Logistics

- The project deadline has been extended due to scheduled system maintenance. See the assignment version 1.3 on UNM learn for details.

Quiz

- Define frequentist inference in opposition to Bayesian inference.
- 5 mins

Regression Analysis

MATTHEW FRICKE

VERSION 1.0 – SEND CORRECTIONTS TO MFRICKE@UNM.EDU

Reversion to the Mean

- *‘‘‘... **while attempting to teach flight instructors that praise is more effective than punishment for promoting skill-learning...one of the most seasoned instructors in the audience raised his hand and made his own short speech..."On many occasions I have praised flight cadets for clean execution of some aerobatic maneuver, and in general when they try it again, they do worse. On the other hand, I have often screamed at cadets for bad execution, and in general they do better the next time. So please don't tell us that reinforcement works and punishment does not, because the opposite is the case."** ...because we tend to reward others when they do well and punish them when they do badly, and because there is regression to the mean, it is part of the human condition that we are statistically punished for rewarding others and rewarded for punishing them.’’’*
- Kahneman, D., 2002, Bank of Sweden "Nobel" Prize Lecture

$P(x)$ x

Reversion to the Mean

Reversion to the mean, is the statistical phenomenon that the greater the deviation of a random variate from its mean, the greater the probability that the next measured variate will deviate less far.

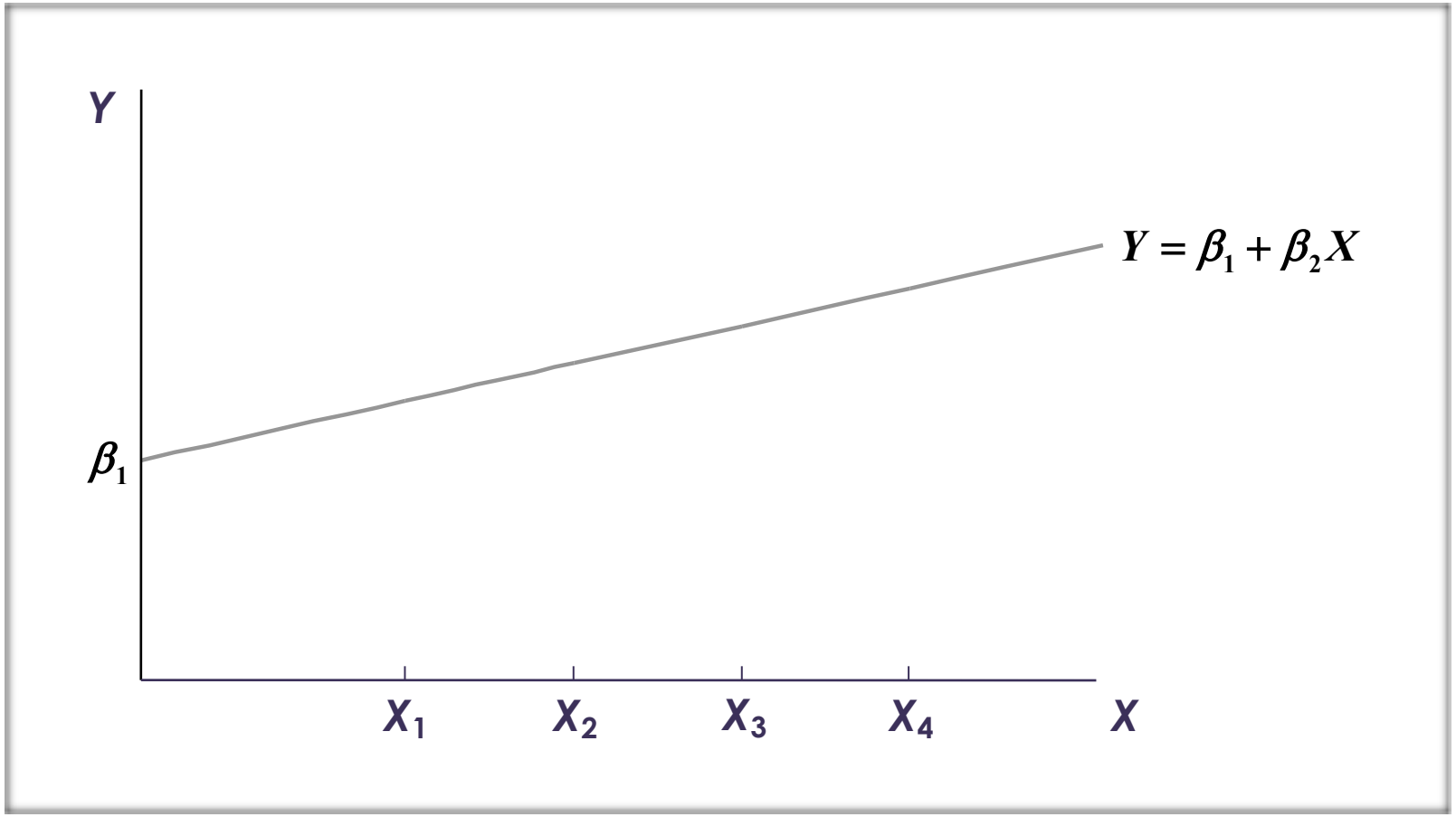
Doesn't this violate the definition of independent events?

But this is a consequence of probability distributions suming to 1 and are non-negative .Thus, as you move away from the mean, the proportion of the distribution that lies closer to the mean than you do increases continuously.

Linear Regression

Suppose that a variable *Y* is a linear function of another variable *X*, with unknown parameters β_1 and β_2 that we wish to estimate.

The line is just a uninformed model so far.



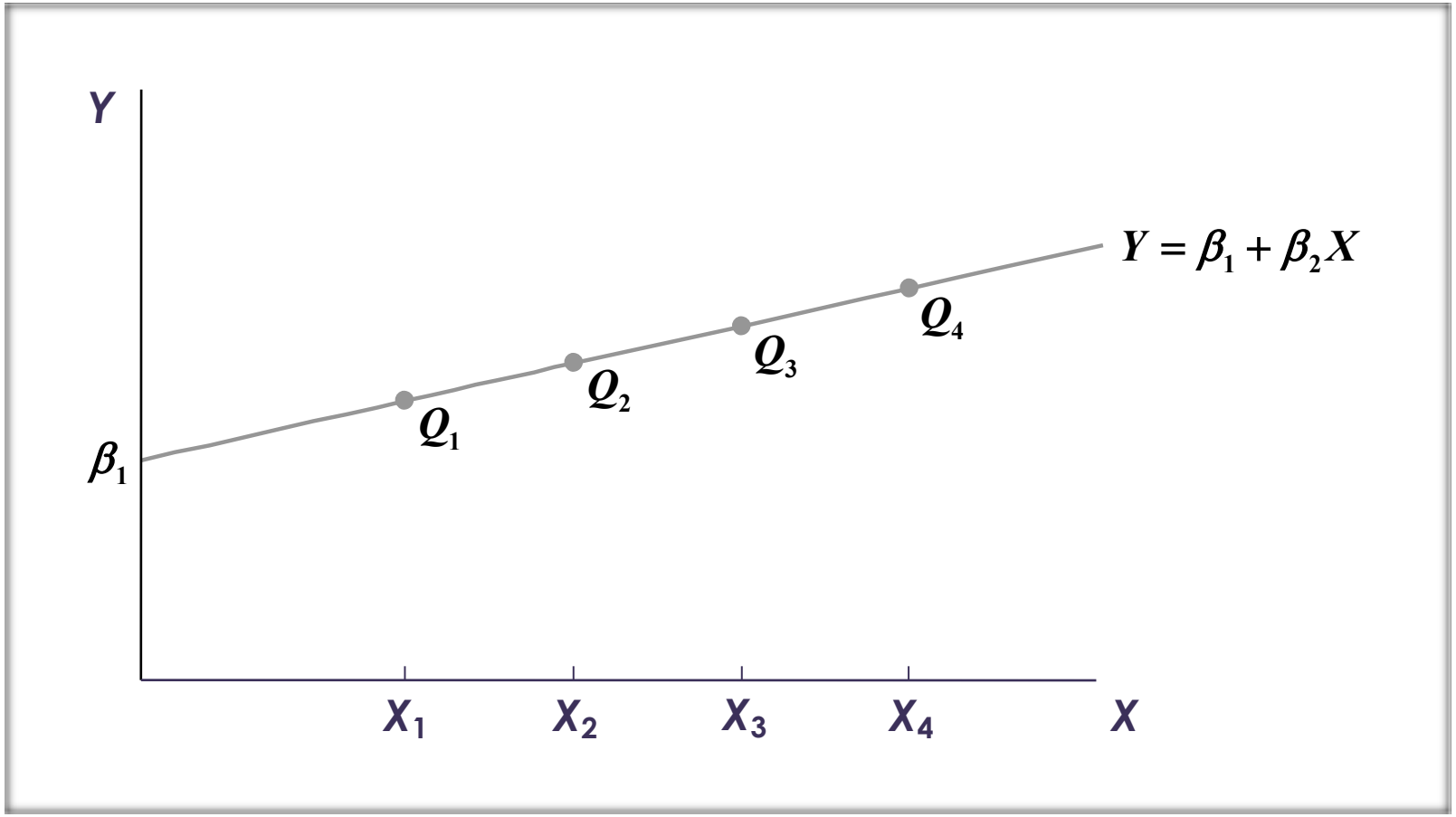
C. Dougherty, *Introduction to Econometrics*, fifth edition 2016, Oxford University Press

Linear Regression

Imagine we have some observations $Q_i = (X_i, Y_i)$. Where X_i is a particular factor level and Y_i is the response.

In this case the coefficients of our model β_1 and β_2 are easy to find.

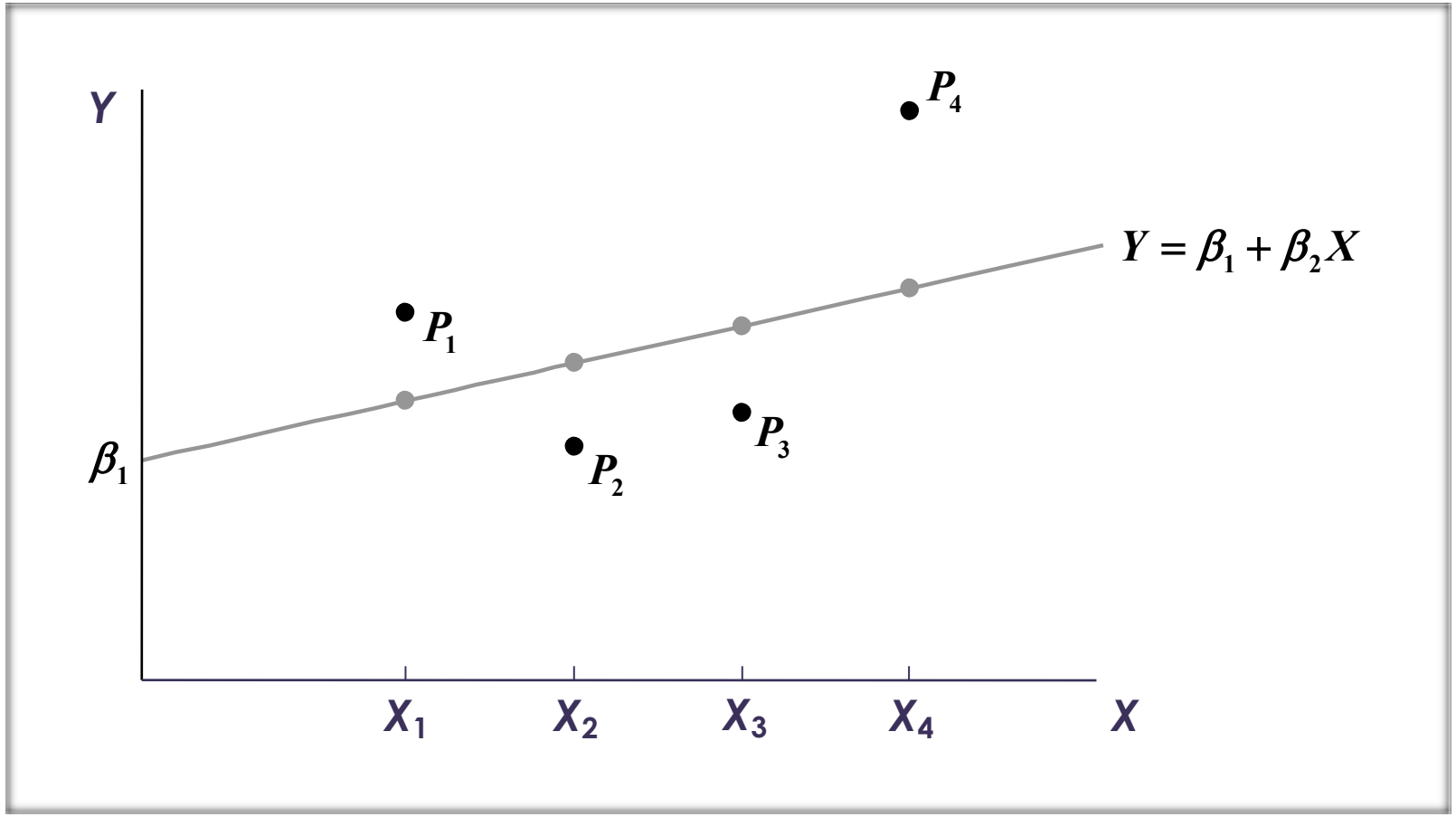
Notice we have defined a 1D response "surface" (ok its just a line).



Linear Regression

But life is rarely that straightforward, $P_i = (X_i, Y_i)$.

We usually have observations that do not lie on a straight line – but where a linear model may still be appropriate.

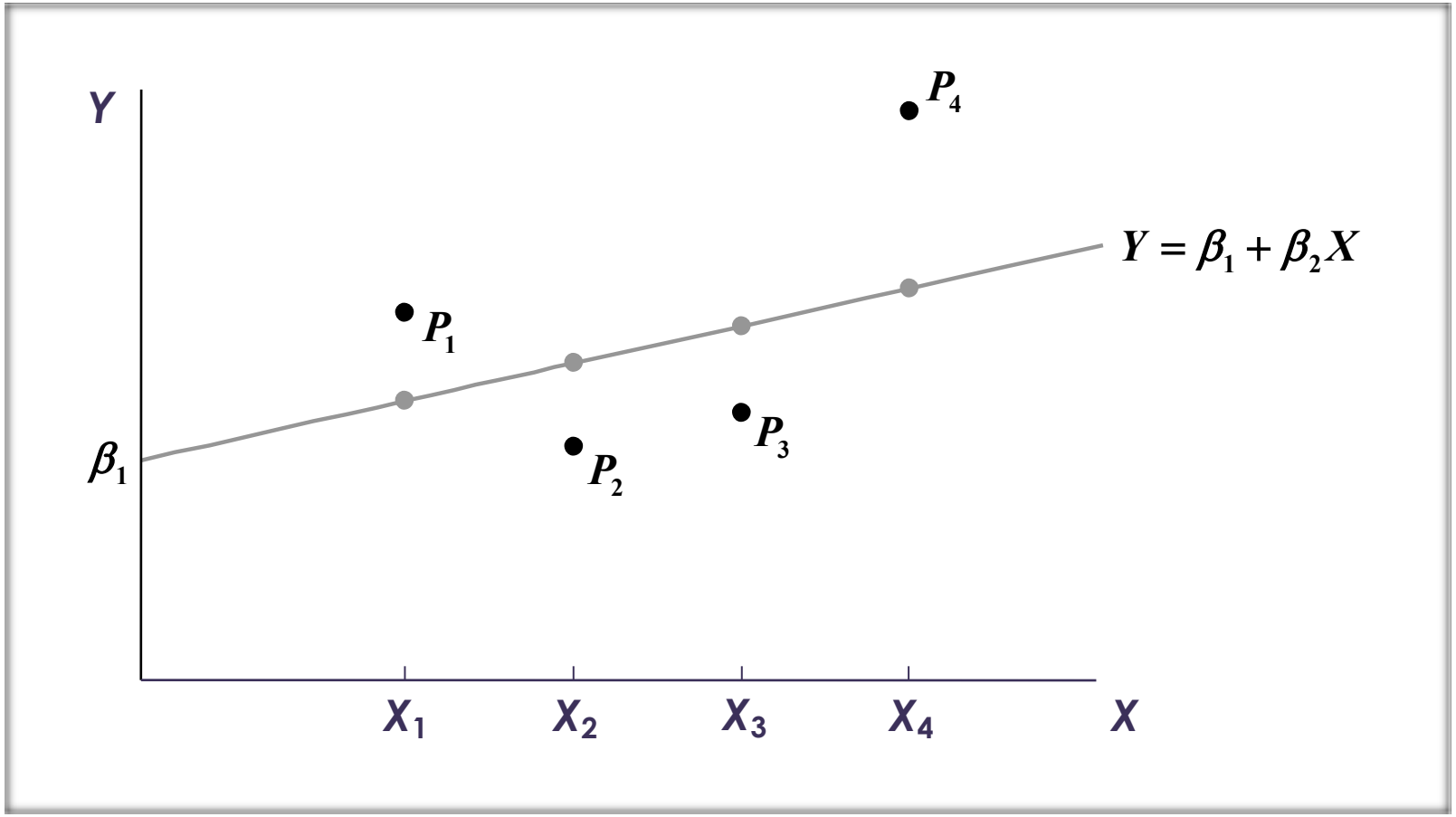


Linear Regression

But life is rarely that straightforward, $P_i = (X_i, Y_i)$.

We usually have observations that do not lie on a straight line – but where a linear model may still be appropriate.

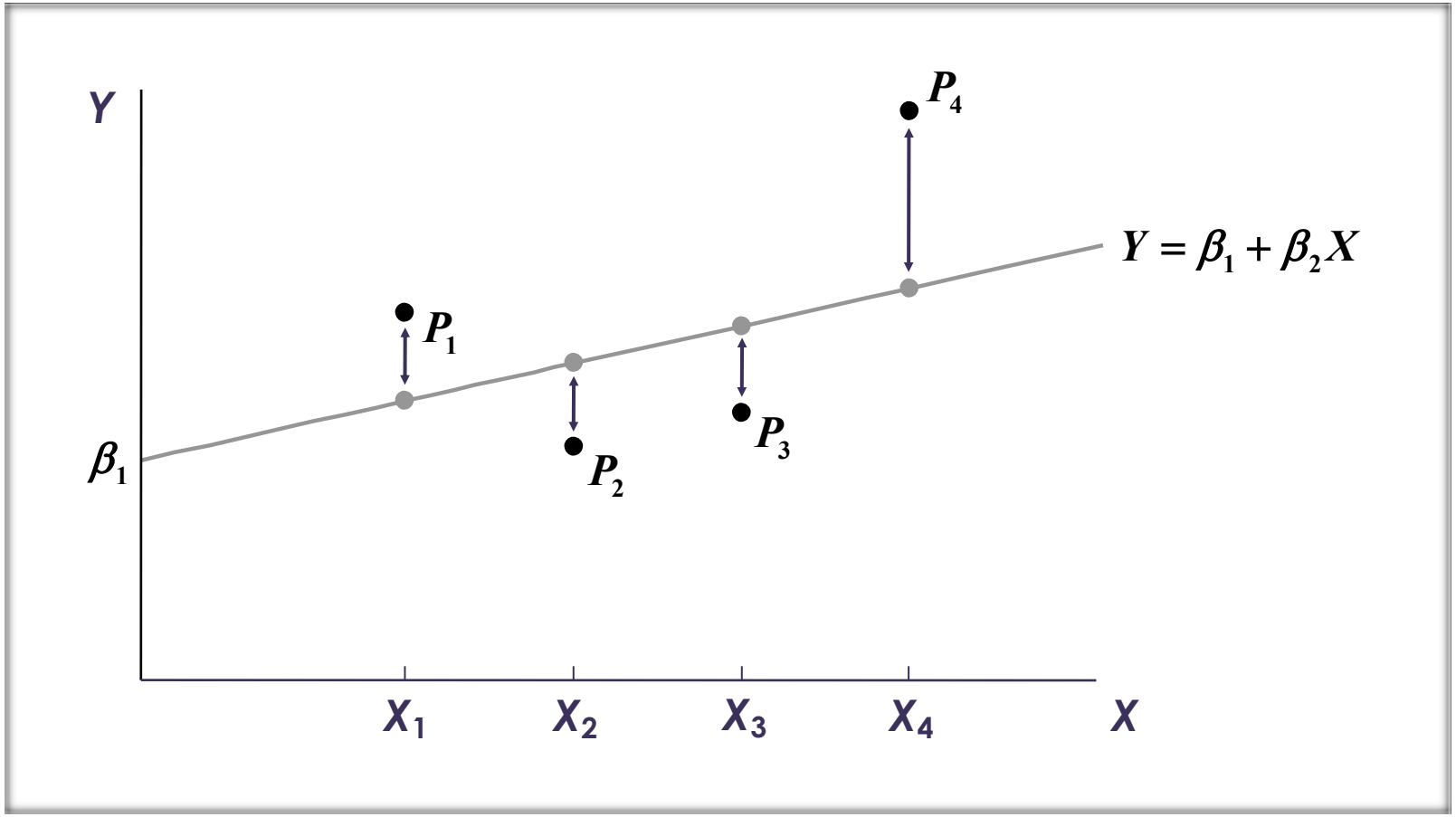
Why might a linear model be appropriate even when the observations are scattered like this?



Linear Regression

Why might a linear model be appropriate even when the observations are scattered like this?

Because we might care about the underlying process, and we recognise that there will be variance in our observations.

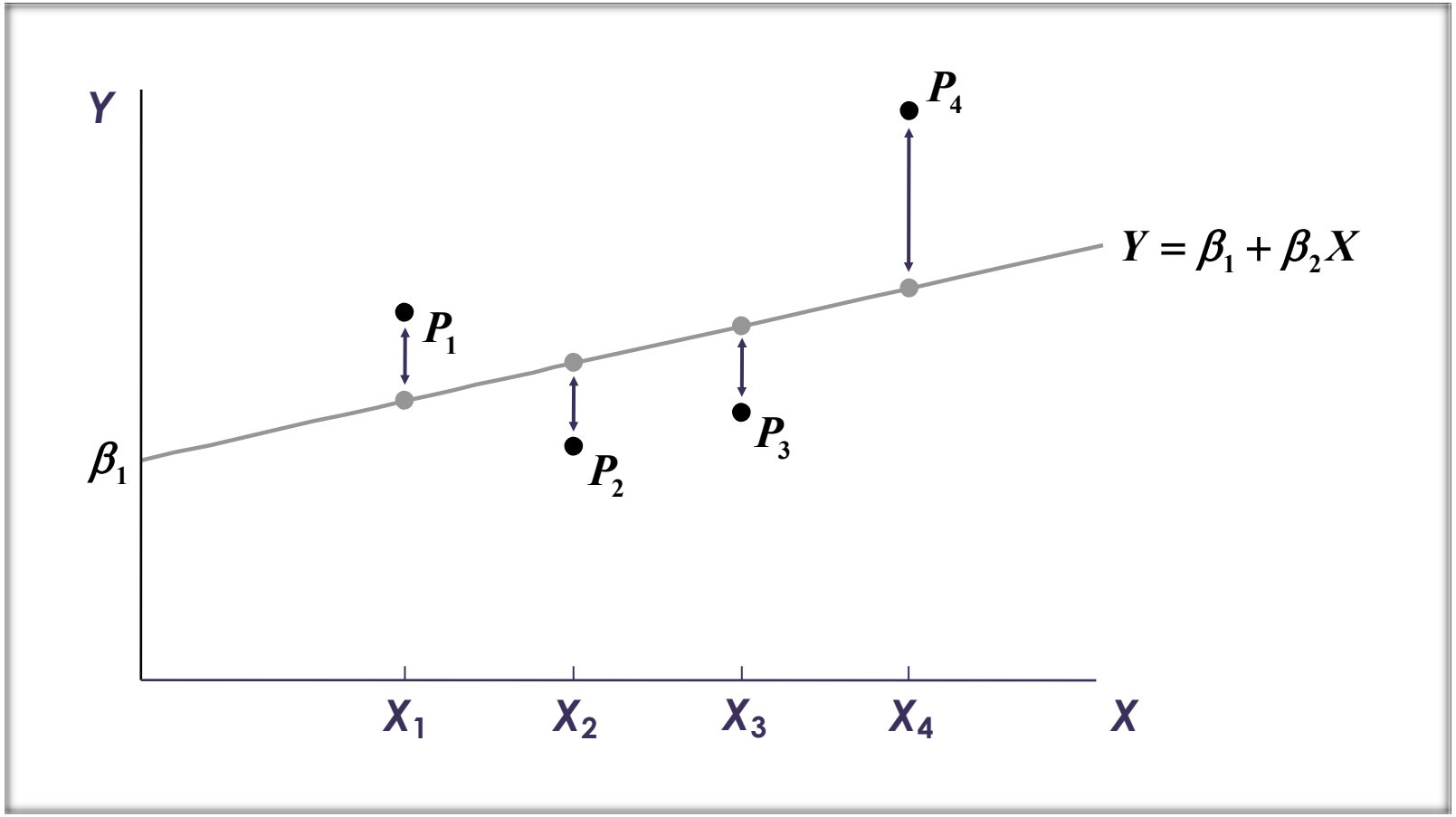


Linear Regression

To be more rigorous we introduce a ***disturbance term****: u*

So now our model of the relationship between X_i and Y_i is:

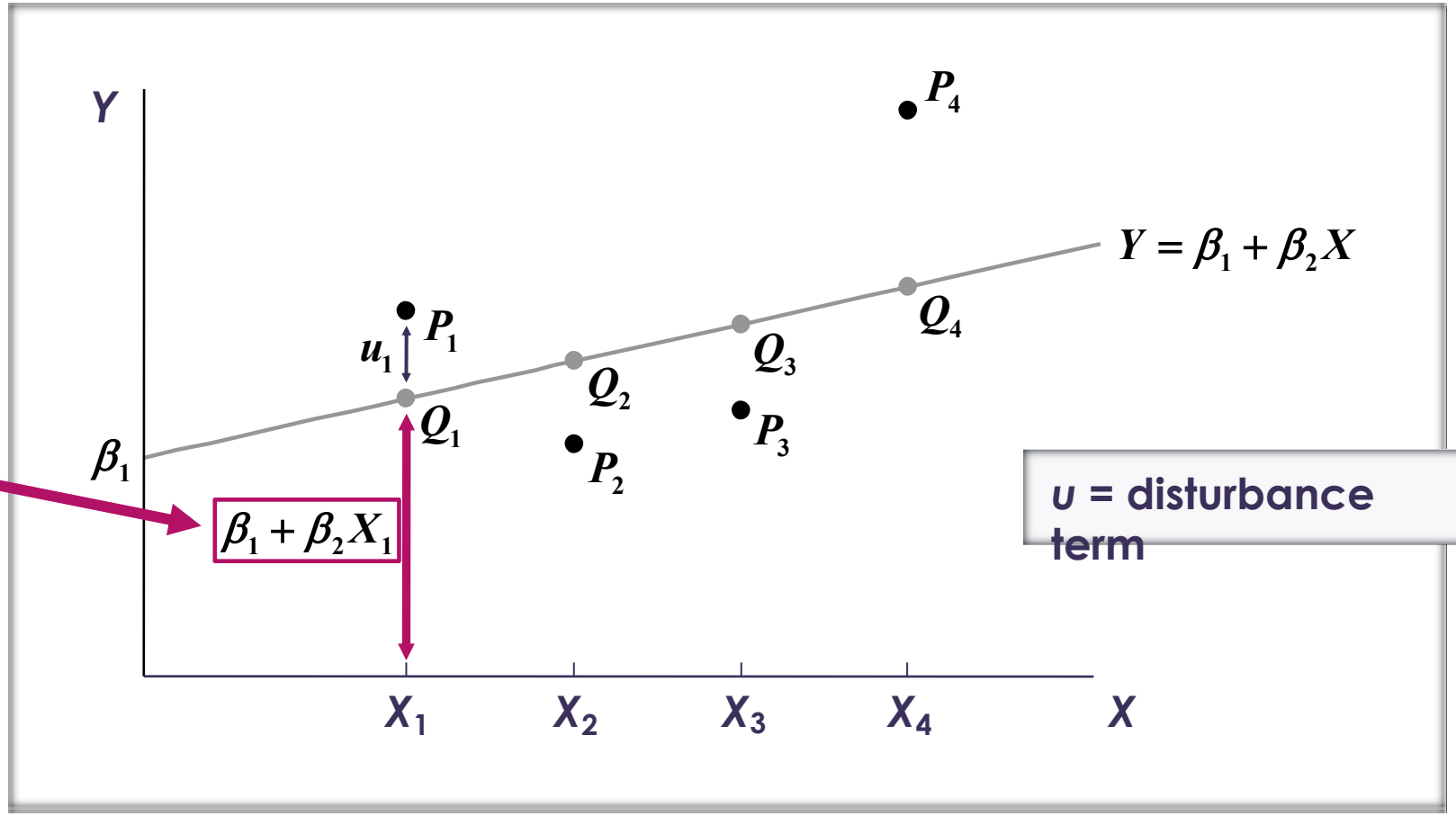
$$Y = \beta_1 + \beta_2 X + u$$



Linear Regression

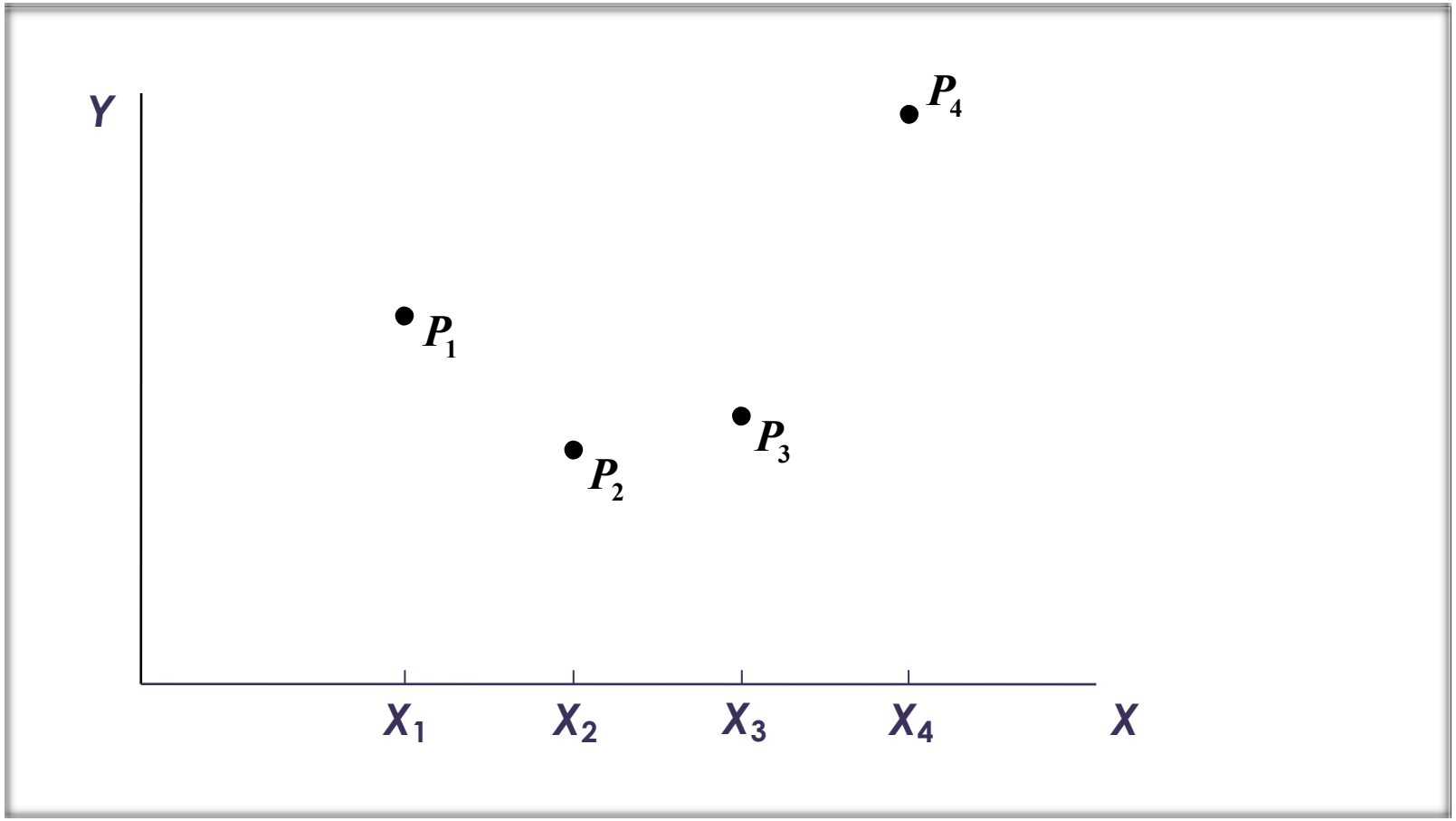
$$Y = \beta_1 + \beta_2 X + u$$

Our model has a deterministic term…



Linear Regression

In practice (your project for example) all you will see are your experimental outcomes P_i.

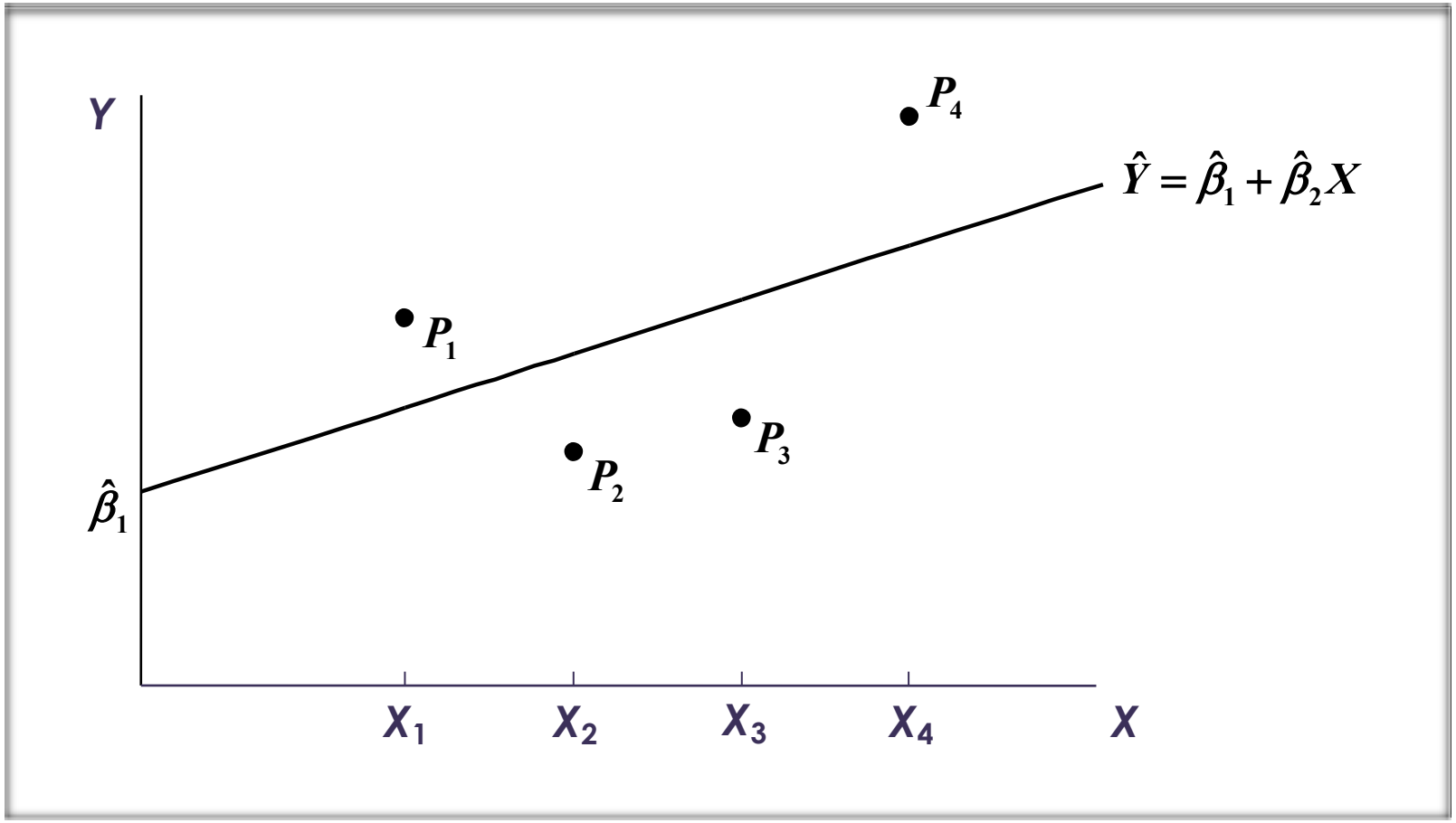


Linear Regression

Obviously we can draw a straight line through these points.

The notation is to write our estimate as

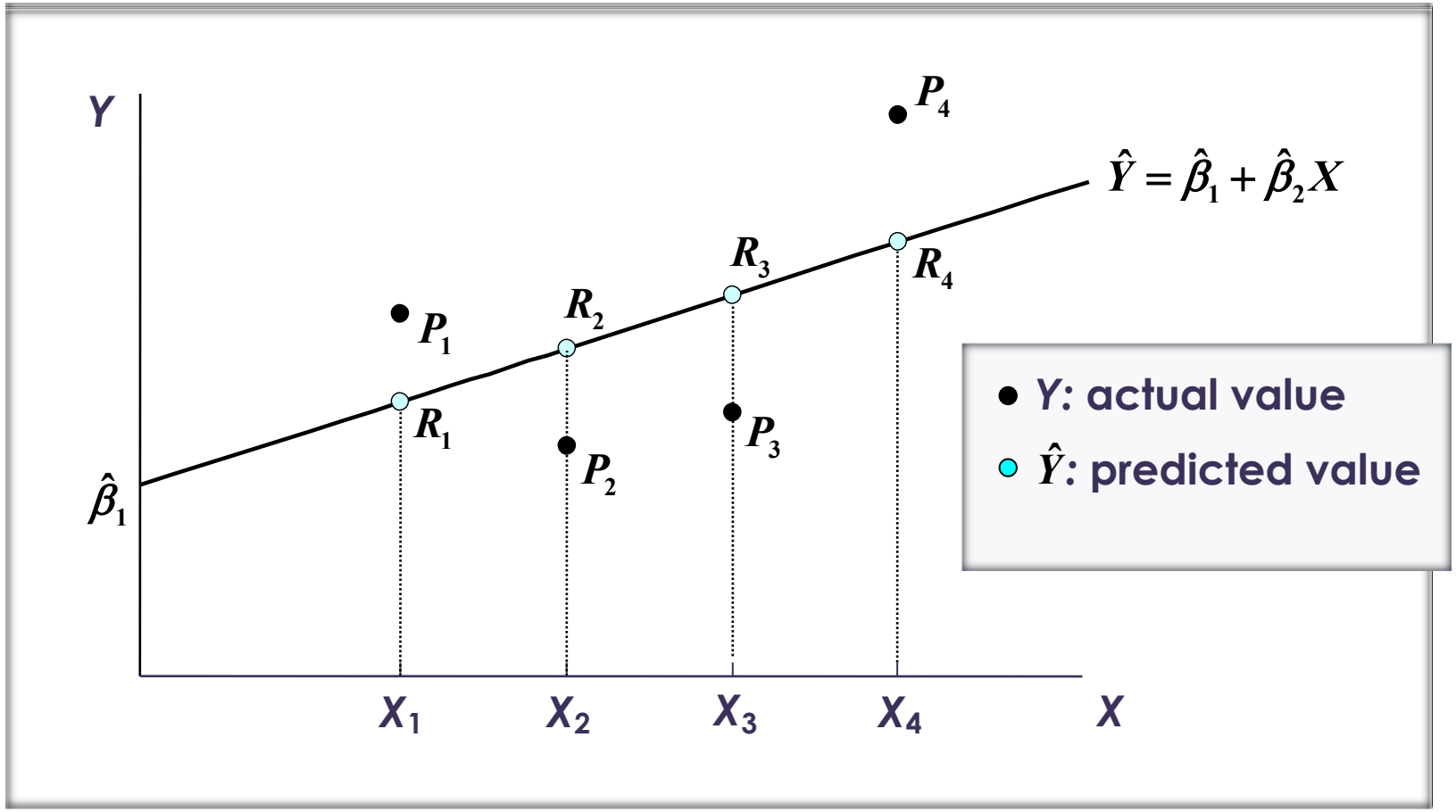
$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$



Linear Regression

*We have now **fit our model** (the line) which predicts values, R_i.*

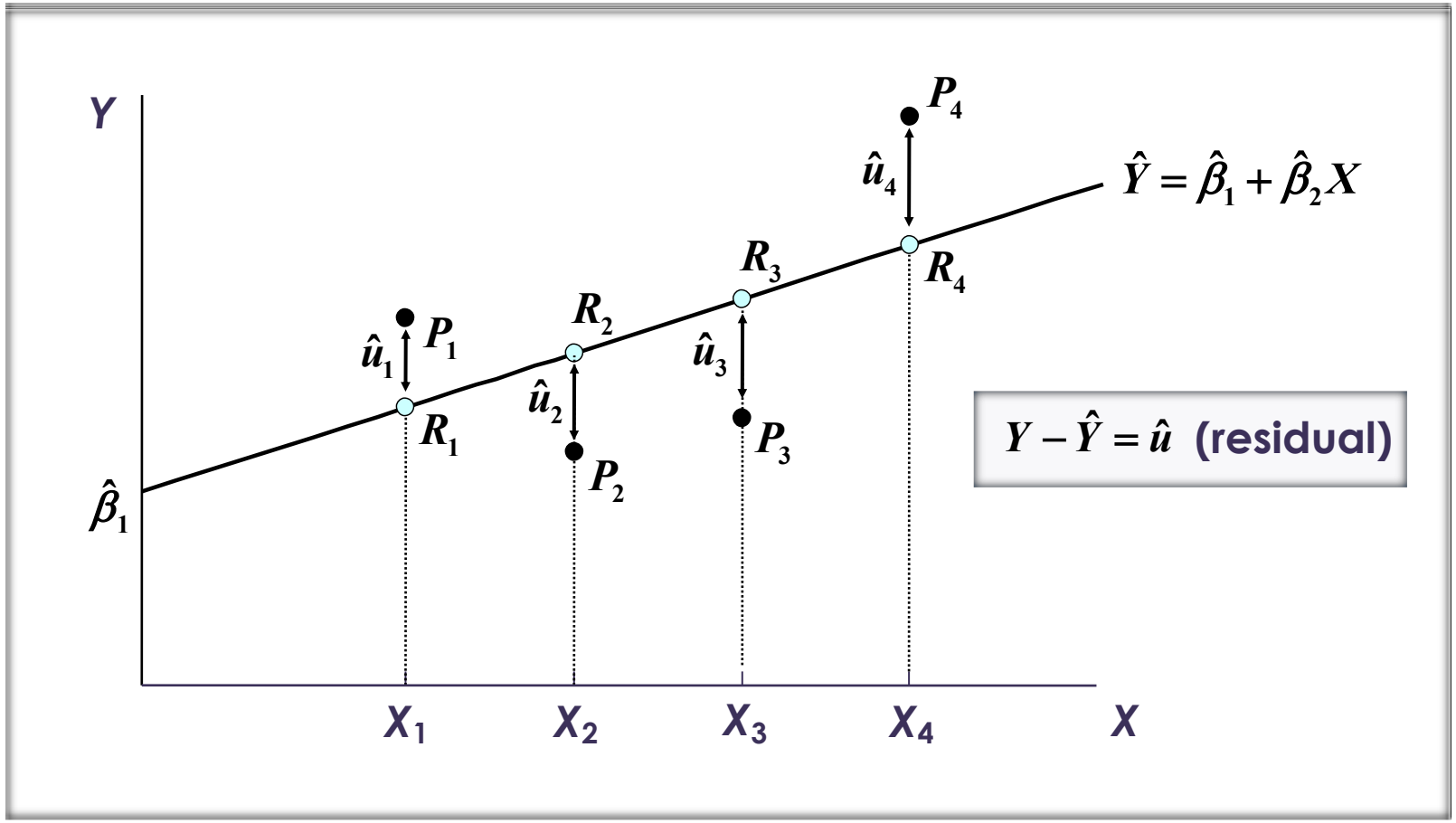
We can start to think about how well it represents the experimental outcomes.



Linear Regression

The difference between a value predicted by the model and the observed value is the ***residual,*** $\hat{u}$

The ***residuals*** *are an estimate of the disturbance* ***term****, but are not the same as the disturbance term.*

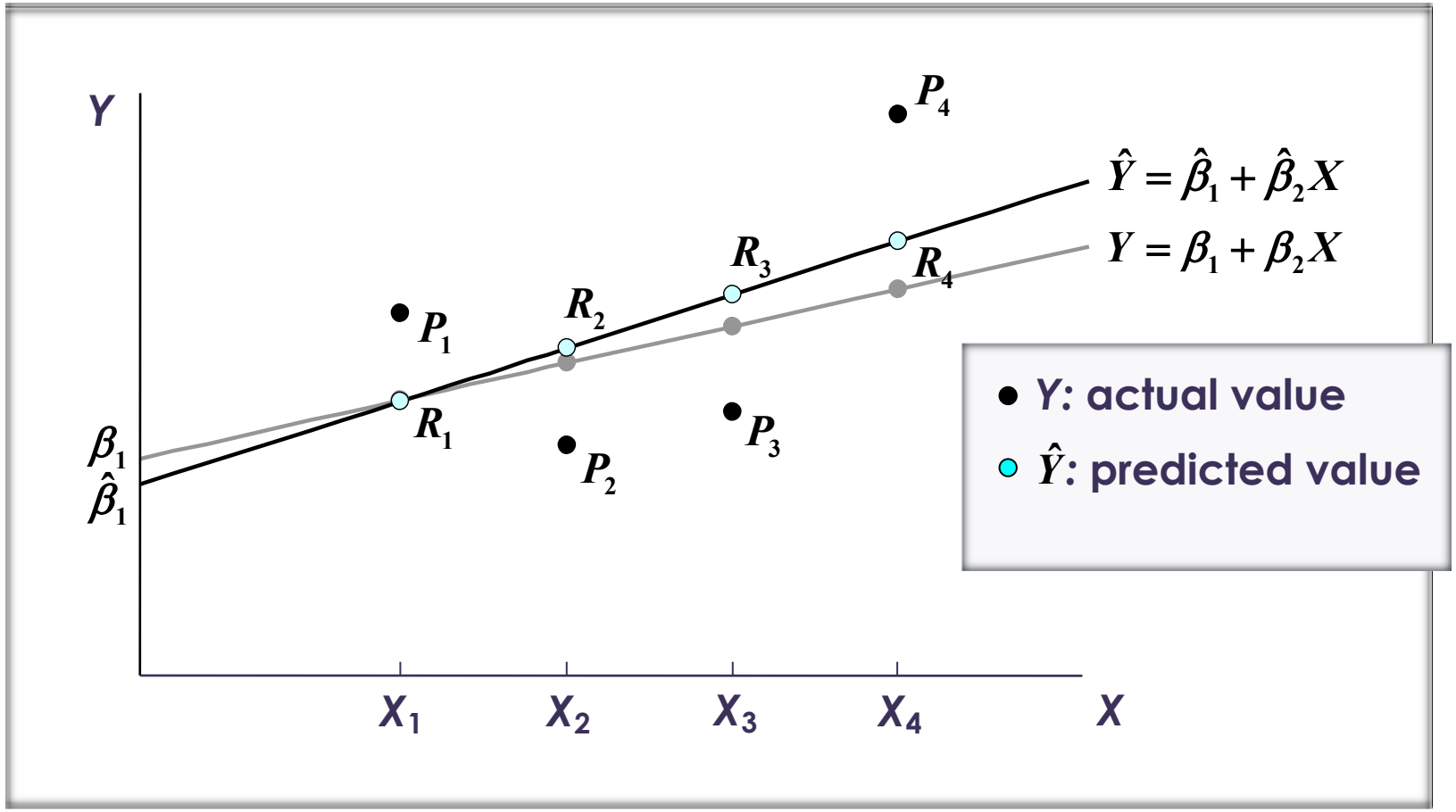


Linear Regression

*The **residuals** are an estimate of the disturbance **term**, but are not the same as the disturbance term.*

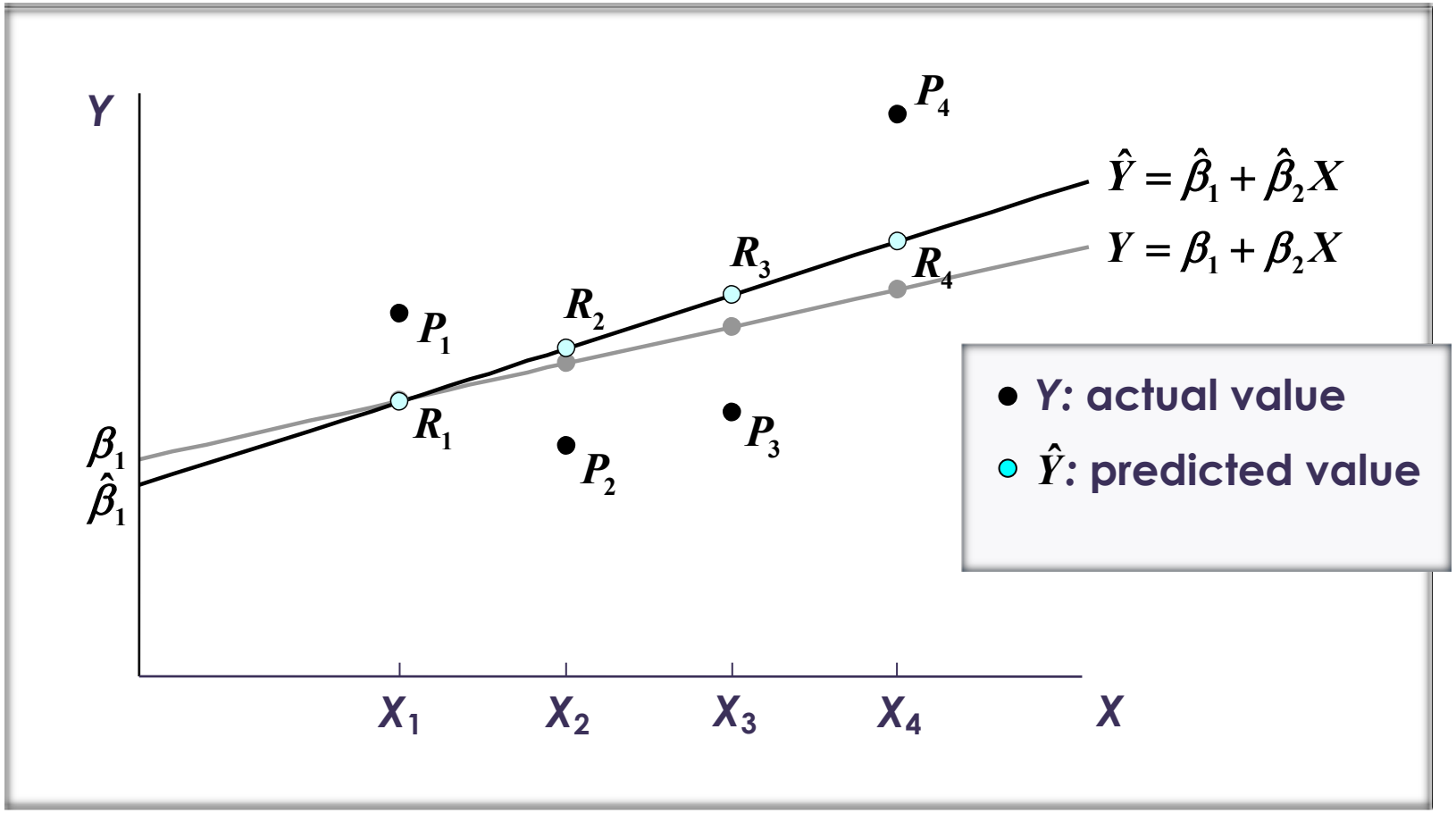
$Y = \beta_1 + \beta_2 X$ is now the deterministic part of the process that generated the observations.

I'm just saying the model fit is wrong because of the disturbance term.



Linear Regression

The better the fit the closer the residuals will approach the disturbance term.

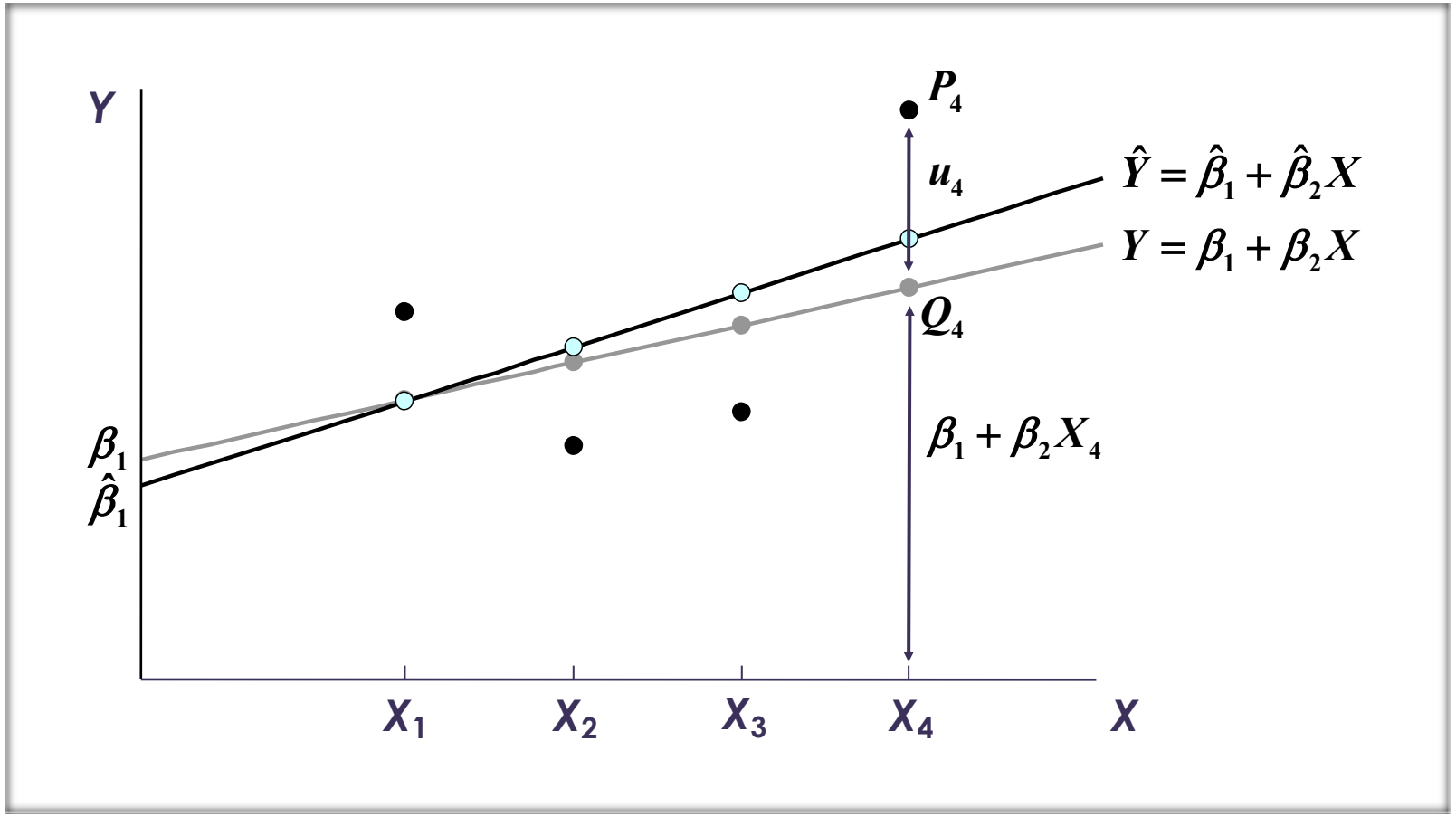


Linear Regression

The better the fit the closer the residuals will approach the disturbance term.

Conceptually we will use the residuals and disturbance term to decompose Y into its deterministic and random components.

Practically we will use the fit line to do this.



Linear Regression

To begin with, we will draw the fitted line so as to minimize the sum of the squares of the residuals, *RSS*. This is described as the least squares criterion.

Least squares criterion:

Minimize RSS (residual sum of squares), where

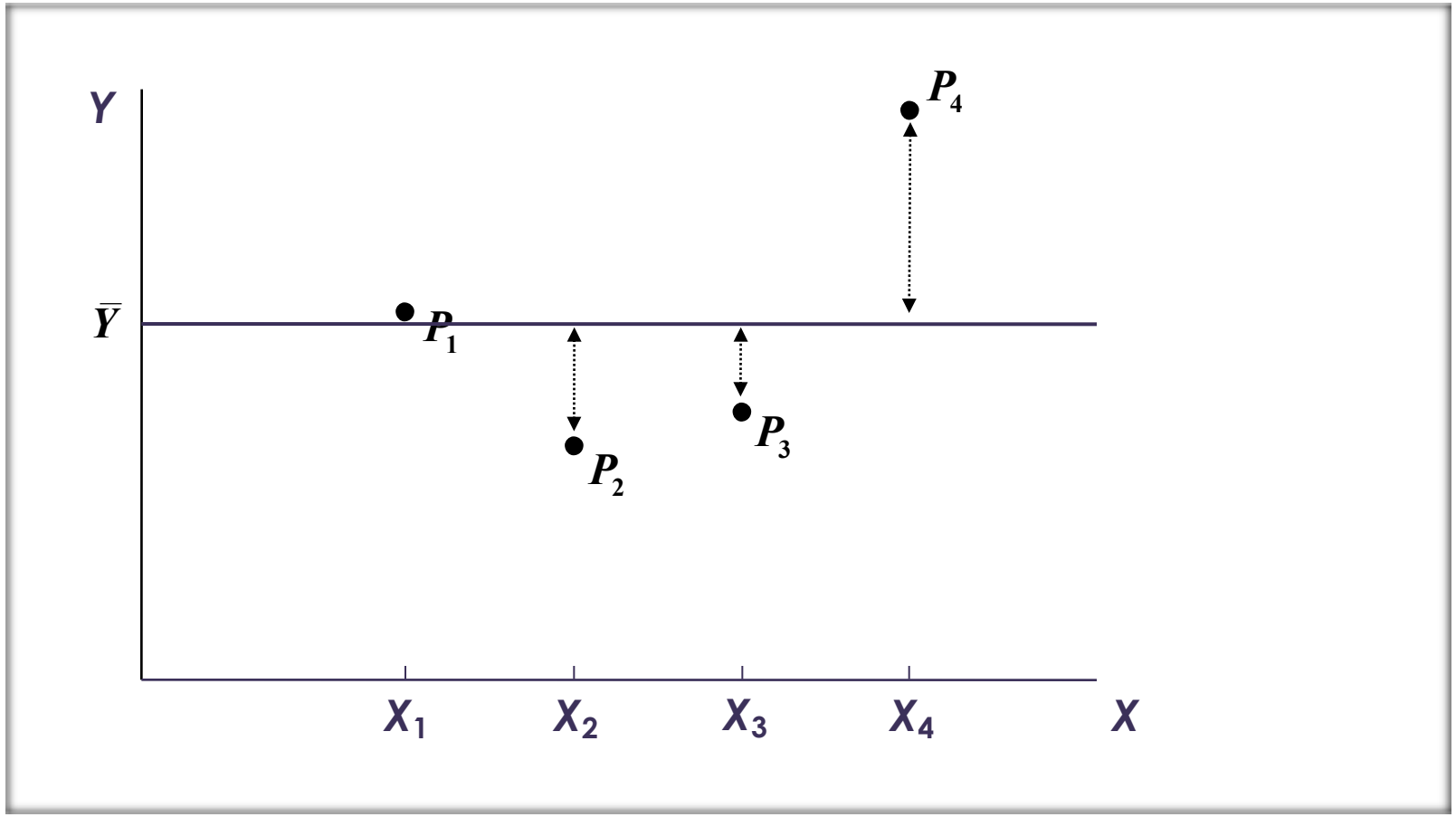
$$RSS = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{u}_1^2 + ... + \hat{u}_n^2$$

Linear Regression

Why minimise the squares instead of the actual residuals?

$$\sum_{i=1}^{n} \hat{u}_i = \hat{u}_1 + \ldots + \hat{u}_n$$

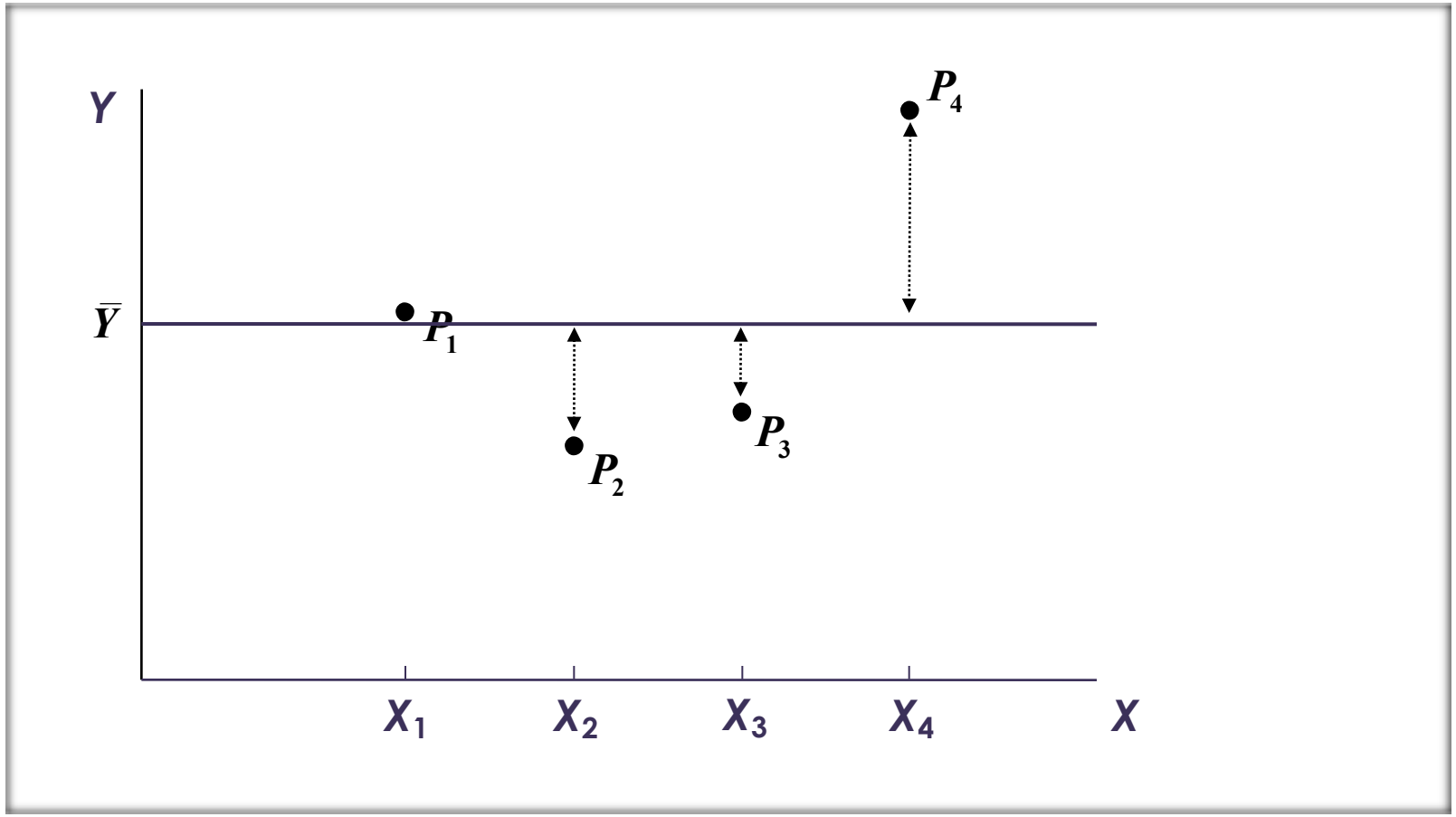
Because then drawing a line through the mean gives you an apparently perfect fit. The sum of residuals would always be zero.



Linear Regression

Why minimise the squares instead of the actual residuals?

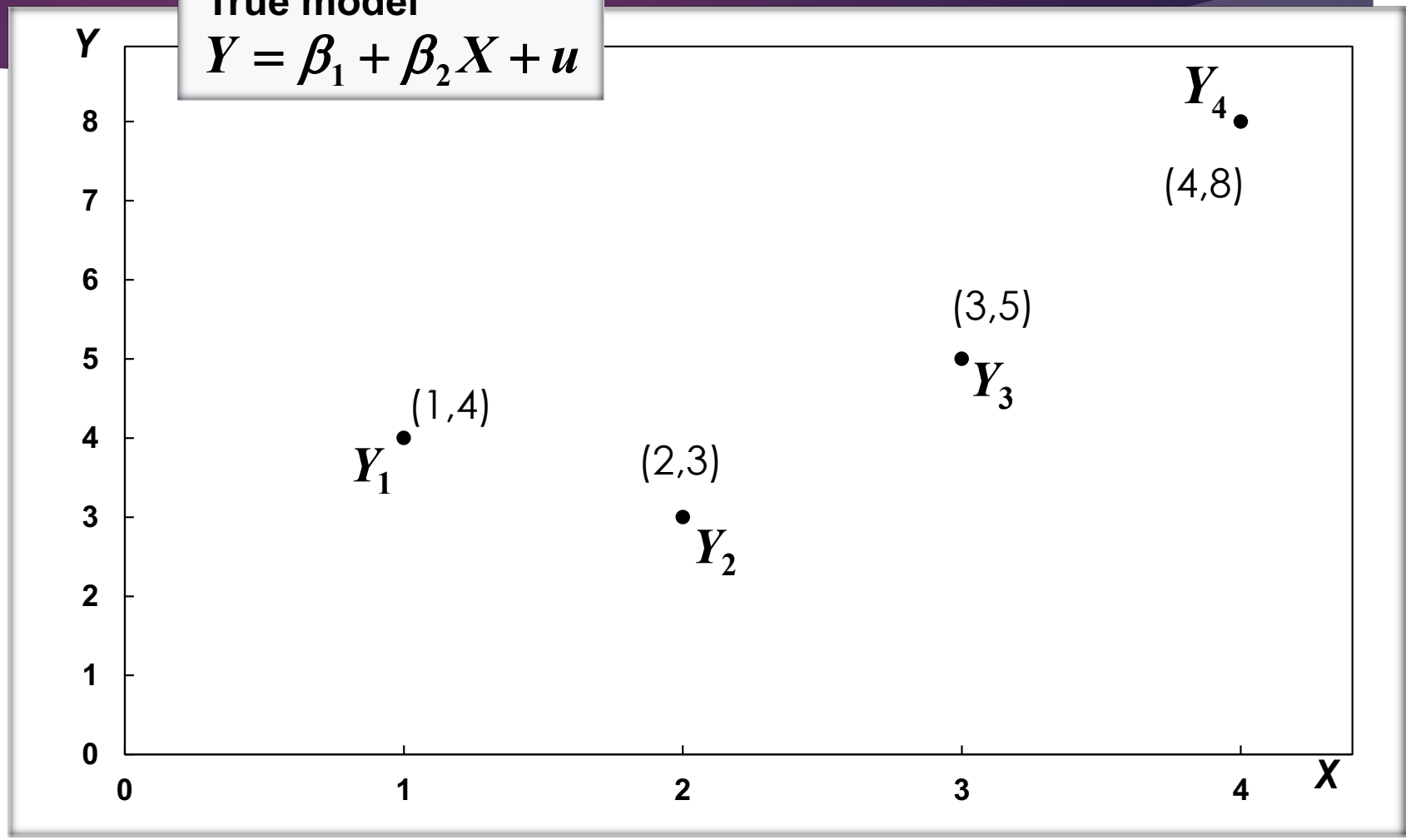
Taking the square ensures all the values being summed are positive so they can't cancel each other.



Linear Regression - Example

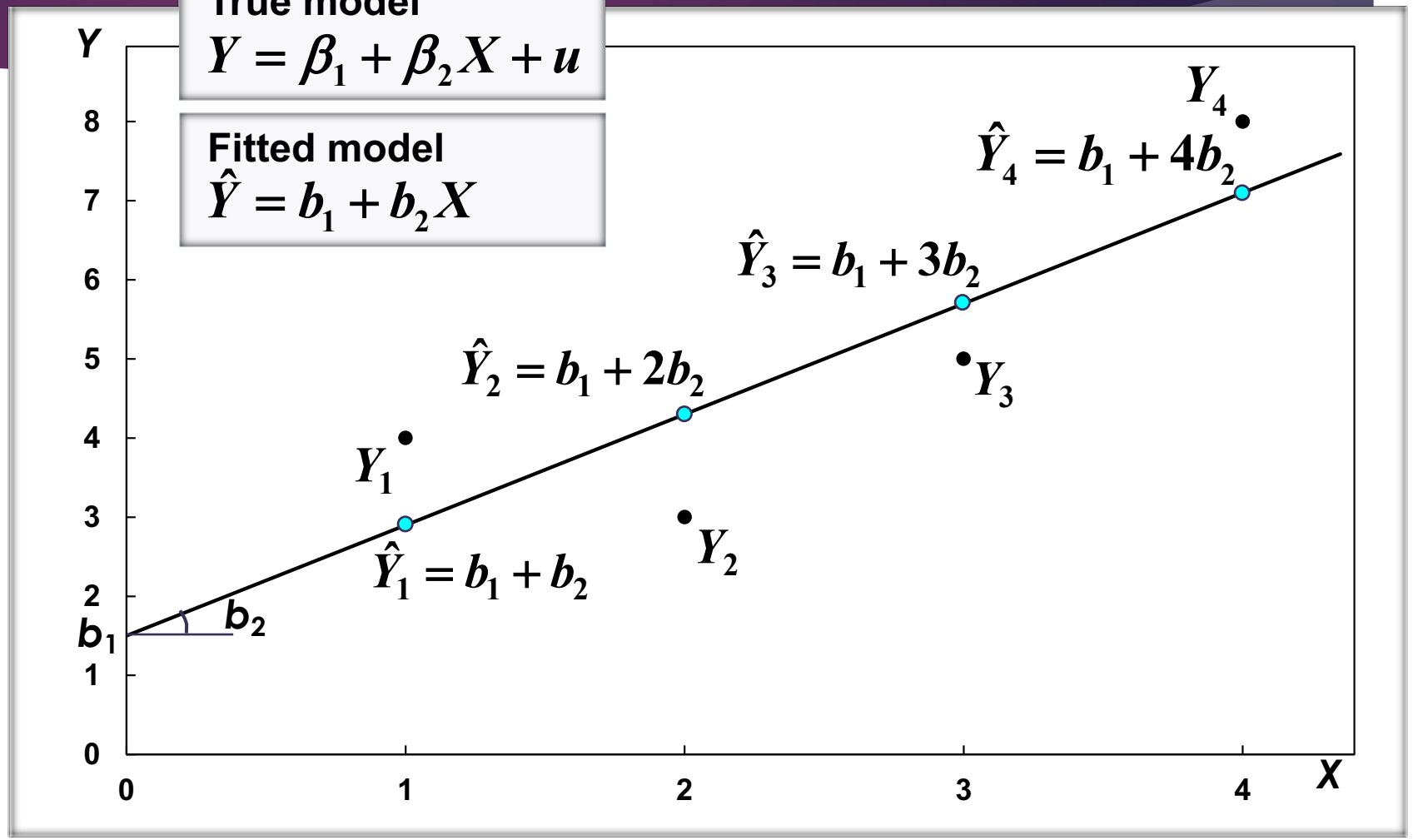
Given a set of factor levels X, and responses Y as shown we will determine the best linear fit.

You have already seen a maximum likelihood way of doing this??



Linear Regression - Example

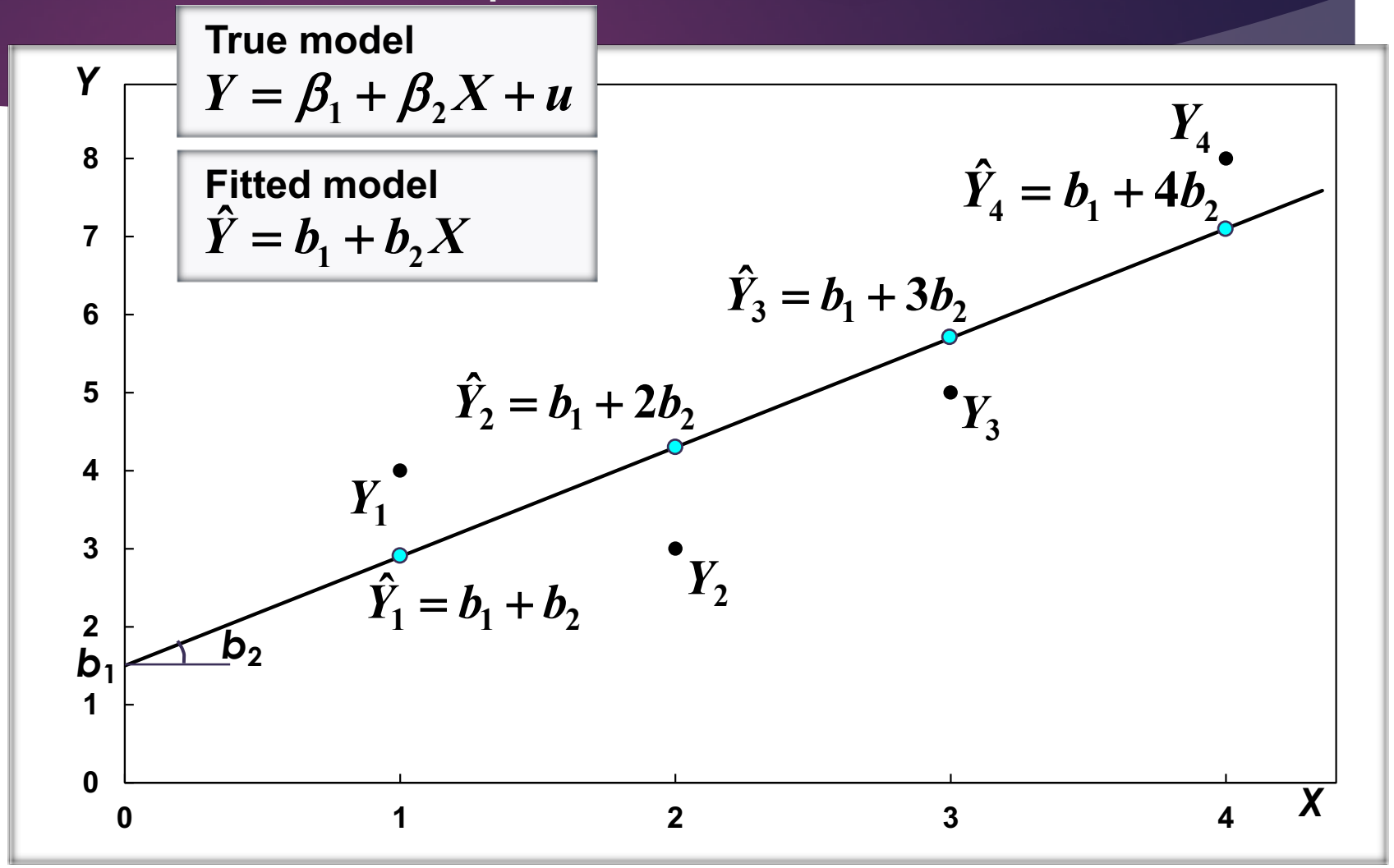
Writing the fitted regression as $\hat{Y} = b_1 + b_2 X$, we will determine the values of b_1 and b_2 that minimize *RSS*, the sum of the squares of the residuals.



Linear Regression - Example

Our residuals will be:

$$\hat{u}_1 = Y_1 - \hat{Y}_1 = 4 - b_1 - b_2$$
$$\hat{u}_2 = Y_2 - \hat{Y}_2 = 3 - b_1 - 2b_2$$
$$\hat{u}_3 = Y_3 - \hat{Y}_3 = 5 - b_1 - 3b_2$$
$$\hat{u}_4 = Y_4 - \hat{Y}_4 = 8 - b_1 - 4b_2$$



Linear Regression - Example

We can write the RSS for our observations as a function of b_1 and b_2.

$$
\begin{aligned}
RSS &= (4-b_1-b_2)^2+(3-b_1-2b_2)^2+(5-b_1-3b_2)^2+(8-b_1-4b_2)^2 \\
&= 16+b_1^2+b_2^2-8b_1-8b_2+2b_1b_2 \\
&\quad +9+b_1^2+4b_2^2-6b_1-12b_2+4b_1b_2 \\
&\quad +25+b_1^2+9b_2^2-10b_1-30b_2+6b_1b_2 \\
&\quad +64+b_1^2+16b_2^2-16b_1-64b_2+8b_1b_2 \\
&= 114+4b_1^2+30b_2^2-40b_1-114b_2+20b_1b_2
\end{aligned}
$$

Linear Regression - Example

We can write the RSS for our observations as a function of b_1 and b_2.

$$
\begin{aligned}
RSS &= (4-b_1-b_2)^2+(3-b_1-2b_2)^2+(5-b_1-3b_2)^2+(8-b_1-4b_2)^2 \\
&= 16+b_1^2+b_2^2-8b_1-8b_2+2b_1b_2 \\
&\quad +9+b_1^2+4b_2^2-6b_1-12b_2+4b_1b_2 \\
&\quad +25+b_1^2+9b_2^2-10b_1-30b_2+6b_1b_2 \\
&\quad +64+b_1^2+16b_2^2-16b_1-64b_2+8b_1b_2 \\
&= 114+4b_1^2+30b_2^2-40b_1-114b_2+20b_1b_2
\end{aligned}
$$

We want to find of b_1 and b_2 such that RSS is minimized. There is a procedure we use a lot whenever we want to minimize or maximize something...

Linear Regression - Example

We take the partial derivatives and apply the first order condition.

$$
\begin{aligned}
RSS &= (4-b_1-b_2)^2+(3-b_1-2b_2)^2+(5-b_1-3b_2)^2+(8-b_1-4b_2)^2 \\
&= 16+b_1^2+b_2^2-8b_1-8b_2+2b_1b_2 \\
&\quad +9+b_1^2+4b_2^2-6b_1-12b_2+4b_1b_2 \\
&\quad +25+b_1^2+9b_2^2-10b_1-30b_2+6b_1b_2 \\
&\quad +64+b_1^2+16b_2^2-16b_1-64b_2+8b_1b_2 \\
&= 114+4b_1^2+30b_2^2-40b_1-114b_2+20b_1b_2
\end{aligned}
$$

$$\frac{\partial RSS}{\partial b_1} = 8b_1+20b_2-40$$

$$\frac{\partial RSS}{\partial b_2} = 20b_1+60b_2-114$$

Linear Regression - Example

We take the partial derivatives and apply the first order condition.

I.e. we set the derivatives to zero and solve.

Notice we are working with our estimates now – no longer candidate b values.

$$
\begin{aligned}
RSS &= (4-b_1-b_2)^2+(3-b_1-2b_2)^2+(5-b_1-3b_2)^2+(8-b_1-4b_2)^2 \\
&= 16+b_1^2+b_2^2-8b_1-8b_2+2b_1b_2 \\
&\quad +9+b_1^2+4b_2^2-6b_1-12b_2+4b_1b_2 \\
&\quad +25+b_1^2+9b_2^2-10b_1-30b_2+6b_1b_2 \\
&\quad +64+b_1^2+16b_2^2-16b_1-64b_2+8b_1b_2 \\
&= 114+4b_1^2+30b_2^2-40b_1-114b_2+20b_1b_2
\end{aligned}
$$

$$
\frac{\partial RSS}{\partial b_1}=0 \quad \Rightarrow \quad 8\hat{\beta}_1+20\hat{\beta}_2-40=0
$$

$$
\frac{\partial RSS}{\partial b_2}=0 \quad \Rightarrow \quad 20\hat{\beta}_1+60\hat{\beta}_2-114=0
$$

Linear Regression - Example

We take the partial derivatives and apply the first order condition.

$$\frac{\partial RSS}{\partial b_1} = 8b_1 + 20b_2 - 40 \qquad \frac{\partial RSS}{\partial b_2} = 20b_1 + 60b_2 - 114$$

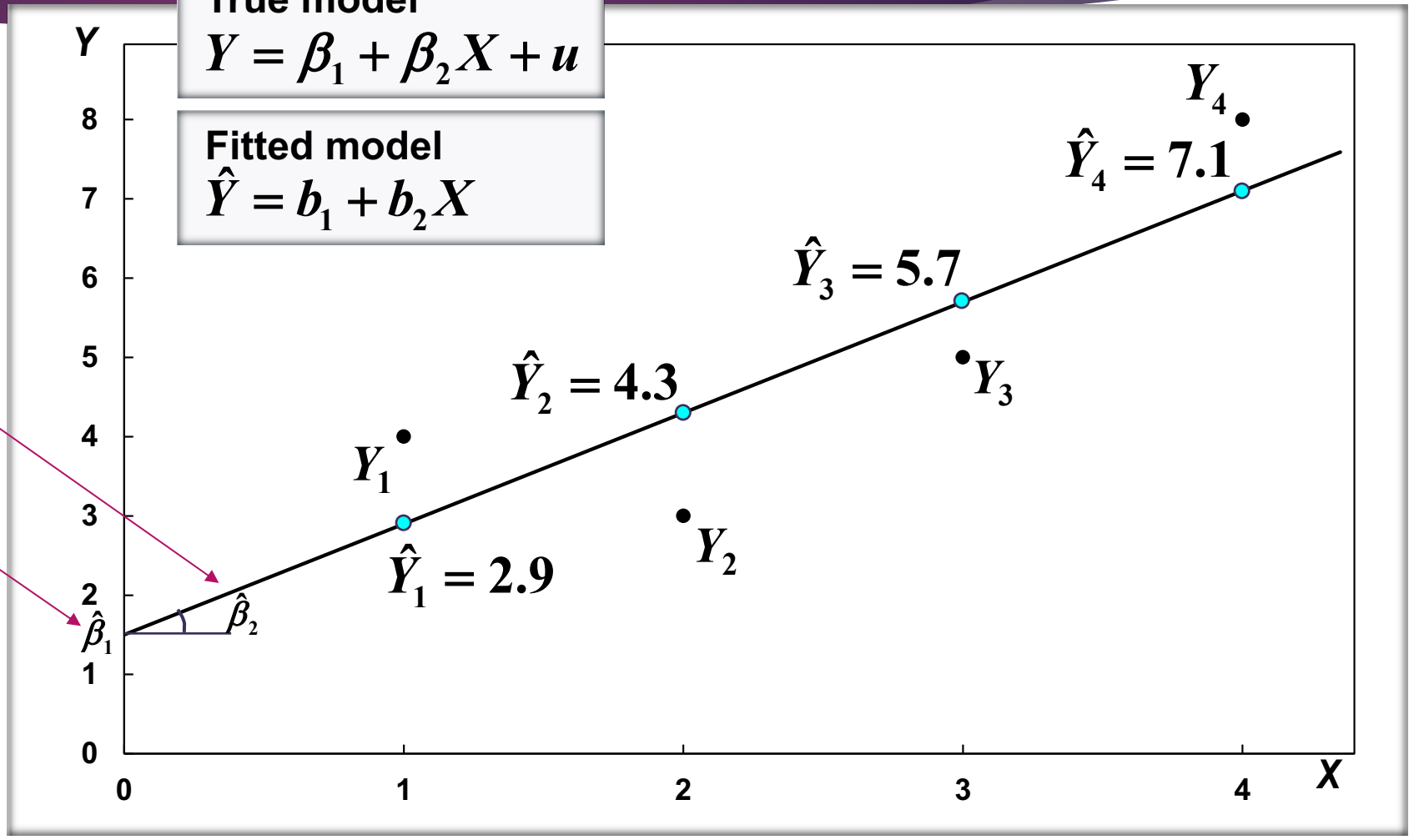
I.e. we set the derivatives to zero and solve.

$$\frac{\partial RSS}{\partial b_1} = \frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad \hat{\beta}_1 = 1.5\,, \quad \hat{\beta}_2 = 1.4$$

Linear Regression - Example

So now we have our slope and intercept for the linear fit.

Easy right!



Linear Regression

Happily we have computers

Linear Regression

Happily we have computers. Let's use MATLAB to do some regressions.

First setup the problem in matrix form.

Again, Y is the response vector and X is the factor level vector.

B are the coefficients that define our model.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Linear Regression

Let's get some data. MATLAB comes with sample datasets.

```
load accidents
x = hwydata(:,14); %Population of states
y = hwydata(:,4); %Accidents per state
format long
```

Linear Regression

And perform a linear regression on β_1 (just β_1 to make a point).

```
load accidents
x = hwydata(:,14); %Population of states
y = hwydata(:,4); %Accidents per state
format long
b1 = x\y
```

Linear Regression

And perform a linear regression on β_1 (just β_1 to make a point).

So now we have the best fit slope.

```
load accidents
x = hwydata(:,14); %Population of states
y = hwydata(:,4); %Accidents per state
format long
b1 = x\y
b1 =

     1.372716735564871e-04
```

Linear Regression

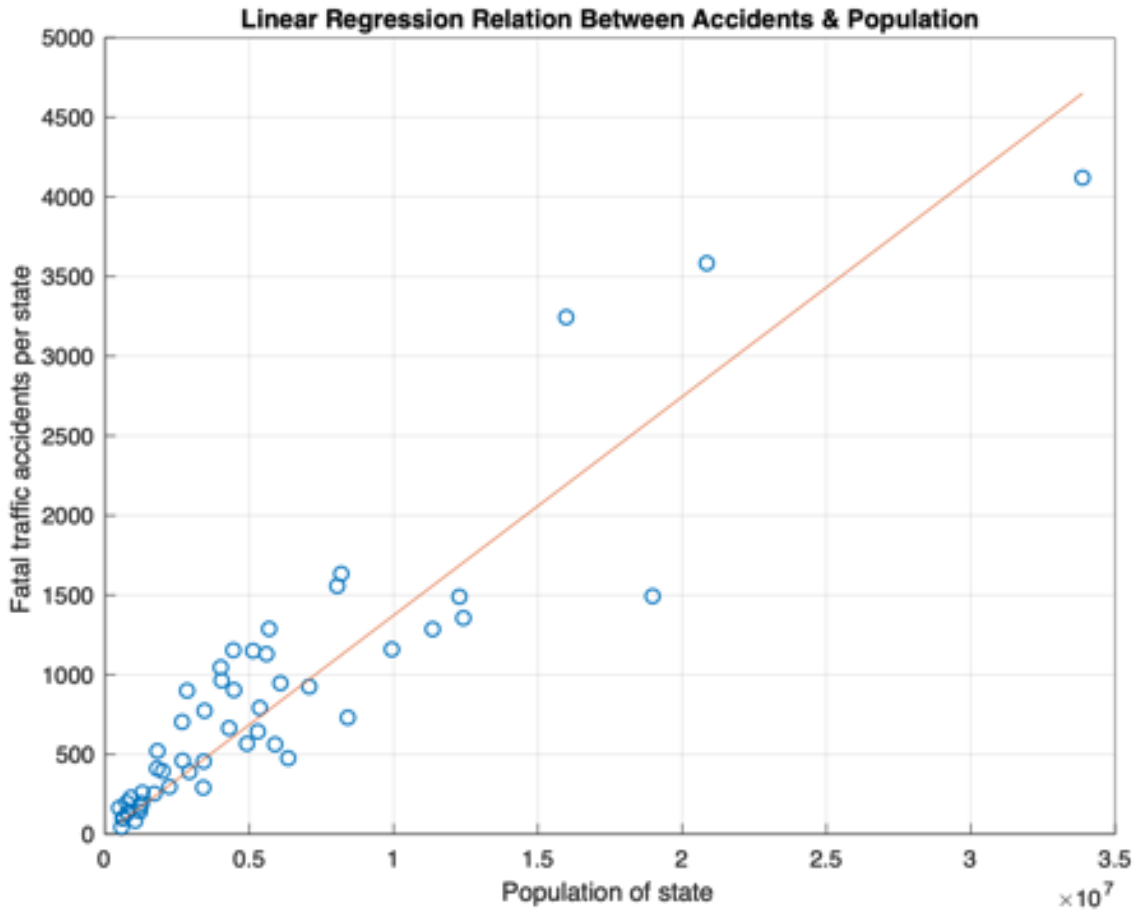
… and can make predictions.

```
b1 =

    1.372716735564871e-04

>> yCalc1 = b1*x;
scatter(x,y)
hold on
plot(x,yCalc1)
xlabel('Population of state')
ylabel('Fatal traffic accidents per state')
title('Linear Regression Relation Between Accidents & Population')
grid on
```

Linear Regression

… and can make predictions.

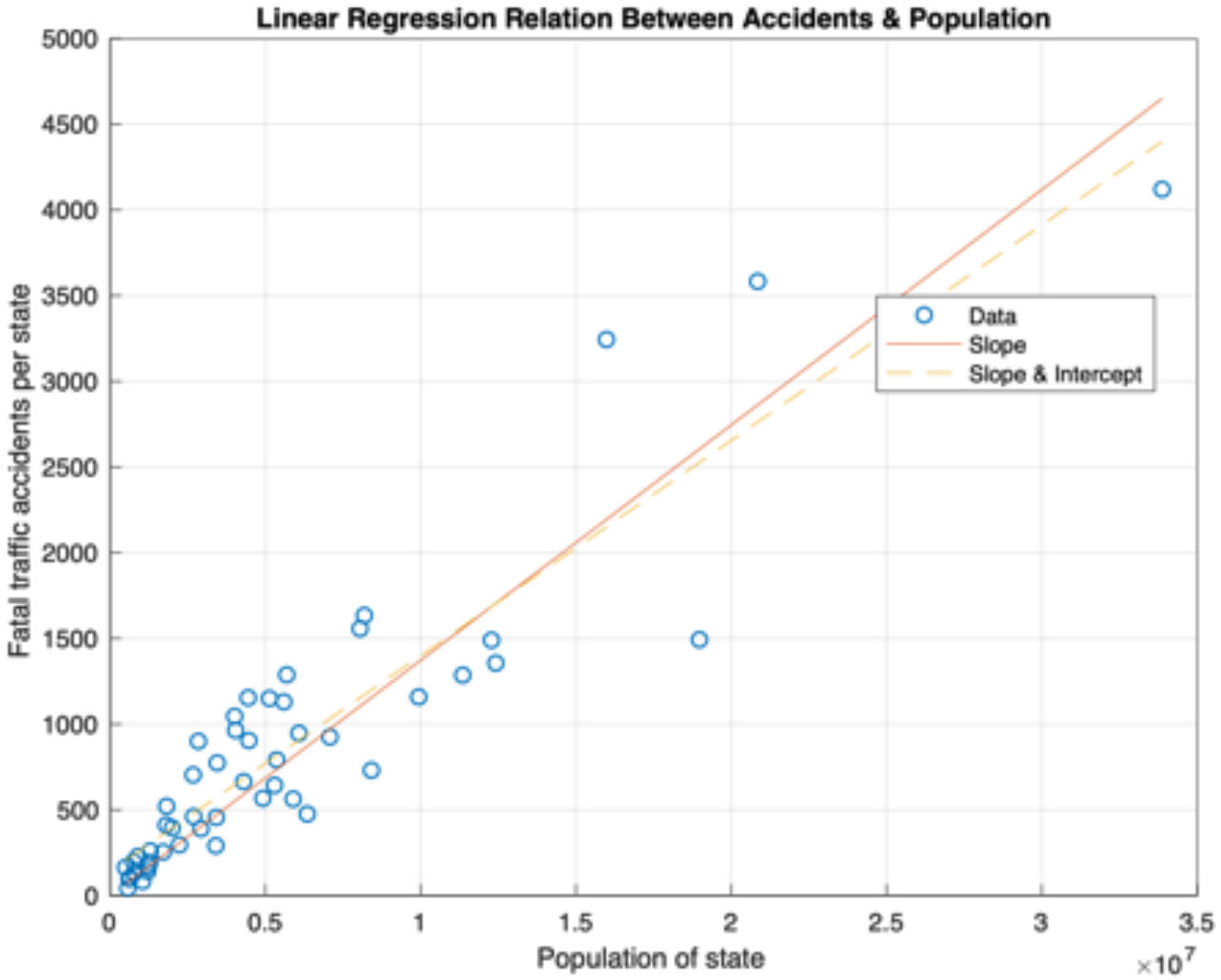
```
b1 =

    1.372716735564871e-04

>> yCalc1 = b1*x;
scatter(x,y)
hold on
plot(x,yCalc1)
xlabel('Population of state')
ylabel('Fatal traffic accidents per state')
title('Linear Regression Relation Between
Accidents & Population')
grid on
```



Linear Regression

Now lets include the intercept…

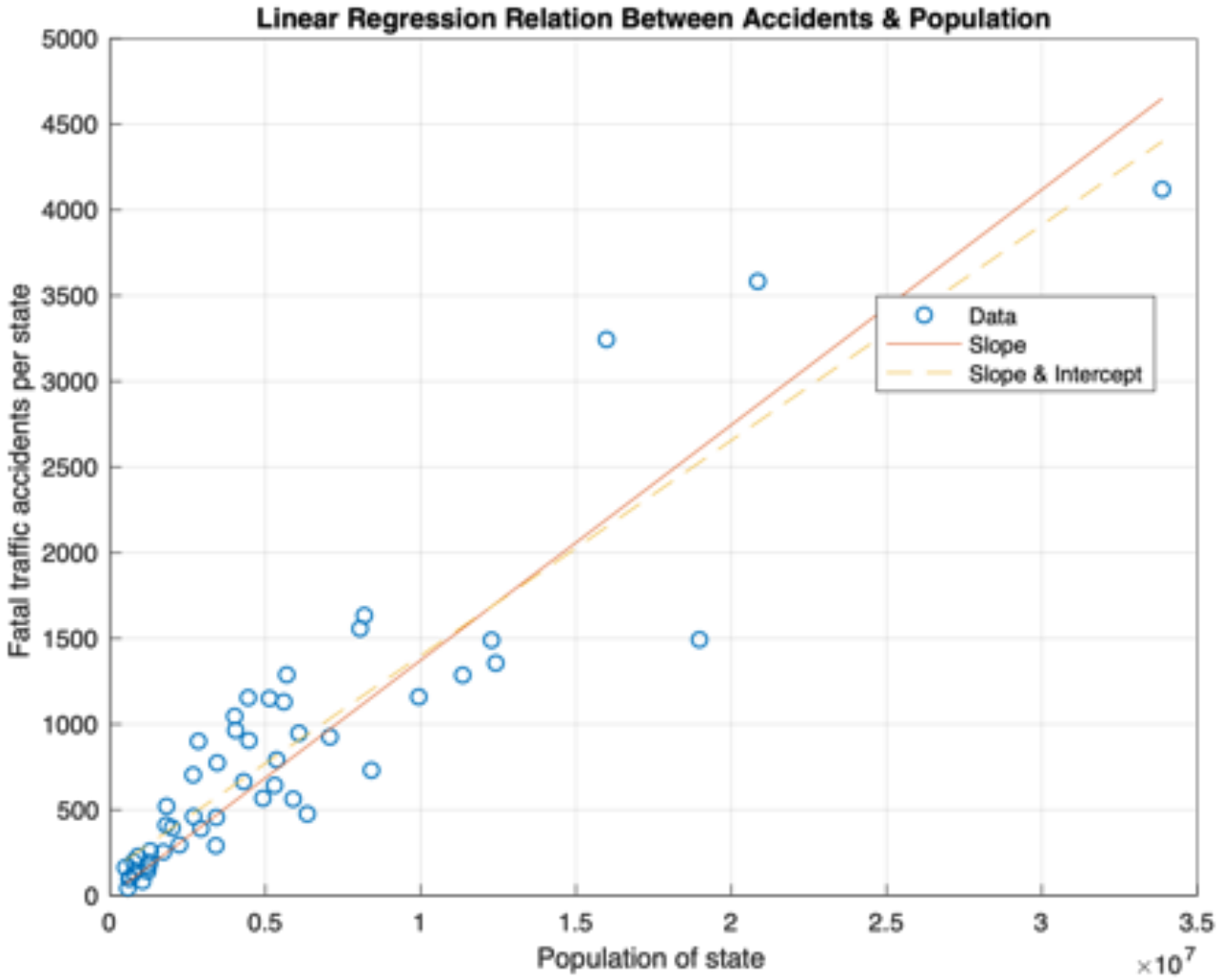
```
X = [ones(length(x),1) x];
b = X\y

b =

   1.0e+02 *

   1.427120171726538
   0.000001256394274

yCalc2 = X*b;
plot(x,yCalc2,'--')
legend('Data','Slope','Slope &
Intercept','Location','best');
```



Linear Regression

Now lets include the intercept.

How can we demonstrate that one fit is better than another?

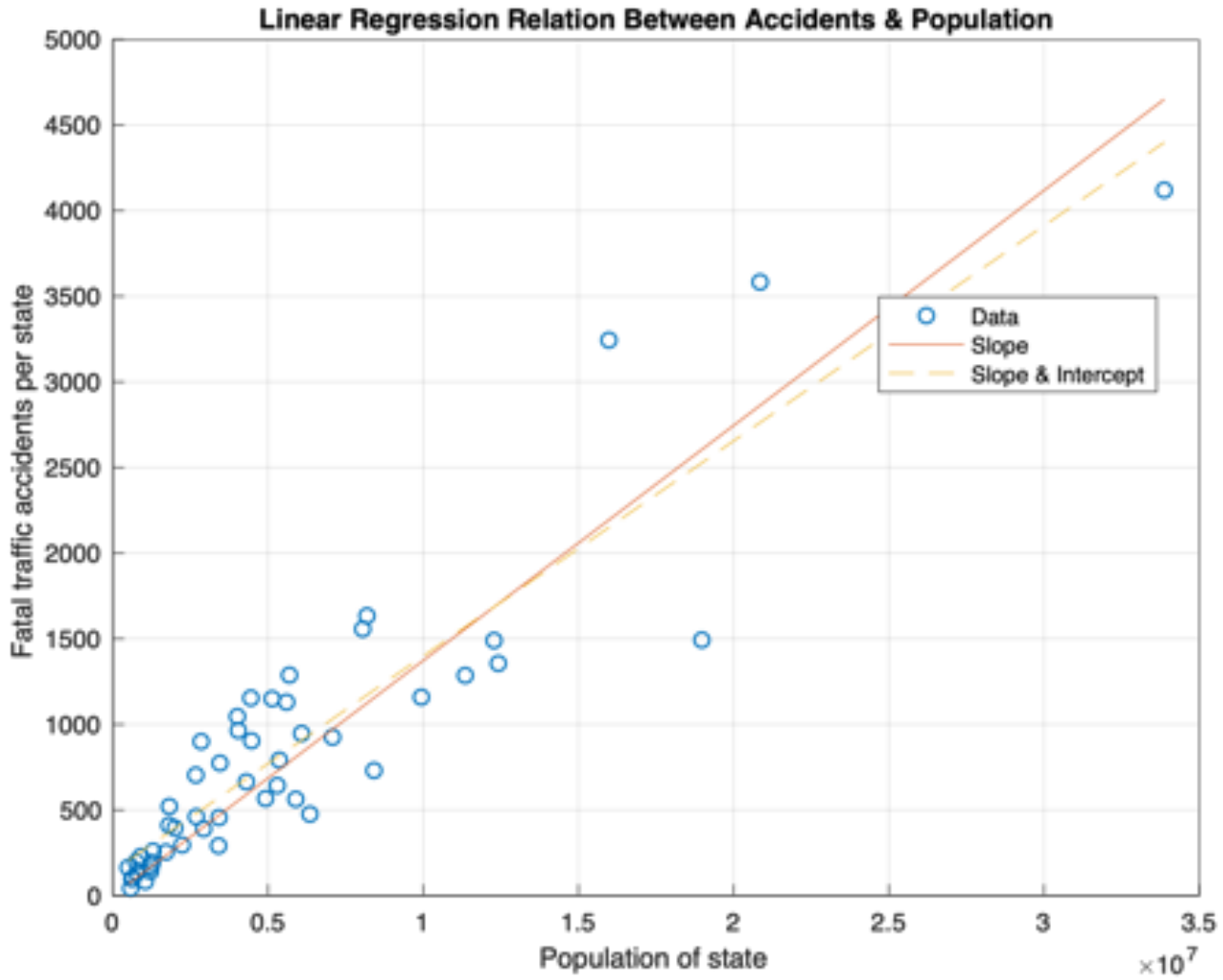


Linear Regression

Now lets include the intercept.

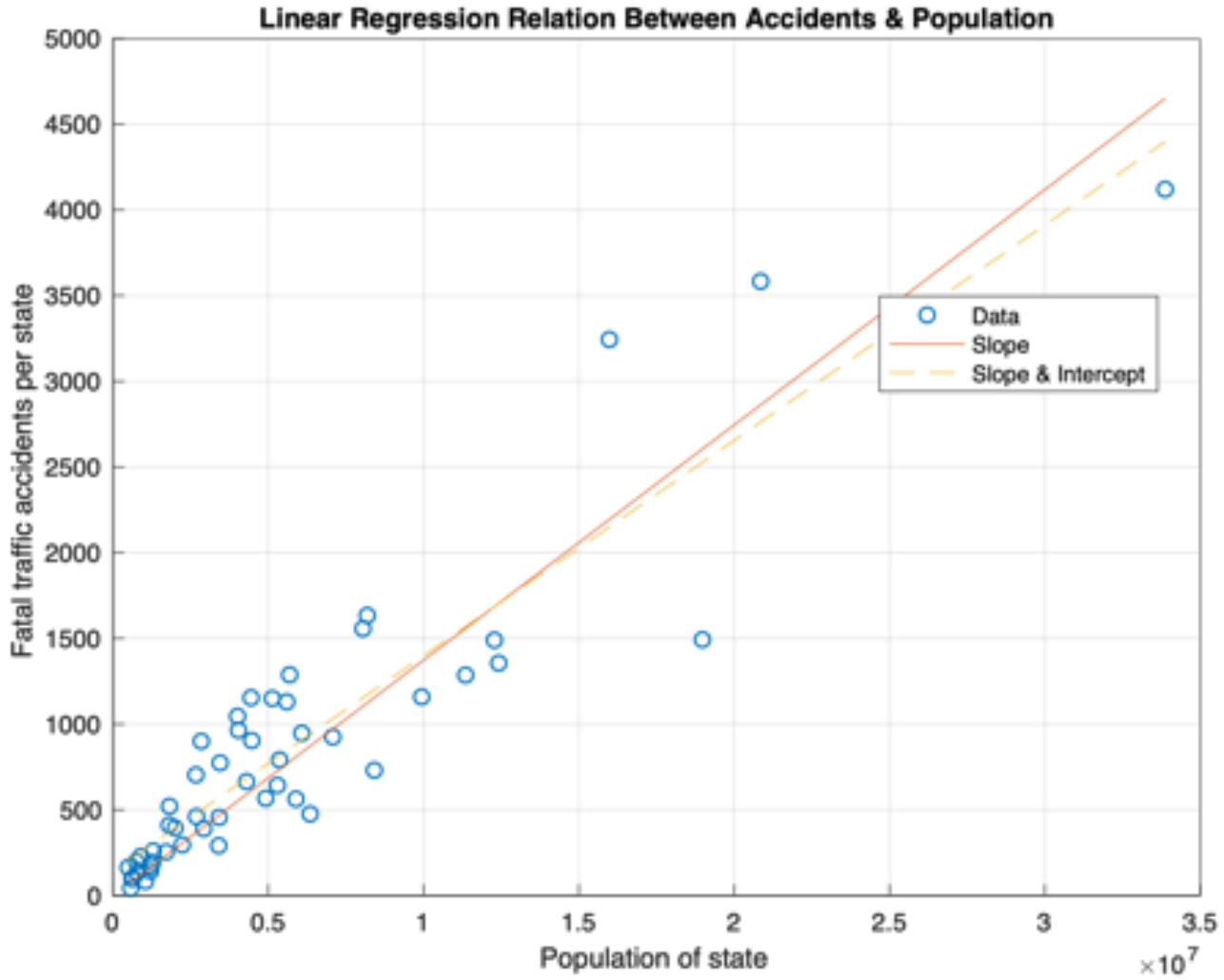
How can we demonstrate that one fit is better than another?

We need a "Goodness of Fit" (GoF) metric.



Linear Regression

… but didn't we just use a goodness of fit measure to fit the model in the first place?

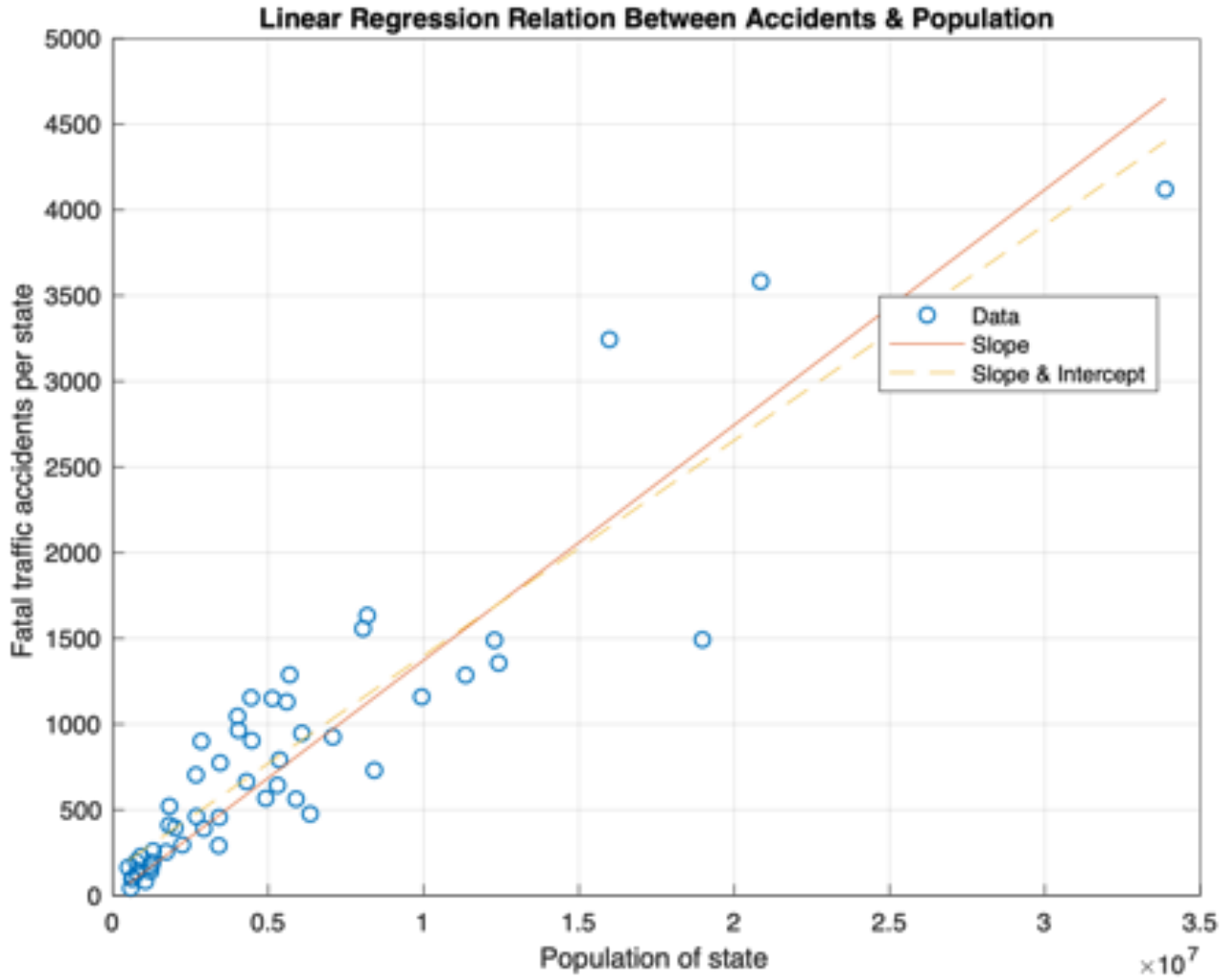


Linear Regression

We want a metric. Something that we can compare across different best fit models.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

This is R^2, notice the numerator is the RSS. But we are normalizing by the total variance.

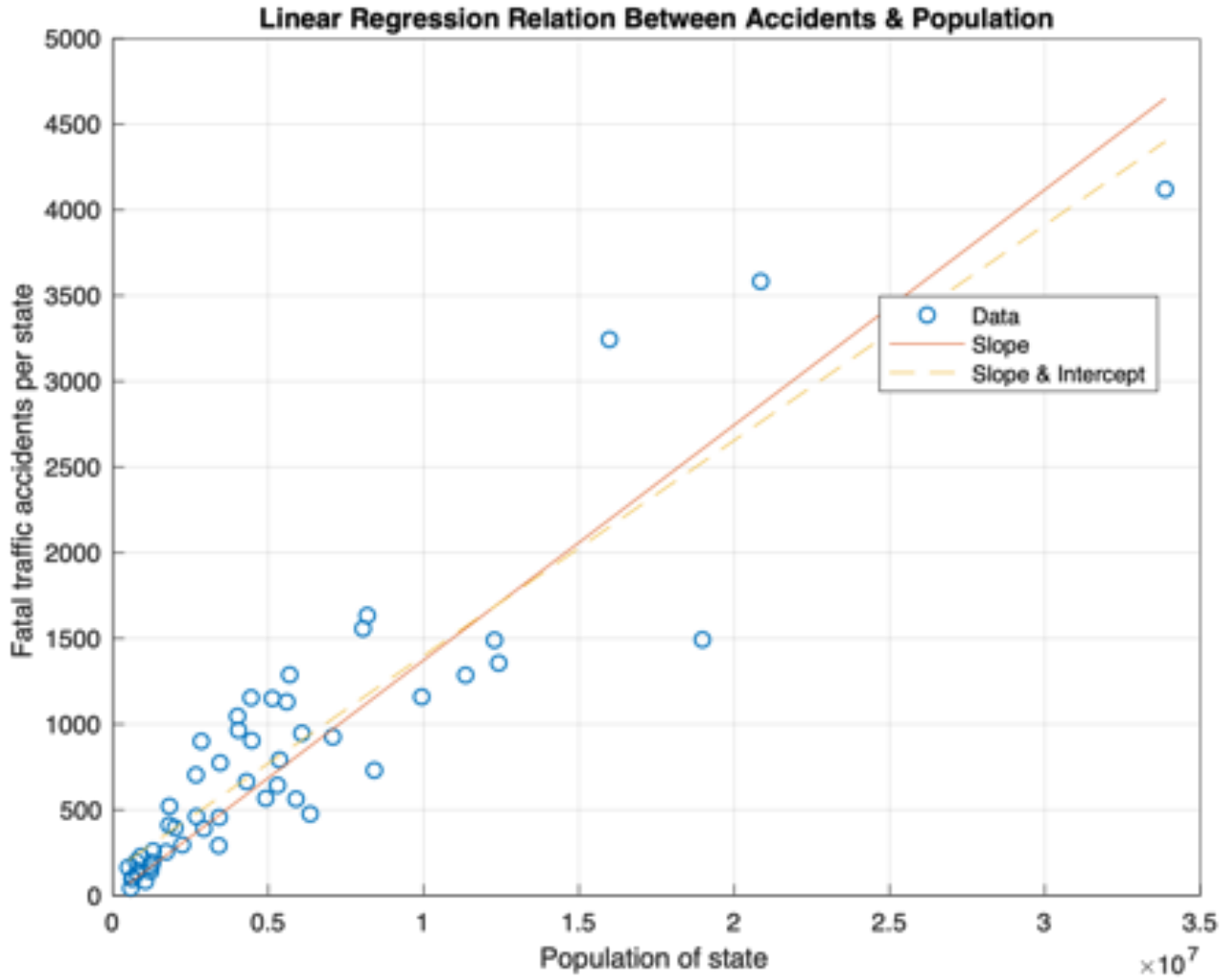


Linear Regression

We want a metric. Something that we can compare across different best fit models.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

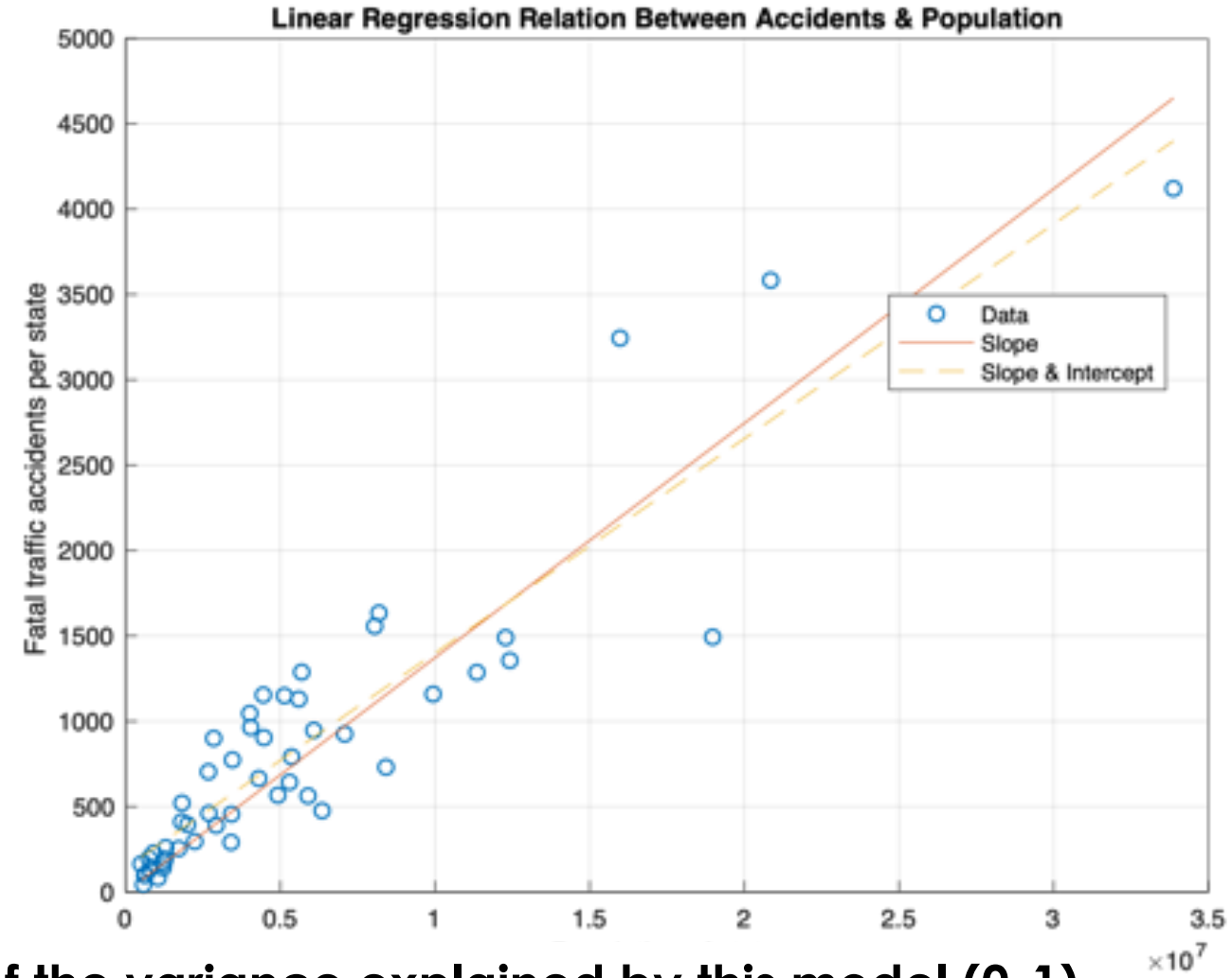
This is R^2, notice the numerator is the ESS (explained sum of squares). But we are normalizing by the total variance (the TSS, total sum of squares).



Linear Regression

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

```
Rsq1 = 1 - sum((y - yCalc1).^2)/sum((y - mean(y)).^2)
Rsq1 = 0.822235650485566
Rsq2 = 1 - sum((y - yCalc2).^2)/sum((y - mean(y)).^2)
Rsq2 = 0.838210531103428
```



We interpret this to mean the proportion of the variance explained by this model (0-1).