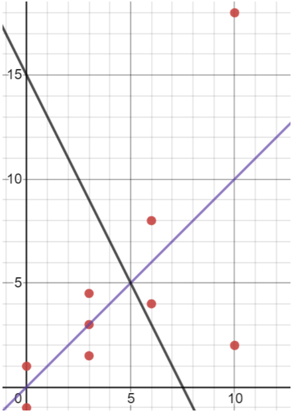
$\in (-1000, 2000)$

CSCI 3022 Intro to Data Science
Regression Inference

• Data
— Line 1
— Line 2

Consider the graph below. Do either of the candidate "best fit" lines violate the 4 big assumptions?



1) Draw a picture
Data + "best fit" line

2) Draw a picture
Errors vs. x-values

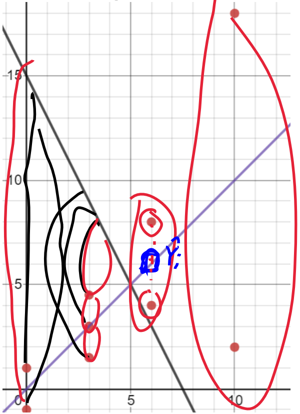
Announcements and Reminders

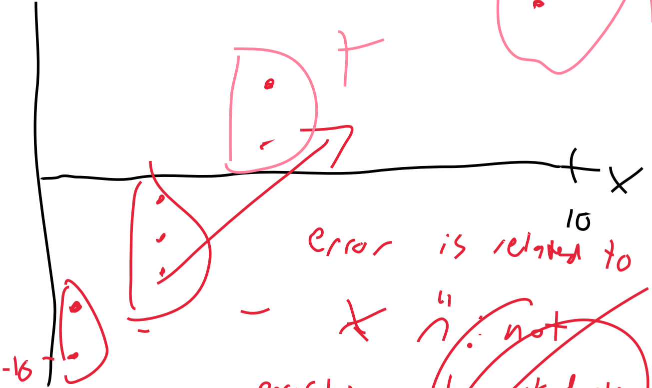
- I forgot to post the minute forms :(. How are you?
- Home stretch! About 4 weeks left, (3 of lecture) with one thing per week! HW 6 and 7 are short and shorter, respectively. Exam 2 follows. Practicum 2 is due at the end of our scheduled "final exam slot."

  Same structure/length as Exam 1!
- We will drop *two* homeworks, not one. Since I said that on a Piazza post. (Oops?... but not really)
- **Notebook day Wednesday, this week**

Do either of the candidate "best fit" lines violate the 4 big assumptions?

One thing to do: plot the errors *as a function of X*:





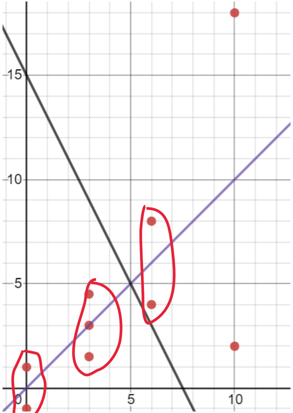
error is related to X

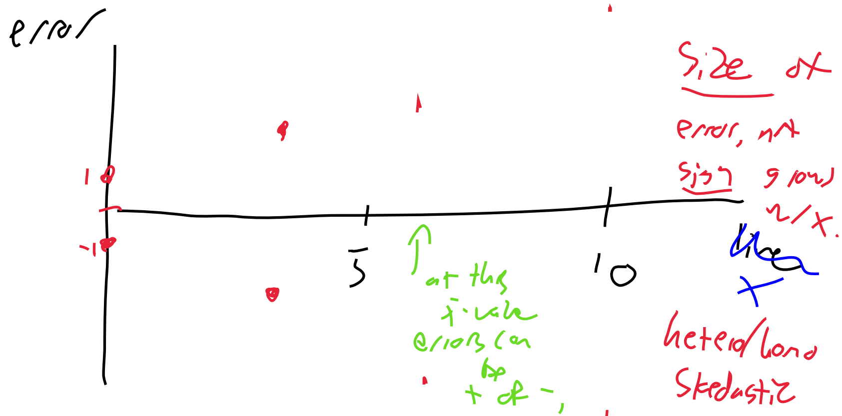
"∴ not ~~homoskedastic~~ -> is heteroskedastic.

error being +/- depends on X: non-independent

Errors are the values given by "data minus estimated line."

Do either of the candidate "best fit" lines violate the 4 big assumptions?
One thing to do: plot the errors *as a function of* X:





Errors are the values given by "data minus estimated line." Black line the errors clump and move up/down as X moves left-right.

but are same size.

Blue line the errors increase in *magnitude* as X goes right.

We've looked at the following test statistics for hypothesis testing.

1. To compare proportions against a baseline or against each other, we use Z-statistics.

$$\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \textbf{ OR } \frac{(\hat{p_1}-\hat{p_2})-\Delta_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}}$$

2. To compare means when the samples are large **or** underlying normal with *known* variances, we also use Z-statistics.

$$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \textbf{ OR } \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \textbf{ OR } \frac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \textbf{ OR } \frac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}}$$

3. To compare means when the samples are small **and** underlying normal, we use t-statistics.

$$t=\frac{\bar{X}-\mu_0}{s/\sqrt{n}} \textbf{ OR } \frac{\left(\bar{X}-\bar{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}}$$

Where we at?

Definition: *Simple Linear Regression* (SLR)

The *Simple Linear Regression* model is a model of the form

With 3 assumptions on ε:

Where we at?

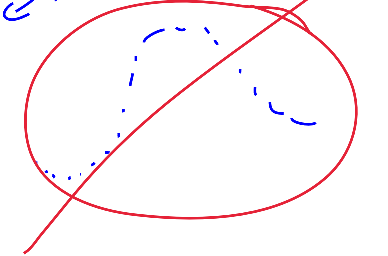
Definition: *Simple Linear Regression* (SLR)

The *Simple Linear Regression* model is a model of the form

is a (straight) line

1.

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



With 3 assumptions on ε:

Where we at?

Definition: *Simple Linear Regression* (SLR)

The *Simple Linear Regression* model is a model of the form

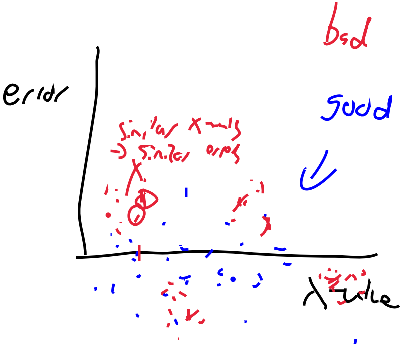
1. $$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

   With 3 assumptions on ε:

2. $$\text{Cov}[\varepsilon_i, \varepsilon_j] = 0 \qquad \forall i, j$$

   Independence of errors

   sign of errors



- no pattern

- no clumping

Where we at?

Definition: *Simple Linear Regression* (SLR)
The *Simple Linear Regression* model is a model of the form

1.
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

With 3 assumptions on ε:

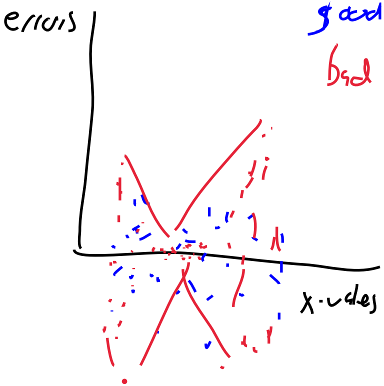
2.
$$\text{Cov}[\varepsilon_i, \varepsilon_j] = 0 \qquad \forall i, j$$

Independence of errors

3.
$$Var(\varepsilon_i) = \sigma^2 \qquad \forall i$$

Homoskedasticity of errors

size of errors (vs. x)



Where we at?

Definition: *Simple Linear Regression* (SLR)

The *Simple Linear Regression* model is a model of the form

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

   With 3 assumptions on ε:

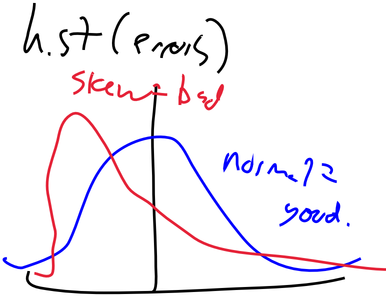
2. $$\text{Cov}[\varepsilon_i, \varepsilon_j] = 0 \qquad \forall i, j$$

   Independence of errors

3. $$Var(\varepsilon_i) = \sigma^2 \qquad \forall i$$

   Homoskedasticity of errors

4. $$\varepsilon_i \sim N(0, 1)$$



Simple Linear Regression Model

best line = least-squares

"min distance from data"

The β estimators in the model are:

1. $\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$ intercept

2. $\hat{\beta}_1 = \frac{Cov[X,Y]}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \boxed{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}$ slope

$\sum_{\text{data}} \underbrace{Y_i}_{\text{data}} - \underbrace{(B_0 + B_1 X_i)}_{\text{line}}$; minimized

Important Terminology:

- x: the independent variable, predictor, or explanatory variable (usually known). x is not random.
- Y: The dependent variable or response variable. For fixed x, Y is random. $Y = B_0 + B_1 x + \varepsilon$ random term ε
- ε: The random deviation or random error term. For fixed x, ε is random. Has variance σ^2. (normal, iid)
- β: the regression coefficients. (GOAL: how does x help predict Y) how do x & y relate?
- r: the *residuals* or observed errors. Used to estimate σ^2.

Y_i : 'true' y-value

$\hat{Y}_i$: our best guess for "Y given x"

Estimating SLR Parameters

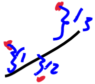
Definitions:

1. The *fitted (or predicted) values* $\hat{Y}_i$ are obtained by plugging in X_i to the equation of the estimated regression line:

   $$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

   estimate $\hat{\beta}_0, \hat{\beta}_1$, draw the resulting line at our original x-values

2. The *residuals* are the differences between the observed and fitted y values:



$$error_i = \overset{\text{error}}{e_i} = \varepsilon_i = \overset{\text{residual}}{r_i} = (\text{data} - \text{estimated line}) = \left(Y_i - \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 X_i)}_{\text{line @ location } x = X_i}\right)$$

Residuals are estimates of the true error. Why?

Estimating SLR Parameters

Definitions:

1. The *fitted (or predicted) values* $\underline{\hat{Y}_i}$ are obtained by plugging in $\underline{\hat{X}_i}$ to the equation of the estimated regression line:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

2. The *residuals* are the differences between the observed and fitted y values:

Residuals are estimates of the true error. Why?

Estimating SLR Parameters

Definitions:

1. The *fitted (or predicted) values* __ are obtained by plugging in __ to the equation of the estimated regression line:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

2. The *residuals* are the differences between the observed and fitted y values:

$$\hat{\varepsilon}_i = r_i = \hat{e_I} = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Residuals are estimates of the true error. Why?

Estimating SLR Parameters

Definitions:

1. The *fitted (or predicted) values* __ are obtained by plugging in __ to the equation of the estimated regression line:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

2. The *residuals* are the differences between the observed and fitted y values:

$$\hat{\varepsilon}_i = r_i = \hat{e}_I = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i$$

σ^2

Residuals are estimates of the true error. Why?
We don't have the true values of $\underbrace{\beta_0, \beta_1}$, so when we estimate them we get variance and error in our estimates.

Estimating SLR Parameters: Results

For a model of the form $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; $\varepsilon \sim N(0, \sigma^2)$

1. $\hat{\beta}_0 =$

2. $\hat{\beta}_1 =$

What happens if $\beta_0 \approx 0$? If $\beta_1 \approx 0$?

Estimating SLR Parameters: Results

For a model of the form $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; $\varepsilon \sim N(0, \sigma^2)$

1. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

2. $\hat{\beta}_1 = \frac{Cov[X,Y]}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

What happens if $\beta_0 \approx 0$? If $\beta_1 \approx 0$?

Estimating SLR Parameters: Results

$y - y_0 = m(x - x_0)$

line through (x_0, y_0), slope m

For a model of the form $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i;\ \varepsilon \sim N(0, \sigma^2)$

1. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$y = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{Y} - \hat{\beta}_1(\bar{X}) + \hat{\beta}_1(x) = \bar{Y} + \hat{\beta}_1(x - \bar{X})$

2. $\hat{\beta}_1 = \frac{Cov[X,Y]}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$

plug in

$y - \bar{Y} = \hat{\beta}_1(x - \bar{X})$

What happens if $\beta_0 \approx 0$? If $\beta_1 \approx 0$?

One result: the regression line goes through $(0, \beta_0)$. It also goes through $\boxed{(\bar{X}, \bar{Y})!}$

middle of data

Estimating SLR Parameters

Definitions:

1. The *fitted (or predicted) values* __ are obtained by plugging in ___ to the equation of the estimated regression line:

2. The *residuals* are the differences between the observed and fitted y values:

Residuals are estimates of the true error. Why?

Estimating SLR Parameters

Definitions:

1. The *fitted (or predicted) values* $\underline{\hat{Y}_i}$ are obtained by plugging in $\underline{\hat{X}_i}$ to the equation of the estimated regression line:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

2. The *residuals* are the differences between the observed and fitted y values:

Residuals are estimates of the true error. Why?

Estimating SLR Parameters

Model!

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Definitions:

1. The *fitted (or predicted) values* __ are obtained by plugging in __ to the equation of the estimated regression line:

estimate $\Rightarrow$ $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\varepsilon}_i$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

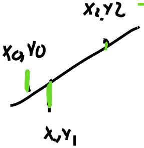
2. The *residuals* are the differences between the observed and fitted y values:

$$\hat{\varepsilon}_i = \boxed{r_i} = \hat{e_I} = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i$$

r

$r = \begin{bmatrix} r_0 \\ r_1 \\ r_2 \end{bmatrix}$

Residuals are estimates of the true error. Why?



Estimating SLR Parameters

Definitions:

1. The *fitted (or predicted) values* __ are obtained by plugging in __ to the equation of the estimated regression line:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

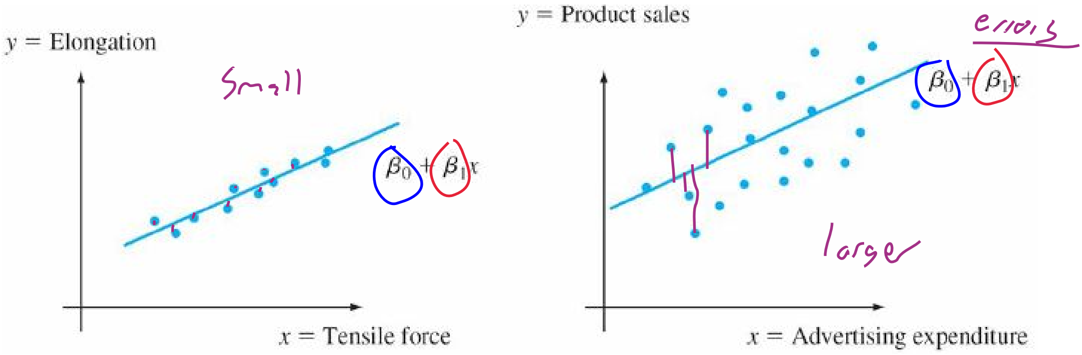
2. The *residuals* are the differences between the observed and fitted y values:

$$\hat{\varepsilon}_i = r_i = \hat{e_I} = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Residuals are estimates of the true error. Why?
We don't have the true values of β_0, β_1, so when we estimate them we get variance and error in our estimates.

Estimating SLR Parameters: $\sigma^2 \approx (\varepsilon)^2$ errors

The parameter σ^2 determines the amount of spread about the true regression line. Two separate examples:



Estimating SLR Parameters: σ^2

variance $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

An estimate of σ^2 will be used in confidence interval formulas and hypothesis testing procedures presented in the next days. Recall that the residual sum of squares or sum of squared errors (SSE) is:

sum of squared errors

$$\text{SSE} = \sum_{i=1}^{n} \left(\epsilon_i \right)^2$$

$$Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X)$$

$r_i = e_i$

$(Y_i - \hat{Y}_i)$

data

fit line

So, our estimate of the variance of the model is like a measure for an average of this summand:

Estimating SLR Parameters: σ^2

An estimate of σ^2 will be used in confidence interval formulas and hypothesis testing procedures presented in the next days. Recall that the residual sum of squares or sum of squared errors (SSE) is:

$$\text{SSE} = \sum (errors)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

So, our estimate of the variance of the model is like a measure for an average of this summand:

Estimating SLR Parameters: σ^2

An estimate of σ^2 will be used in confidence interval formulas and hypothesis testing procedures presented in the next days. Recall that the residual sum of squares or sum of squared errors (SSE) is:

$$\text{SSE}=\sum (errors)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

So, our estimate of the variance of the model is like a measure for an average of this summand:

$$\hat{\sigma^2} = \frac{SSE}{n-2}$$

Wait, what? Why the $n-2$??

not n!!!

Estimating SLR Parameters: σ^2

An estimate of σ^2 will be used in confidence interval formulas and hypothesis testing procedures presented in the next days. Recall that the residual sum of squares or sum of squared errors (SSE) is:

$$\text{SSE} = \sum (errors)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

So, our estimate of the variance of the model is like a measure for an average of this summand:

$$\hat{\sigma}^2 = \frac{SSE}{n-2}$$

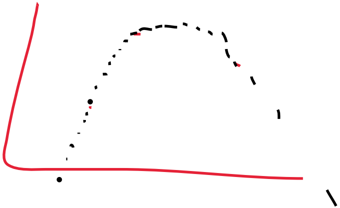
thing #1 $\hat{\beta}_0$

thing #2 $\hat{\beta}_1$

thing #3

Wait, what? Why the $n-2$??
These are again *degrees of freedom*.

Degrees of Freedom Intuition



Suppose you have 3 (random) points on the XY plane.

1. Can you draw a line through them?
2. Can you draw a parabola through them?

   Yes, 1 of 'em

3. Can you draw a cubic function through them?

   $Y = ax^3 + bx^2 + cx + d$: could be any any of a, b, c, d, still has a solution

4. Can you draw a quartic function through them?

Degrees of Freedom Intuition

Suppose you have 3 (random) points on the XY plane.

1. Can you draw a line through them?
   It's very unlikely. In fact, for truly random (normal) points, this result has probability zero!

2. Can you draw a parabola through them?
   Yes, but there's only one such parabola.

3. Can you draw a cubic function through them?
   Yes. Not only that, you could choose *any one* of a, b, c, d in the $ax^3 + bx^2 + cx + d = 0$ and then solve for the others. You have **one degree of freedom**.

4. Can you draw a quartic function through them?
   Yes. Not only that, you could choose *any two* of a, b, c, d, e in the $ax^4 + bx^3 + cx^2 + dx + e = 0$ and then solve for the others. You have **two degrees of freedom**.

Degrees of Freedom

The takeaway?

One property of mathematical estimation: the more you estimate, the more you risk *overfitting*. In this model we've estimated **2** "means" ($\hat{\beta}_0$,$\hat{\beta}_1$) before we got to σ, which "costs" us two degrees of freedom.

The more we estimate, the less options - degrees of freedom - we get for the remaining terms.

Estimating SLR Parameters: σ^2

Some properties of our estimate:

1. The divisor $n-2$ in is the number of degrees of freedom (df) associated with SSE and $\hat{\sigma}^2$.

   $\hat{\beta}_0, \hat{\beta}_1$

2. This is because to obtain $\hat{\sigma}^2$, two parameters must first be estimated, which results in a loss of 2 df.

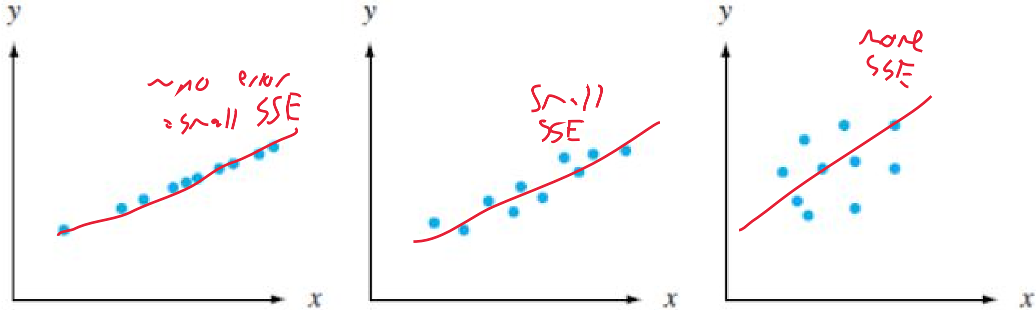
3. Replacing each y_i in the formula for $\hat{\sigma}^2$ by the r.v. Y_i gives a random variable.

   Y values: center at the $\beta_0 + \beta_1 x$ $+ \varepsilon$ $\mathcal{N}(0, \sigma^2)$

4. It can be shown that the r.v. $\hat{\sigma}^2$ is an unbiased estimator for σ^2.

The Coefficient of Determination

The residual sum of squares SSR can be interpreted as a measure of how much variation in y is left unexplained by the model—that is, how much cannot be attributed to a linear relationship. In the first plot, $SSE = 0$, and there is no unexplained variation, whereas unexplained variation is small for second, and large for the third plot.



Picturing Sums of Squares

The goodness-of-fit of a regressive model is often decomposed into three components based on squared deviations. These are:

1. SSE: Sum of squared errors: (vertical) distances from the regression line to the data values.

   size of errors ⇒ "unexplained variability"

2. SST: Sum of squares, total: total deviation in Y. Looks like $Var[Y]$.

   "total variability"

3. SSR: Sum of squares of regression line: the amount of variability tied to the model.

   regressed

   "captured variability"

Picturing Sums of Squares

The goodness-of-fit of a regressive model is often decomposed into three components based on squared deviations. These are:

1. SSE: Sum of squared errors: (vertical) distances from the regression line to the data values.
$$\sum_i \left(\hat{Y} - Y_i\right)^2$$
2. SST: Sum of squares, total: total deviation in Y. Looks like $Var[Y]$.
$$\sum_i \left(Y_i - \bar{Y}\right)^2$$
3. SSR: Sum of squares of regression line: the amount of variability tied to the model.
$$\sum_i \left(\hat{Y}_i - \bar{Y}\right)^2$$

Picturing Sums of Squares

SSR: diff between "sloped line" & "flat line $\bar{y}$"

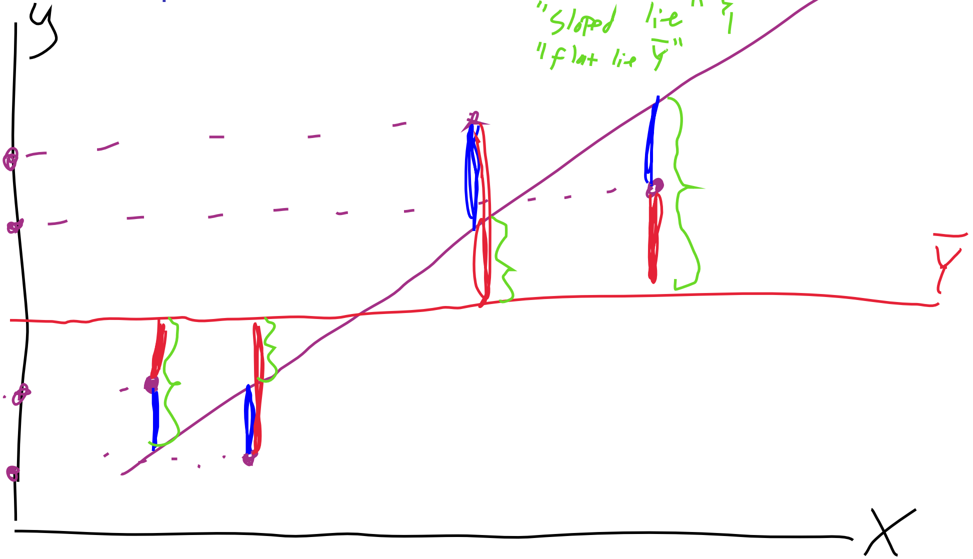
SSE: errors

$\sum (\ )^2$.

SST: total y variability

$\sum (\ )^2$

$\sum \{^2$



The Coefficient of Determination

The sum of squared deviations about the least squares line is smaller than the sum of squared deviations about any other line, i.e. $SSE < SST$ unless the horizontal line itself is the least squares line.

The ratio SSE/SST is the proportion of total variation that cannot be explained by the simple linear regression model. The coefficient of determination is:

This coefficient is a number between 0 and 1 and is the *proportion of observed y variation explained by the model*.

The Coefficient of Determination

The sum of squared deviations about the least squares line is smaller than the sum of squared deviations about any other line, i.e. $SSE < SST$ unless the horizontal line itself is the least squares line.

The ratio SSE/SST is the proportion of total variation that cannot be explained by the simple linear regression model. The coefficient of determination is:

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$$

This coefficient is a number between 0 and 1 and is the *proportion of observed y variation explained by the model*.

The Coefficient of Determination

Again, R^2 is the proportion of observed y variation explained by the model.

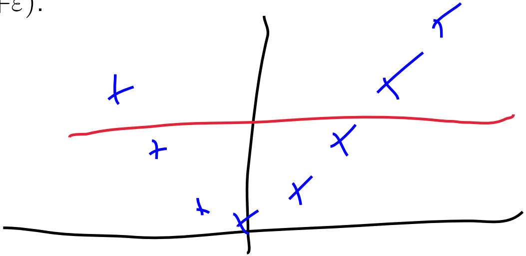
The higher the value of R^2, the more successful is the simple linear regression model in explaining y variation, assuming the linear model is correct.

The Coefficient of Determination

Again, R^2 is the proportion of observed y variation explained by the model.

The higher the value of R^2, the more successful is the simple linear regression model in explaining y variation, assuming the linear model is correct.

Crucially, R^2 is a measure of *linear* dependence between X and Y. If $R^2 = 0$, X and Y may still be related! Ex: $Y = X^2(+\varepsilon)$.



$R^2 = 0 \ldots$

but X does affect Y

Inferences about Parameters

The parameters in SLR have distributions. From these distributions, we can conduct hypothesis tests (e.g., __________), compute confidence intervals, etc.

Distributions:

Inferences about Parameters

The parameters in SLR have distributions. From these distributions, we can conduct hypothesis tests (e.g., $\underline{H_0 : \beta_1 = 0}$), compute confidence intervals, etc.

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}; \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$

Distributions:

Inferences about Parameters

The parameters in SLR have distributions. From these distributions, we can conduct hypothesis tests (e.g., $\underline{H_0 : \beta_1 = 0}$), compute confidence intervals, etc.

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}; \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}$

Distributions:

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\left(X_i - \bar{X}\right)^2}\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\left(X_i - \bar{X}\right)^2}\right)$$

... but of course, we don't know σ^2, so we estimate with $SSE/(n-2)$.

Inferences about Parameters

Confidence Intervals: The CIs for regression are two-sided, and because $\varepsilon \sim N(0, \sigma^2)$, we may use t statistics. Since we have written down the variances of the βs, we can also write down their standard errors:

$$s.e.(\hat{\beta}_0) = \sigma\sqrt{\frac{1}{n}\frac{\bar{X}^2}{\left(X_i - \bar{X}\right)^2}}; \qquad s.e.(\hat{\beta}_1) = \sigma\sqrt{\frac{1}{\left(X_i - \bar{X}\right)^2}}$$

These lead to CIs of

where we replace σ with the estimate $s = \frac{SSE}{n-2}$

Inferences about Parameters

Confidence Intervals: The CIs for regression are two-sided, and because $\varepsilon \sim N(0, \sigma^2)$, we may use t statistics. Since we have written down the variances of the βs, we can also write down their standard errors:

$$s.e.(\hat{\beta}_0) = \sigma\sqrt{\frac{1}{n}\frac{\bar{X}^2}{\left(X_i - \bar{X}\right)^2}}; \qquad s.e.(\hat{\beta}_1) = \sigma\sqrt{\frac{1}{\left(X_i - \bar{X}\right)^2}}$$

These lead to CIs of

$$\beta_i \in (\hat{\beta}_i \pm t_{\alpha/2,n-2} \cdot s.e.(\hat{\beta}_i))$$

where we replace σ with the estimate $s = \frac{SSE}{n-2}$

Inferences about Parameters

Confidence Intervals: The CIs for regression are two-sided, and because $\varepsilon \sim N(0, \sigma^2)$, we may use t statistics. Since we have written down the variances of the βs, we can also write down their standard errors:

$$s.e.(\hat{\beta}_0) = \sigma\sqrt{\frac{1}{n}\frac{\bar{X}^2}{\left(X_i - \bar{X}\right)^2}}; \qquad s.e.(\hat{\beta}_1) = \sigma\sqrt{\frac{1}{\left(X_i - \bar{X}\right)^2}}$$

These lead to CIs of

$$\beta_i \in (\hat{\beta}_i \pm t_{\alpha/2,n-2} \cdot s.e.(\hat{\beta}_i))$$

where we replace σ with the estimate $s = \frac{SSE}{n-2}$

Tests then result from comparing $t = \frac{\hat{\beta}_i}{s.e.(\hat{\beta}_i)}$ to the corresponding critical t values for a one or two-tailed test.

Inferences about Y

There are more types on confidence intervals we may care about!

1. Last slide was how to perform inference on the **parameters** of the *line* β. We also might care about inference on values of Y!
2. A **confidence band** is how sure we are about the mean of Y at specific values of X, or $E[Y|X]$.
3. A **prediction band** is how we estimate the distribution of new Y observations at specific values of X. It's the same as the confidence band, but also includes our estimate for ε. This is also known as a *forecast*.

Idea: If we want to **guess** the *average* $y = \beta_0 + \beta_1 x$, (for a specified x) we have to combine our uncertainties for the βs. If we want to describe *all* the y's for a single value of x, we also would need to include the uncertainty $s^2 \approx \sigma^2$ that accompanies ε.
See: nb accompanying lecture: `SLR Prediction and Confidence`

The usual inference:

The most common inference for linear regression is to answer the question "Does x affect y?" This is a hypothesis test asking about the value of the *slope* of the regression line. We have a CI for this of

$$\beta_1 \pm t_{\alpha/2,n-2} \cdot s.e.(\hat{\beta}_i))$$

where

$$s.e.(\hat{\beta}_1) = \sigma \sqrt{\frac{1}{(X_i - \bar{X})^2}}$$

The corresponding hypothesis test is $t = \frac{\hat{\beta}_i}{s.e.(\hat{\beta}_i)}$ with $n = 2$ degrees of freedom.

Two big things to notice

1. The error grows as σ grows: noisy/random data is harder to estimate.
2. The denominator *looks a lot* like the "standard deviation of x." We get more **confident** in our estimates if the predictor variable locations are spread out!

Daily Recap

Today we learned

1. Regression Inference!

Moving forward:

- nb day Friday

Next time in lecture:

- More Regression! More predictor!