# Line Fitting and Regression

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

# Outline

Least Squares Simple Regression

Least Squares Multiple Regression

Least Squares Orthogonal Regression

Experimental Design

Discussion

Homework

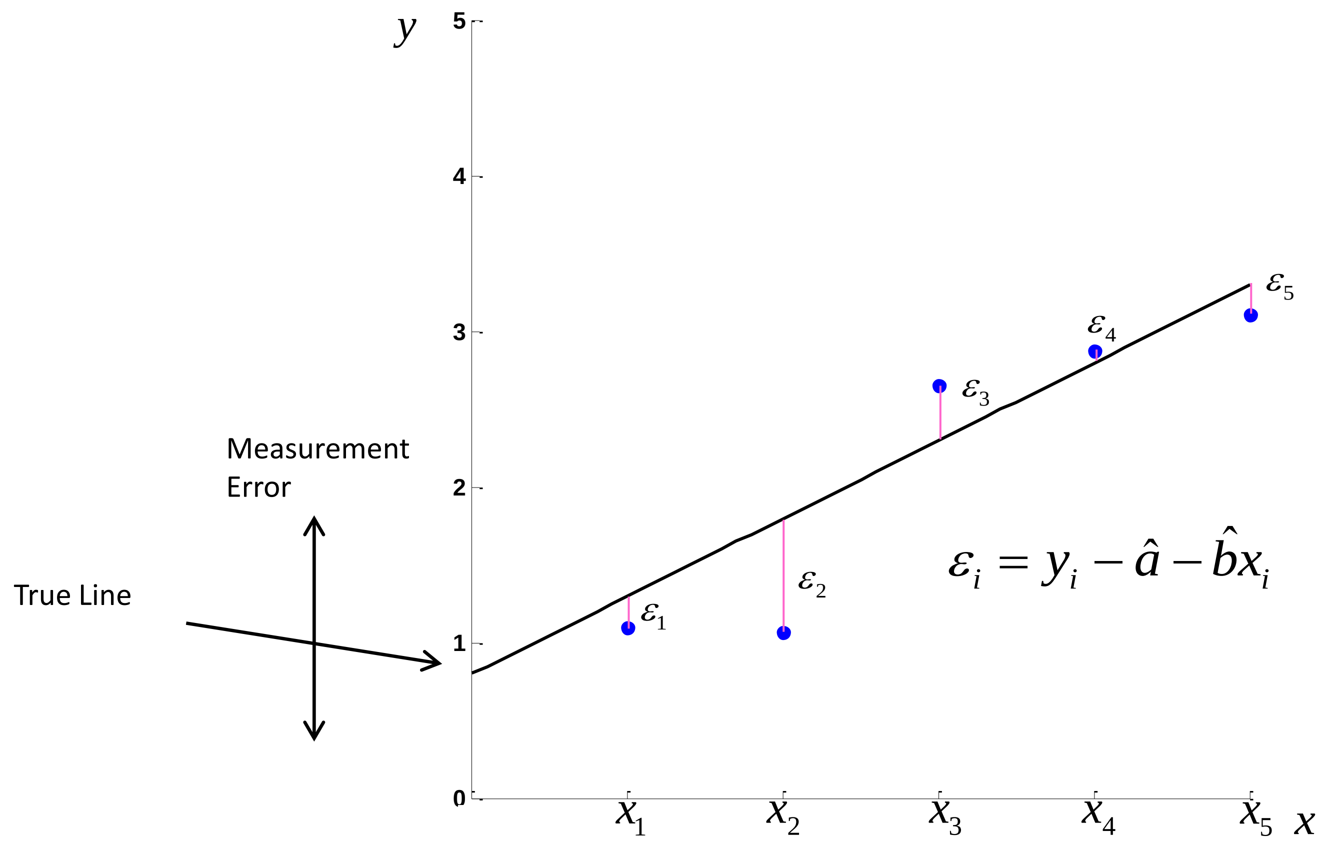
# Least Squares Simple Regression

For LSR we have points $(x_1,y_1),\ldots,(x_n,y_n)$ and wish to find the "best" fit line $y = \hat{a} + \hat{b}x$ to the data.

The "best" line defined as minimizing sum of squared errors.

$$Q = \sum_{i=1}^{n} \varepsilon_i^2$$

$$Q = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$



# Least Squares Simple Regression

Define $Q=\sum_{i=1}^{n}(y_i-a-bx_i)^2$ as the “score” function to be

minimized to obtain the optimal $(a,b)$ denoted as $(\hat{a},\hat{b})$.

What we want to do is find the values of $(a,b)$ that minimize $Q$.
The values $(a,b)$ that minimize $Q$ are the optimal values are $(\hat{a},\hat{b})$.

Find $(\hat{a},\hat{b})$ that minimize $\sum_{i=1}^{n}(y_i-a-bx_i)^2$ wrt $(a,b)$.

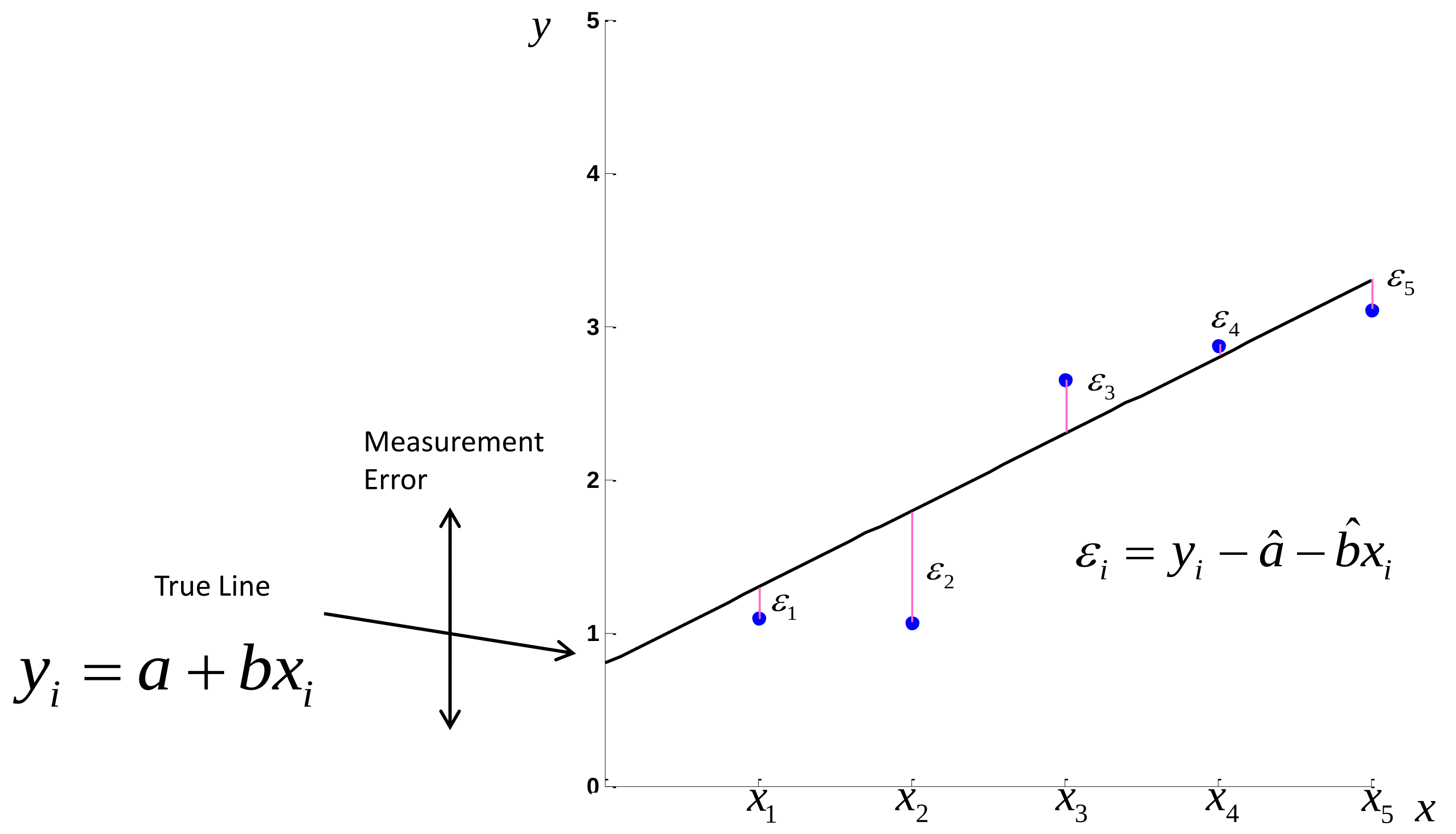
# Least Squares Simple Regression

Differentiating $Q$ wrt $a$, then $b$, then set = 0

$$Q = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \qquad i = 1, \ldots, n$$

$$\left.\frac{\partial Q}{\partial a}\right|_{\hat{a},\hat{b}} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1) = 0$$

$$\left.\frac{\partial Q}{\partial b}\right|_{\hat{a},\hat{b}} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-x_i) = 0$$



# Least Squares Simple Regression

$$s_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

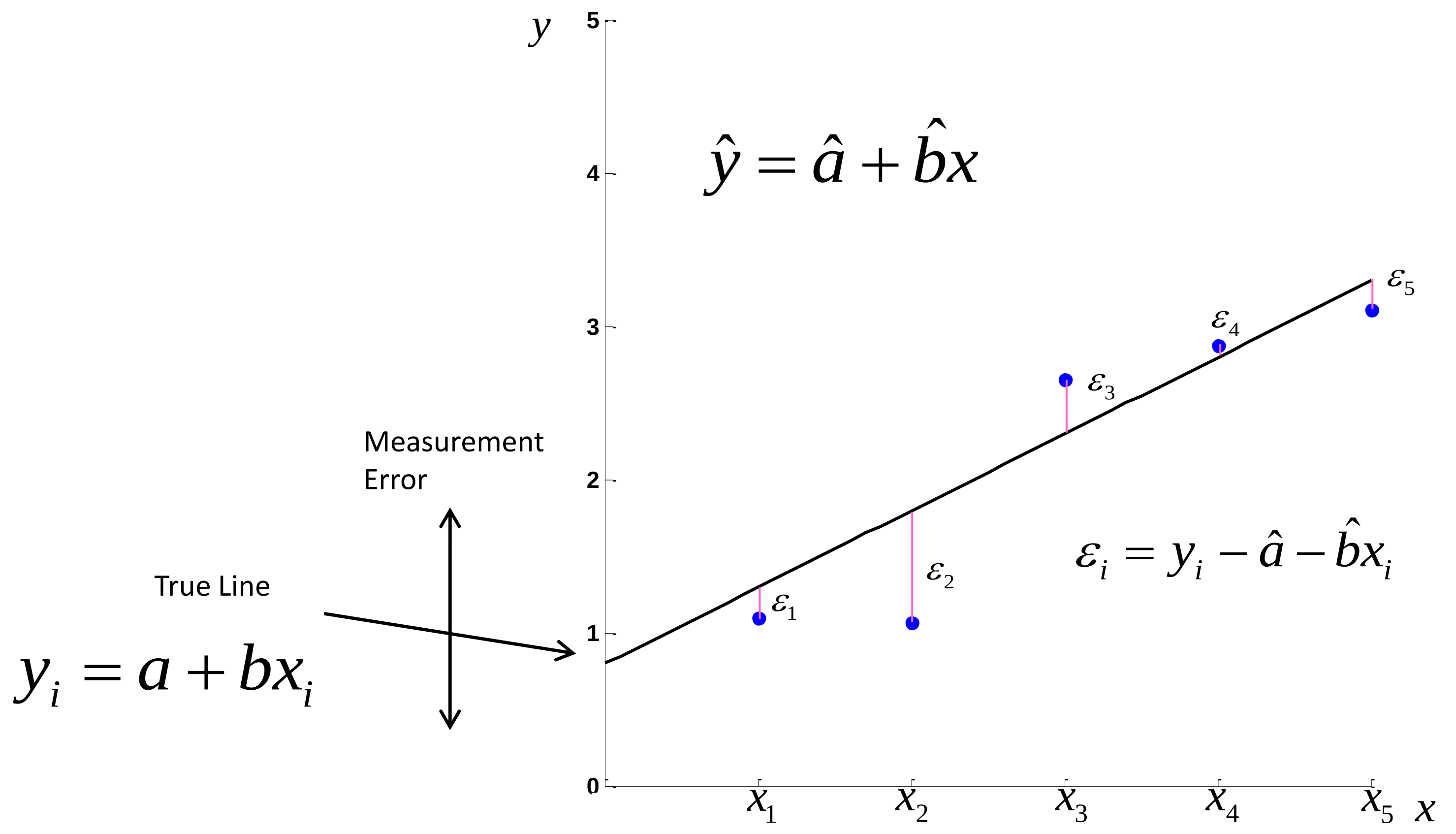
Differentiating $Q$ wrt $a$, then $b$, then set = 0

$$Q=\sum_{i=1}^{n}(y_i-a-bx_i)^2 \qquad i=1,\ldots,n$$

$$\left.\frac{\partial Q}{\partial a}\right|_{\hat{a},\hat{b}} = \sum_{i=1}^{n}2(y_i-a-bx_i)(-1)=0$$

$$\left.\frac{\partial Q}{\partial b}\right|_{\hat{a},\hat{b}} = \sum_{i=1}^{n}2(y_i-a-bx_i)(-x_i)=0$$

$$\hat{b}=\frac{S_{xy}}{S_{xx}} \qquad \hat{a}=\bar{y}-\hat{b}\bar{x}$$



# Least Squares Simple Regression

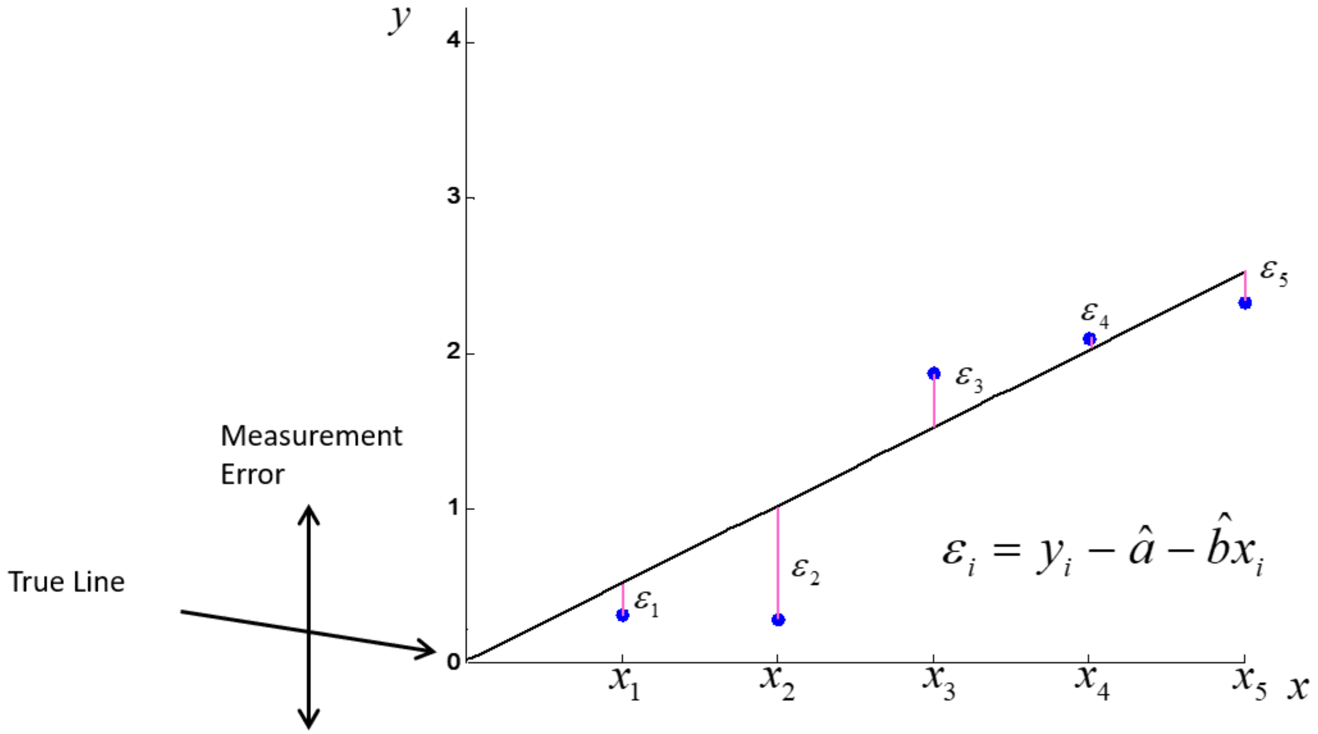
Because of the scientific application, it may be known that the $y$-intercept should truly be zero.

$$y_i = \beta_1 x_i + \varepsilon_i$$

This is known as regression through the origin

Minimize: $Q = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$ to get

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$ Compare to $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}.$$




$y_i =$

# Least Squares Simple Regression

The least squares estimation score function

$$Q=\sum_{i=1}^{n}(y_i-a-bx_i)^2 \quad i=1,\ldots,n$$

is equivalently represented as

$$Q=(y-X\beta)'(y-X\beta) \quad \text{where}$$

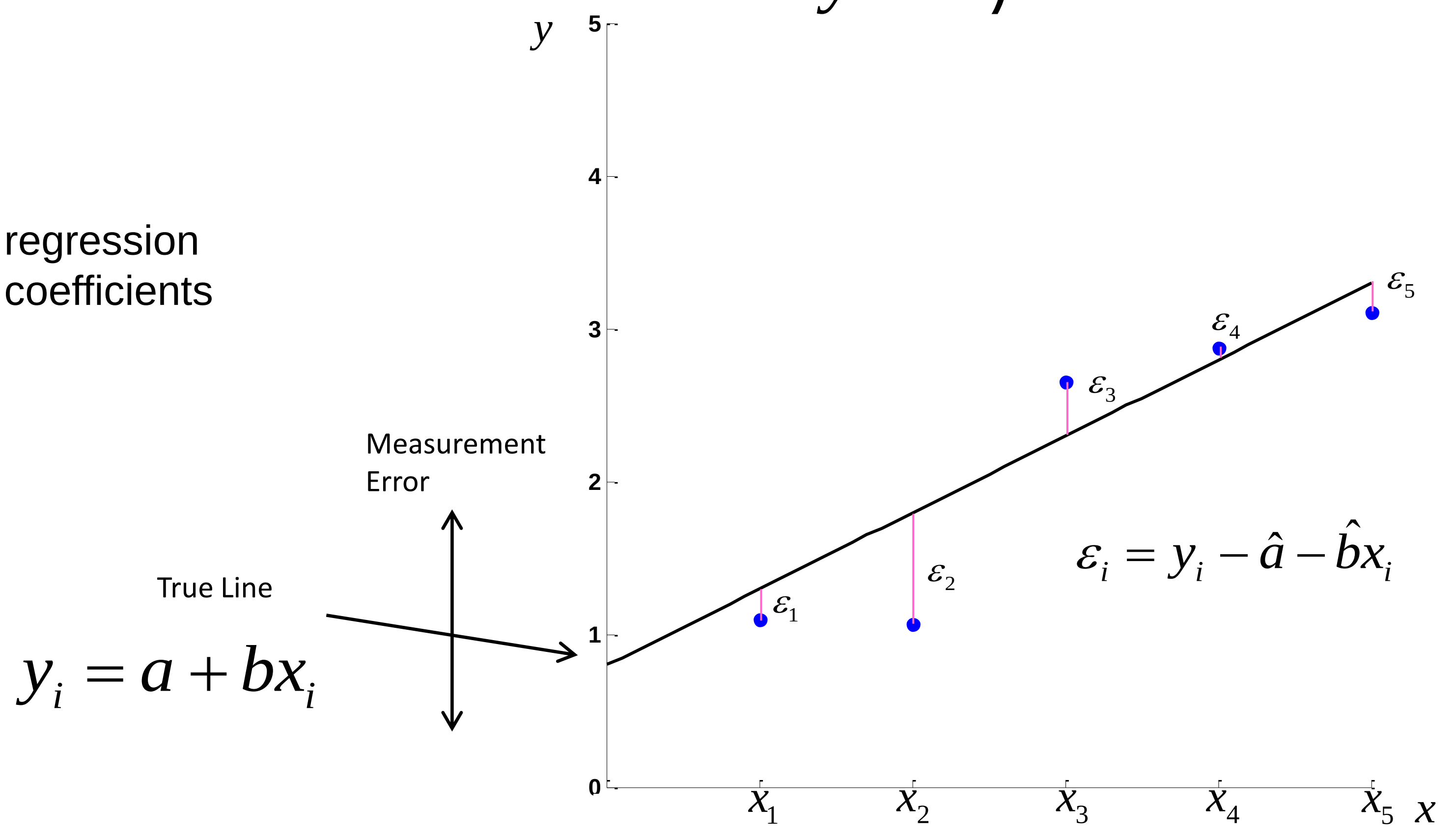
design matrix

regression coefficients

measured $y$ data

$$\underset{n\times 1}{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{n\times 2}{X}=\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \underset{2\times 1}{\beta}=\begin{pmatrix} a \\ b \end{pmatrix}.$$

$$y=X\beta+\varepsilon$$



# Least Squares Simple Regression

We don't need to take the derivative of $Q$ wrt $\beta$ (although we could).

We can write with algebra

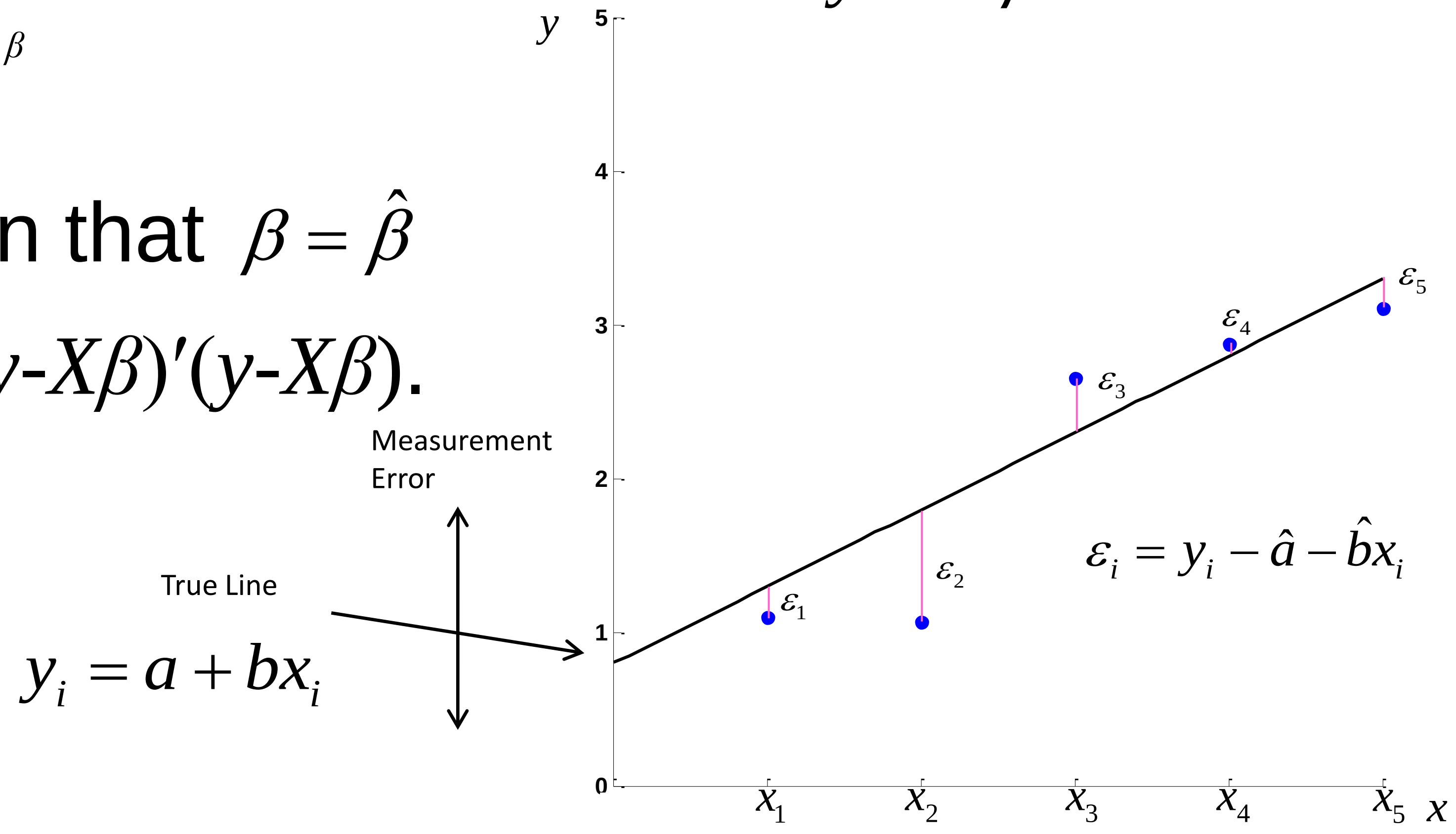
$$(y - X\beta)'(y - X\beta) = \underbrace{(y - X\hat{\beta})'(y - X\hat{\beta})}_{\text{does not depend on } \beta} + (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})$$

add and subtract $X\hat{\beta}$

$$y = X\beta + \varepsilon$$

invertible

where $\hat{\beta} = (X'X)^{-1}X'y$ . It can be seen that $\beta = \hat{\beta}$ minimizes $Q$ because it minimizes $(y\text{-}X\beta)'(y\text{-}X\beta)$.



# Least Squares Simple Regression

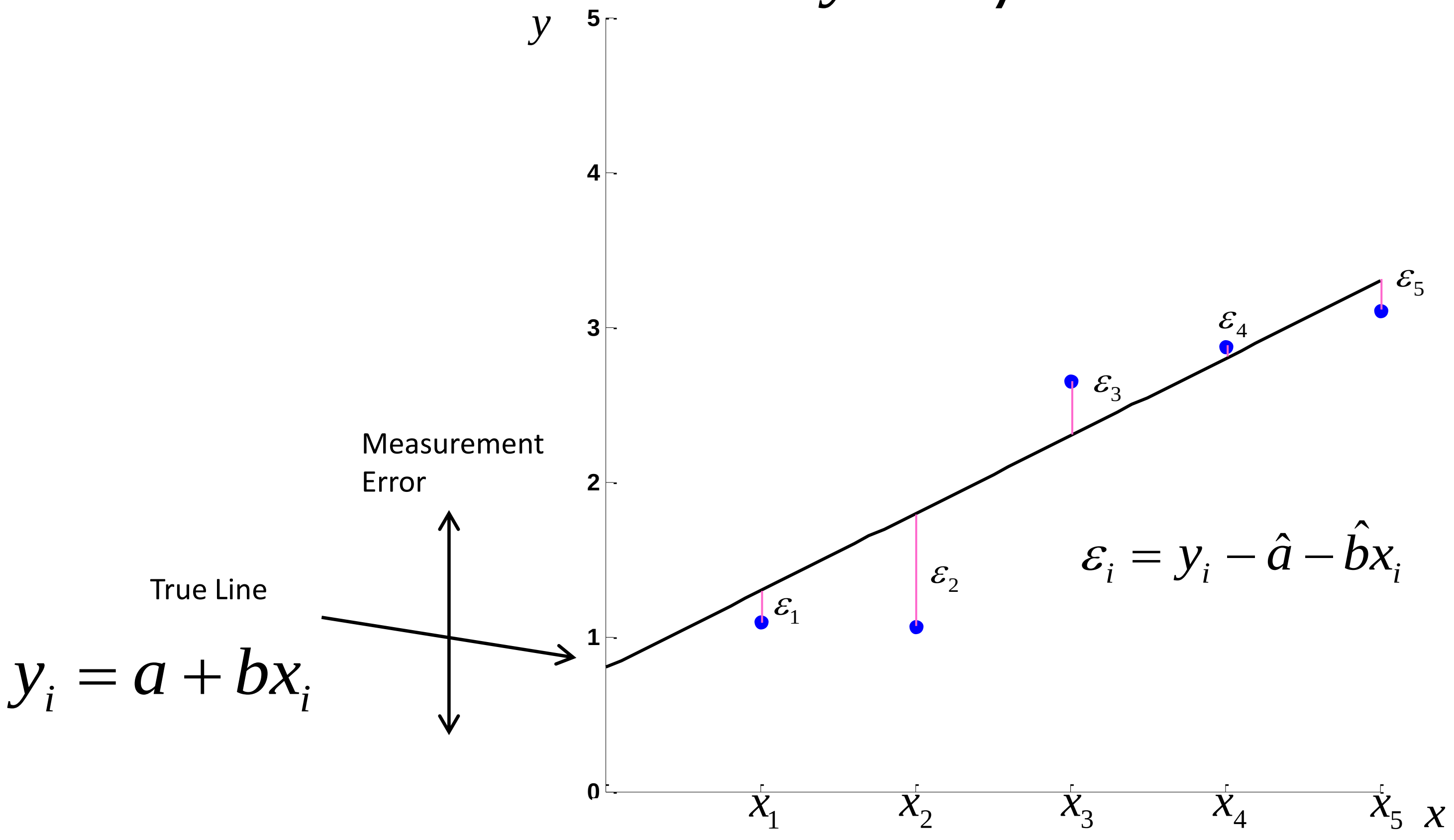
Generate simulated data by adding random noise to a noiseless line.

Let $y_i = a + bx_i + \varepsilon_i$ ,

where $\varepsilon_i \sim N(0,\sigma^2)$, $i=1,\ldots,n$

are independent.

$$y = X\beta + \varepsilon$$



# Least Squares Simple Regression

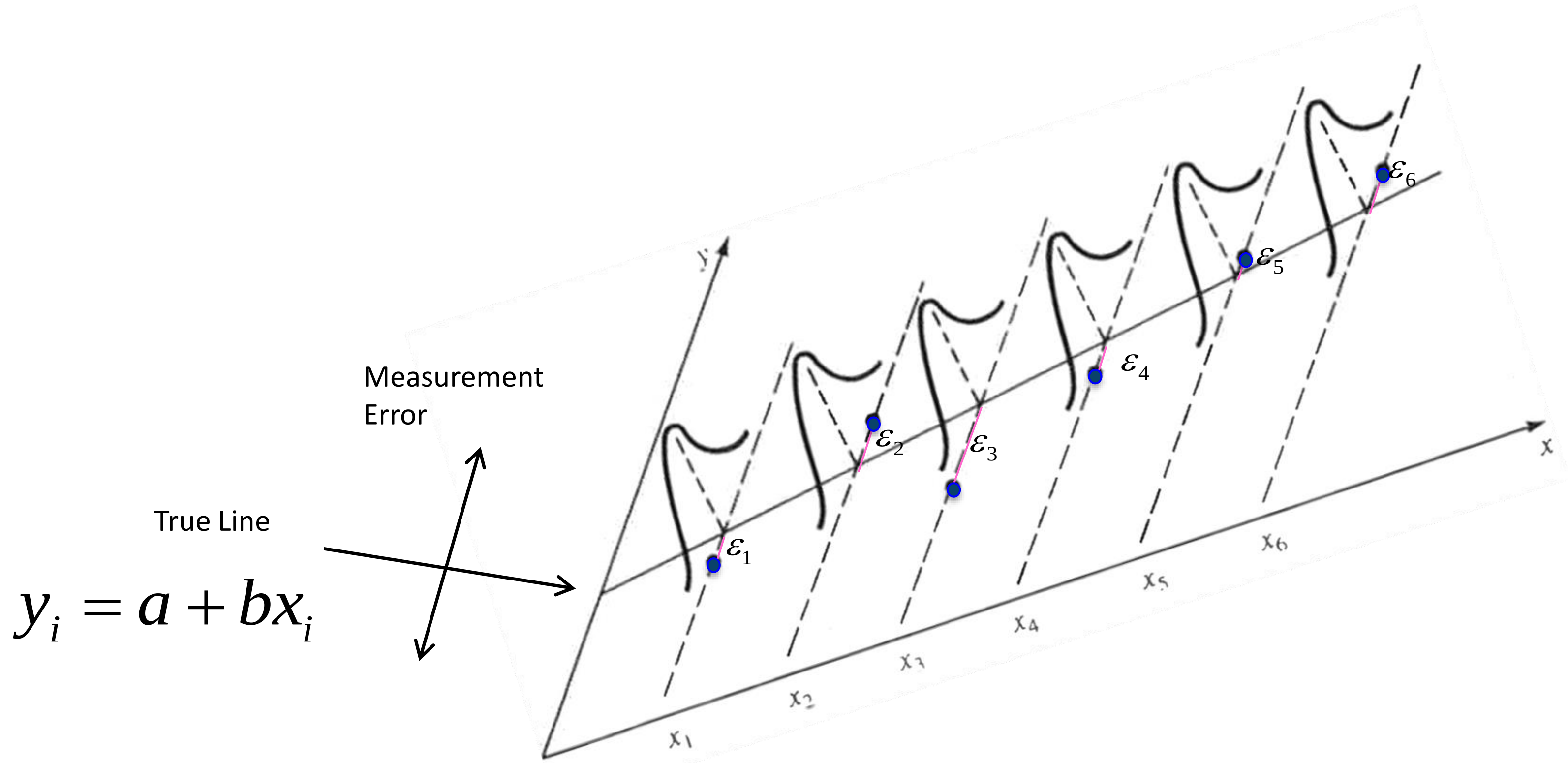
Generate simulated data by adding random noise to a noiseless line.

Let $y_i = a + bx_i + \varepsilon_i$ ,

$$y = X\beta + \varepsilon$$

where $\varepsilon_i \sim N(0,\sigma^2)$, $i=1,\ldots,n$

are independent.



# Least Squares Simple Regression

Let $a=0$, $b=1$, and $\sigma=1$ in $y_i=a+bx_i+\varepsilon_i$ with $\varepsilon_i \sim N(0,\sigma^2)$, $i=1,\ldots,n$.

Generate $y$ values to go along with $x=1,2,3,4$.

$$y \quad = a \quad e_n + b \quad x + \quad \varepsilon$$

$$\begin{bmatrix} 1.5377 \\ 3.8339 \\ 0.7412 \\ 4.8622 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0.5377 \\ 1.8339 \\ -2.2588 \\ 0.8622 \end{bmatrix}$$

```
rng('default')

n=4; a=0; b=1; sigma=1;
x=[1,2,3,4]';
e=sigma*randn(n,1);
y=a+b*x+e;

sumX=sum(x); sumX2=sum(x.*x);
sumY=sum(y); sumXY=sum(x.*y);

bhat=(n*sumXY-sumX*sumY)/(n*sumX2-sumX^2)
ahat=sumY/n-bhat*sumX/n
```

# Least Squares Simple Regression

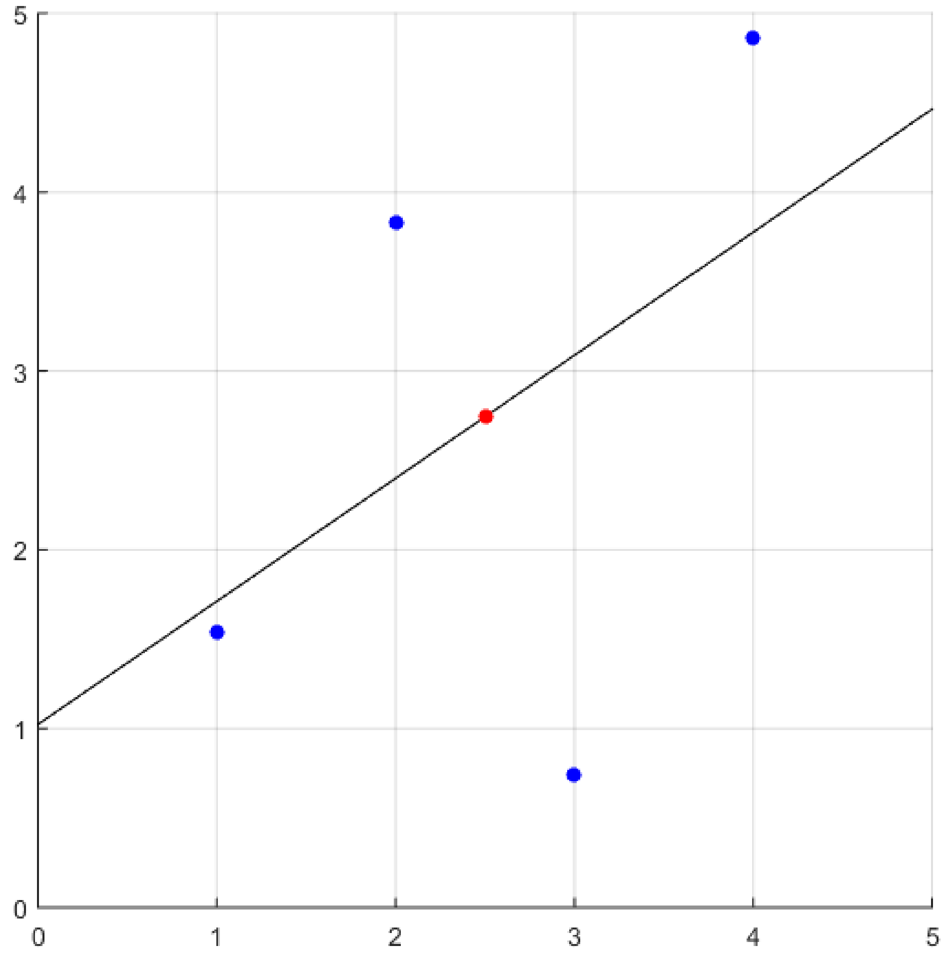
$a$=0, $b$=1, and $\sigma$=1

$x$=1,2,3,4.

$$y_i = a + bx_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

```
figure;
line([0,5],[ahat,ahat+bhat*5],'color','k')
hold on
scatter(x,y,'bo','filled')
scatter(sumX/n,sumY/n,'ro','filled')
grid on, axis square
xlim([0,5]), ylim([0,5])
set(gca,'xtick',(0:5)),set(gca,'ytick',(0:5))
```



# Least Squares Simple Regression

As previously noted, the least squares regression

$y_i = a + bx_i + \varepsilon_i$ can be written as

$$\underset{n\times 1}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \underset{n\times 2}{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}$$

$$y \quad = a \quad e_n + b \quad x \quad + \qquad \varepsilon$$

$$\begin{bmatrix} 1.5377 \\ 3.8339 \\ 0.7412 \\ 4.8622 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0.5377 \\ 1.8339 \\ -2.2588 \\ 0.8622 \end{bmatrix}$$

$$\underset{2\times 1}{\beta} = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\underset{2\times 1}{\hat{\beta}} = (X'X)^{-1}X'y$$

# Least Squares Simple Regression

As it turns out (not shown here, take MSSC 5780),

$$y_i = a + bx_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

$$\underset{2\times 1}{\hat{\beta}} = (X'X)^{-1}X'y$$

$$\underset{n\times 1}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \underset{n\times 2}{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}$$

$$E(\underset{2\times 1}{\hat{\beta}}) = \beta \qquad \operatorname{cov}(\underset{2\times 2}{\hat{\beta}}) = \sigma^2(X'X)^{-1} = \sigma^2 W$$

$$\underset{n\times 2}{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \qquad \sigma^2 \underset{2\times 2}{(X'X)^{-1}} = \sigma^2 \begin{bmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{bmatrix}$$

$$\underset{2\times 1}{\beta} = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

# Least Squares Simple Regression

$a$=0, $b$=1, and σ=1

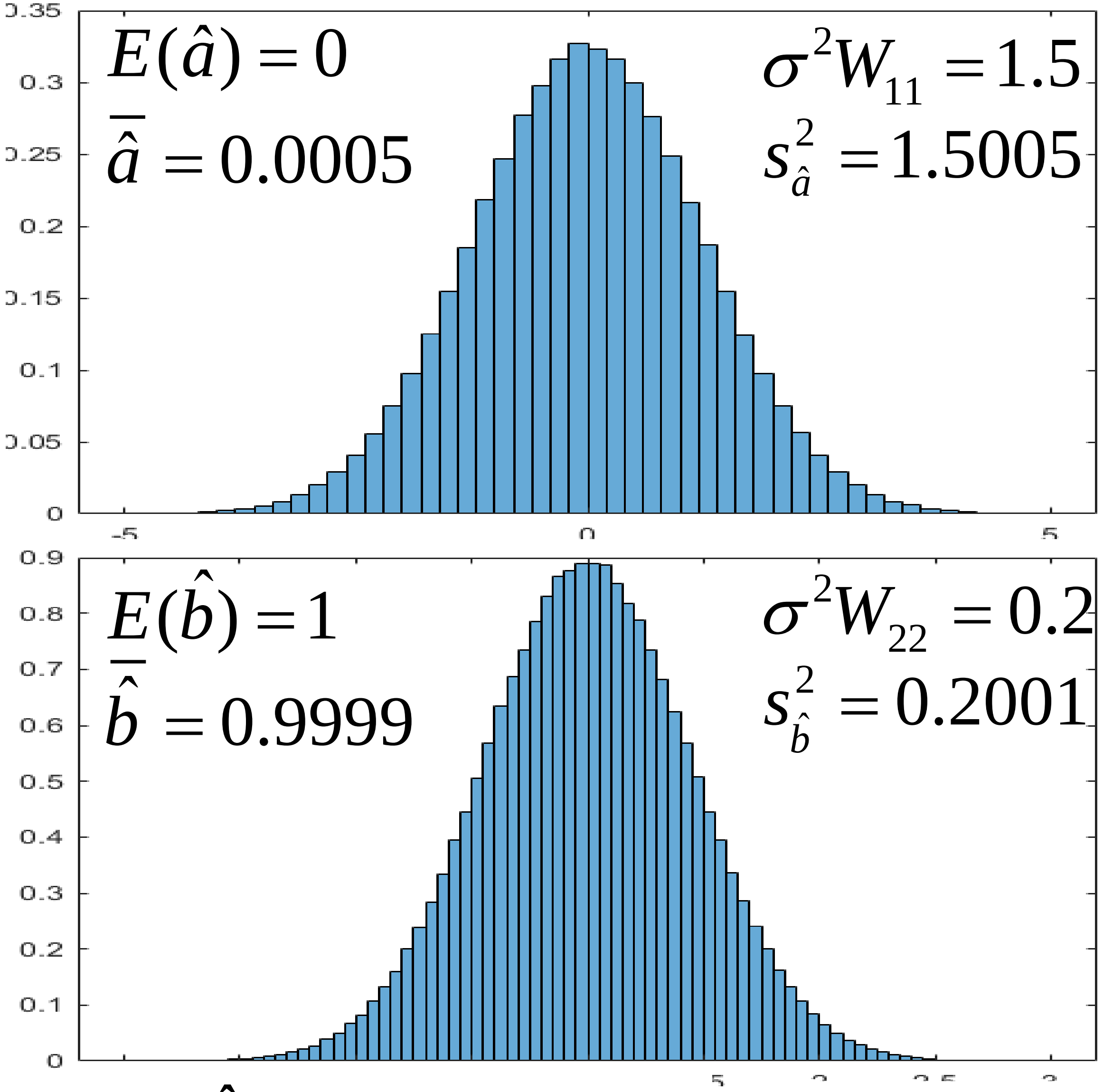
$x$=1,2,3,4.

## Repeated $10^6$ times

$$\underset{(q+1)\times 1}{\hat{\beta}} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

$$\beta = (a,b)' \qquad W = (X'X)^{-1}$$

```
num=10^6; a=0;b=1; sigma=1;
x=[1,2,3,4]'; n=4;
mu=a+b*x';, X=[ones(n,1),x];
y=sigma*randn(num,n)...
+ones(num,1)*mu;
betahat=inv(X'*X)*X'*y';
figure; histogram(betahat(1,:),(-5:.2:5)','normalization','pdf')
figure; histogram(betahat(2,:),(-1:.05:3)','normalization','pdf')
betabar=mean(betahat,2)
covbetahat=cov(betahat')
```




$$\mathrm{cov}(\hat{a},\hat{b}) = \sigma^2 W_{12} = -0.5 \qquad s_{\hat{a}\hat{b}} = -0.5003$$

# Least Squares Multiple Regression

More generally, we can have a multiple regression model

$$\underset{n\times 1}{y} = X\beta + \varepsilon \qquad \underset{n\times 1}{\varepsilon} \sim N(0,\sigma^2 I_n)$$

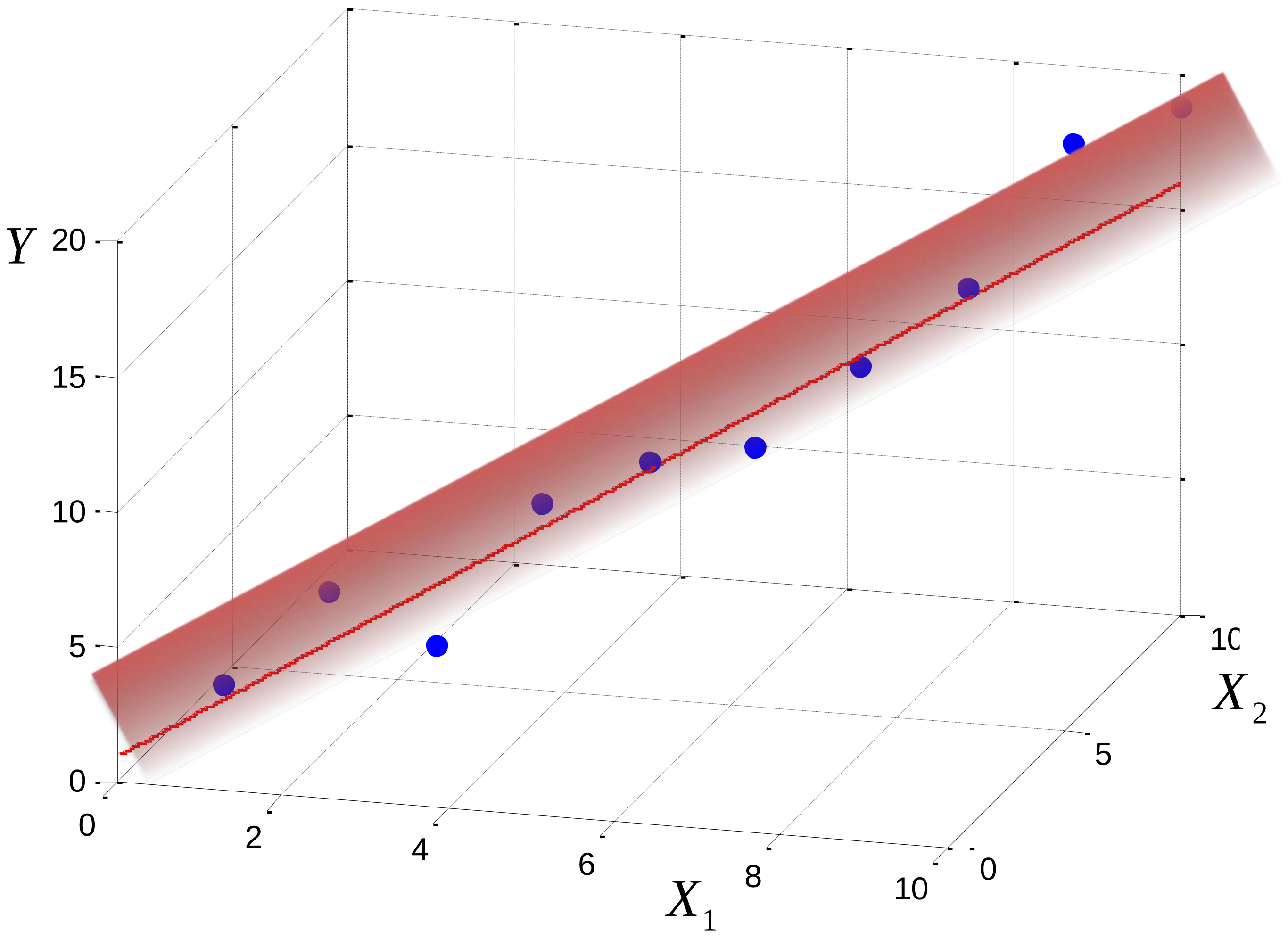
$$\underset{n\times 1}{y} = \overbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}^{\text{measured data}} \qquad \underset{n\times(q+1)}{X} = \overbrace{\begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ 1 & x_{21} & & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{pmatrix}}^{\text{design matrix}} \qquad \underset{(q+1)\times 1}{\beta} = \overbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}}^{\text{regression coefficients}} \qquad \underset{n\times 1}{\varepsilon} = \overbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}^{\text{measurement error}}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq} + \varepsilon_i$$

# Least Squares Multiple Regression

The points are considered to be $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq} + \varepsilon_i$

$$i = 1, \ldots, n$$




How do we estimate the coefficients?

# Least Squares Multiple Regression

$$s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n-q-1}$$

The MLEs are the same,

$$\underset{(q+1)\times 1}{\hat{\beta}} = (X'X)^{-1}X'y \quad \text{and} \quad \underset{1\times 1}{\hat{\sigma}^2} = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}) \; .$$

In addition,

$$\underset{(q+1)\times 1}{\hat{\beta}} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right) \qquad n\frac{\hat{\sigma}^2}{\sigma^2} = \frac{(n-q-1)s^2}{\sigma^2} \sim \chi^2(n-q-1)$$

$$\underbrace{(y - X\beta)'(y - X\beta)}_{\chi^2(n)} = \underbrace{(y - X\hat{\beta})'(y - X\hat{\beta})}_{\chi^2(n-q-1)} + \underbrace{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}_{\chi^2(q+1)}$$

could + & - $X\hat{\beta}$

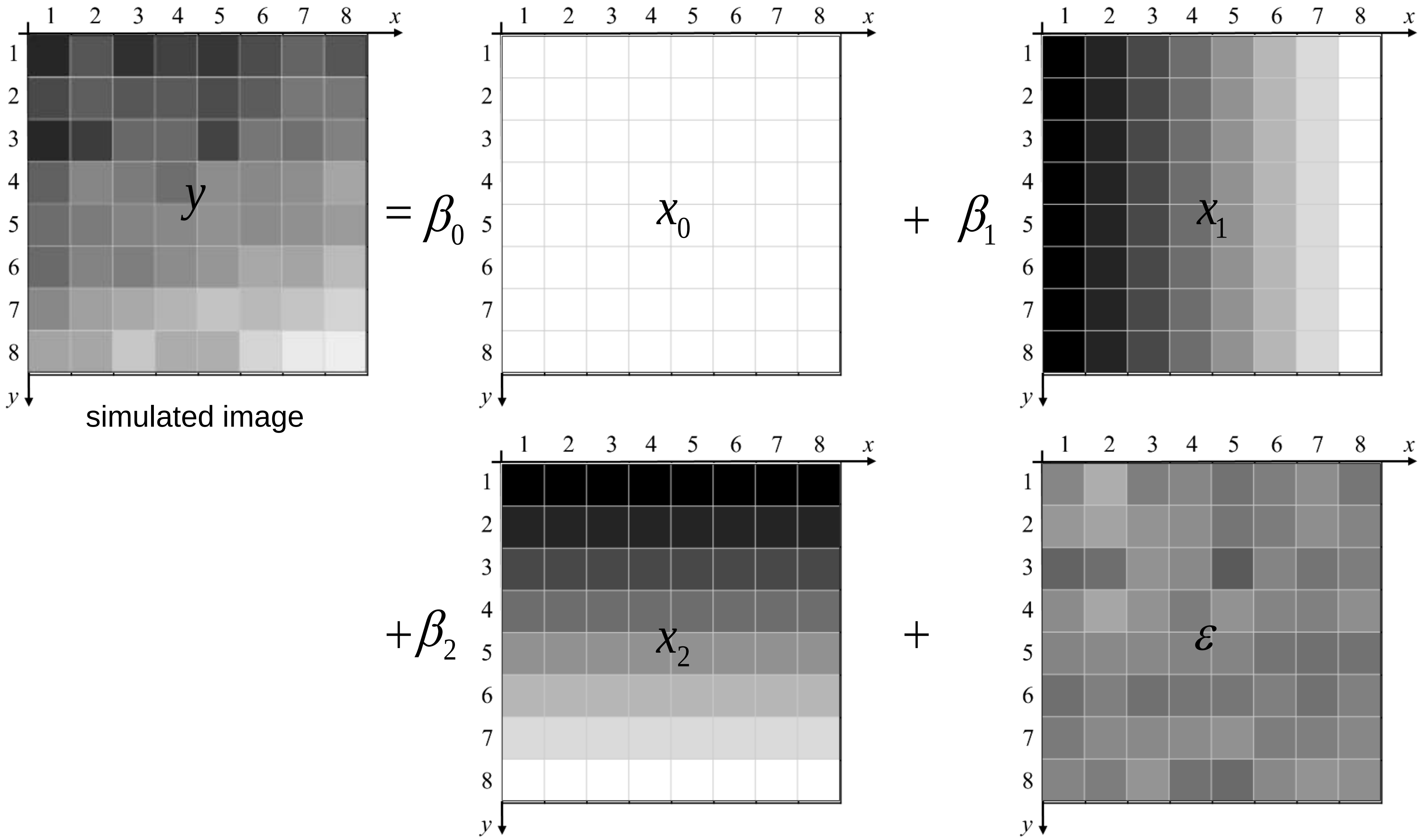
This means we should use a denominator of $n$-$q$-1 for unbiased estimator of $\sigma^2$.

independent

# Least Squares Multiple Regression

$$y = X\beta + \varepsilon$$

**Example:** We can use regression to fit a surface to $(x,y)$ data.

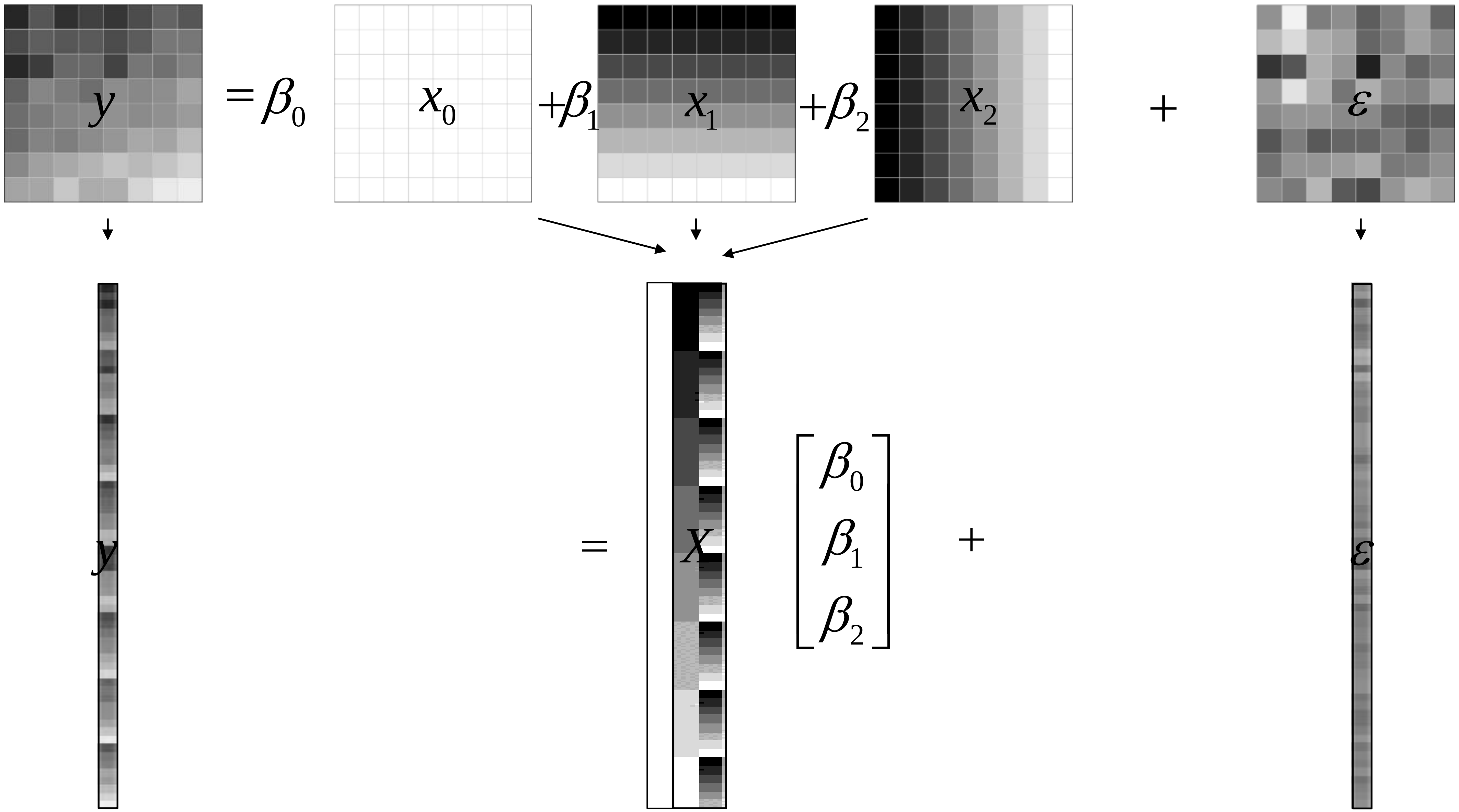



simulated image

# Least Squares Multiple Regression

$$y = X\beta + \varepsilon$$

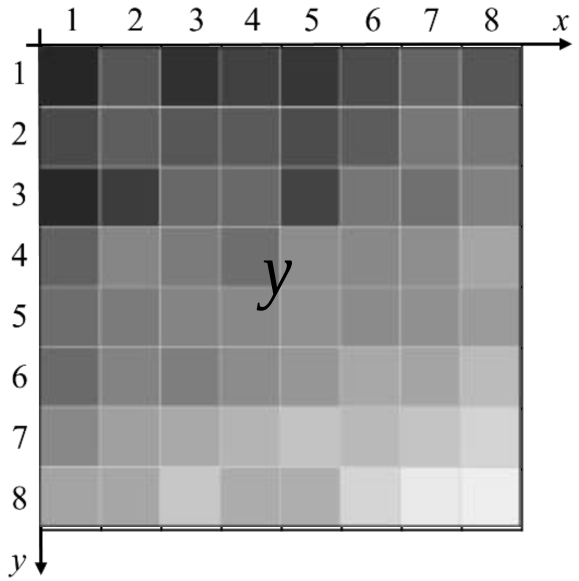
**Example:** We can use regression to fit a surface to $(x,y)$ data.



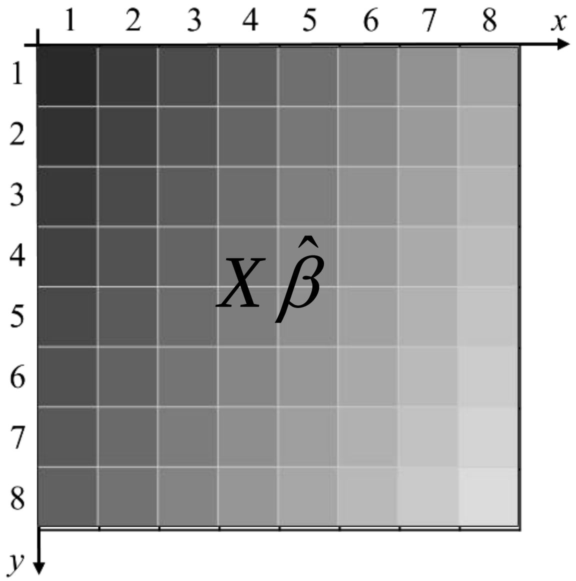
# Least Squares Multiple Regression

$$y = X\beta + \varepsilon$$

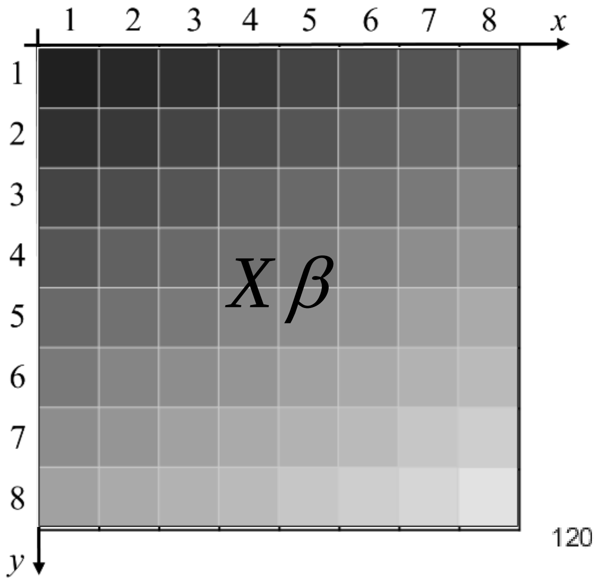
**Example:** We can use regression to fit a surface to $(x,y)$ data.



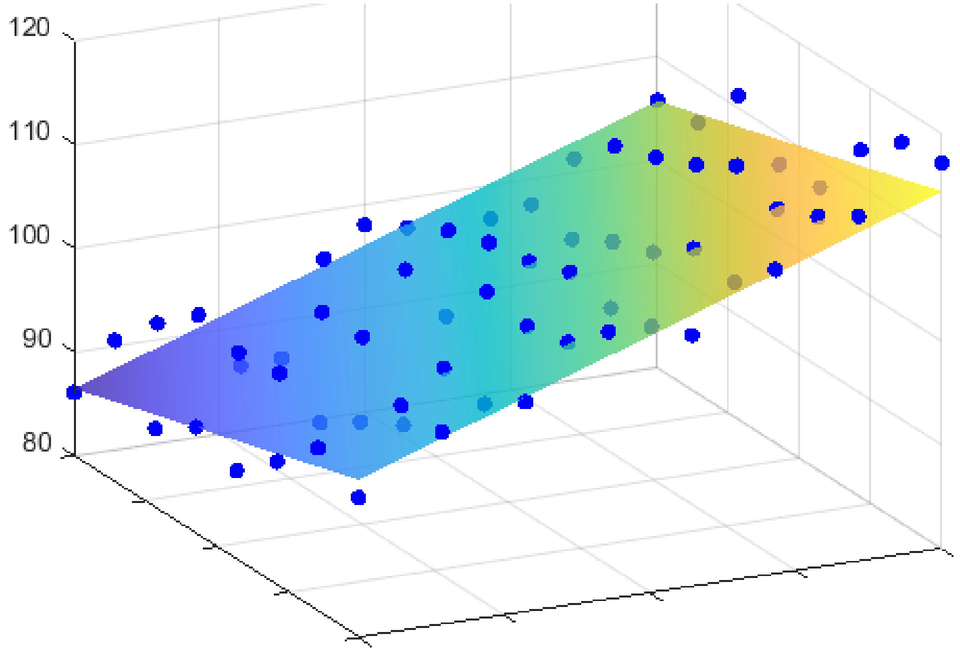

Observed




Estimated




True



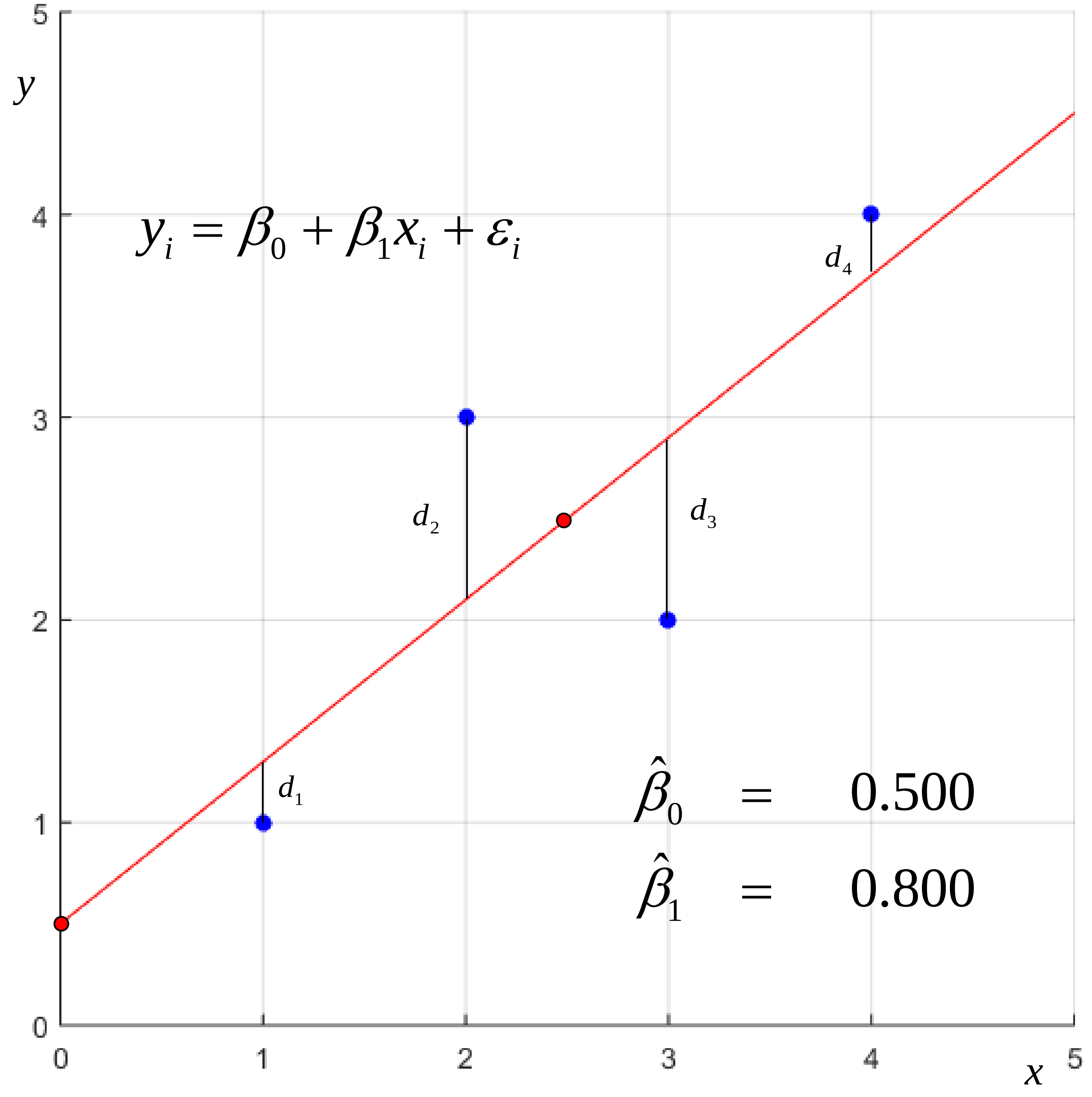

$$\hat{\beta} = (X'X)^{-1}X'y$$

# Least Squares Orthogonal Regression

If we have random error in $y$
$x$ as independent variable
$y$ as dependent variable
and minimized the sum of squared vertical distances from the point to the line.

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \qquad \hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

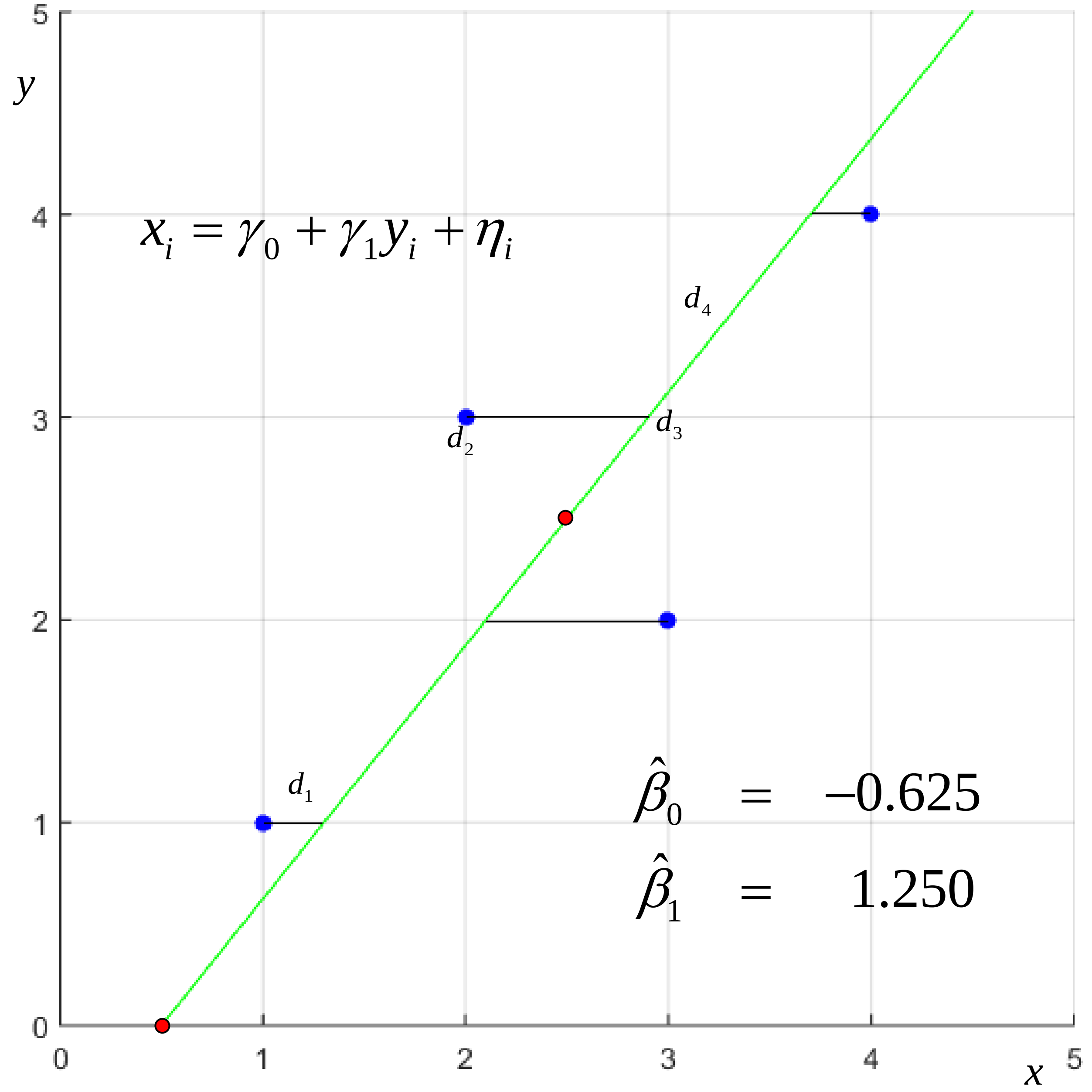
($x$,$y$) pairs: (1,1),(3,2),(2,3),(4,4)



# Least Squares Orthogonal Regression

If we have random error in $x$
$y$ as independent variable
$x$ as dependent variable
and minimized the sum of squared horizontal distances from the point to the line.

$(x,y)$ pairs: (1,1),(3,2),(2,3),(4,4)




$$\hat{\gamma}_1 = \frac{s_{xy}}{s_{yy}} \qquad \hat{\gamma}_0 = \bar{x} - \bar{y}\hat{\gamma}_1$$

$$\hat{\beta}_1 = 1/\hat{\gamma}_0 \qquad \hat{\beta}_0 = -\hat{\gamma}_1/\hat{\gamma}_0$$

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

# Least Squares Orthogonal Regression

When there is measurement error in both $x$ and $y$, we model this as $y_i = y_i^* + \varepsilon_i$ and $x_i = x_i^* + \eta_i$ where $y_i$ and $x_i$ are observed noisy data, $y_i^* = \beta_0 + \beta_1 x_i$ and $x_i^*$ are true unobserved values, $\varepsilon_i \sim N(0, \sigma_y^2)$ and $\eta_i \sim N(0, \sigma_x^2)$ .

If we specify that $\sigma_y^2 = \sigma_x^2 = \sigma^2$, then we have an orthogonal regression.

# Least Squares Orthogonal Regression

In orthogonal regression, $y_i$ and $x_i$ are observed while $y_i^*$ and $x_i^*$ are true unobserved values.

$$y_i = y_i^* + \varepsilon_i$$
$$x_i = x_i^* + \eta_i$$
$$i = 1, \ldots, n$$

Differentiating $Q$ wrt $y_i^*$, $x_i^*$, and $\lambda_i$, then set = 0

$$Q = \sum_{i=1}^{n}\left[(y_i - y_i^*)^2 + (x_i - x_i^*)^2\right] - 2\sum_{i=1}^{n}\lambda_i(y_i^* - \beta_0 - \beta_1 x_i^*)$$

Lagrange Multiplier

$$\left.\frac{\partial Q}{\partial x_i^*}\right|_{\hat{x}_i^*, \hat{y}_i^*, \hat{\lambda}_i} = 0 \qquad \left.\frac{\partial Q}{\partial y_i^*}\right|_{\hat{x}_i^*, \hat{y}_i^*, \hat{\lambda}_i} = 0 \qquad \left.\frac{\partial Q}{\partial \lambda_i}\right|_{\hat{x}_i^*, \hat{y}_i^*, \hat{\lambda}_i} = 0$$

obtain solution to orthogonal regression.

# Least Squares Orthogonal Regression

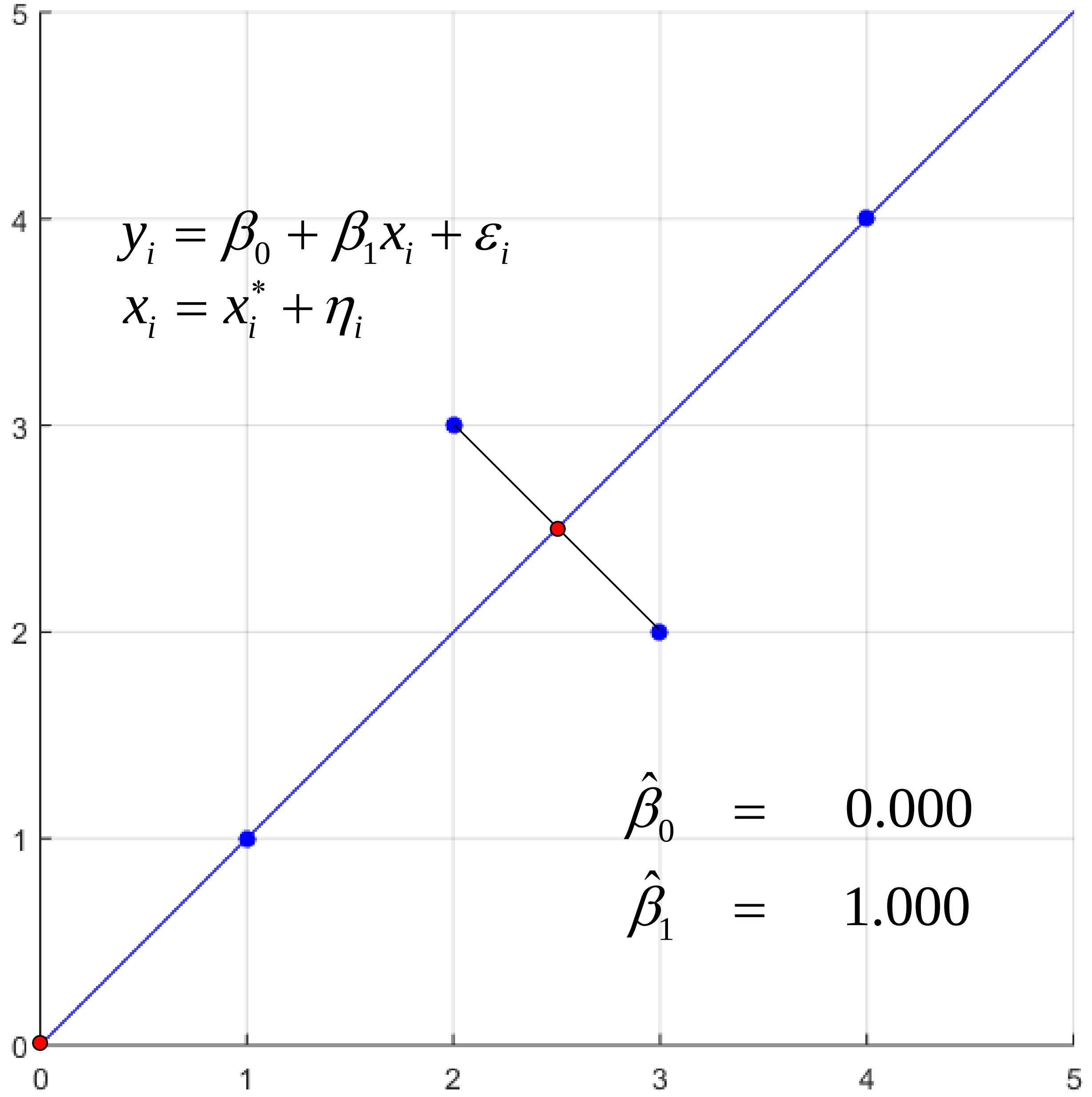
$y$ as dependent variable
$x$ as independent variable
and minimized the sum of squared orthogonal distances from the point to the line.

$$\hat{\beta}_1 = \frac{(s_{yy} - s_{xx}) + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

$$\hat{x}_i^* = x_i + \frac{\hat{\beta}_1}{\hat{\beta}_1^2 + 1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

($x$,$y$) pairs: (1,1),(3,2),(2,3),(4,4)




Adcock. Annals of Mathematics, 5:53-54,1878.
Deming. Statistical Adjustments of Data, 1943.

# Least Squares Orthogonal Regression

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \qquad \hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

---

$$\hat{\gamma}_1 = \frac{s_{xy}}{s_{yy}} \qquad \hat{\gamma}_0 = \bar{x} - \bar{y}\hat{\gamma}_1$$

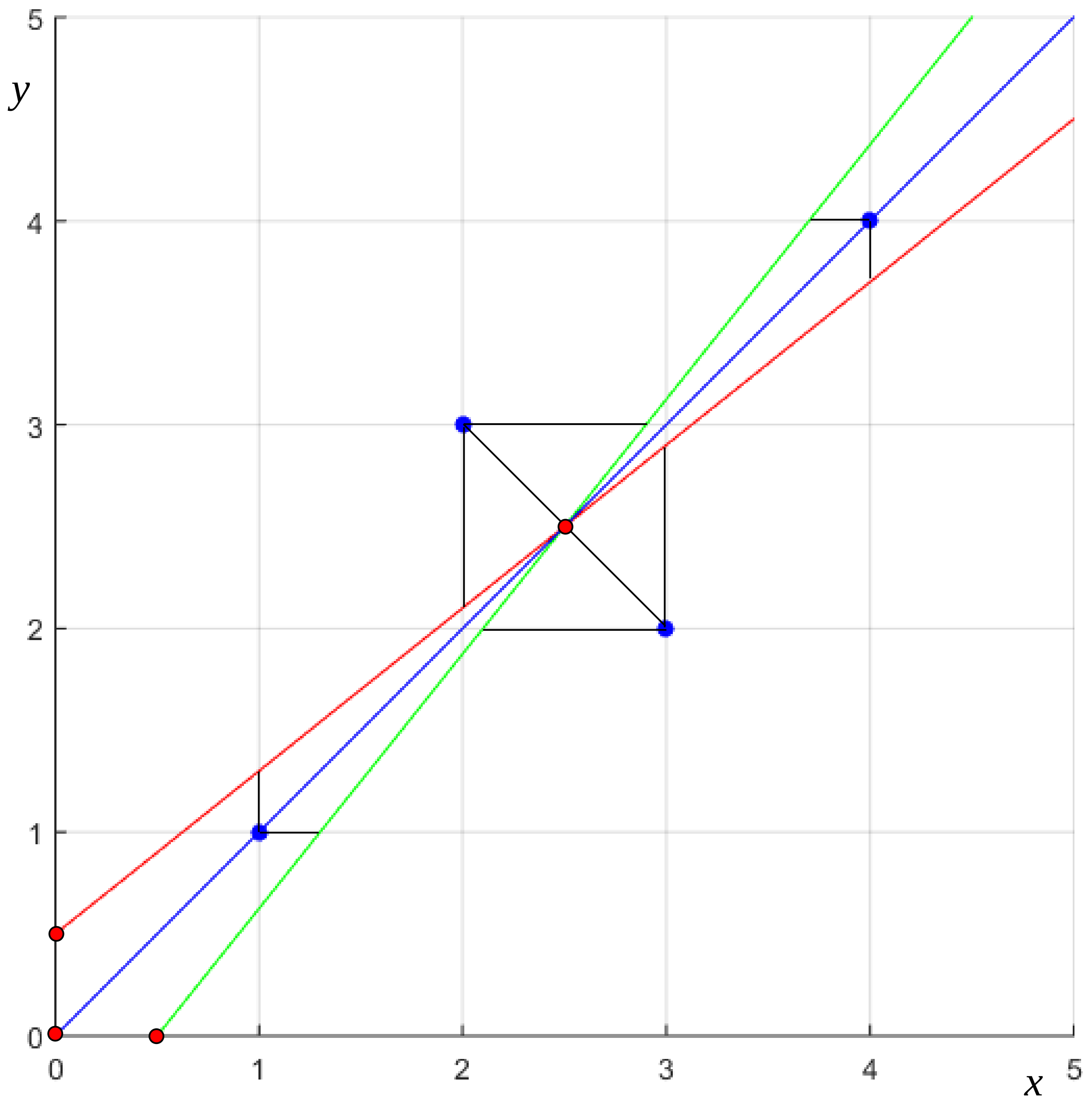
$$\hat{\beta}_1 = 1/\hat{\gamma}_0 \qquad \hat{\beta}_0 = -\hat{\gamma}_1/\hat{\gamma}_0$$

---

$$\hat{\beta}_1 = \frac{(s_{yy} - s_{xx}) + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}}$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

$$\hat{x}_i^* = x_i + \frac{\hat{\beta}_1}{\hat{\beta}_1^2 + 1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$



# Experimental Design - The Beginning

**Example:**

For my lab experiment I know that there is a linear relationship between my independent variable $x$ and dependent variable $y$.
I can select $x$ to be any value between $x_{min}$ and $x_{max}$.

**Option 1:** Spread out $x$'s
Select the $x$ values at every $\Delta x=1/(x_{max}-x_{min})$.

**Option 2:** Clump the $x$'s
Select $n/2$ at $x_{min}$ and select $n/2$ at $x_{max}$.

$$t=\frac{\hat{\beta}_1}{\sqrt{s^2/S_{xx}}} \qquad s^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2 \qquad S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

# Experimental Design - The Beginning

**Example:** $\beta_0$=10, $\beta_1$=1, σ=.5 $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

For my lab experiment I know that there is a linear relationship.

**Option 1:** Spread out $x$'s

$$y \quad = \beta_0 + \beta_1 \; x \; + \; \varepsilon$$

$$\begin{bmatrix} 11.2688 \\ 12.9169 \\ 11.8706 \\ 14.4311 \\ 15.1594 \\ 15.3462 \\ 16.7832 \\ 18.1713 \\ 20.7892 \\ 21.3847 \end{bmatrix} = 10 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{bmatrix} + \begin{bmatrix} 0.2688 \\ 0.9169 \\ -1.1294 \\ 0.4311 \\ 0.1594 \\ -0.6538 \\ -0.2168 \\ 0.1713 \\ 1.7892 \\ 1.3847 \end{bmatrix}$$

**Option 2:** Clump the $x$'s

$$y \quad = \beta_0 + \beta_1 \; x \; + \; \varepsilon$$

$$\begin{bmatrix} 11.2688 \\ 11.9169 \\ 9.8706 \\ 11.4311 \\ 11.1594 \\ 19.3462 \\ 19.7832 \\ 20.1713 \\ 21.7892 \\ 21.3847 \end{bmatrix} = 10 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \end{bmatrix} + \begin{bmatrix} 0.2688 \\ 0.9169 \\ -1.1294 \\ 0.4311 \\ 0.1594 \\ -0.6538 \\ -0.2168 \\ 0.1713 \\ 1.7892 \\ 1.3847 \end{bmatrix}$$

Same Error Added

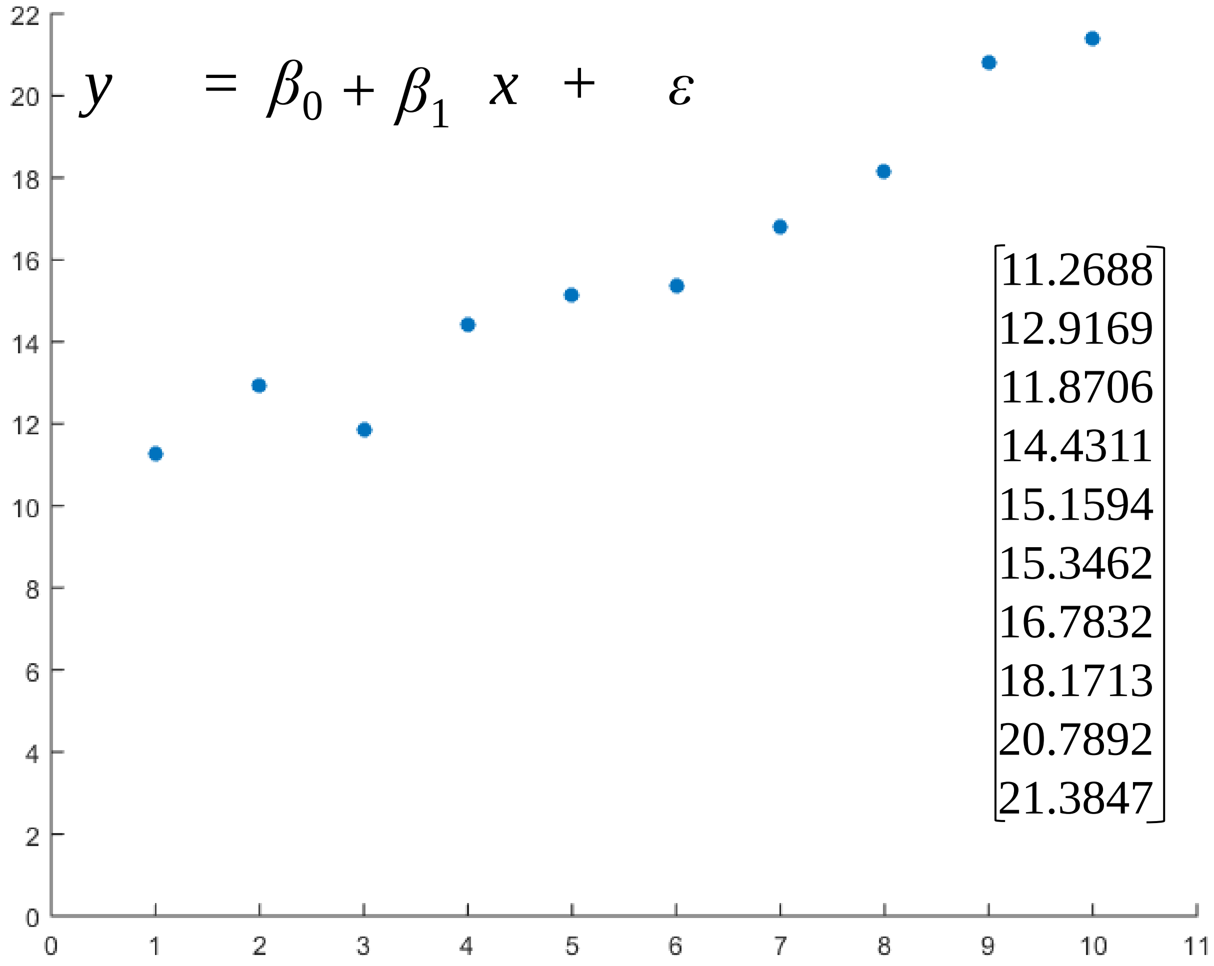
# Experimental Design - The Beginning

**Example:** $\beta_0$=10, $\beta_1$=1, σ=.5

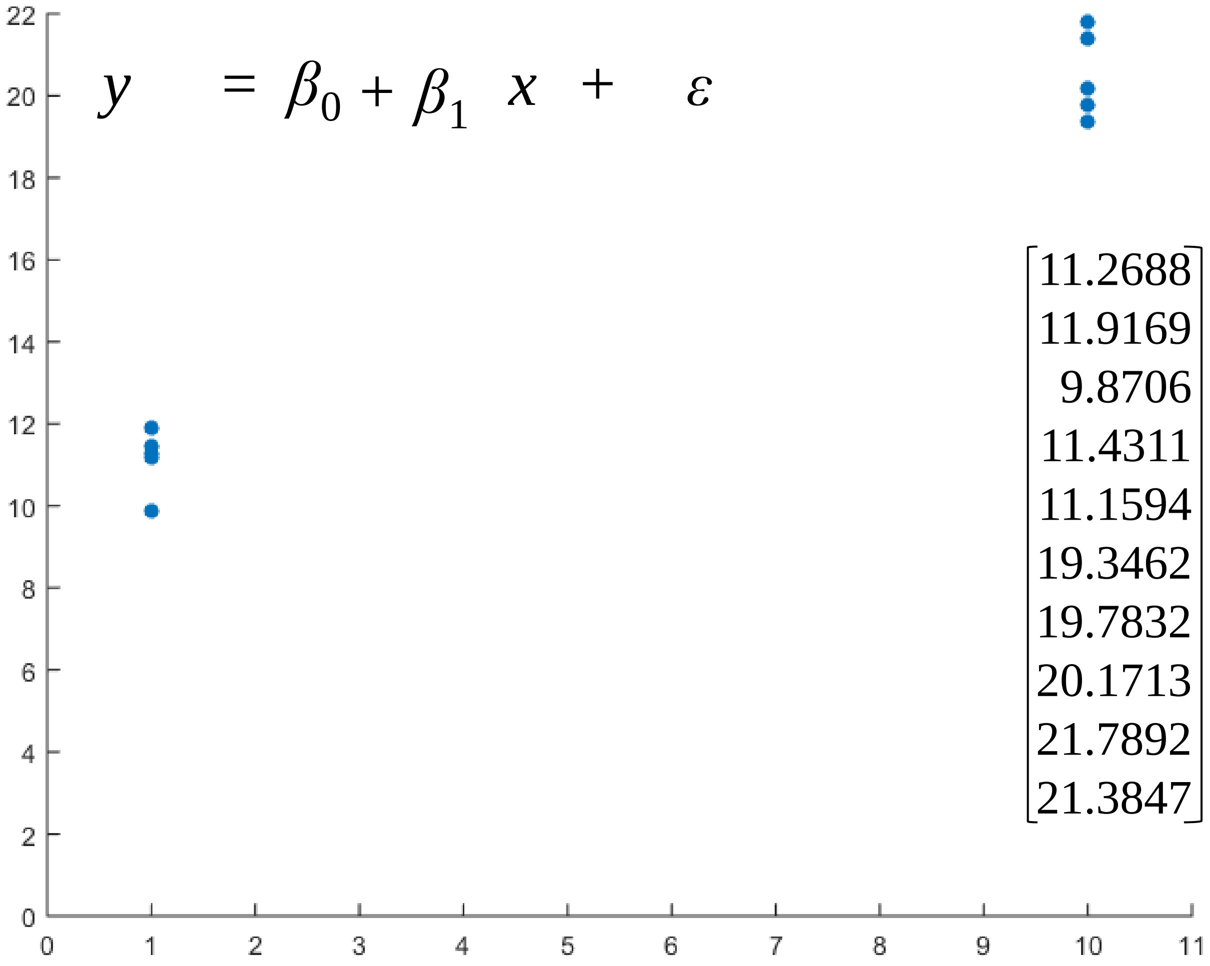
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

For my lab experiment I know that there is a linear relationship.

**Option 1:** Spread out $x$'s




**Option 2:** Clump the $x$'s



# Experimental Design - The Beginning

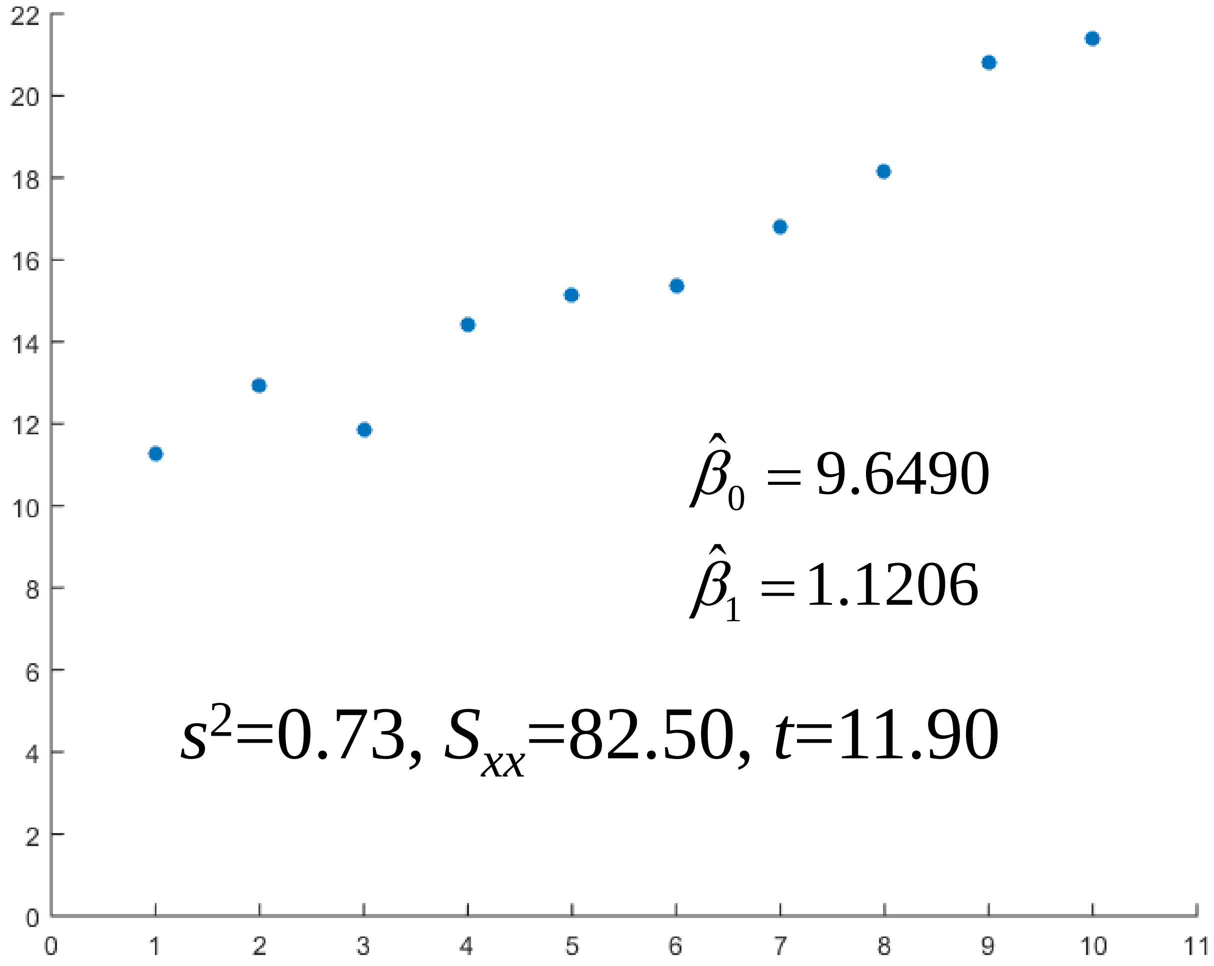
**Example:** $\beta_0$=10, $\beta_1$=1, σ=.5

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

For my lab experiment I know that there is a linear relationship.

**Option 1:** Spread out $x$'s

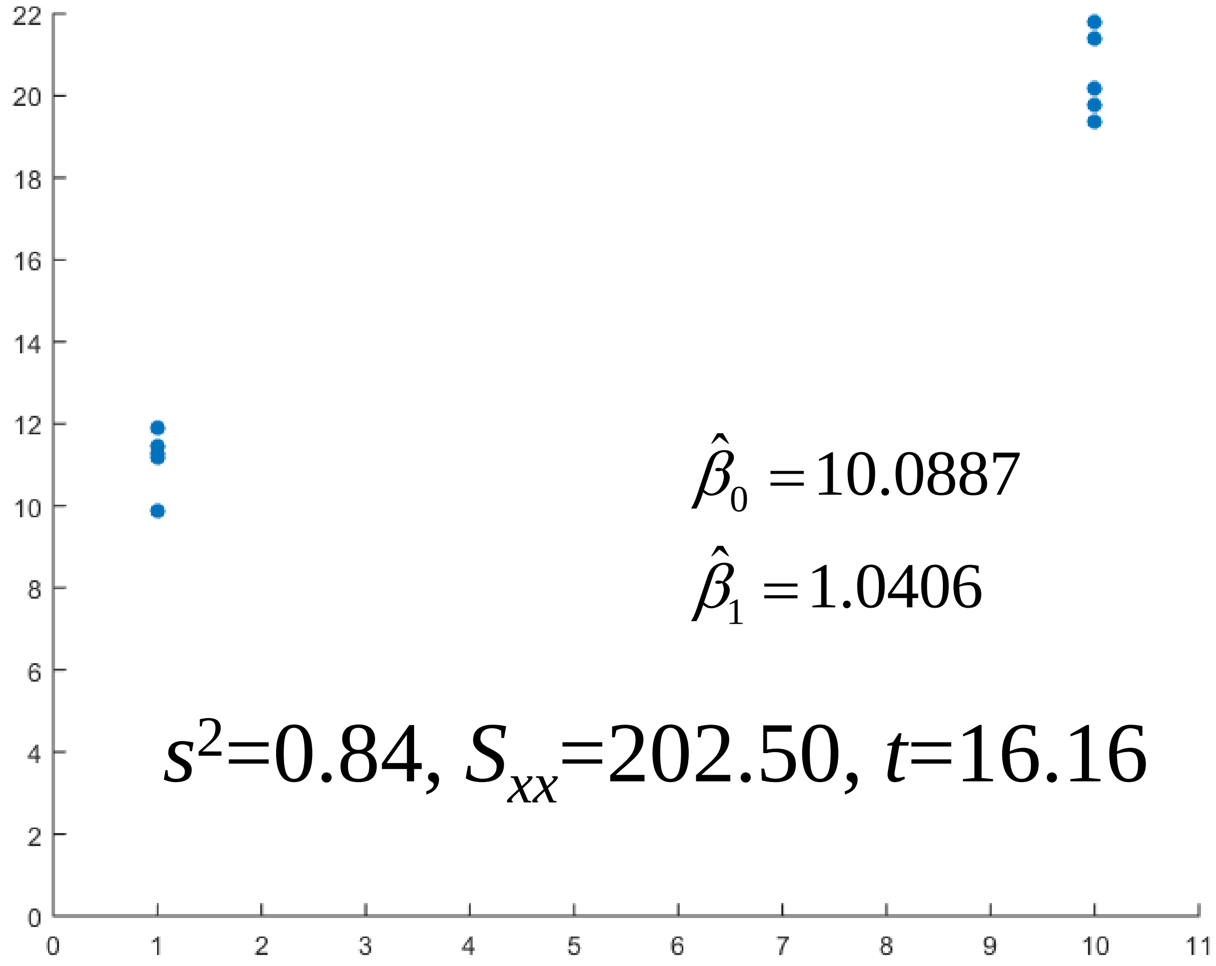



$$\hat{\beta}_0 = 9.6490$$

$$\hat{\beta}_1 = 1.1206$$

$s^2$=0.73, $S_{xx}$=82.50, $t$=11.90

**Option 2:** Clump the $x$'s




$$\hat{\beta}_0 = 10.0887$$

$$\hat{\beta}_1 = 1.0406$$

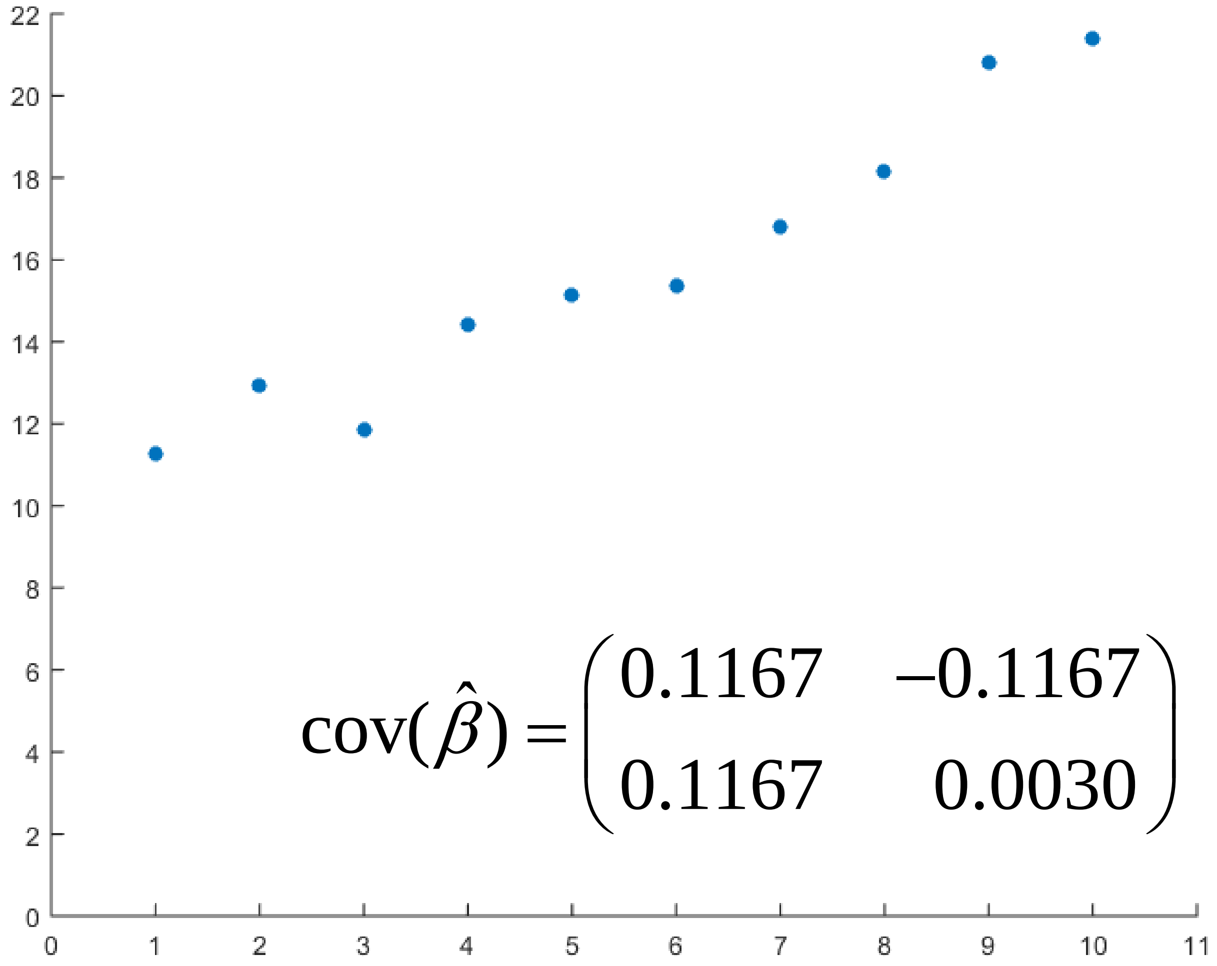
$s^2$=0.84, $S_{xx}$=202.50, $t$=16.16

# Experimental Design - The Beginning

**Example:** $\beta_0$=10, $\beta_1$=1, σ=.5 $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$

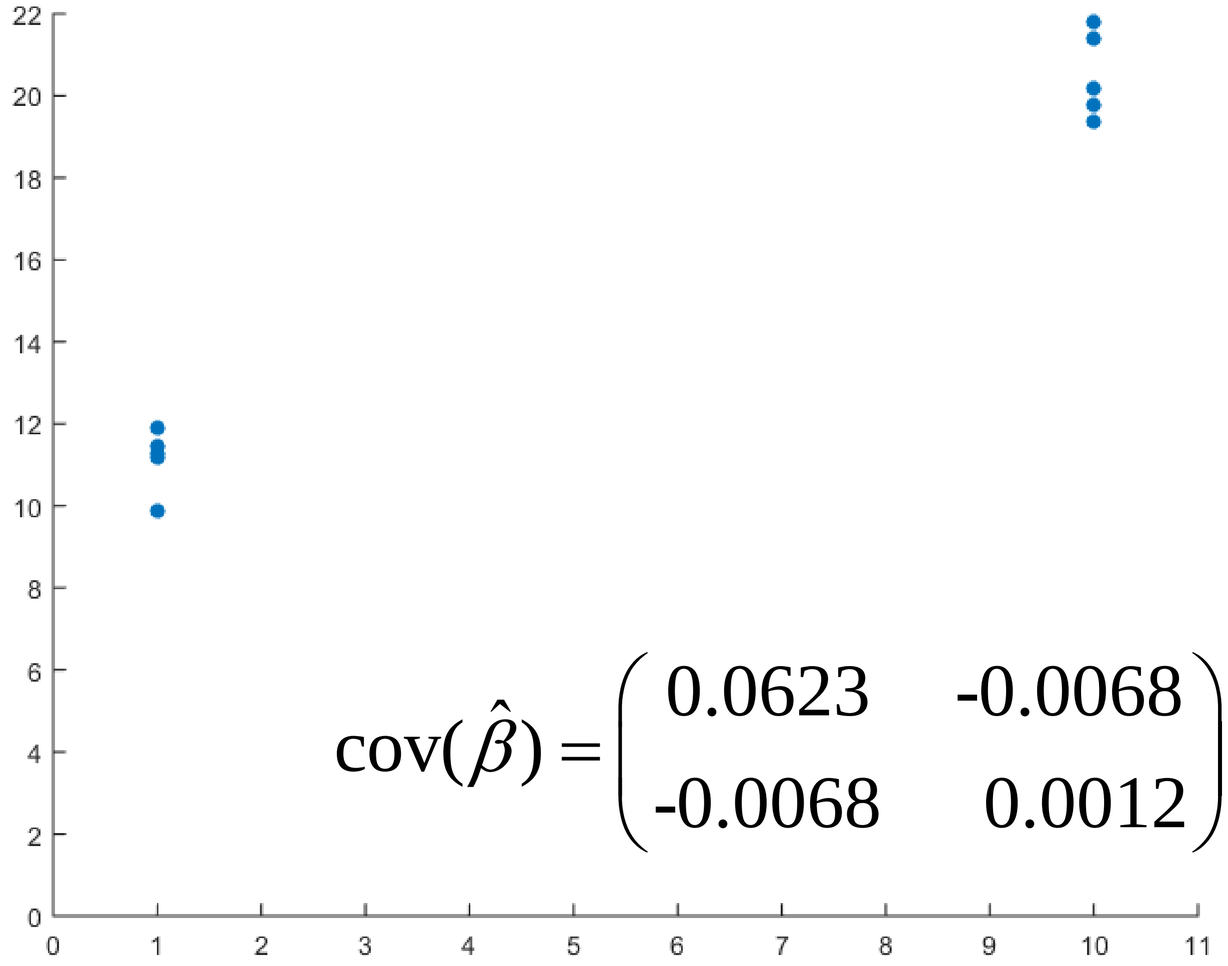
For my lab experiment I know that there is a linear relationship.

**Option 1:** Spread out $x$'s




$$\text{cov}(\hat{\beta}) = \begin{pmatrix} 0.1167 & -0.1167 \\ 0.1167 & 0.0030 \end{pmatrix}$$

**Option 2:** Clump the $x$'s




$$\text{cov}(\hat{\beta}) = \begin{pmatrix} 0.0623 & -0.0068 \\ -0.0068 & 0.0012 \end{pmatrix}$$

# Discussion

## Questions?

# Homework 6

1. For orthogonal regression, let $\beta_0=0$, $\beta_1=1$, $\sigma_x=1$, $\sigma_y=1$.

   Generate $10^6$ random lines. $x^*=[1,2,3,4]$.

   For each set of 4 data points, get estimates $\hat{\beta}_0$ , $\hat{\beta}_1$ , $\hat{x}_i^*$

   $$\hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i^*)^2 \quad \text{and} \quad \hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x}_i^*)^2 .$$

   Plot histograms of estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{x}_i^*$, $\hat{\sigma}_y^2$, $\hat{\sigma}_x^2$ .

   Compute means and variances of estimates.

   Using same data, repeat for ordinary least squares for $\hat{\beta}_0$ , $\hat{\beta}_1$ , $s^2$.

   Compare results.

# Homework 6

2. Create your own solution for the surface fitting example.

   Create a larger image with higher order polynomial true curvature and with noise.
   Fit a higher order surface to it.
   Present true, noisy, and estimated image surfaces.

## Homework 6

3. For the experimental design regression, repeat the two case line simulations $10^6$ times.

   For each simulation for each case, estimate $\hat{\beta}_0$ , $\hat{\beta}_1$ , and $s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1)^2$ .

   Plot histograms of estimates.

   Compute means & variances of estimated values.

   Compute & make histograms for the $s^2(X'X)^{-1}$ .