Linear Models

BIOE 210

A simple linear system

Let's solve the linear system

$$y = \beta x$$

When $y = 2.4$ and $x = 2$. What is the value of β?

A simple linear system

Let's solve the linear system

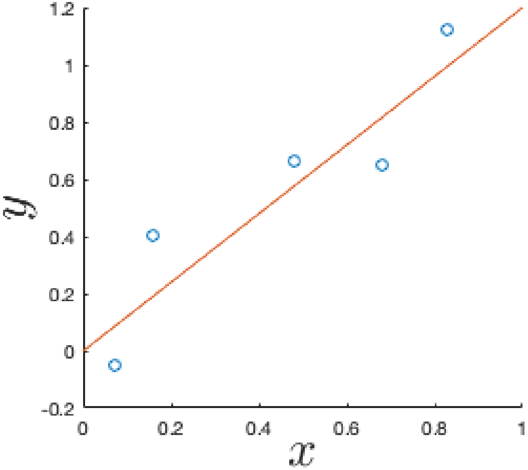
$$y = \beta x$$

When $y = 2.4$ and $x = 2$. What is the value of β?

$$\beta = y/x = 2.4/2 = 1.2$$

A *noisy* linear system

x^{obs}	y^{obs}
0.07	-0.05
0.16	0.40
0.48	0.66
0.68	0.65
0.83	1.12



Each point gives a bad estimate of β

x^{obs}	y^{obs}
0.07	-0.05
0.16	0.40
0.48	0.66
0.68	0.65
0.83	1.12

1. $\beta = -0.05/0.07 = -0.7$
2. $\beta = 0.40/0.16 = 2.5$
3. $\beta = 0.66/0.48 = 1.3$
4. $\beta = 0.65/0.68 = 0.9$
5. $\beta = 1.12/0.83 = 1.3$

Each point gives a bad estimate of β

x^{obs}	y^{obs}
0.07	-0.05
0.16	0.40
0.48	0.66
0.68	0.65
0.83	1.12

1. $\beta = -0.05/0.07 = -0.7$
2. $\beta = 0.40/0.16 = 2.5$
3. $\beta = 0.66/0.48 = 1.3$
4. $\beta = 0.65/0.68 = 0.9$
5. $\beta = 1.12/0.83 = 1.3$

Average $\beta = 1.09$, not the true value of 1.2.
Using *linear regression* we can use the same data to estimate $\beta = 1.21$.

Quantifying Error

The goal of linear regression is to minimize the error between the model and the observed data.

Quantifying Error

The goal of linear regression is to minimize the error between the model and the observed data.

Error has two properties:

1. $\text{Error}(y^{\text{obs}}, y^{\text{pred}}) \geq 0$
2. $\text{Error}(y^{\text{obs}}, y^{\text{pred}}) = 0 \iff y^{\text{obs}} = y^{\text{pred}}$

Quantifying Error

The goal of linear regression is to minimize the error between the model and the observed data.

Error has two properties:

1. $\mathrm{Error}(y^{\mathrm{obs}}, y^{\mathrm{pred}}) \geq 0$
2. $\mathrm{Error}(y^{\mathrm{obs}}, y^{\mathrm{pred}}) = 0 \iff y^{\mathrm{obs}} = y^{\mathrm{pred}}$

There are many ways to define error:

1. $\mathrm{Error}(y^{\mathrm{obs}}, y^{\mathrm{pred}}) = |y^{\mathrm{obs}} - y^{\mathrm{pred}}|$
2. $\mathrm{Error}(y^{\mathrm{obs}}, y^{\mathrm{pred}}) = (y^{\mathrm{obs}} - y^{\mathrm{pred}})^2$

Why choose the squared error?

- Quadratic functions are continuously differentiable.
- The squared error fights against outliers.
- Minimizing squared error has a unique solution.

Fitting Linear Models

1. Choose a model that you think explains the relationship between inputs (x) and outputs (y). The models should contains unknown parameters (β) that you will fit to a set of observations.
2. Find values for β by minimizing the total error between the observed outputs (y^{obs}) and the outputs predicted from the model

   $$\min_{\beta} \sum_{y^{\mathrm{obs}}} \left(y^{\mathrm{obs}} - y^{\mathrm{pred}} \right)^2$$

   Substitute the model you selected in Step 1 in place of y^{pred} in the above minimization.
3. Minimize the function by taking the derivative of the sum squared error and setting it equal to zero. Solve for the unknown parameters β.

Fitting a single parameter linear model

The simplest linear model has only a single parameter

$$y^{\text{pred}} = \beta_0$$

Fitting a single parameter linear model

The simplest linear model has only a single parameter

$$y^{\mathrm{pred}} = \beta_0$$

Assume we have n observations $(x^{\mathrm{obs}}, y^{\mathrm{obs}})$. The total error is

$$\min_{\beta_0} \sum_{i=1}^{n} \left(y_i^{\mathrm{obs}} - y_i^{\mathrm{pred}} \right)^2$$

Fitting a single parameter linear model

The simplest linear model has only a single parameter

$$y^{\text{pred}} = \beta_0$$

Assume we have n observations $(x^{\text{obs}}, y^{\text{obs}})$. The total error is

$$\min_{\beta_0} \sum_{i=1}^{n} \left(y_i^{\text{obs}} - y_i^{\text{pred}} \right)^2$$

Substituting our model we have

$$\min_{\beta_0} \sum_{i=1}^{n} \left(y_i^{\text{obs}} - \beta_0 \right)^2$$

Solving for the unknown parameter β_0

$$\frac{d}{d\beta_0}\left(\sum_{i=1}^{n}(y_i^{\text{obs}} - \beta_0)^2\right) = 0$$

$$\sum_{i=1}^{n}\left(\frac{d}{d\beta_0}(y_i^{\text{obs}} - \beta_0)^2\right) = 0$$

$$\sum_{i=1}^{n}\left(2(y_i^{\text{obs}} - \beta_0)(-1)\right) = 0$$

$$-2\sum_{i=1}^{n}(y_i^{\text{obs}} - \beta_0) = 0$$

$$\sum_{i=1}^{n} y_i^{\text{obs}} - \sum_{i=1}^{n}\beta_0 = 0$$

$$\sum_{i=1}^{n} y_i^{\text{obs}} - n\beta_0 = 0$$

Final solution for the single parameter model

For the simple linear model

$$y^{\mathrm{pred}} = \beta_0$$

the least-squares estimate of the parameter β_0 is

$$\beta_0 = \frac{1}{n} \sum_{i=1}^{n} y_i^{\mathrm{obs}}$$

Final solution for the single parameter model

For the simple linear model

$$y^{\mathrm{pred}} = \beta_0$$

the least-squares estimate of the parameter β_0 is

$$\beta_0 = \frac{1}{n} \sum_{i=1}^{n} y_i^{\mathrm{obs}}$$

This is the mean! So that's where the mean comes from.

The two parameter linear model

The next simplest model has two unknown parameters

$$y^{\text{pred}} = \beta_0 + \beta_1 x$$

The two parameter linear model

The next simplest model has two unknown parameters

$$y^{\mathrm{pred}} = \beta_0 + \beta_1 x$$

How much more work does it take to fit this model? Stop the video and work through §14.2.

Well, that got out of hand quickly.

We can derive estimates for any type of linear model. But the amount of math scales cubicly with the number of parameters (why?).

Instead, let's turn to a matrix formalism for linear models.

A matrix formalism for linear models

Let's write out one equation for each observation of the model $y = \beta_0 + \beta_1 x$.

$$
\begin{aligned}
-0.05 &= \beta_0 + 0.07\beta_1 + \epsilon_1 \\
0.40 &= \beta_0 + 0.16\beta_1 + \epsilon_2 \\
0.66 &= \beta_0 + 0.48\beta_1 + \epsilon_3 \\
0.65 &= \beta_0 + 0.68\beta_1 + \epsilon_4 \\
1.12 &= \beta_0 + 0.83\beta_1 + \epsilon_5
\end{aligned}
$$

A matrix formalism for linear models

Let's write out one equation for each observation of the model $y = \beta_0 + \beta_1 x$.

$$
\begin{aligned}
-0.05 &= \beta_0 + 0.07\beta_1 + \epsilon_1 \\
0.40 &= \beta_0 + 0.16\beta_1 + \epsilon_2 \\
0.66 &= \beta_0 + 0.48\beta_1 + \epsilon_3 \\
0.65 &= \beta_0 + 0.68\beta_1 + \epsilon_4 \\
1.12 &= \beta_0 + 0.83\beta_1 + \epsilon_5
\end{aligned}
$$

$$
\begin{pmatrix} -0.05 \\ 0.40 \\ 0.66 \\ 0.65 \\ 1.12 \end{pmatrix} = \begin{pmatrix} 1 & 0.07 \\ 1 & 0.16 \\ 1 & 0.48 \\ 1 & 0.68 \\ 1 & 0.83 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}
$$

A matrix formalism for linear models

Let's write out one equation for each observation of the model $y = \beta_0 + \beta_1 x$.

$$\begin{aligned}
-0.05 &= \beta_0 + 0.07\beta_1 + \epsilon_1 \\
0.40 &= \beta_0 + 0.16\beta_1 + \epsilon_2 \\
0.66 &= \beta_0 + 0.48\beta_1 + \epsilon_3 \\
0.65 &= \beta_0 + 0.68\beta_1 + \epsilon_4 \\
1.12 &= \beta_0 + 0.83\beta_1 + \epsilon_5
\end{aligned}$$

$$\begin{pmatrix} -0.05 \\ 0.40 \\ 0.66 \\ 0.65 \\ 1.12 \end{pmatrix} = \begin{pmatrix} 1 & 0.07 \\ 1 & 0.16 \\ 1 & 0.48 \\ 1 & 0.68 \\ 1 & 0.83 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Solving the linear system

A few points about $\mathbf{y} = \mathbf{X}\beta + \epsilon$:

- The unknowns are β, not $\mathbf{X}$.
- The coefficient matrix $\mathbf{X}$ is called the *design matrix*.
- The design matrix $\mathbf{X}$ is rarely square.

Solving the linear system

A few points about $\mathbf{y} = \mathbf{X}\beta + \epsilon$:

- The unknowns are β, not $\mathbf{X}$.
- The coefficient matrix $\mathbf{X}$ is called the *design matrix*.
- The design matrix $\mathbf{X}$ is rarely square.

The solution to this system that minimizes the errors in ϵ is

$$\beta = \mathbf{X}^{+}\mathbf{y}$$

where $\mathbf{X}^{+}$ is the *pseudoinverse* of $\mathbf{X}$.

For next time

- Follow through the end of Chapter 14 to see how we calculate and use the pseudoinverse.
- Next time we will demonstrate how to formulate and solve more complex linear models.