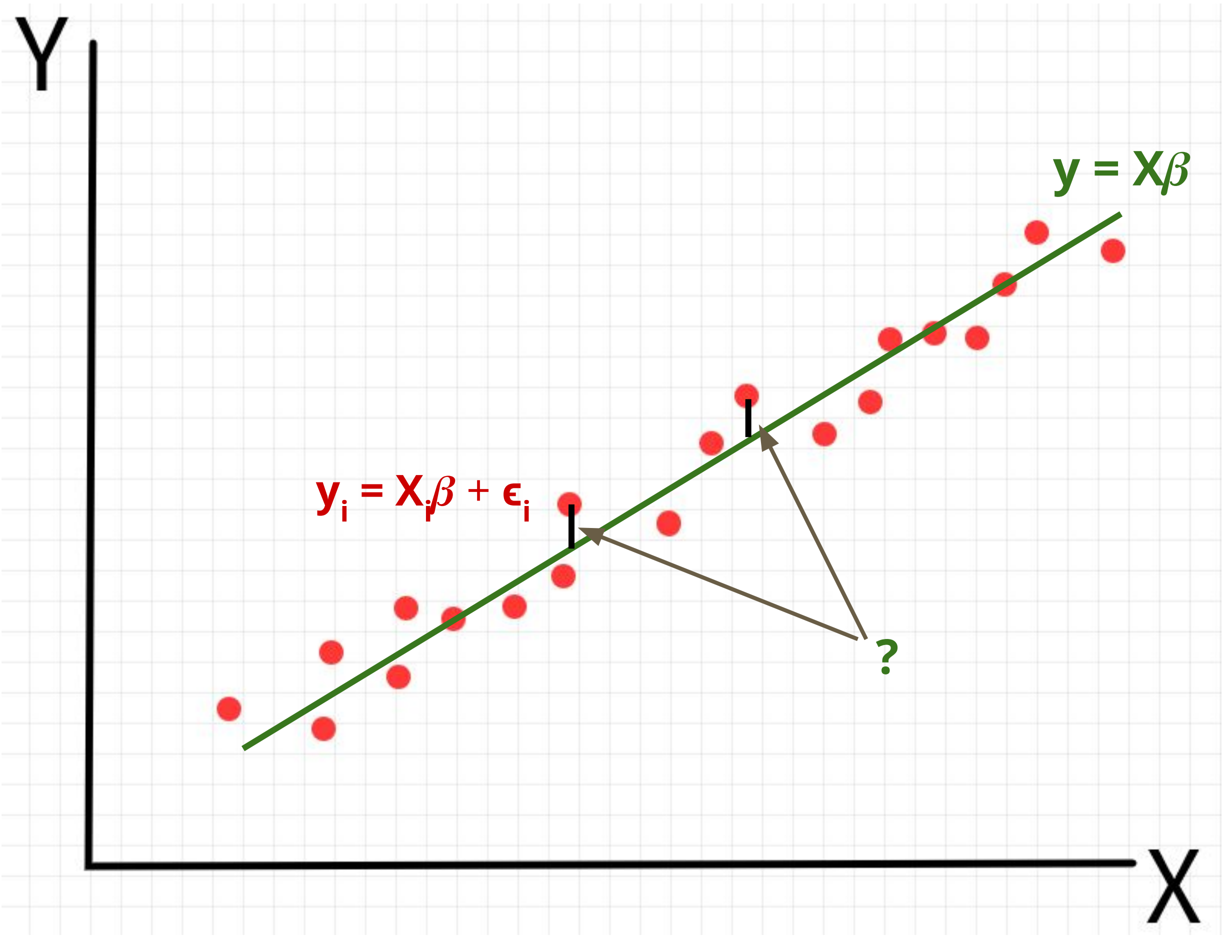
Let's label the diagram

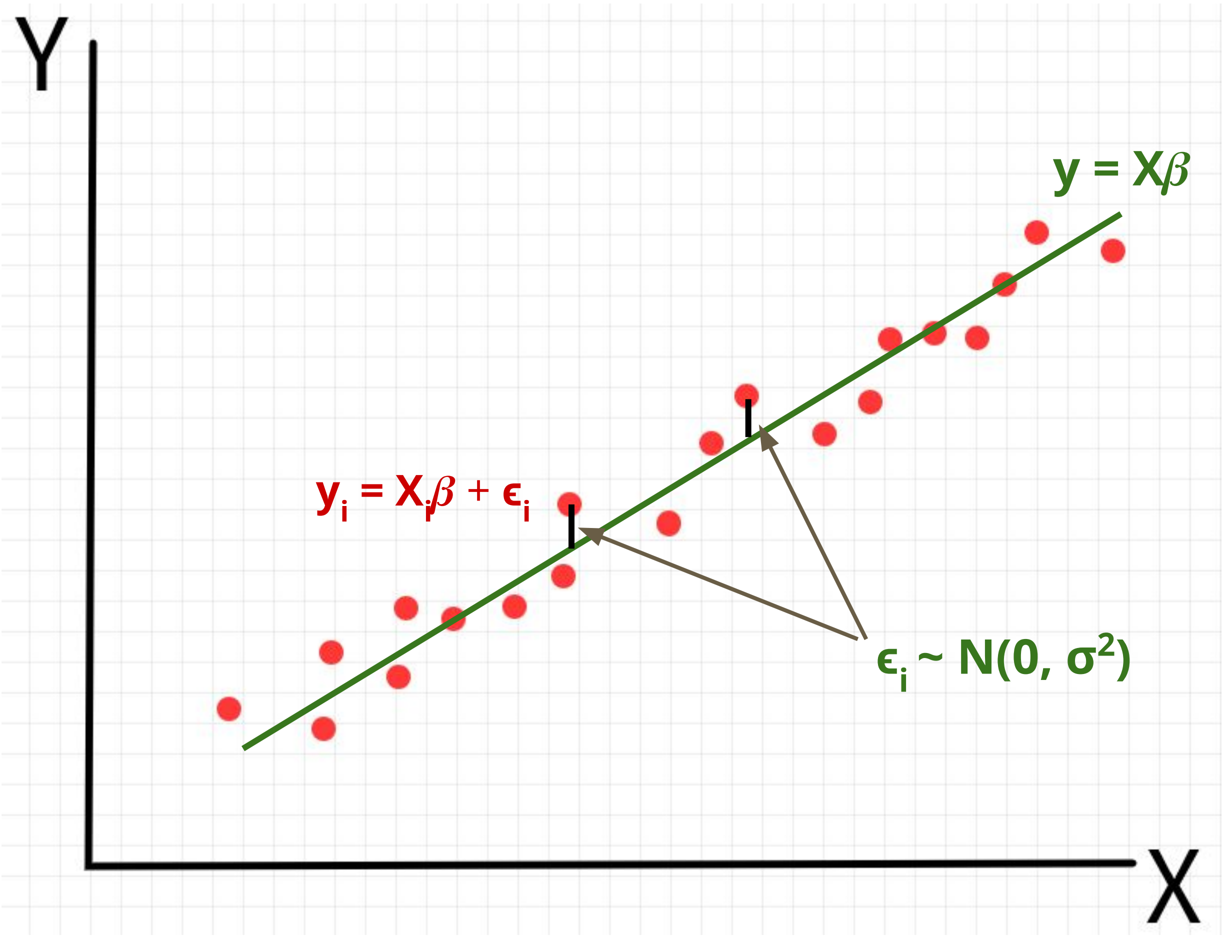

Y
?
X

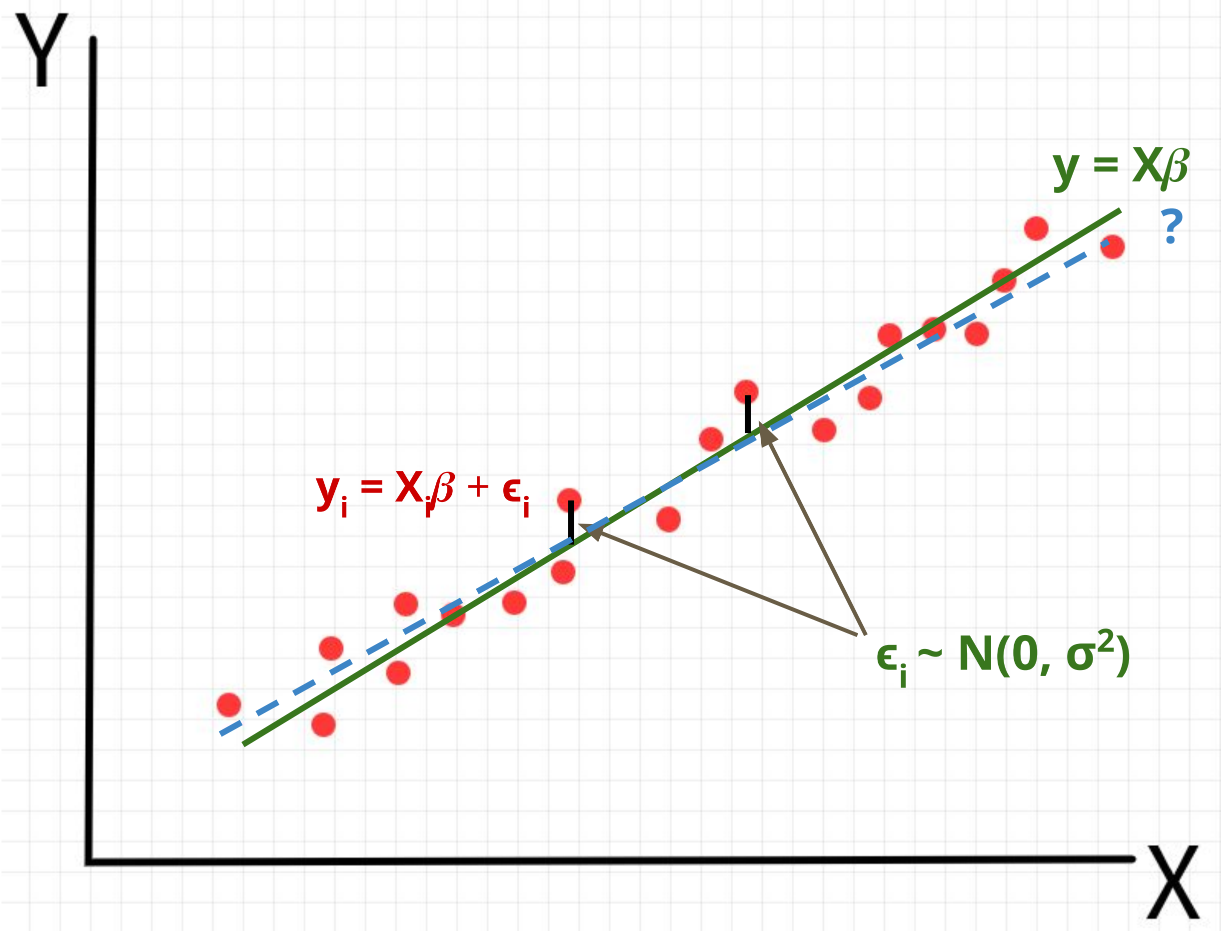
Y
$y = X\beta$
X


Y
y = Xβ
?
X
x_i

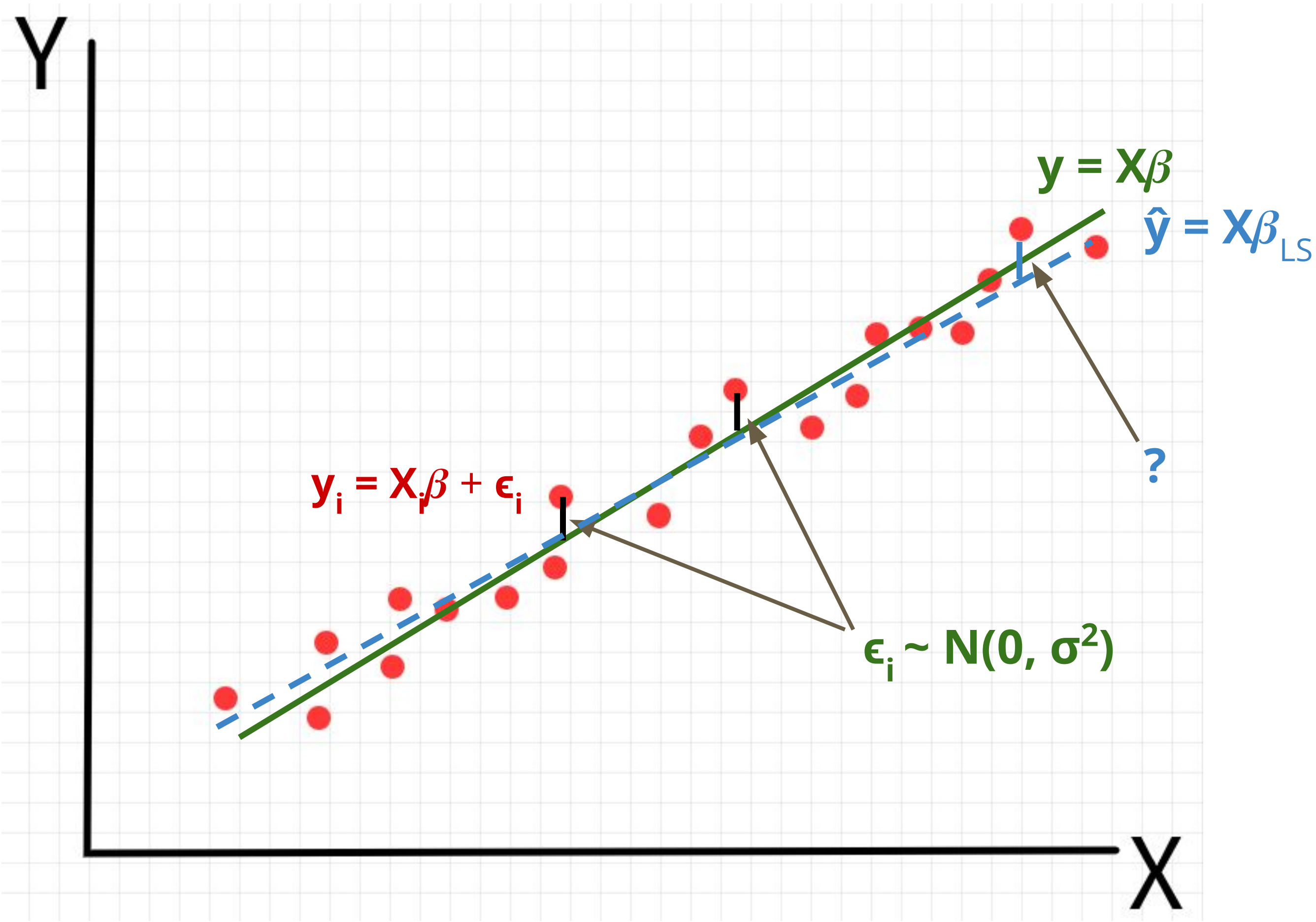

Y
y = Xβ
yi = Xiβ + ϵi
X
xi


Y
y = Xβ
y_i = X_iβ + ϵ_i
?
X

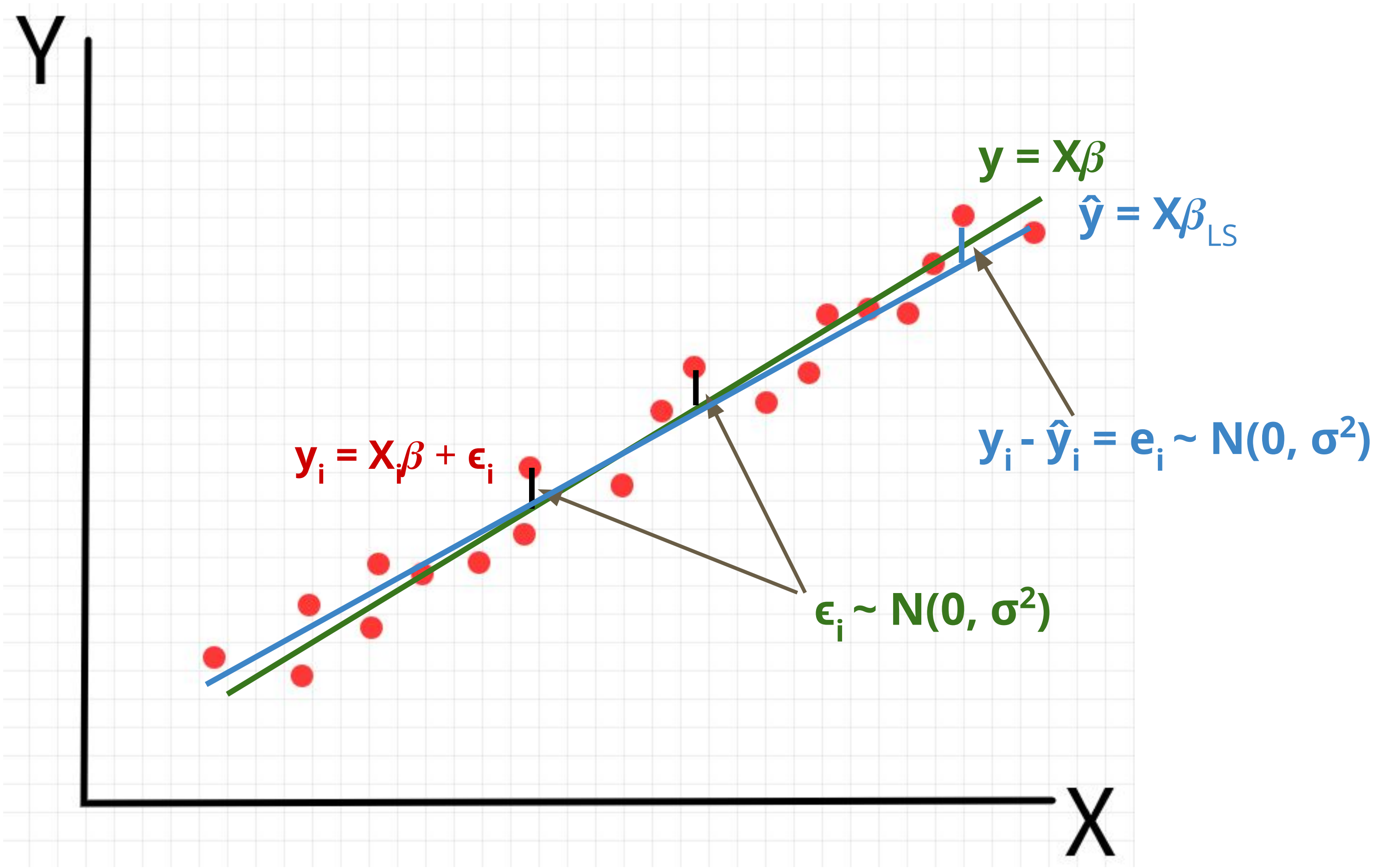

Y
y = Xβ
y_i = X_iβ + ϵ_i
ϵ_i ~ N(0, σ^2)
X

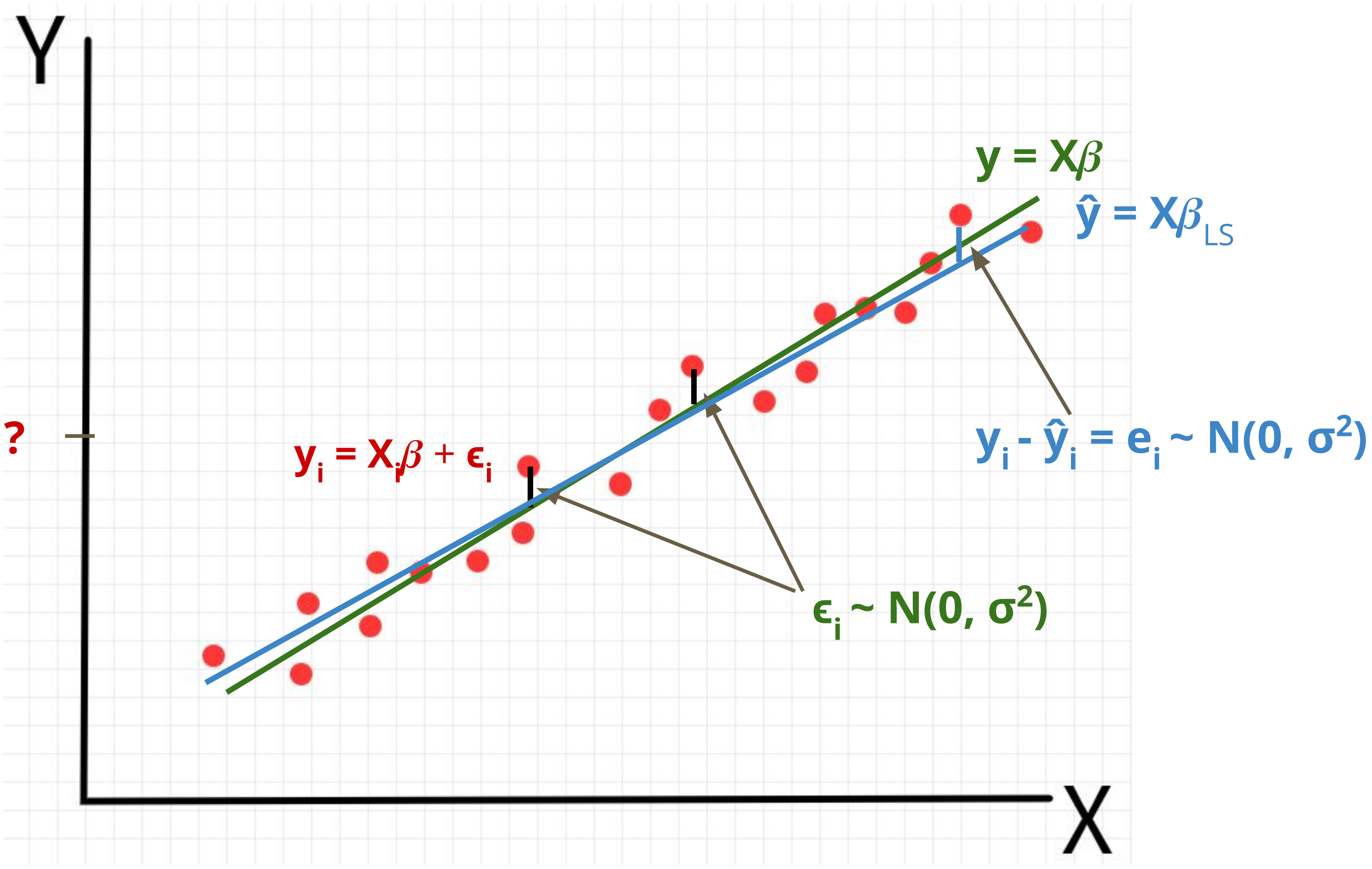


Y
X
y = Xβ
?
yi = Xiβ + ϵi
ϵi ~ N(0, σ2)

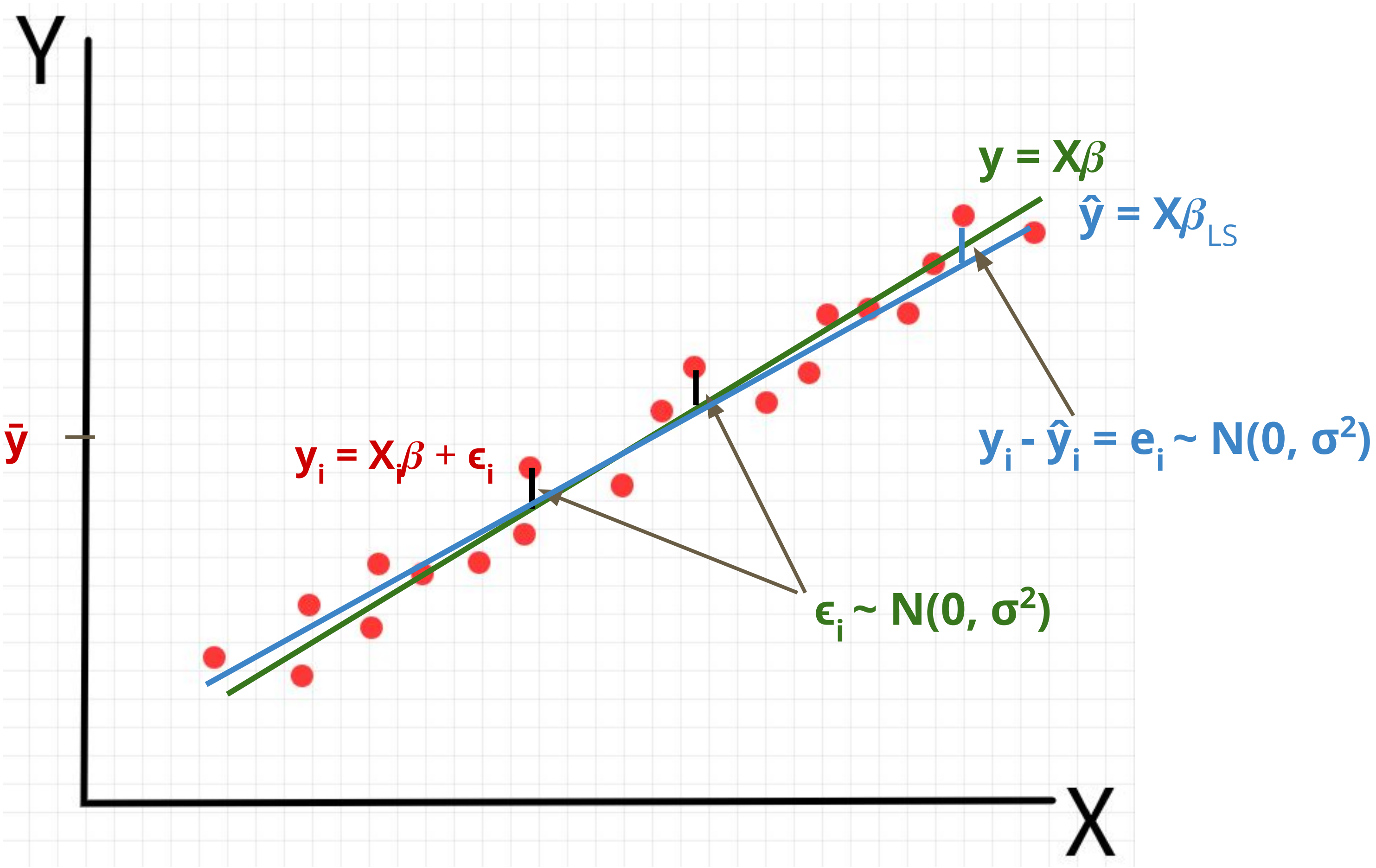


Y
X
$y = X\beta$
$\hat{y} = X\beta_{LS}$
$y_i = X_i\beta + \epsilon_i$
$\epsilon_i \sim N(0, \sigma^2)$
?


Y
X
$y = X\beta$
$\hat{y} = X\beta_{LS}$
$y_i - \hat{y}_i = e_i \sim N(0, \sigma^2)$
$y_i = X_i\beta + \epsilon_i$
$\epsilon_i \sim N(0, \sigma^2)$



Y
X
?
$y = X\beta$
$\hat{y} = X\beta_{LS}$
$y_i - \hat{y}_i = e_i \sim N(0, \sigma^2)$
$y_i = X_i\beta + \epsilon_i$
$\epsilon_i \sim N(0, \sigma^2)$



Y
X
ȳ
y = Xβ
ŷ = Xβ_LS
y_i = X_iβ + ϵ_i
y_i - ŷ_i = e_i ~ N(0, σ²)
ϵ_i ~ N(0, σ²)

How do we know a linear model is applicable?

Is it true that $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\sigma}^2)$?

How do we know a linear model is applicable?

Is it true that $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim \mathbf{N(0, \sigma^2)}$?

If this is true then: $\mathbf{y_i - \hat{y}_i = e_i \sim N(0, \sigma^2)}$

Can we check this?

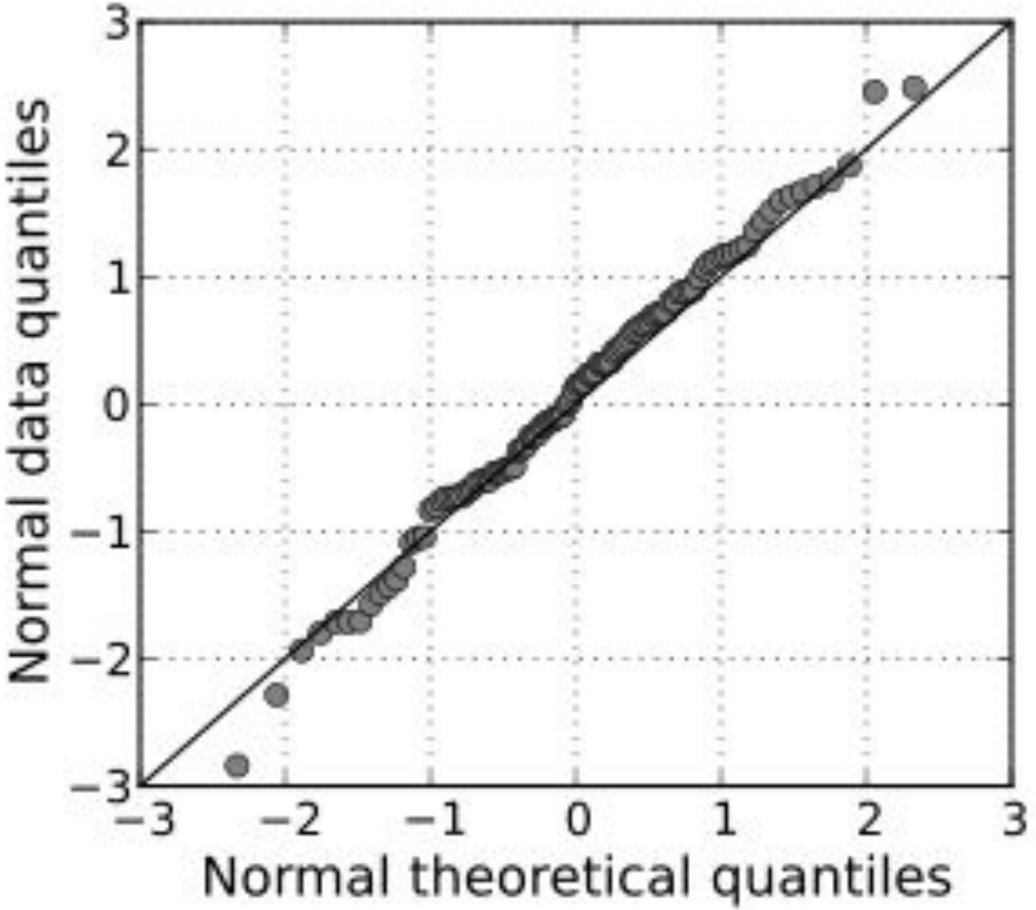
QQ plot

Quantiles are the values for which a particular % of values are contained below it.

For example the 50% quantile of a N(0,1) distribution is 0 since 50% of samples would be contained below 0 were you to sample a large number of times.

Forall quantiles q, if sample.q == known_distribution.q then they have the same distribution.

QQ plot



Normal data quantiles
Normal theoretical quantiles
3
2
1
0
−1
−2
−3

demo

Evaluating Our Regression Model

Some Notation:

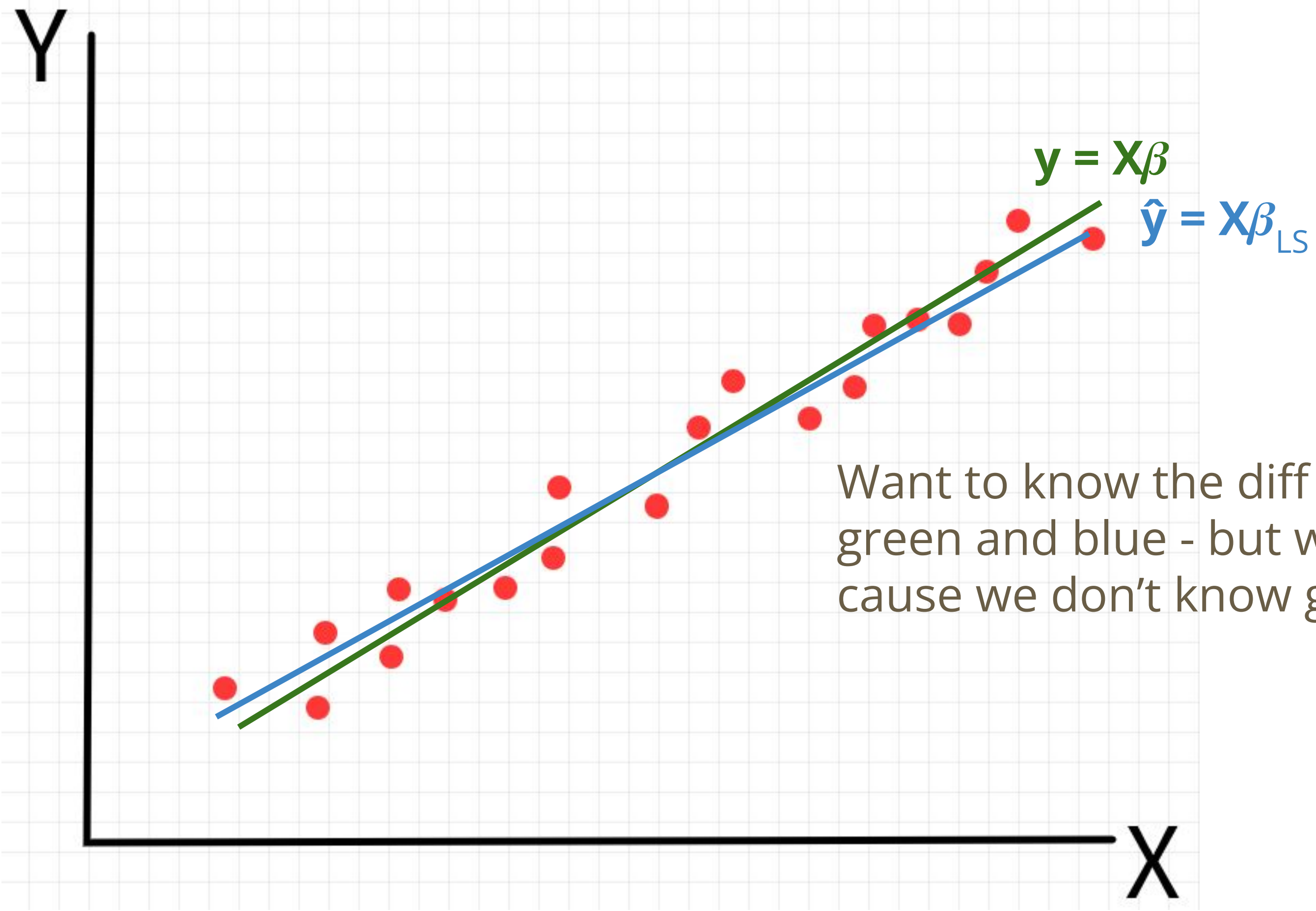
$\mathbf{y_i}$ is the "true" value from our data set (i.e. $\mathbf{x_i}\boldsymbol{\beta} + \boldsymbol{\epsilon}_\mathbf{i}$)

$\mathbf{\hat{y}_i}$ is the estimate of y_i from our model (i.e. $\mathbf{x_i}\boldsymbol{\beta}_\mathbf{LS}$)

$\mathbf{\bar{y}}$ is the sample mean all $\mathbf{y_i}$

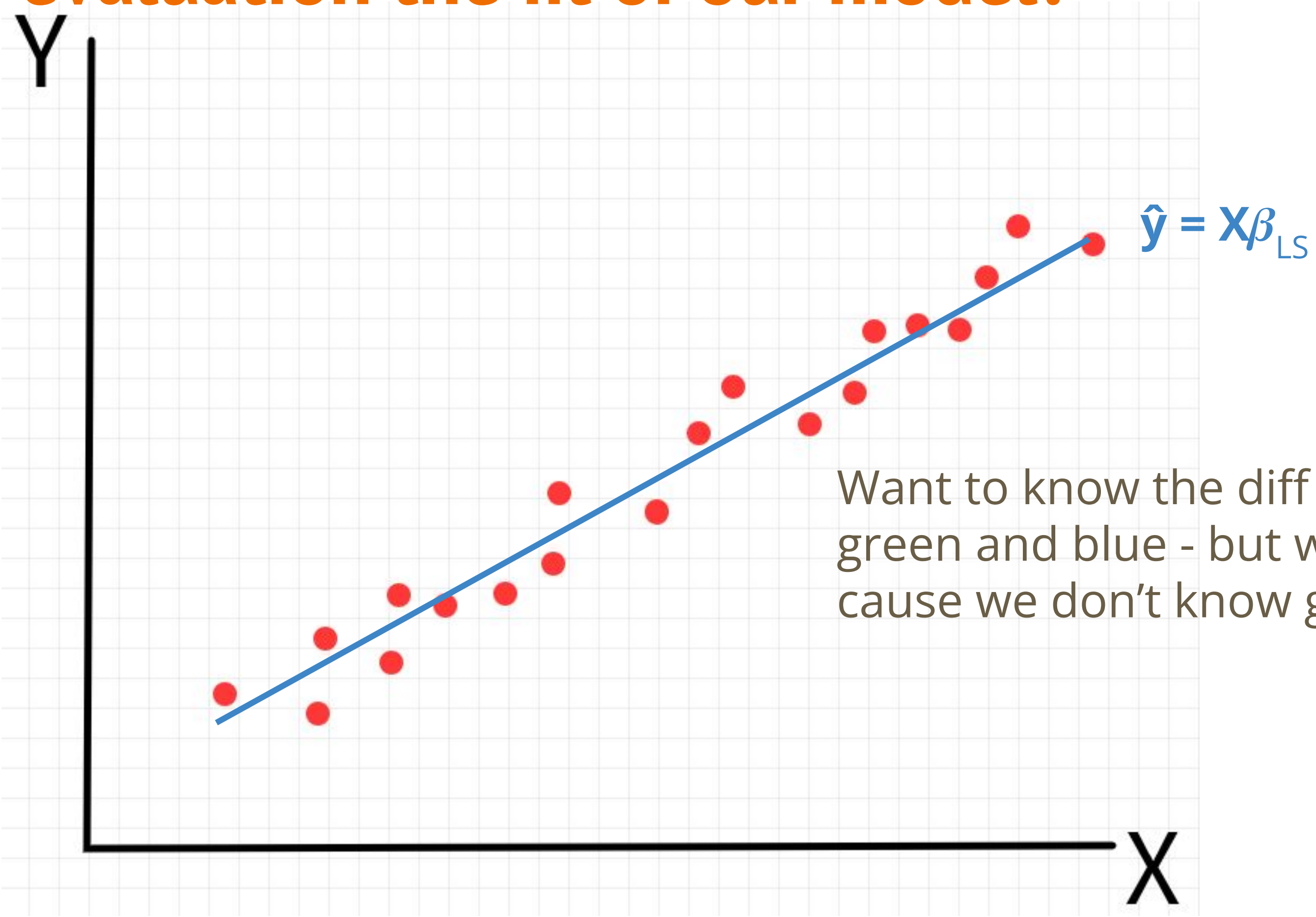
$\mathbf{y_i} - \mathbf{\hat{y}_i}$ are the estimates of $\boldsymbol{\epsilon}_\mathbf{i}$ and are referred to as residuals

Metric for evaluation the fit of our model?



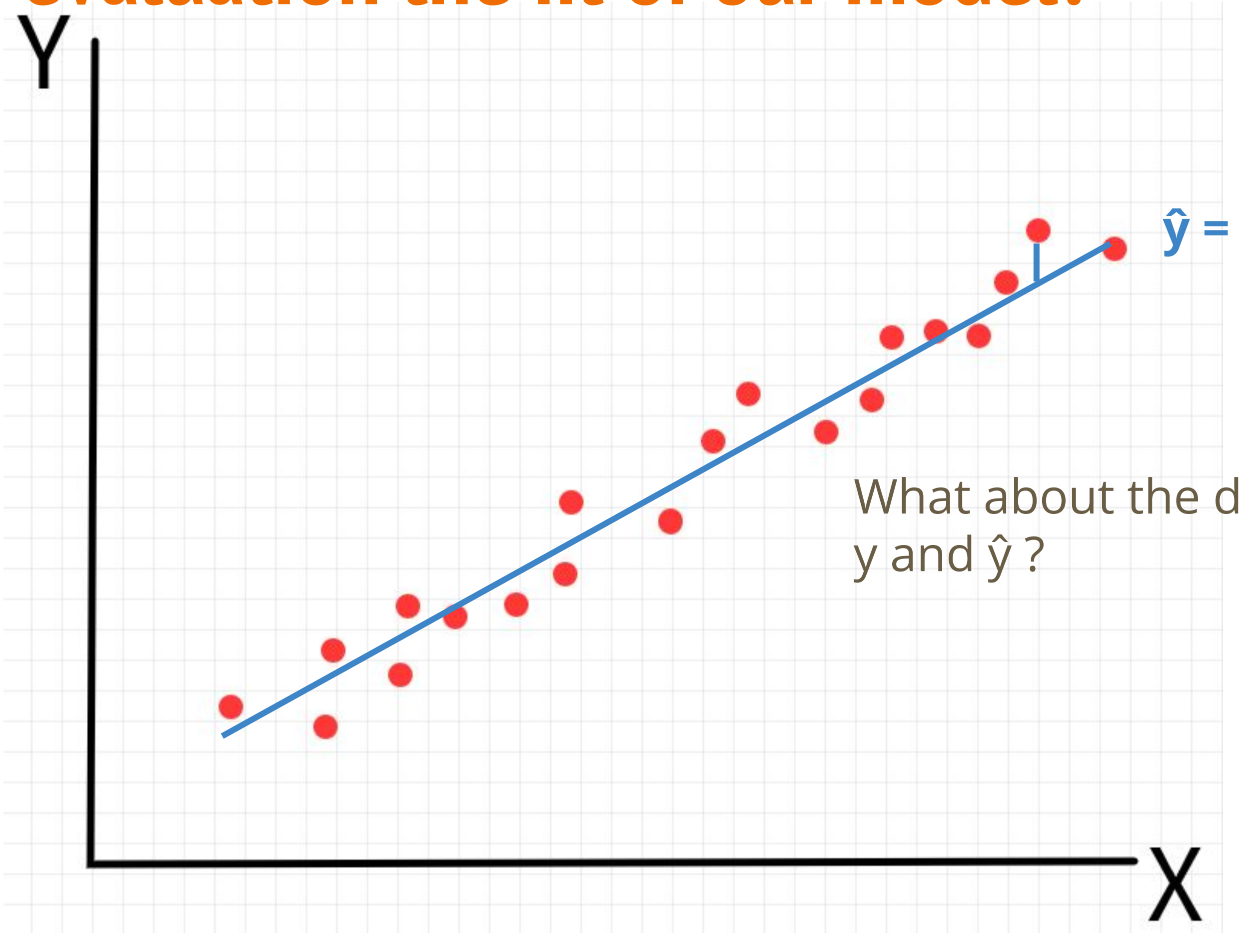
Want to know the diff between green and blue - but we can't cause we don't know green

Metric for evaluation the fit of our model?



Want to know the diff between green and blue - but we can't cause we don't know green

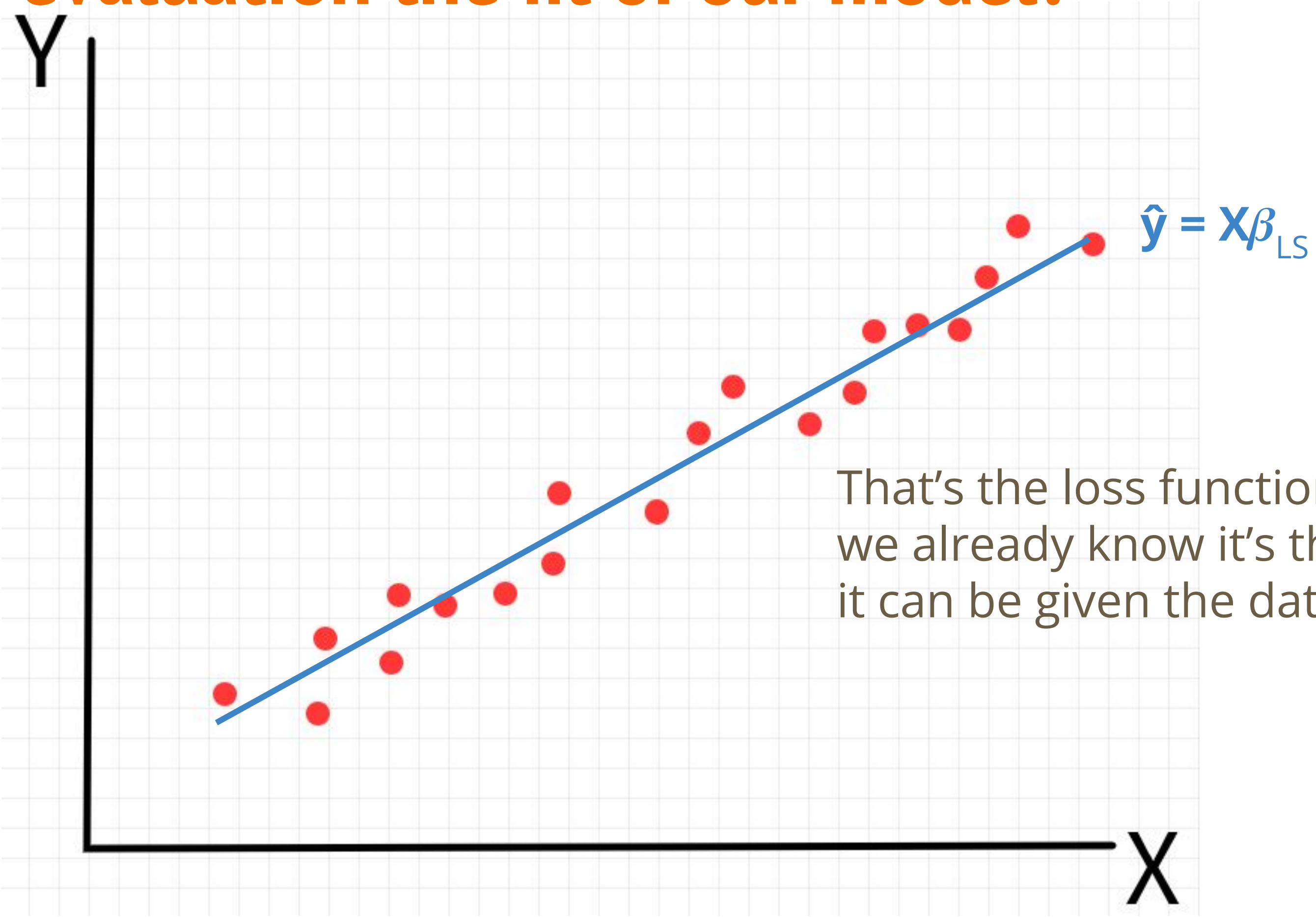
Metric for evaluation the fit of our model?



$\hat{\mathbf{y}} = \mathbf{X}\beta_{LS}$

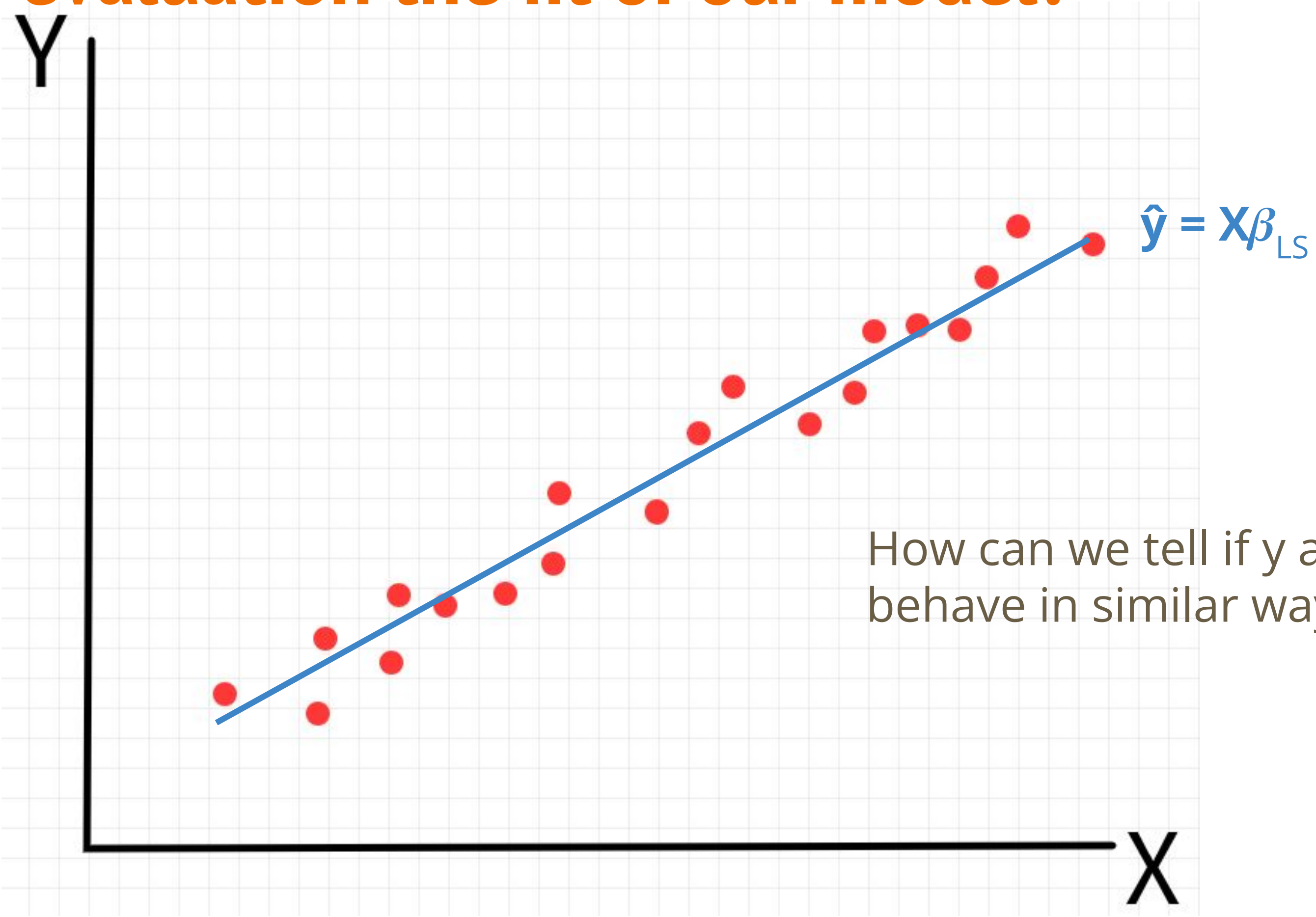
What about the diff between y and ŷ ?

Metric for evaluation the fit of our model?



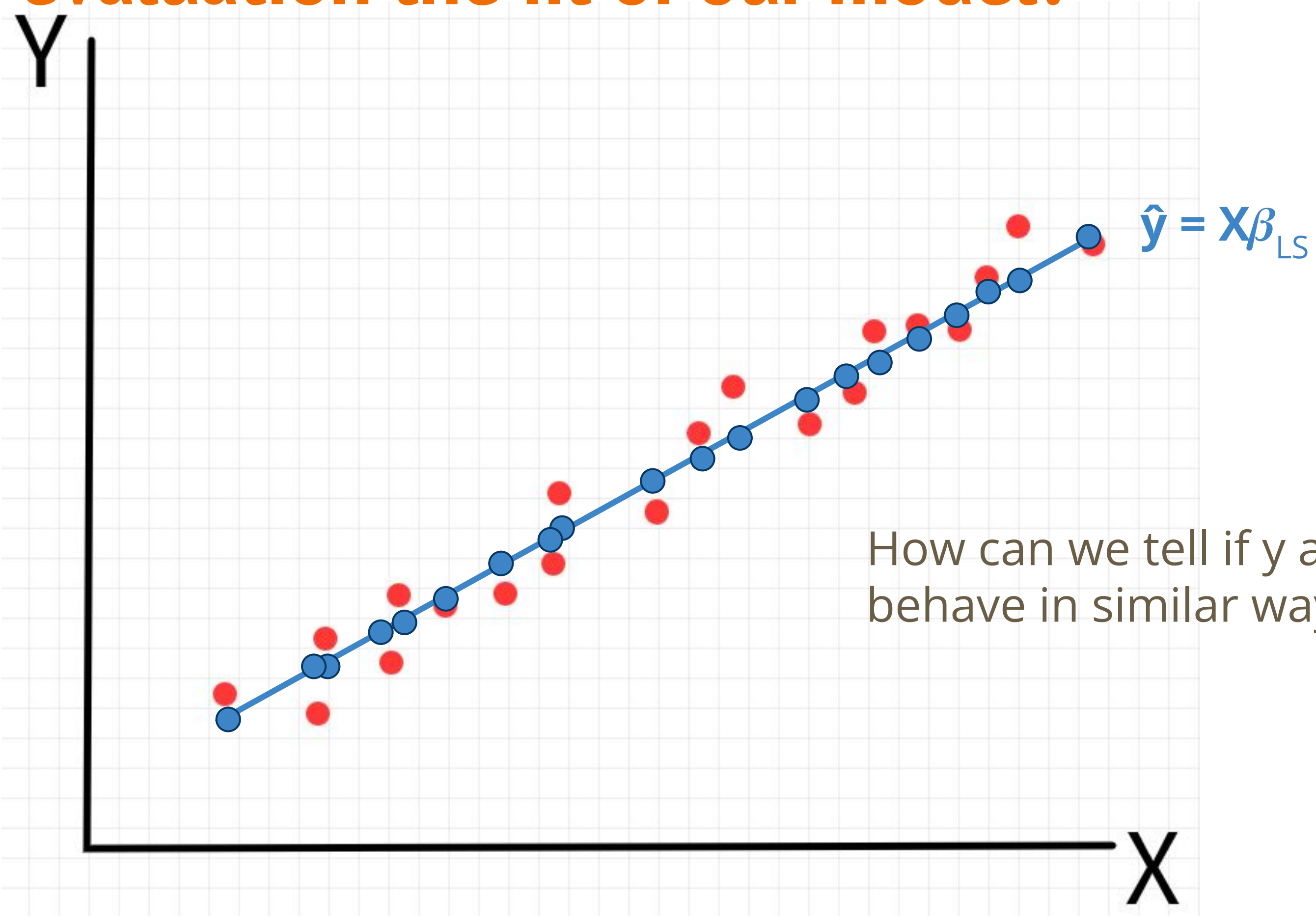
That’s the loss function and we already know it’s the best it can be given the data…

Metric for evaluation the fit of our model?



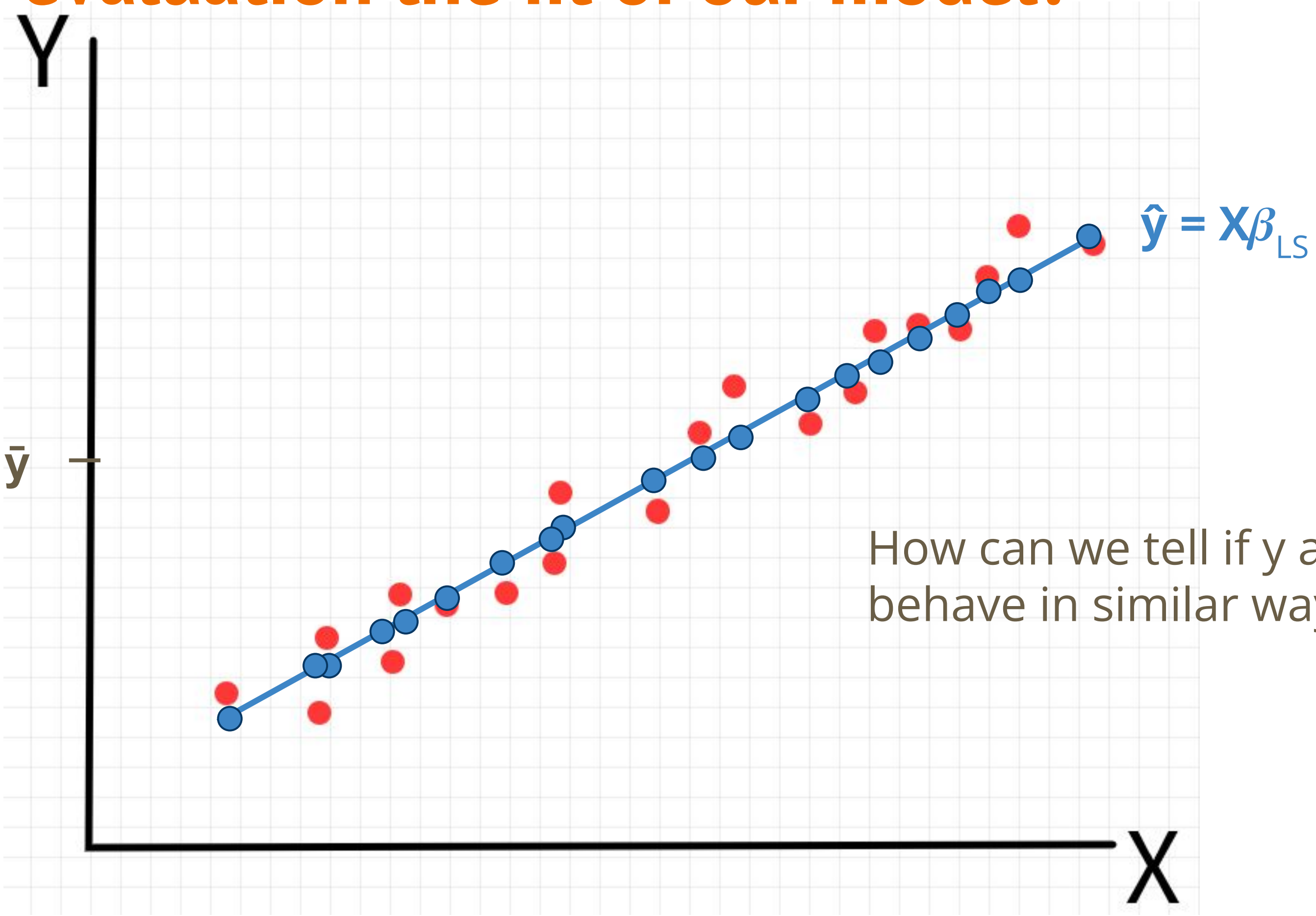
How can we tell if y and ŷ behave in similar ways?

Metric for evaluation the fit of our model?



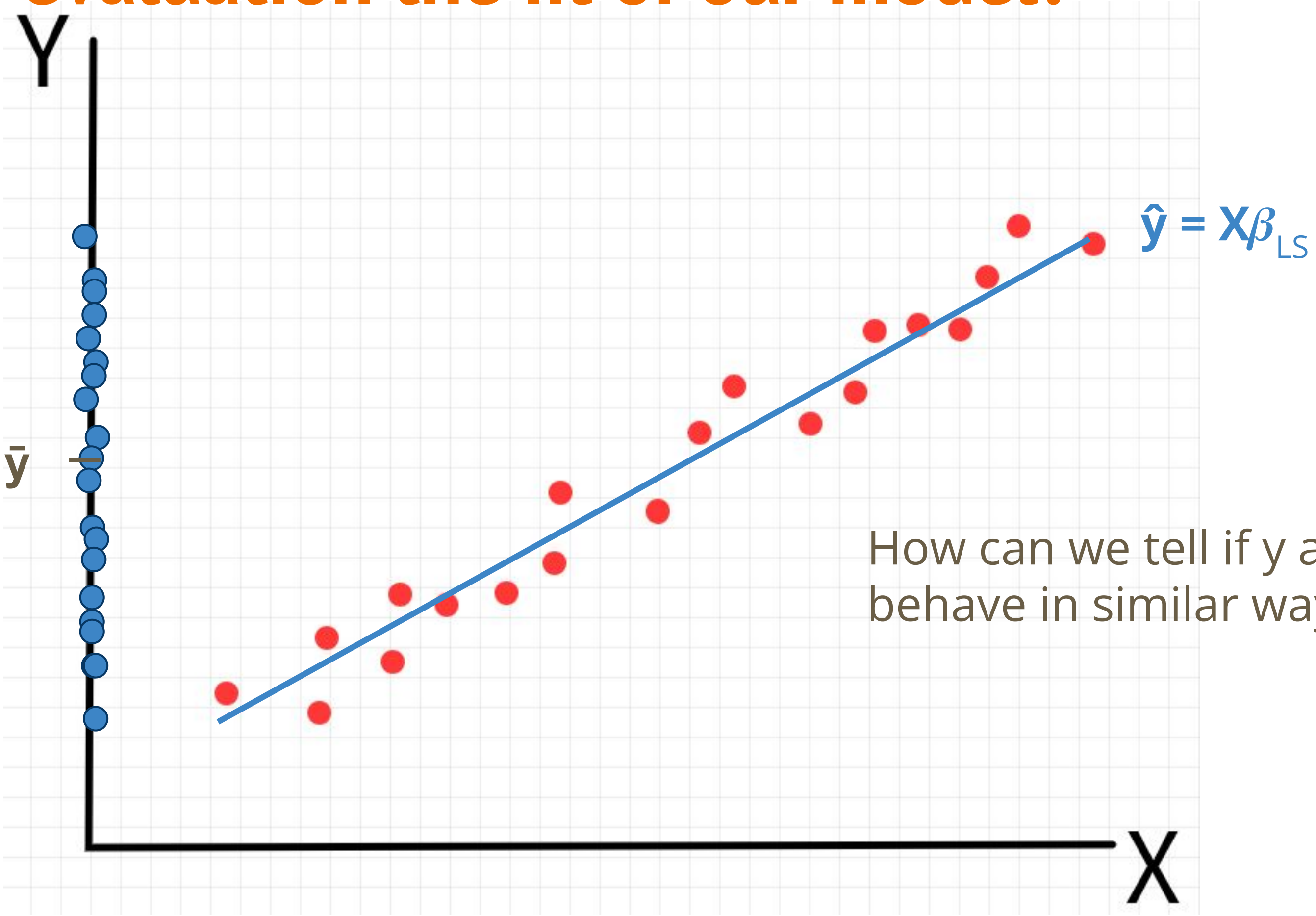
How can we tell if y and ŷ behave in similar ways?

Metric for evaluation the fit of our model?



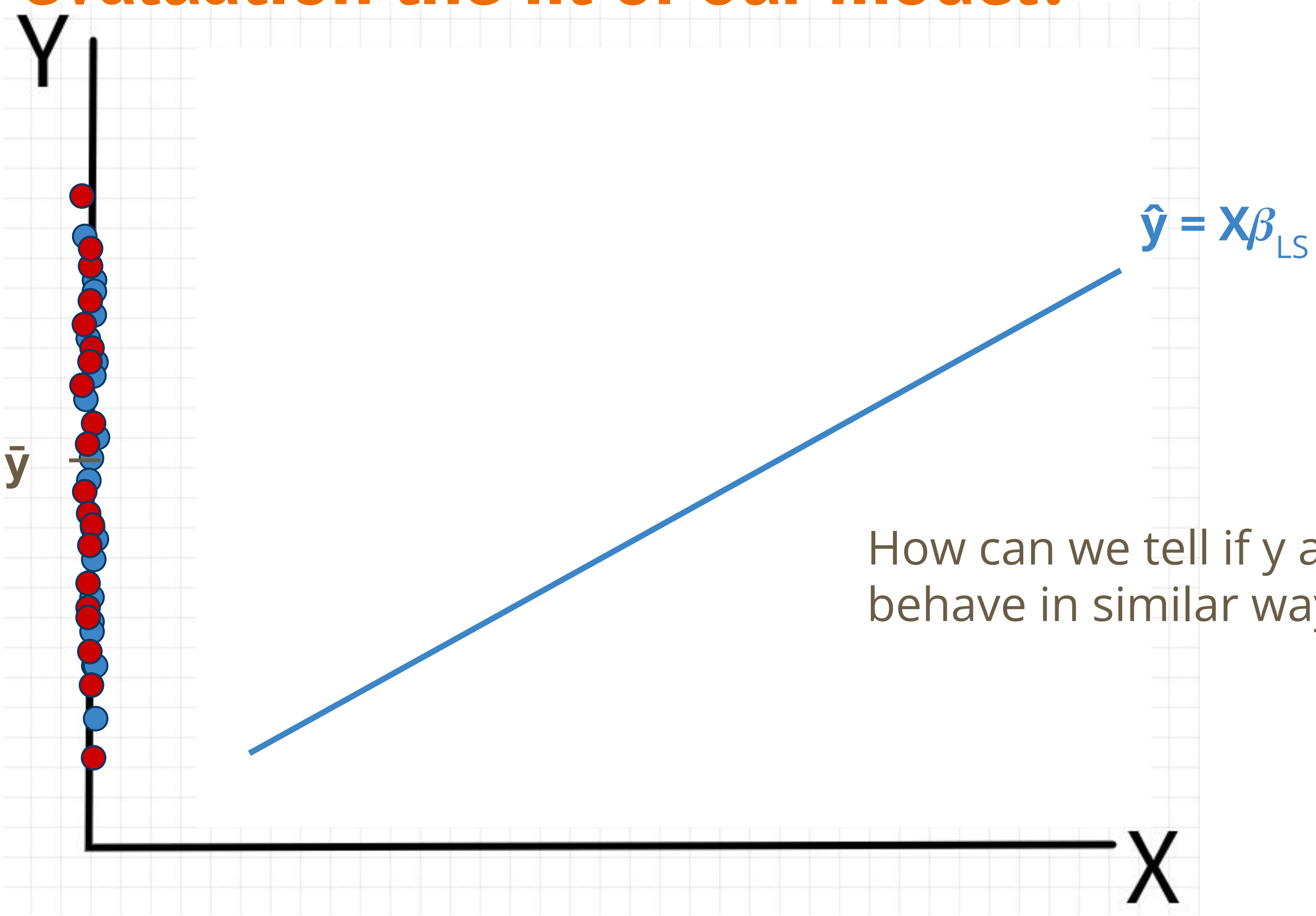
How can we tell if y and ŷ behave in similar ways?

Metric for evaluation the fit of our model?



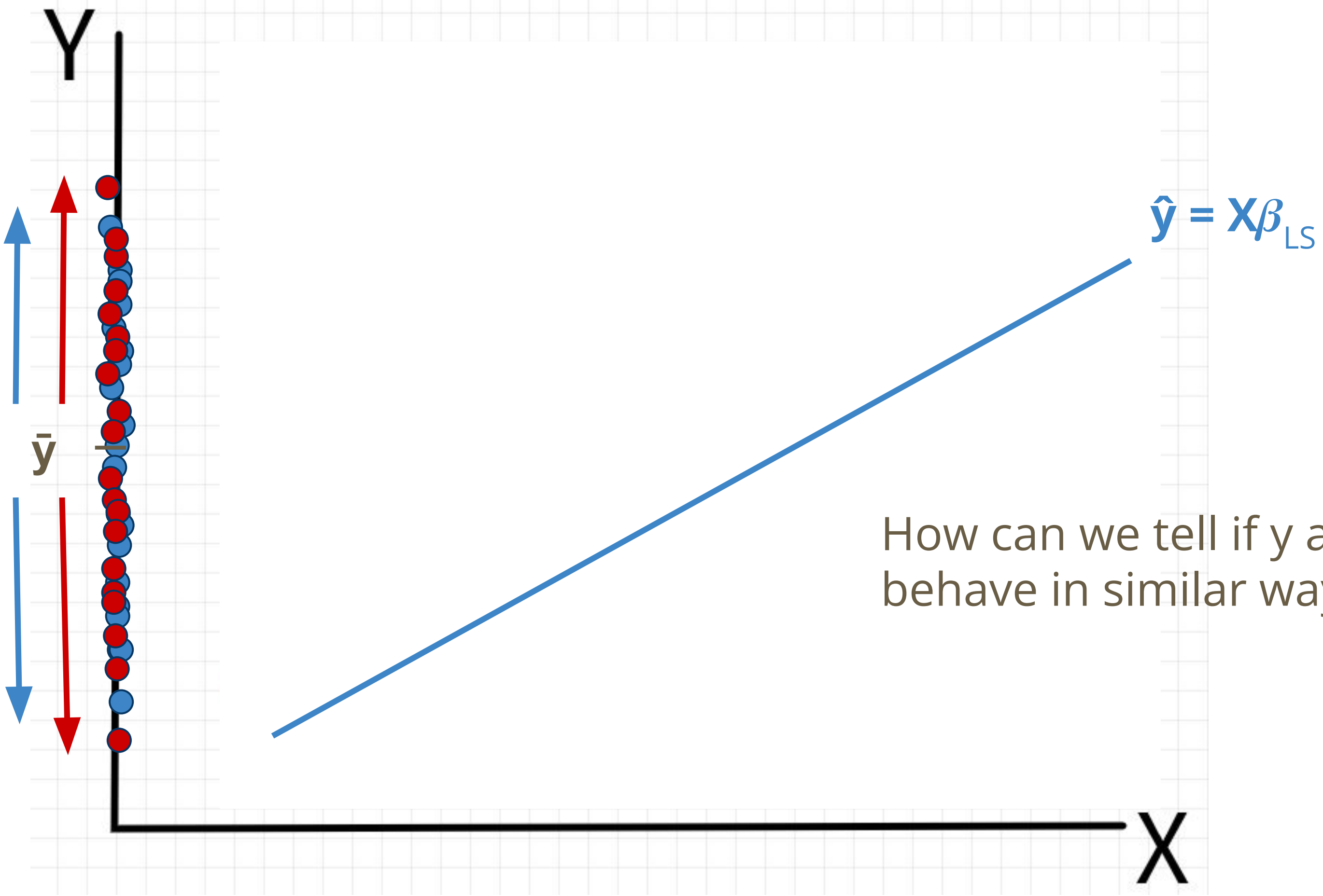
How can we tell if y and ŷ behave in similar ways?

Metric for evaluation the fit of our model?



How can we tell if y and ŷ behave in similar ways?

Metric for evaluation the fit of our model?



How can we tell if y and ŷ behave in similar ways?

Metric for evaluation the fit of our model?

Is the value of the loss function sufficient? i.e.

$$||y - X\beta||_2^2 = \sum_i^n (y_i - \hat{y}_i)^2$$

Evaluating Our Regression Model

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

This is a measure of the spread of y_i around the mean of y

Evaluating Our Regression Model

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

This is a measure of the spread of y_i around the mean of y

$$ESS = \sum_{i}^{n} (\hat{y_i} - \bar{y})^2$$

This is a measure of the spread of our model's estimates of y_i around the mean of y

Evaluating Our Regression Model

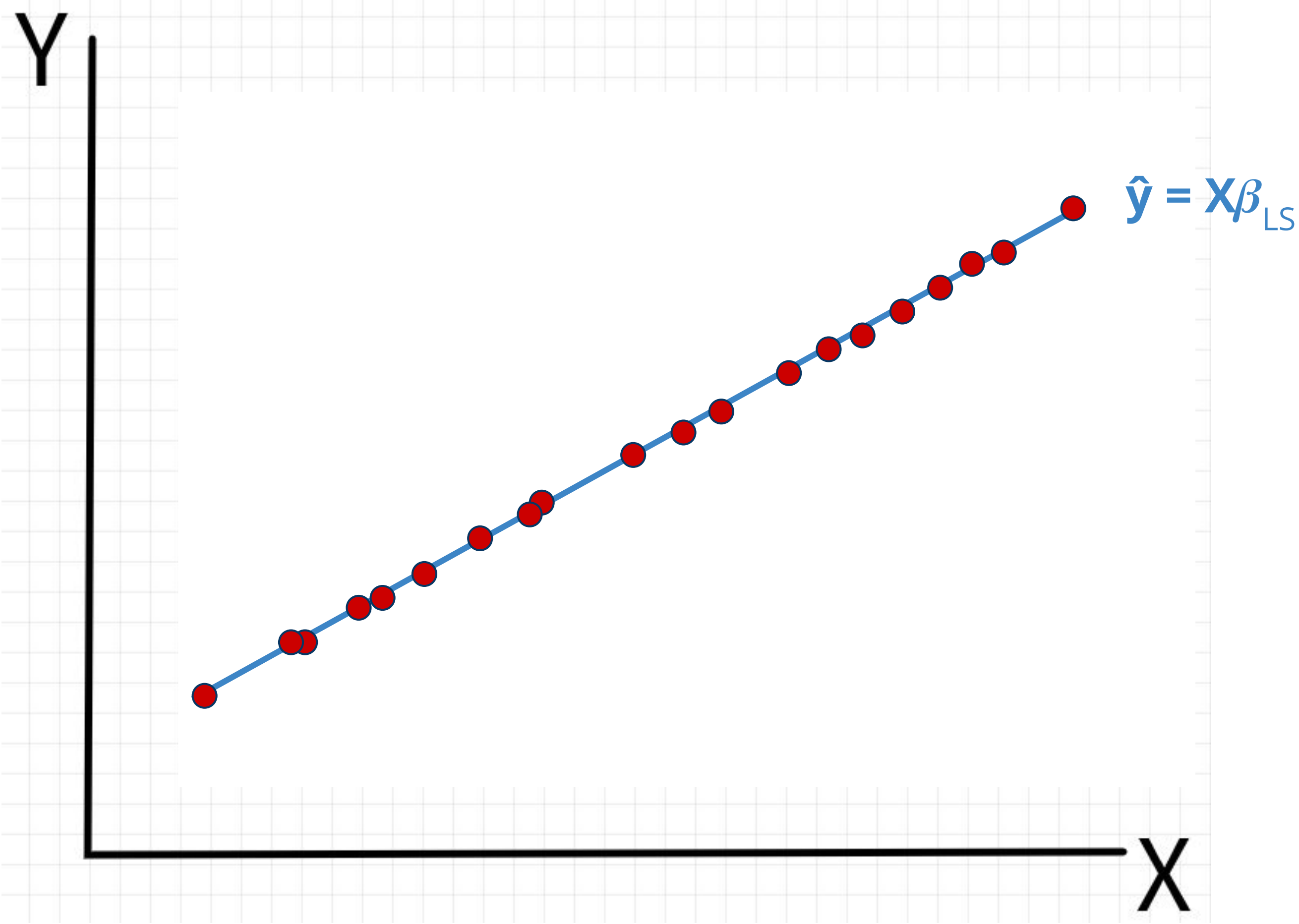
$$TSS = \sum_{i}^{n} (y_i - \bar{y})^2$$

$$R^2 = \frac{ESS}{TSS}$$

$$ESS = \sum_{i}^{n} (\hat{y}_i - \bar{y})^2$$

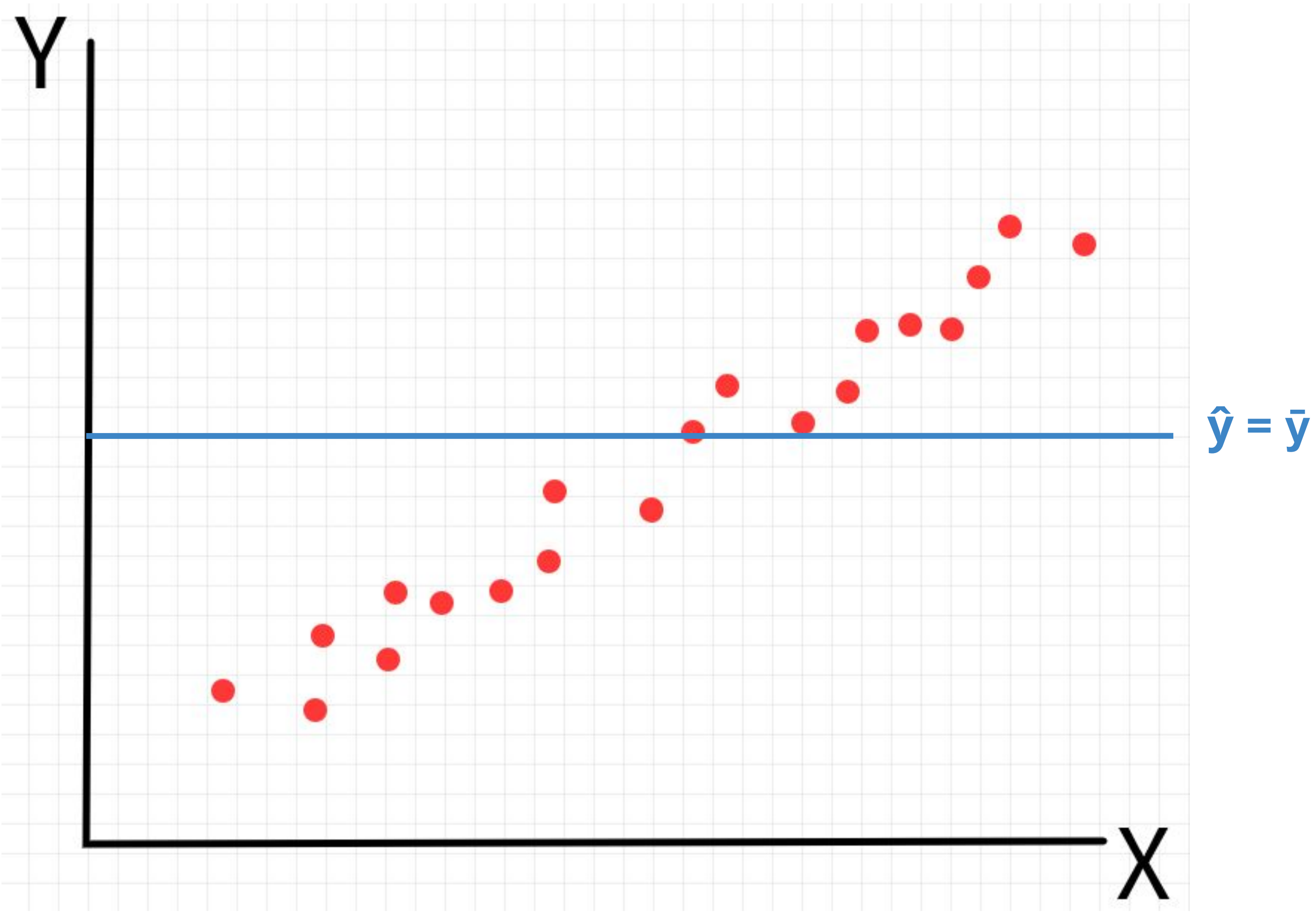
R^2 measures the fraction of variance that is explained by ŷ (our model)

$R^2 = 1$



$R^2 = 0$

$R^2 = 0$



Evaluating our Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.840
Model:                            OLS   Adj. R-squared:                  0.836
Method:                 Least Squares   F-statistic:                     254.1
Date:                Sun, 20 Mar 2022   Prob (F-statistic):           2.72e-39
Time:                        11:36:16   Log-Likelihood:                -482.37
No. Observations:                 100   AIC:                             970.7
Df Residuals:                      97   BIC:                             978.5
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.1912      3.162      0.693      0.490      -4.085       8.467
x1            29.3912      3.274      8.977      0.000      22.893      35.889
x2            78.1391      3.594     21.741      0.000      71.006      85.272
==============================================================================
Omnibus:                        1.279   Durbin-Watson:                   1.824
Prob(Omnibus):                  0.527   Jarque-Bera (JB):                1.065
Skew:                           0.253   Prob(JB):                        0.587
Kurtosis:                       2.999   Cond. No.                         1.38
==============================================================================
```

Evaluating our Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.840
Model:                            OLS   Adj. R-squared:                  0.836
Method:                 Least Squares   F-statistic:                     254.1
Date:                Sun, 20 Mar 2022   Prob (F-statistic):           2.72e-39
Time:                        11:36:16   Log-Likelihood:                -482.37
No. Observations:                 100   AIC:                             970.7
Df Residuals:                      97   BIC:                             978.5
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.1912      3.162      0.693      0.490      -4.085       8.467
x1            29.3912      3.274      8.977      0.000      22.893      35.889
x2            78.1391      3.594     21.741      0.000      71.006      85.272
==============================================================================
Omnibus:                        1.279   Durbin-Watson:                   1.824
Prob(Omnibus):                  0.527   Jarque-Bera (JB):                1.065
Skew:                           0.253   Prob(JB):                        0.587
Kurtosis:                       2.999   Cond. No.                         1.38
==============================================================================
```

Evaluating our Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.840
Model:                            OLS   Adj. R-squared:                  0.836
Method:                 Least Squares   F-statistic:                     254.1
Date:                Sun, 20 Mar 2022   Prob (F-statistic):           2.72e-39
Time:                        11:36:16   Log-Likelihood:                -482.37
No. Observations:                 100   AIC:                             970.7
Df Residuals:                      97   BIC:                             978.5
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.1912      3.162      0.693      0.490      -4.085       8.467
x1            29.3912      3.274      8.977      0.000      22.893      35.889
x2            78.1391      3.594     21.741      0.000      71.006      85.272
==============================================================================
Omnibus:                        1.279   Durbin-Watson:                   1.824
Prob(Omnibus):                  0.527   Jarque-Bera (JB):                1.065
Skew:                           0.253   Prob(JB):                        0.587
Kurtosis:                       2.999   Cond. No.                         1.38
==============================================================================
```

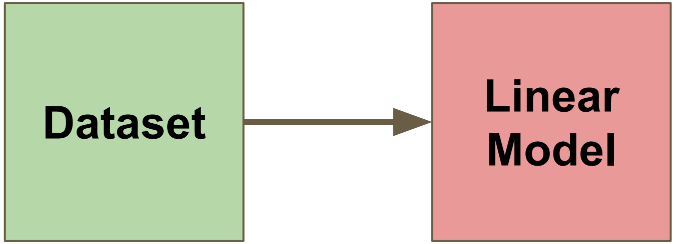
Evaluating our Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.840
Model:                            OLS   Adj. R-squared:                  0.836
Method:                 Least Squares   F-statistic:                     254.1
Date:                Sun, 20 Mar 2022   Prob (F-statistic):           2.72e-39
Time:                        11:36:16   Log-Likelihood:                -482.37
No. Observations:                 100   AIC:                             970.7
Df Residuals:                      97   BIC:                             978.5
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.1912      3.162      0.693      0.490      -4.085       8.467
x1            29.3912      3.274      8.977      0.000      22.893      35.889
x2            78.1391      3.594     21.741      0.000      71.006      85.272
==============================================================================
Omnibus:                        1.279   Durbin-Watson:                   1.824
Prob(Omnibus):                  0.527   Jarque-Bera (JB):                1.065
Skew:                           0.253   Prob(JB):                        0.587
Kurtosis:                       2.999   Cond. No.                         1.38
==============================================================================
```

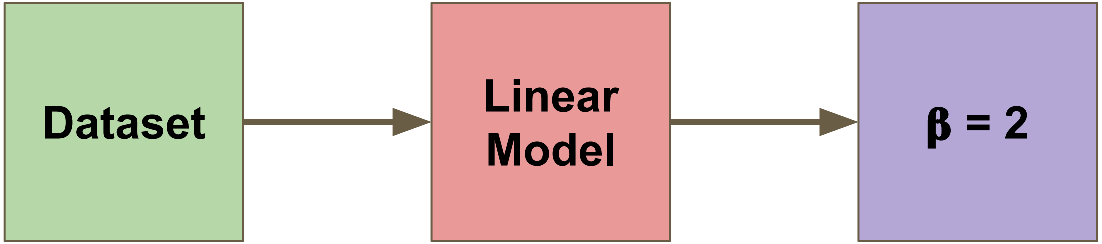
Hypothesis Testing

Dataset

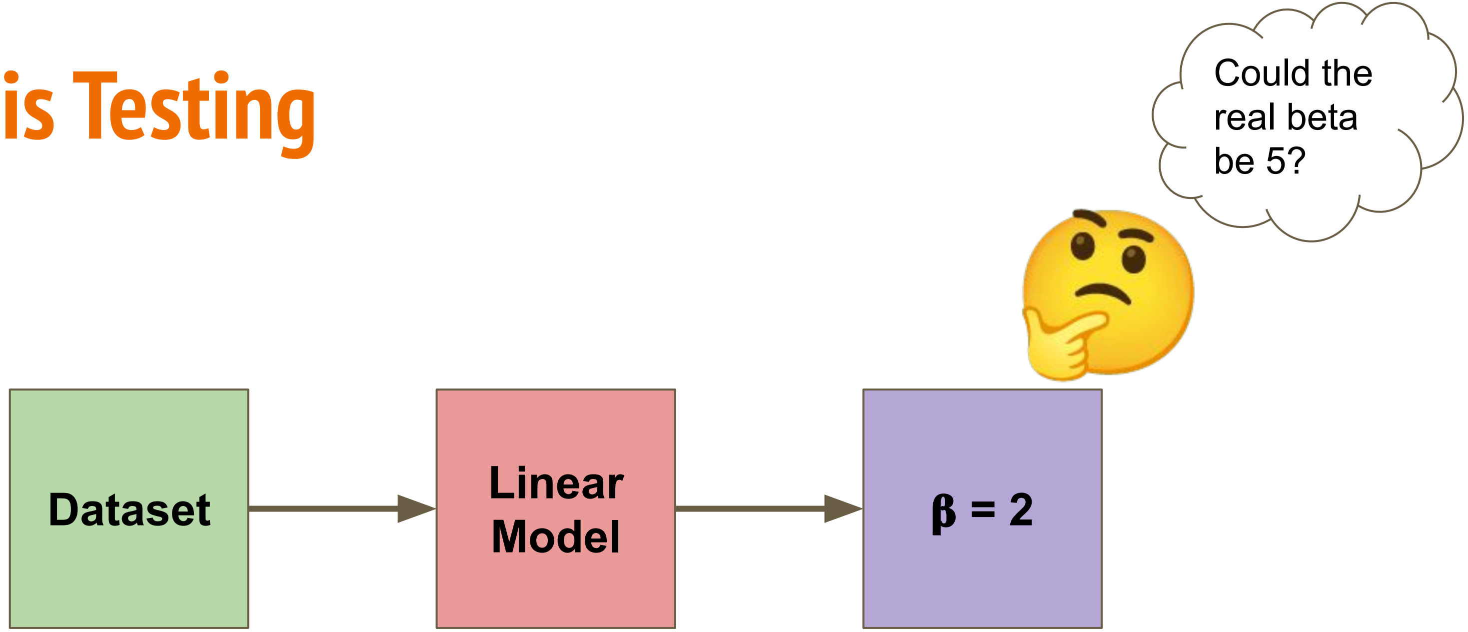
Hypothesis Testing

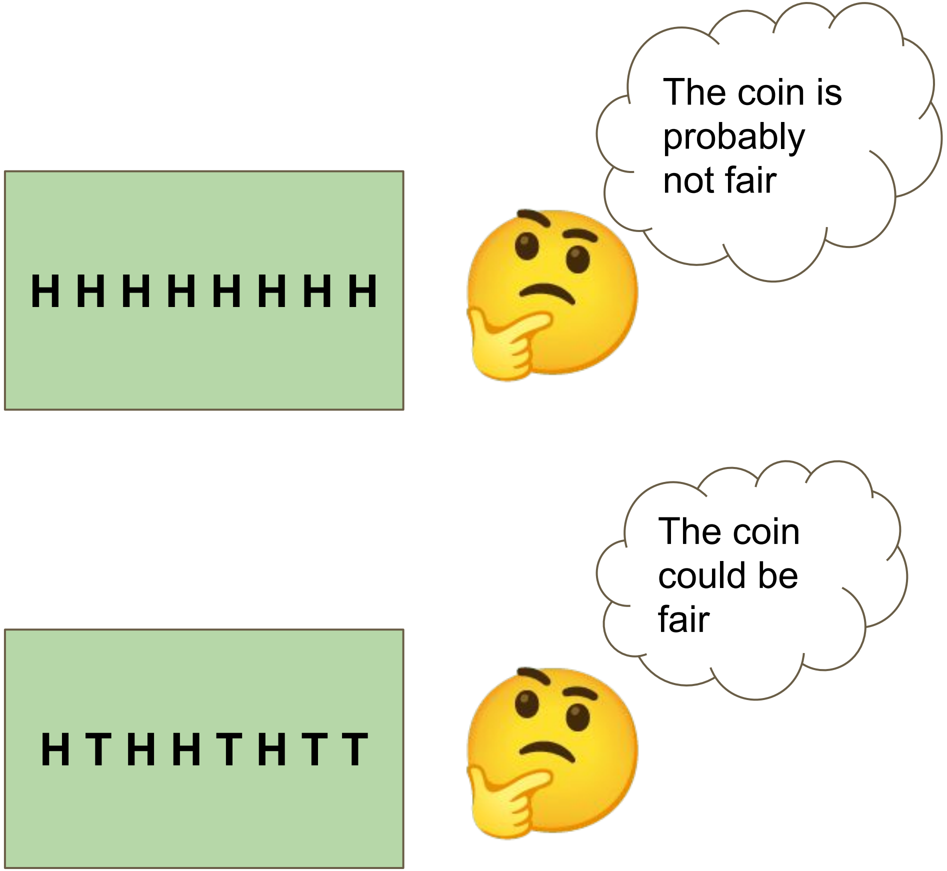


Hypothesis Testing

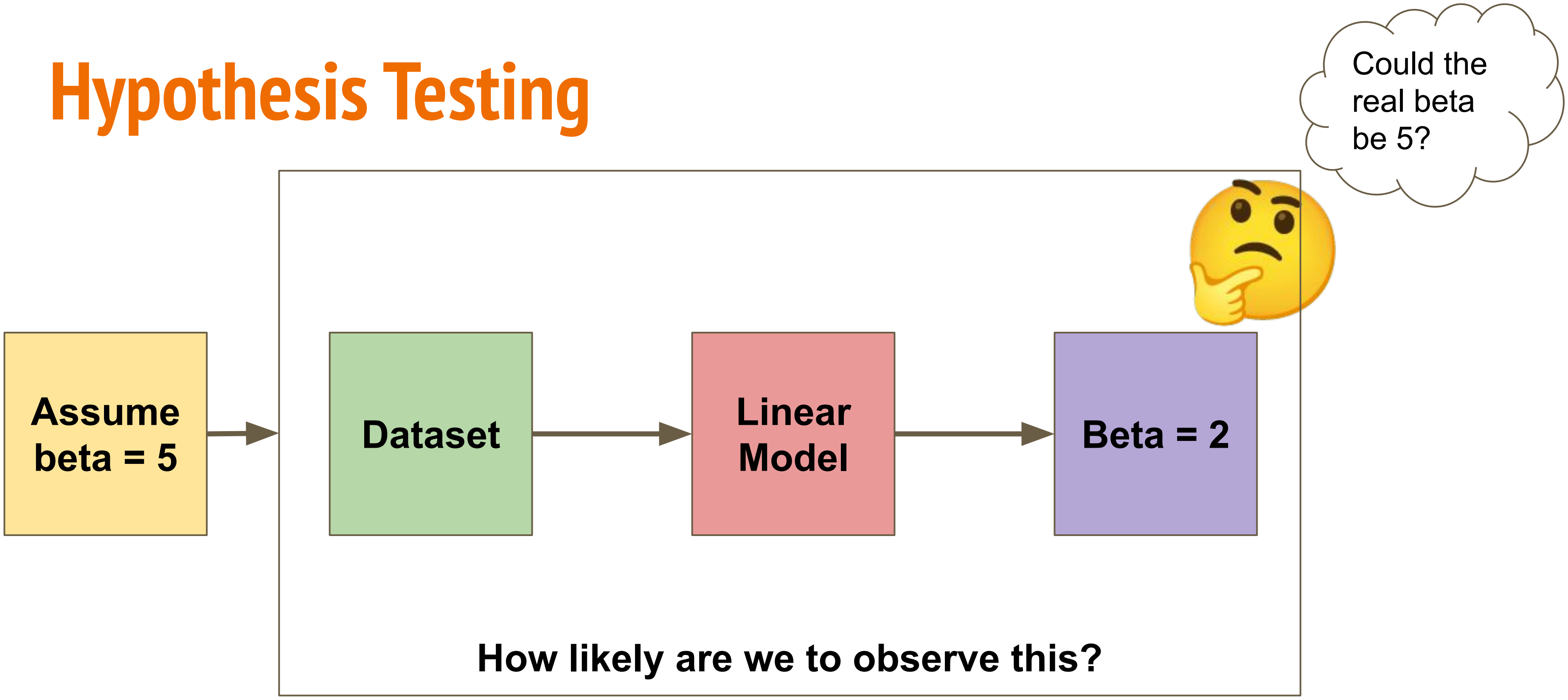


Hypothesis Testing

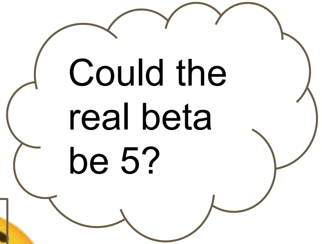


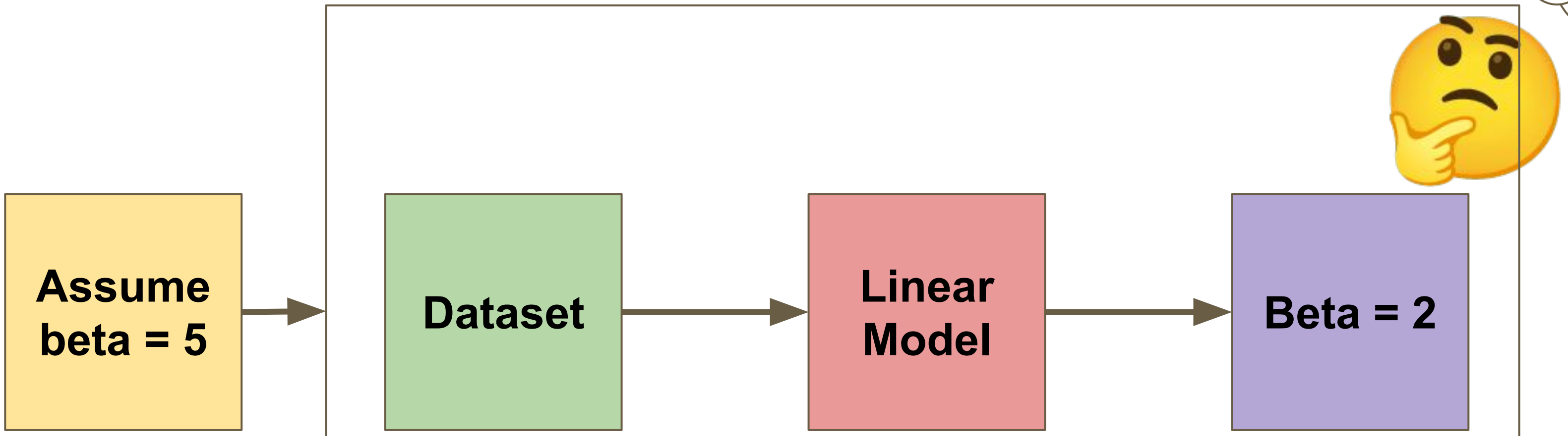

H H H H H H H H
The coin is probably not fair
H T H H T H T T
The coin could be fair

Hypothesis Testing



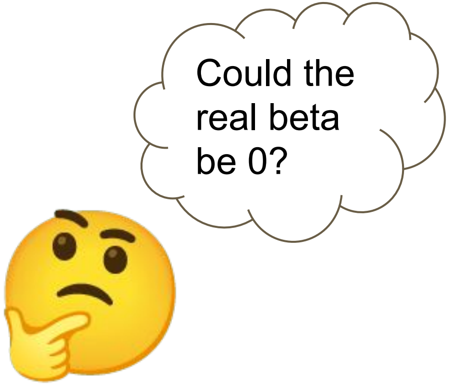
Hypothesis Testing





Is there enough evidence that we can reject that assumption / hypothesis?

worksheet



Hypothesis Testing

Each parameter of an independent variable **x** has an associated confidence interval and t-value + p-value.

If the parameter / coefficient is not significantly distinguishable from 0 then we cannot assume that there is a significant linear relationship between that independent variable and the observations **y** (i.e. if the interval includes 0 or if the p-value is too large)

Hypothesis Test

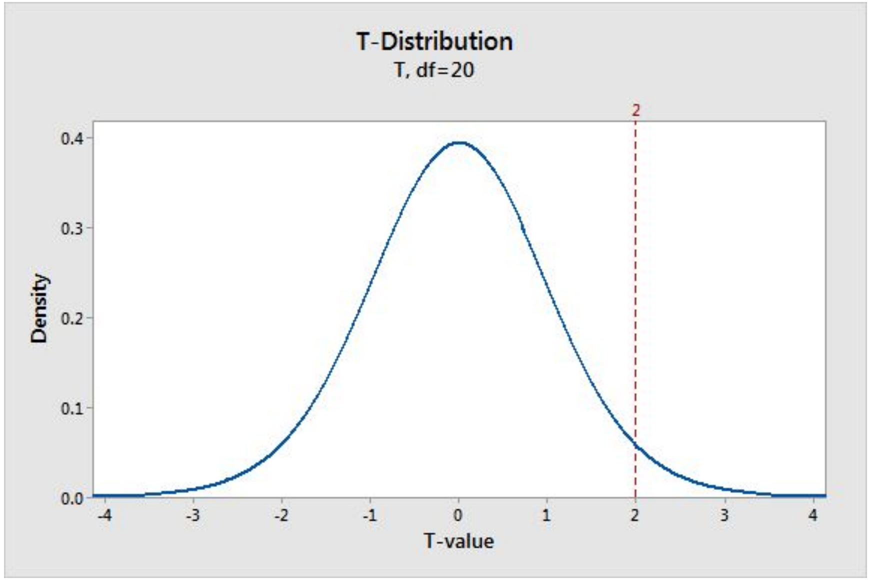
We want to know if there is evidence to reject the hypothesis $H0 : \beta = 0$ (i.e. that there is no linear relation between X and Y) using the information from β hat.

We want to know the largest probability of obtaining the data observed, under the assumption that the null hypothesis is correct.

How do we obtain that probability?

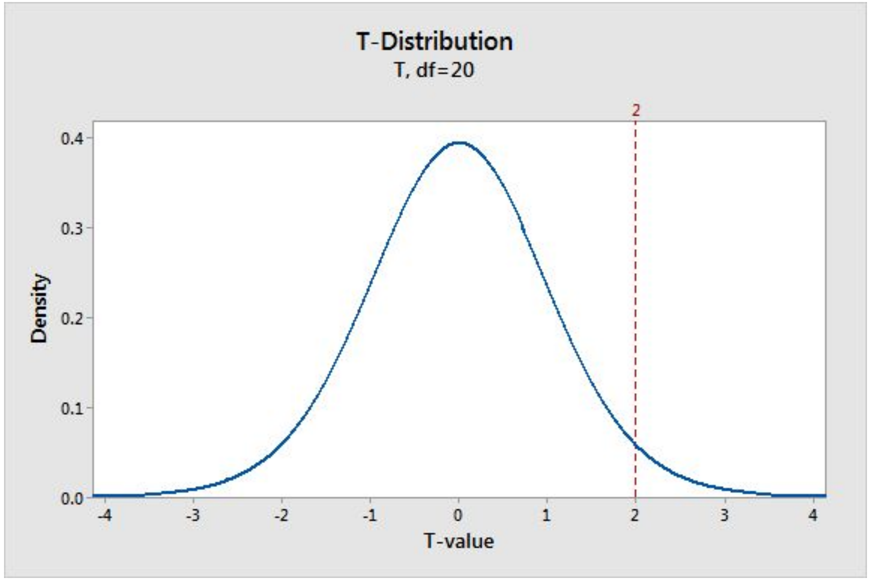
Hypothesis Test

Under the null hypothesis what should be the distribution of the normalized estimates? T-distribution (parametrized by the sample size)



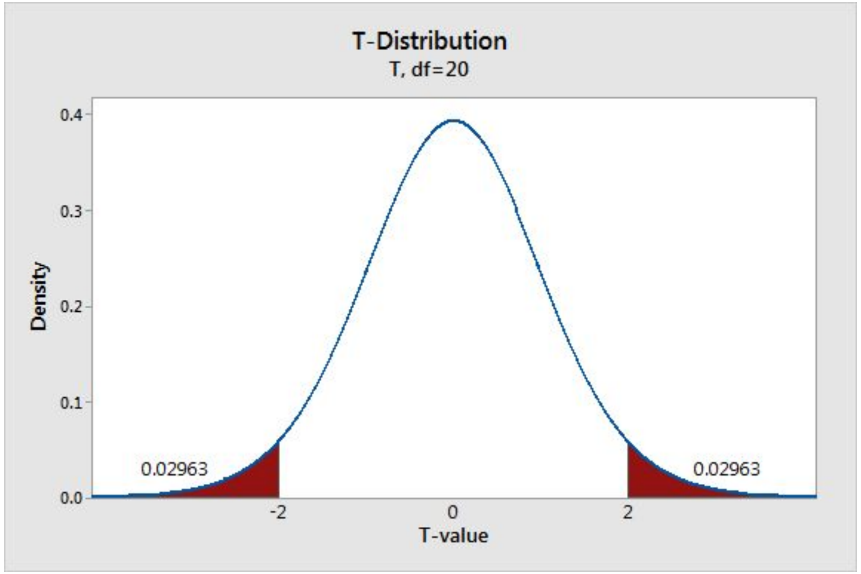
Hypothesis Test

We can then compute the t-value that corresponds to the sample we observed.



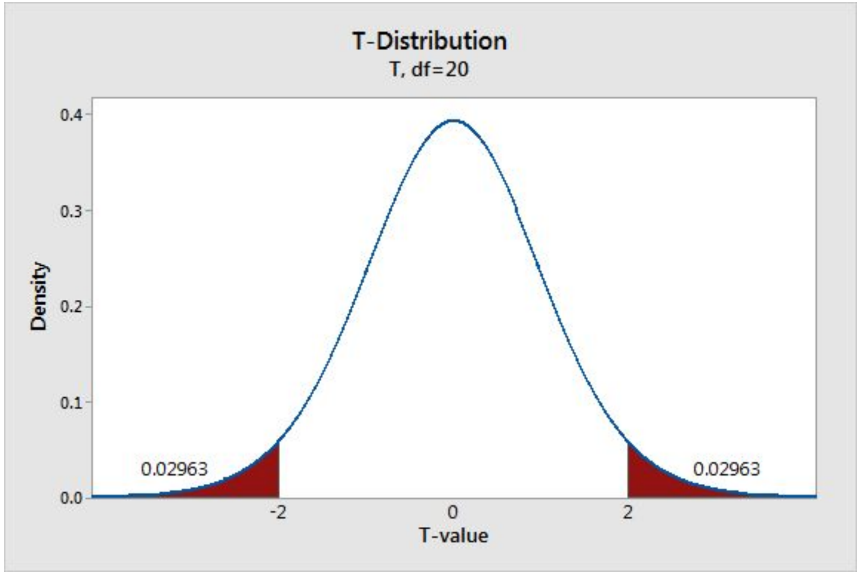
Hypothesis Test

And then compute the probability of observing estimates of β at least as extreme as the one observed. (i.e. trying to find evidence against H0)



Hypothesis Test

This probability is called a p-value.



Hypothesis Test

A p-value **smaller than a given threshold** would mean the data was unlikely to be observed under H0 so we can reject the hypothesis H0. If not, then we lack the evidence to reject H0.

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.1912	3.162	0.693	0.490	-4.085	8.467
x1	29.3912	3.274	8.977	0.000	22.893	35.889
x2	78.1391	3.594	21.741	0.000	71.006	85.272

Hypothesis Test

Which parameters should we not include in our linear model?

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.1912	3.162	0.693	0.490	-4.085	8.467
x1	29.3912	3.274	8.977	0.000	22.893	35.889
x2	78.1391	3.594	21.741	0.000	71.006	85.272

Evaluating our Regression Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.840
Model:                            OLS   Adj. R-squared:                  0.836
Method:                 Least Squares   F-statistic:                     254.1
Date:                Sun, 20 Mar 2022   Prob (F-statistic):           2.72e-39
Time:                        11:36:16   Log-Likelihood:                -482.37
No. Observations:                 100   AIC:                             970.7
Df Residuals:                      97   BIC:                             978.5
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.1912      3.162      0.693      0.490      -4.085       8.467
x1            29.3912      3.274      8.977      0.000      22.893      35.889
x2            78.1391      3.594     21.741      0.000      71.006      85.272
==============================================================================
Omnibus:                        1.279   Durbin-Watson:                   1.824
Prob(Omnibus):                  0.527   Jarque-Bera (JB):                1.065
Skew:                           0.253   Prob(JB):                        0.587
Kurtosis:                       2.999   Cond. No.                         1.38
==============================================================================
```

Confidence Intervals

Goal: for a given confidence level (let's say 90%), construct an interval around an estimate such that, if the estimation process were repeated indefinitely, the interval would contain the true value (that the estimate is estimating) 90% of the time.

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.1912	3.162	0.693	0.490	-4.085	8.467
x1	29.3912	3.274	8.977	0.000	22.893	35.889
x2	78.1391	3.594	21.741	0.000	71.006	85.272

Z-values

These are the number of standard deviations from the mean of a N(0,1) distribution required in order to contain a specific % of values were you to sample a large number of times.

To find the .95 z-value (the value z such that 95% of the observations lie within z standard deviations of the mean (μ ± z * σ)) you need to solve:

$$\int_{-z}^{z} \frac{1}{2\pi} e^{-\frac{1}{2}x^2} dx = .95$$

Z-values

The .95 z-value is 1.96.

This means 95% of observations from a $N(\mu, \sigma)$ lie within 1.96 standard deviations of the mean ($\mu \pm 1.960 * \sigma$)

If we get a sample from a $N(\mu, \sigma)$ of size n, how would we create a confidence interval around the estimated mean?

Confidence Intervals

How do we build a confidence interval?

Assume $\mathbf{Y_i \sim N(5, 25)}$, for $\mathbf{1 \leq i \leq 100}$ and $\mathbf{y_i = \mu + \epsilon}$ where $\mathbf{\epsilon \sim N(0, 25)}$. Then the Least Squares estimator of $\boldsymbol{\mu}$ ($\boldsymbol{\mu_{LS}}$) is

the sample mean $\mathbf{\bar{y}}$

What is the 95% confidence interval for $\boldsymbol{\mu_{LS}}$?

$$\begin{aligned} CI_{.95} &= [\bar{y} - 1.96 \times SE(\mu_{LS}), \bar{y} + 1.96 \times SE(\mu_{LS})] \\ &= [\bar{y} - 1.96 \times .5, \bar{y} + 1.96 \times .5] \end{aligned}$$

$$\begin{aligned} SE(\mu_{LS}) &= \sigma_{\epsilon} / \sqrt{n} \\ &= 5 / \sqrt{100} \\ &= .5 \end{aligned}$$

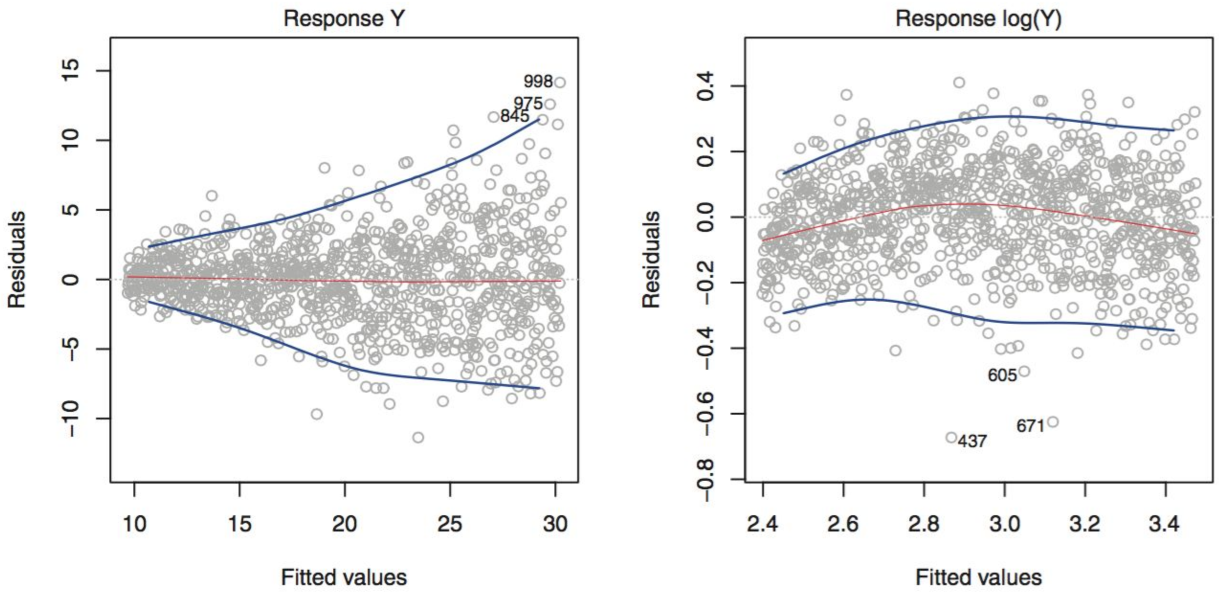
Z-value for 95% Confidence Interval

Checking our Assumptions

1. Normal Distribution?
2. Constant Variance?

Constant Variance

One of our assumptions was that our noise had constant variance. How can we verify this?
We can plot residuals (noise estimates) for each fitted value $\hat{\mathbf{y}}_i$



Extending our Linear Model

Changing the assumptions we made can drastically change the problem we are solving. A few ways to extend the linear model:

1. Non-constant variance - used in WLS (weighted least squares)
2. Distribution of error is not Normal - used in GLM (generalized linear models)