UCI Paul Merage School of Business

Leadership for a Digitally Driven World™

MFIN 290: Financial Econometrics

Lecture 2-1

Lecture Notes

- I recommend you look at the notes before lecture, take notes on a copy during, and then revisit the slides after lecture.

Last time

- Matrices
- Matrix Manipulation
- Matrix Multiplication
- Matrix Differentiation – Linear and Quadratic Forms

This time

- Ordinary Least Squares
- Gauss Markov Assumptions/Properties of OLS
- Hypothesis Testing

- Textbook references: 2.3-2.9, 5.1-5.5

Ordinary Least Squares

- We have n observations of a dependent variable y
- n observations of k different features for k independent variables $\mathrm{X} = [x_k]$

- And a linear relationship between them such that:

- $\begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1K} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nK} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \dots \\ \epsilon_n \end{bmatrix}$

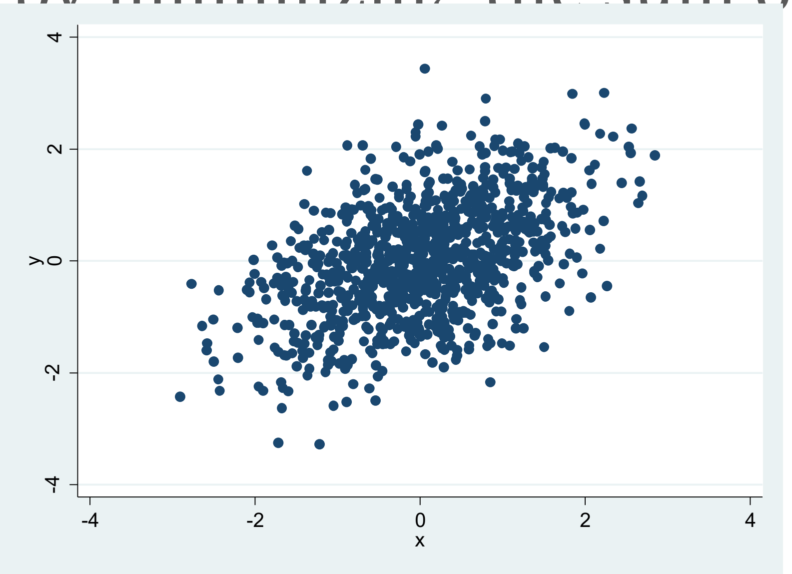
- Q: What does the dimension of β have to be here?

OLS

- OLS is an estimator that seeks to find the "best" value for β, which we call $\hat{\beta}$ or b.

- It finds the best set of coefficients by minimizing the sum of squared residuals:

- $\sum \epsilon_i^2 = \epsilon'\epsilon = (y - x\beta)'(y - x\beta)$



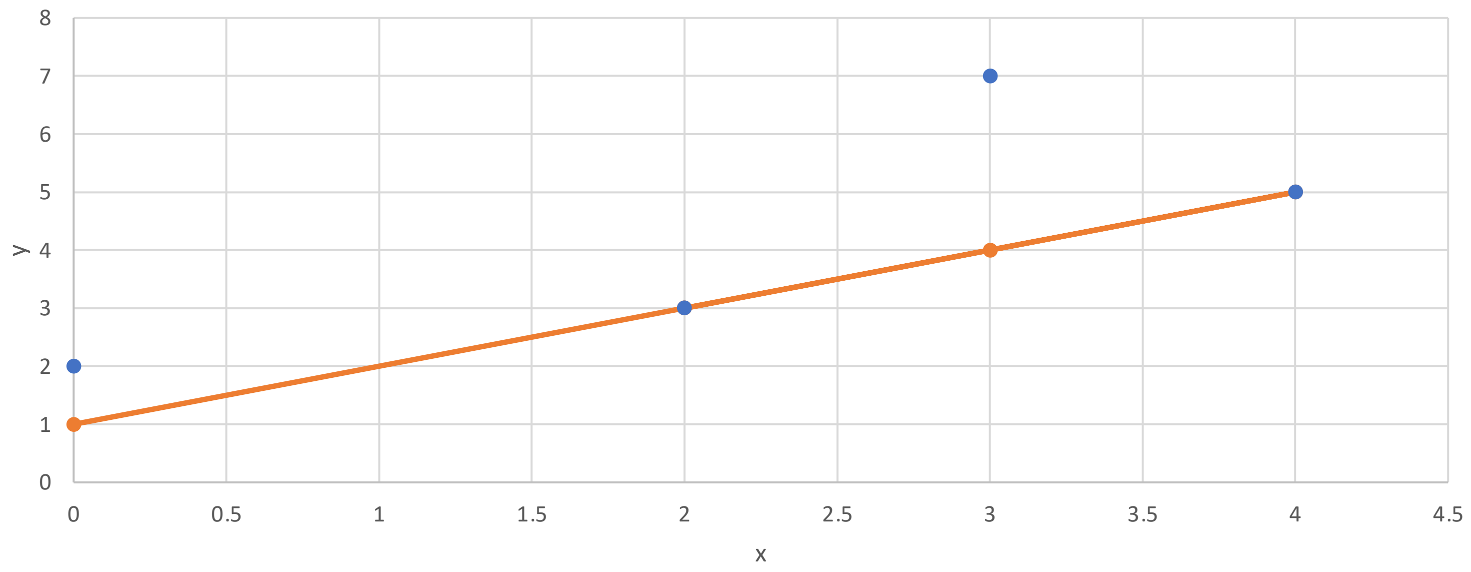
Finding the “right” $\beta's$

- We can square each of the errors and add them up.
- For any $\hat{\beta}$ values, there will be specific $\hat{y}$ values and specific $\hat{\epsilon}$

	y	intercept	x	$\hat{y}(\beta_1 = \beta_2 = 1)$	$\hat{\epsilon} = y - \hat{y}$	$\hat{\epsilon}^2$
	3	1	2	3	0	0
	5	1	4	5	0	0
	7	1	3	4	3	9
	2	1	0	1	1	1
Total						**10**

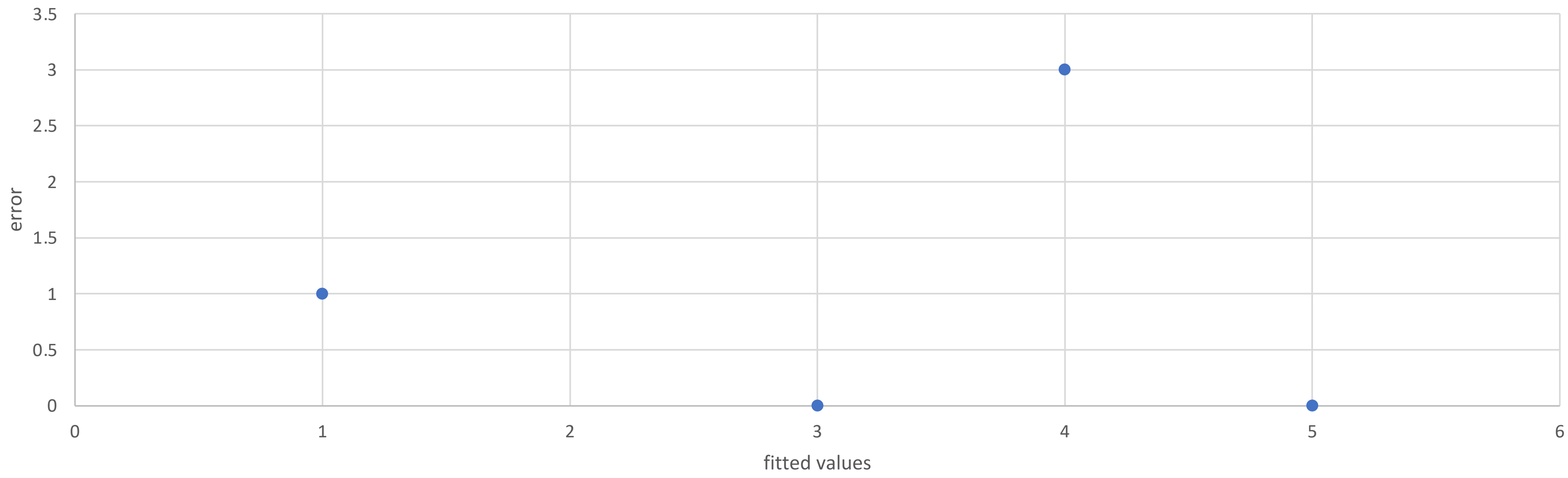
Finding the "right" $\beta's$

- Sometimes it can be easier to visually inspect the relationship of the dependent variable and the independent variable(s)



Finding the “right” $\beta's$

- Sometimes it can be easier to visually inspect the relationship of the fitted values and the residuals… What’s missing here?



Finding the “right” $\beta's$

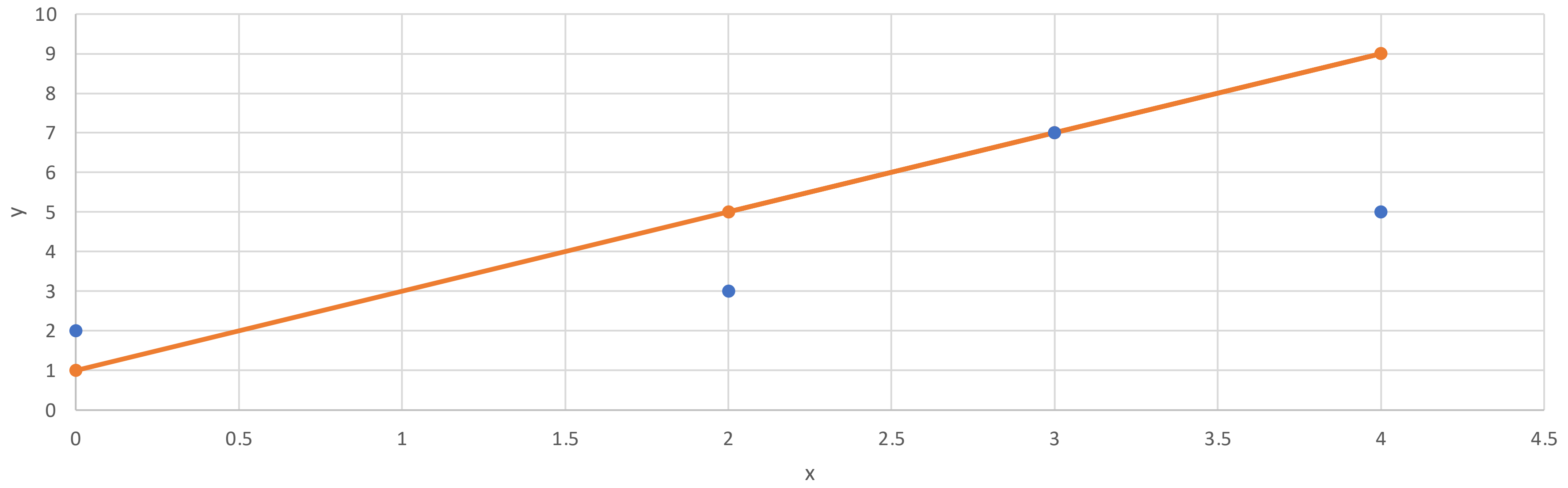
- Least squares betas give a smaller sum of squared errors… hence the name

Note that the mean of the errors is zero

	y	intercept	x	$\hat{y}(\beta_1 = 2, \beta_2 = 1)$	$\hat{\epsilon} = y - \hat{y}$	$\hat{\epsilon}^2$
	3	1	2	4	-1	1
	5	1	4	6	-1	1
	7	1	3	5	2	4
	2	1	0	2	0	0
Total						**6**

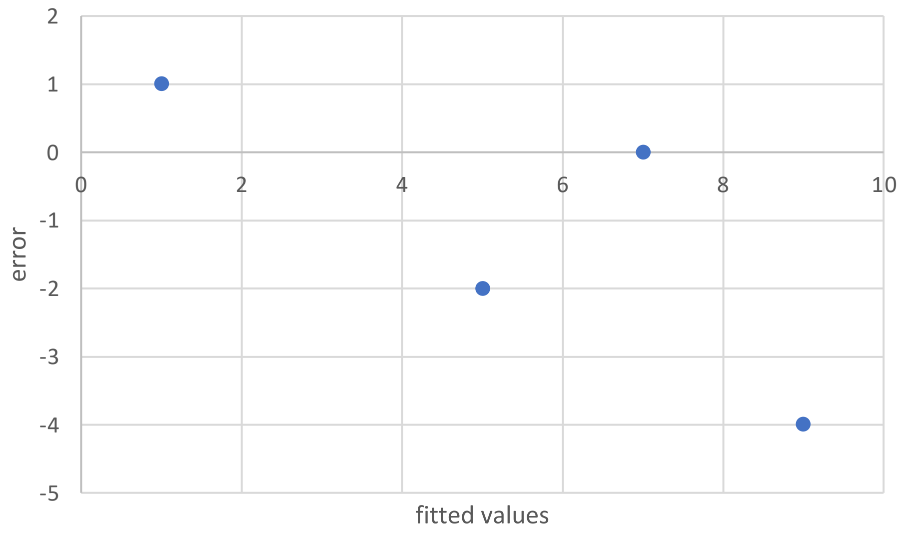
Finding the "right" $\beta's$

- Sometimes it can be easier to visually inspect the relationship of the dependent variable and the independent variable(s)



Finding the “right” $\beta's$

- Sometimes it can be easier to visually inspect the relationship of the dependent variable and the independent variable(s)



First: Single Variable OLS

- Single variable, y and x both mean zero:
- *=> both mean zero implies that the intercept is zero…have a single coefficient here*

- $y_i = x_i\beta + e$
- $\hat{\beta} = \ argmin \ \sum_{i=1}^{N} e_i^2$
- $= argmin \sum_{i=1}^{N}(y_i - x_i\beta)^2$

- $\sum_{i=1}^{N} e_i^2$ is the sum of squared residuals.

First: Single Variable OLS

- $= argmin \sum_{i=1}^{N}(y_i - x_i\beta)^2$

- FOC:
- $2\sum_{i=1}^{N}(y_i - x_i\beta)x_i = 0$
- $\sum_{i=1}^{N} x_i y_i - \beta \sum_{i=1}^{N} x_i{}^2 = 0$
- $\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i y_i)}{\sum_{i=1}^{N}(x_i x_i)} = \frac{NCov(x,y)}{NVar(x)} = \quad Cov(x,y)/\,Var(x)$

Multivariate/Matrix OLS

$y = x\beta + \epsilon$

- y is n x 1
- X is n x k. If we have an intercept, this includes a column of 1's.
- ϵ is n x 1

- β is a k x 1 vector of coefficients

Multivariate/Matrix OLS

- Sum of squared residuals
- $= \sum_{i=1}^{N} \epsilon_i^2$
- $= \epsilon'\epsilon$
- $= (y - x\beta)'(y - x\beta)$
- … Factor this out (remember (AB)’ = B’A’)
- $y'y - \beta'x'y - y'x\beta + \beta'x'x\beta$

OLS

- $(y - x\beta)'(y - x\beta)$ ... Factor this out (remember (AB)' = B'A')
- $y'y - \beta'x'y - y'x\beta + \beta'x'x\,\beta$
- Differentiate w.r.t. β:
- $\dfrac{\partial y'y}{\partial \beta} = 0$
- $\dfrac{\partial \beta'x'y}{\partial \beta} = x'y$ since $\dfrac{\partial \beta' A}{\partial \beta} = A$
- $\dfrac{\partial y'x\beta}{\partial \beta} = (y'x)' = x'y$ since $\dfrac{\partial A\beta}{\partial \beta} = A'$
- $\dfrac{\partial \beta'x'x\,\beta}{\partial \beta} = 2x'x\beta$ since $\dfrac{\partial \beta' A\beta}{\partial \beta} = 2A\beta$

OLS

- $y'y - \beta'x'y - y'x\beta + \beta'x'x\ \beta$
- $FOC\ \beta$:
  $0x'y - x'y - 2x'xB$
- $2x'(y - x\beta) = 0$

- Note that this is equivalent to $X'\epsilon\ =\ 0$ from before since $y = x\beta + \epsilon \Rightarrow \epsilon = y - x\beta$
- $2x'y = 2x'x\beta$ …

- $b = (x'x)^{-1}x'y$

OLS – Assumed Background

- There are several concepts from undergraduate econometrics that we may leverage in this class. If you haven't seen these concepts or topics before, please consult Hill/Lim or ask me for resources.

1. *Frisch Waugh Theorem*
2. R^2 *and the interpretation as percentage of variance in y explained by X*
3. *Including a constant term* ⇔ *Demeaning the regression.*
4. *Independence of the estimates of* $\hat{\beta}$ *and estimates of* σ^2

-

OLS – Assumptions

1) Linearity: the "true" model is linear and as estimated.

- $Y = X\beta + \varepsilon$ is actually the data generating process

- 2) If X is an n x k matrix of regressors, X has full rank (the columns are linearly independent).
  - This ensures that $(X'X)^{-1}$ is defined in our formula for the solution

OLS – Assumptions

- 3) $E[\varepsilon|X] = 0$ (called exogeneity/identification)

  - NOTE: No matter what regression you run, your estimated residuals will have this property. This is quite literally how the betas are identified: get the coefficients by setting $E[\varepsilon|X] = 0$ and solving for the values that make that true (see the FOCs above).

- 4) $Var(\varepsilon) = E[\varepsilon\varepsilon'] = \sigma^2 I.$ Homoscedasticity/"Sphericality"/no autocorrelation.

OLS – Assumptions

- 4) $Var(\varepsilon) = E[\varepsilon\varepsilon'] = \sigma^2 I$. Homoscedasticity/"Sphericality"/no autocorrelation.

$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \cdots \\ \varepsilon_n \end{bmatrix}$… what are the elements of the n x n matrix $\varepsilon\varepsilon'$?

OLS – Assumptions

- 4) $Var(\varepsilon) = E[\varepsilon\varepsilon'] = \sigma^2 I.$ Homoscedasticity/”Sphericality”/no autocorrelation.
- Called “Variance – Covariance” Matrix. In this case, the covariance of every error observation with every other error observation.
  - Awfully nice that it only has one parameter….

What does $\sigma^2 I$ look like?

What is the dimension of $\varepsilon\varepsilon'$ if ε is n x 1?

OLS – Assumptions

- 4) $Var(\varepsilon) = E[\varepsilon\varepsilon'] = \sigma^2 I.$ Homoscedasticity/"Sphericality"/no autocorrelation.
- Called "Variance – Covariance" Matrix. In this case of every error observation with every other error observation.
- N rows x N columns: $\begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}$
- This is going to be symmetric… we would we have *many* free parameters otherwise!
- How many? $\frac{(N-1)^2}{2}$ = half of the matrix ex diagonal, plus the diagonal $= \frac{(N-1)^2}{2} + N$

Example: CAPM

$$y_{it} = r_f + \beta_i(r_{mt} - r_f)$$

An asset's return (y_{it}) is related to the riskfree rate (r_f), the market rate of return (r_{mt}) and n asset-specific a measure of the sensitivity to the market portfolio β_i

Example: CAPM

Stata. Dataset capm4n (on campus site)

We are going to estimate

$$excessretIBM_t = \beta_0 + \beta_1 excessretSP500_t + e_t$$

Where excess return = returns less the risk free rate

Commands:

Insheet = bring in file as a dataset (use with .csv)

gen = generate new variables
reg = regress y on x

Example: CAPM

```
. gen ibm_excess = ibm-riskfree

. gen mkt_excess = mkt - riskfree

. reg ibm_excess   mkt_excess

      Source |       SS           df       MS      Number of obs   =       132
-------------+----------------------------------   F(1, 130)       =     88.32
       Model |  .432585943         1  .432585943   Prob > F        =    0.0000
    Residual |  .636721741       130  .00489786    R-squared       =    0.4045
-------------+----------------------------------   Adj R-squared   =    0.4000
       Total |  1.06930768       131  .008162654   Root MSE        =    .06998

------------------------------------------------------------------------------
  ibm_excess |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  mkt_excess |   1.188208   .1264327     9.40   0.000     .9380763     1.43834
       _cons |   .0058513   .0060914     0.96   0.339    -.0061999    .0179024
------------------------------------------------------------------------------
```

Properties of OLS - Unbiased

- $E(b|X) = \beta$
- $E\left((X'X)^{-1}X'y|X\right) = E\left((X'X)^{-1}X'(X\beta + \epsilon)|X\right)$

  here, we are using β when we substitute, not $\hat{\beta}$ => the true value
- $E\left((X'X)^{-1}X'(X\beta + \epsilon)\right) = E(\beta) + E\left((X'X)^{-1}X'\epsilon|X\right) = \beta$
- This used

1. the fact that the model was correct (linearity) to substitute $X\beta + \epsilon$ (with the true beta!) for Y
2. that X was full rank (i.e., $(X'X)^{-1}$ exists) and
3. that $E[\varepsilon|X] = 0$.

- Didn't use homoscedasticity or sphericality at all… so why are we making this strong assumption?? Will return to this….

What does Unbiasedness Mean?

- We can say that the least squares estimation procedure (or the least squares estimator) is unbiased
- The property of unbiasedness is about the average values of $\widehat{\beta}$ if many samples, of the same size, were to be drawn from the same population

- If we took the averages of estimates from many samples, these averages would approach the true parameter values

- Unbiasedness does not say that an estimate from any one sample is close to the true parameter value
- This is an important slide to understand.

Properties of OLS - Variance

- $var(b|X) = E\left((b-\beta)(b-\beta)'|X\right)$

- $b = (X'X)^{-1}X'y$
- $= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon$ (i.e., plug in for y)
- $= I\beta + (X'X)^{-1}X'\epsilon$
- $= \beta + (X'X)^{-1}X'\epsilon$

- Call this $\beta + A\epsilon$

Properties of OLS - Variance

- $var(b|X) = E\big((b-\beta)(b-\beta)'|X\big)$

- $b = (X'X)^{-1}X'y = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon$ (plug in for y)
- Call this $\beta + A\epsilon$
- $\Rightarrow b - \beta = \beta + A\epsilon - \beta = A\epsilon$
- $E\big((b-\beta)(b-\beta)'\big) = \mathrm{E}\big((A\epsilon)(A\epsilon)'|X\big)$
- $= E(A\epsilon\epsilon'A'|X)$
- $= AE(\epsilon\epsilon')\mathrm{A}' = \mathrm{A}\sigma^2 I\, A' = \sigma^2 AA'$

Properties of OLS - Variance

- $= AE(\epsilon\epsilon')A' = \sigma^2 AA'$
- Plug back in for $A = (X'X)^{-1}X'$
- $= \sigma^2((X'X)^{-1}X')\left((X'X)^{-1}X'\right)' = (X'X)^{-1}X'X(X'X)^{-1}$
- $= \sigma^2(X'X)^{-1}$

- Here’s where we use sphericality… we also used unbiasedness in our definition of variance…

Properties of OLS - BLUE

Under these assumptions, the OLS estimator is the minimum variance linear unbiased estimator… often referred to as the best linear unbiased estimator, or BLUE.

-

Properties of OLS - BLUE

Under these assumptions, the OLS estimator is the minimum variance linear unbiased estimator… often referred to as the best linear unbiased estimator, or BLUE.

That is, any other linear, unbiased estimator has variance at least as large as OLS. Proof relies on sphericality (and unbiasedness), is in the text, and is worth reviewing.

This means we may have superior *non-linear* estimators, or may have lower variance biased estimators… just setting $\hat{\beta} = 5$ has variance zero, for example.

There is occasionally a bias- variance tradeoff that we may be willing to make. This comes up a lot in machine learning…

-

OLS and Normality

Often, the Gauss Markov assumptions are presented with an additional assumption that the errors are normally distributed.

As we saw, this isn't required for BLUE or for unbiasedness, but it does (when combined with an estimate of σ^2) allow us to construct confidence intervals.

If $\epsilon \sim N(0, \sigma^2)$, then $\hat{\beta} - \beta = A\epsilon \sim N(0, \sigma^2(X'X)^{-1})$

The square root of the diagonal elements of $\sigma^2(X'X)^{-1}$ give you the standard errors of each estimate in $\hat{\beta}$

-

How to Estimate σ^2

$$
\begin{aligned}
\mathrm{Var}(e_i) &= \sigma^2 = E\left[(e_i - E[e_i])^2\right] \\
&= E\left[e_i^2\right] - \underbrace{E[e_i]^2}_{=0} \\
&= E\left[e_i^2\right]
\end{aligned}
$$

Will use a two variable example to make this concrete…

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

Naïve approach

Since the "expectation" is an average value we might consider estimating σ^2 as the average of the squared errors, namely:

$$\sigma^2 = \frac{1}{N}\sum e_i^2$$

Well, we already have an estimate of e_i to leverage:

$$\widehat{\sigma^2} = \frac{1}{N}\sum\left(y_i - \widehat{\beta_1} - \widehat{\beta_2}x_i\right)^2$$

Naïve approach

It turns out that this is biased!

You have to adjust slightly:

$$\widehat{\sigma^2} = s^2 = \frac{1}{N-2}\sum\left(y_i - \widehat{\beta_1} - \widehat{\beta_2}x_i\right)^2$$

N-2 is due to the fact that we've already estimated two parameters.

Intuition: If you tell me any of the N-2 data points, and the two estimates of $\widehat{\beta_1}$ & $\widehat{\beta_2}$, I can recover the two missing pieces of data. In the general model with K regressors, this adjustment is N-K

What does Unbiasedness Mean?

- Going to show this with MATLAB code. lecture2_variance.m

- Four steps:

1. Set up a matrices to store data
2. Simulate random data according to a KNOWN model
3. Estimate Coefficients, store them
4. Repeat and summarize results

MATLAB: lecture2_variance.m

```
N=100;
K = 1000;

x = zeros(N,K);
epsilon = zeros(N,K);
y = zeros(N,K);
beta_hat = zeros(2,K);
e_hat = zeros(N,K);
s_2 = zeros(1,K);
theoretical_var = zeros(1,K);

for i = 1:K
        x(:,i) = mvnrnd(0,1,N);
        intercept = ones(N,1);
        eps(:,i) = mvnrnd(0,1,N);
        y(:,i) = 1 + 2*x(:,i) + eps(:,i);

        X = [intercept x(:,i)];

        beta_hat(:,i) = inv(X'*X)*X'*y(:,i);
        e_hat(:,i) = y(:,i)-X*beta_hat(:,i);
        s_2(i) = (e_hat(:,i)'*e_hat(:,i))/(N-2);
        var_temp= s_2(i)*inv(X'*X);
        theoretical_var(i) = var_temp(2,2);

end

var(beta_hat(2,:))          % Simulated
mean(theoretical_var)       % Average analytical Estimate

var_temp                    % Single analytical estimate
```

- Draw 100 observations from a data generating process

$$y = 1 + 2x + \epsilon$$

- Estimate OLS coefficients, their variance, store them
- Repeat 1000 times

Show:

the simulated variance of estimates over the 1000 runs

The avg estimate of the theoretical/analytical form variance

A single estimate

What does Unbiasedness Mean?


```
>> lecture2_variance

ans =

    0.0106


ans =

    0.0104


var_temp =

    0.0094    0.0000
    0.0000    0.0070
```

With repeated samples, the distribution of estimated coefficients (if this is the right data generating process) will be normal with variance given by $s^2 * (X'X)^{-1}$

Confidence Intervals

Often, the Gauss Markov assumptions are presented with an additional assumption that the errors are normally distributed.

This isn't required for BLUE or for unbiasedness, but it does (when combined with an estimate of σ^2) allow us to construct confidence intervals in finite samples.

If $\epsilon \sim N(0, \sigma^2)$ and is i.i.d. (independent and identically distributed), then $\hat{\beta} - \beta = A\epsilon \sim N(0, \sigma^2(X'X)^{-1})$

We know what this distribution looks like!

What's the problem here?

▪

Distribution of $\hat{\beta}$

In practice, it's very unlikely you will know σ^2

Actually, would be kind of strange to know the variance of the residuals, but not the coefficient that produces those residuals… not impossible, but certainly odd.

Much more likely that we can estimate σ^2 with s^2

This creates a t-statistic

Distribution of $\hat{\beta}$

$(\widehat{\beta_2} - \beta_2)/\sqrt{(\hat{v}(\widehat{\beta_2}))} \sim t(N\text{-}K)$

Where $\hat{v}(\widehat{\beta_2})$ is the estimated variance of $\widehat{\beta_2}$

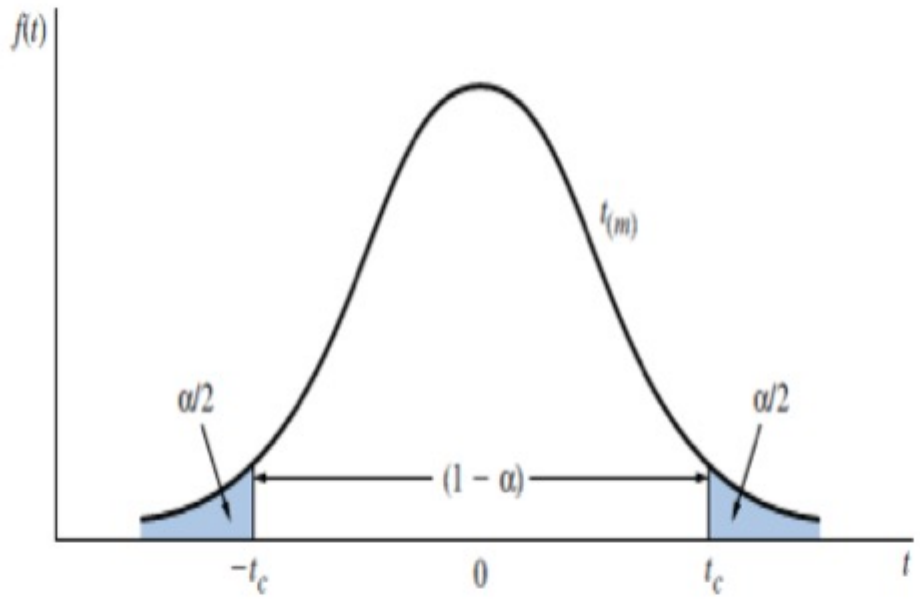
Assuming the model has K right hand side variables (hence N-K degrees of freedom on the T statistic)

Hypothesis Testing

Idea for all hypothesis testing is to transform the estimator or the data into something that has a known distribution under the null hypothesis.

Then we can compare the value of what we see from the estimator/from the data to the distribution of values that it should have if the null hypothesis was true to get a quantitative sense of how likely that is.

T Distribution

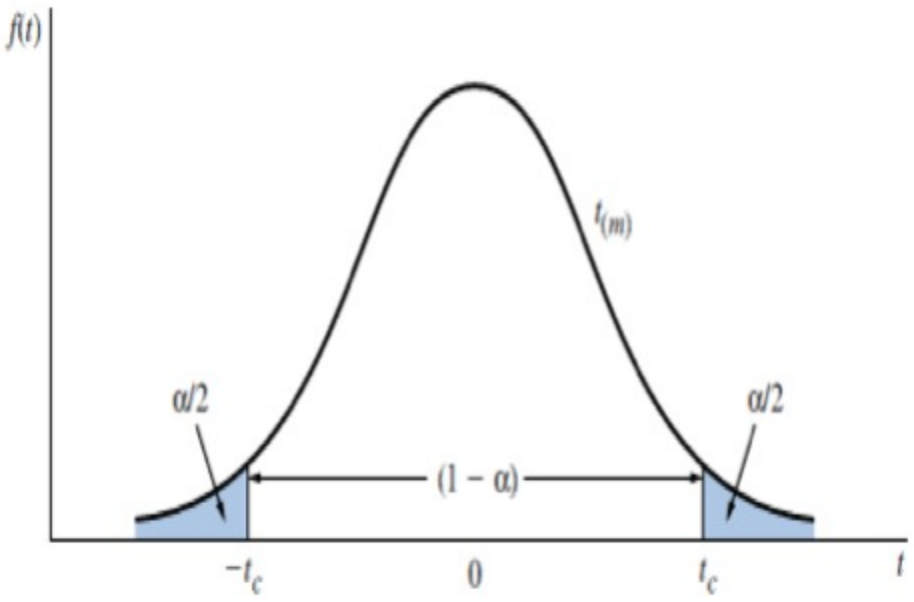


The t-distribution is a bell shaped curve centered at zero

It looks like the standard normal distribution, except it is more spread out, with a larger variance and thicker tails

The shape of the t-distribution is controlled by a single parameter called the degrees of freedom, often abbreviated as “df”

T Distribution



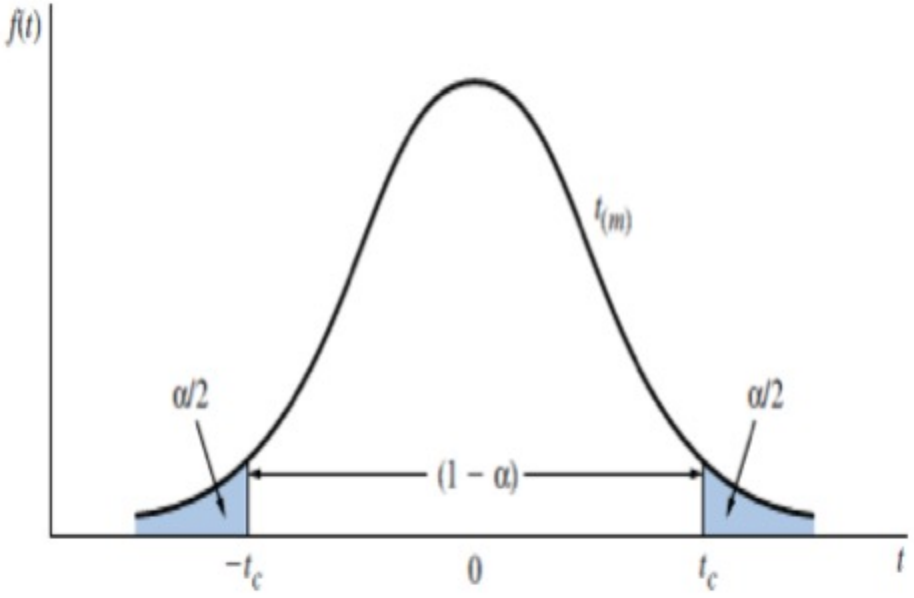
We can find a "critical value" from a t-distribution such that

$$\Pr(t \geq t_c) = \Pr(t \leq -t_c) = \frac{\alpha}{2}$$

where α is a probability often taken to be $\alpha = 0.01$, or $\alpha = 0.05$. This happens to give t stats very near 3 and 2 respectively…

In principle, the right probability should depend on the research and question. This is not always carefully thought through…

T Distribution



The critical value t_c for degrees of freedom m is the percentile value

$$t(1-\alpha/2, m)$$

Each shaded "tail" area contains $\alpha/2$ of the probability, so that $1-\alpha$ of the probability is contained in the center portion

MATLAB: Lecture2_OLS.m

```
N=100;
x = mvnrnd(0,1,N);
intercept = ones(N,1);
eps = mvnrnd(0,1,N);
y = 1 + 2*x + eps;
X = [intercept x];
beta_hat = inv(X'*X)*X'*y;
e_hat = y-X*beta_hat;
s_2 = (e_hat'*e_hat)/(N-2);
var_beta_hat = s_2*inv(X'*X);
t_stat = beta_hat./diag(sqrt(var_beta_hat));
```

What does this code do?

T Distribution

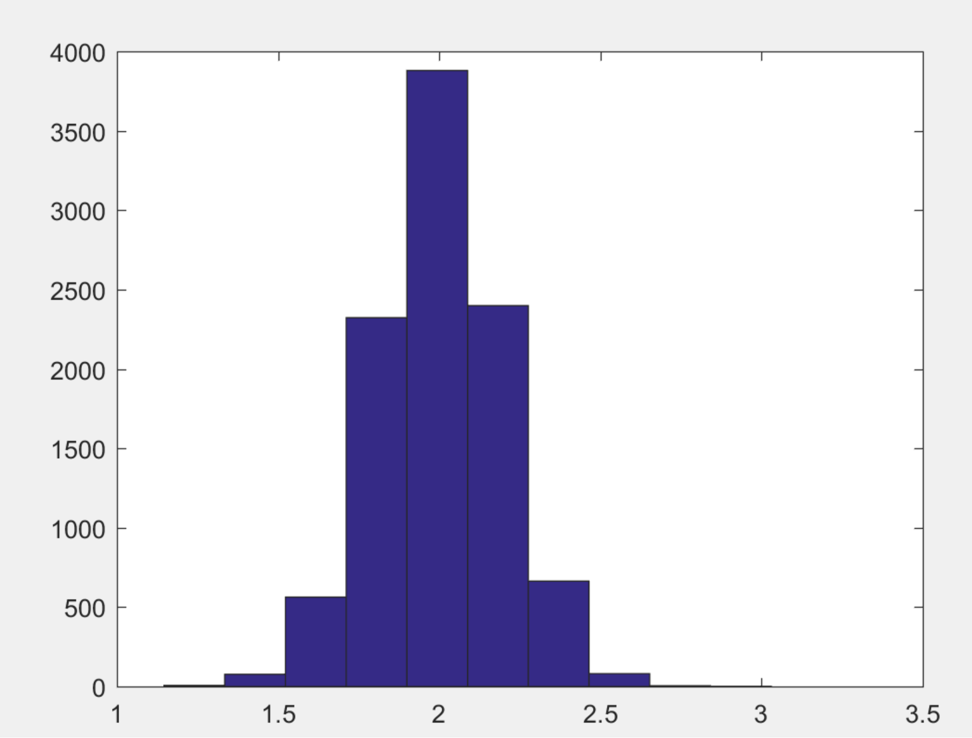
Repeat that code K times, store coefficients in 2 x K matrix “beta_hat”

(See lecture2_variance.m for full code)

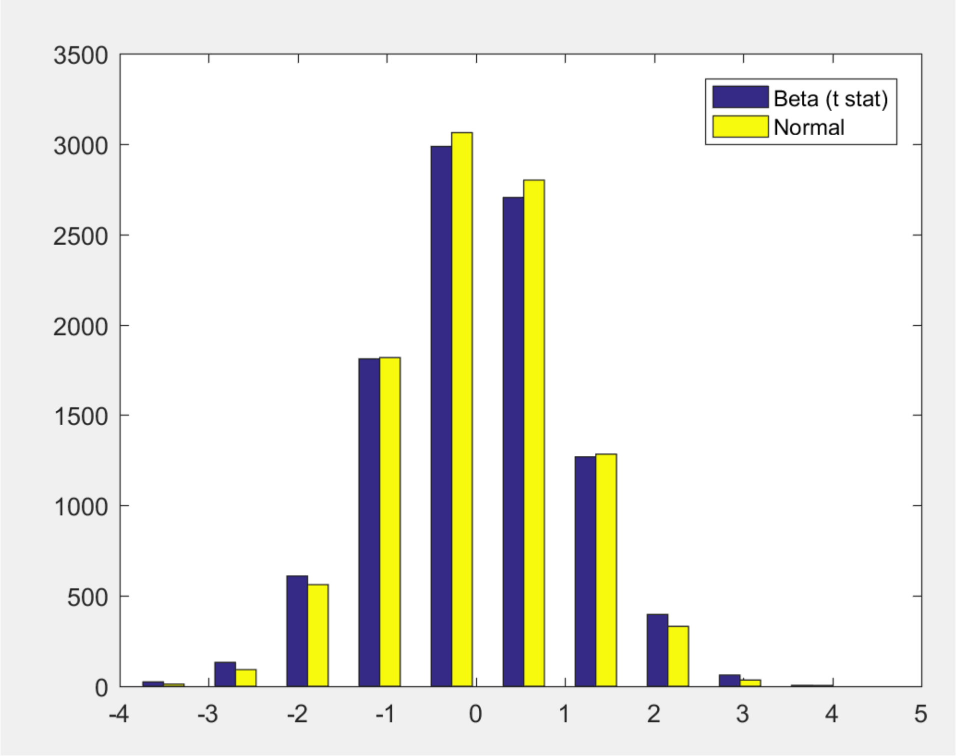
```
hist(beta_hat(2,:))
Z = mvnrnd(0,1,K);
hist([(beta_hat(2,:)-2)./sqrt(simulated_var);Z']')
legend('Beta (t stat)','Normal')
```

Transform our simulated betas to a t stat by subtracting off the mean (here I use the true value), and divide by the square root of the estimated variance each time (s^2)

T Distribution



Simulated betas: N = 30, K = 10,000



Simulated betas transformed into t stat vs. a Normal

Confidence Intervals

Given the critical values $t_c, -t_c$, we can make the following equivalent statements for each estimated beta (I use $\widehat{\beta_2}$ below):

$$Pr(-t_c \leq (\widehat{\beta_2} - \beta_2)/\sqrt{(\hat{v}(\widehat{\beta_2}))} \sim \leq t_c) = 1 - \alpha$$

$$Pr(\widehat{\beta_2} - t_c\sqrt{(\hat{v}(\widehat{\beta_2})} \leq \beta_2 \leq \widehat{\beta_2} + t_c\sqrt{(\hat{v}(\widehat{\beta_2}))} = 1 - \alpha$$

There is a $1 - \alpha$ probability that $\widehat{\beta_2}$ is within $t_c\sqrt{(\hat{v}(\widehat{\beta_2})}$ standard deviations of the true β_2

- Q: In matrix notation, what's the variance/distribution of a linear combination of betas $= c\hat{\beta}$? ... Hint: $\mathrm{Var}(c\hat{\beta}) = c\ \mathrm{var}(\hat{\beta})\ c'$ ...

Interpretation

If we repeated the estimation procedure on data with a data generating process given by

$$y = X\beta + \epsilon$$

We would see $\widehat{\beta_2}$ is within $t_c\sqrt{(\hat{v}(\widehat{\beta_2})}$ standard deviations of β_2 α percent of the time.

Often, you will see $\alpha = 5\%$ as a sort of default. There is NOTHING special about this value.

When "confidence intervals" are discussed, remember that our <u>confidence is in the procedure used to construct the interval estimate; it is not in any one interval estimate calculated from a sample of data</u>

Interpretation

Is $\beta_2$actually in the interval?

- We do not know, and we will never know
- What we do know is that when the procedure we used is applied to many random samples of data from the same population, then$1-\alpha\%$ of all the interval estimates constructed using this procedure will contain the true parameter

Interpretation

Is β_2 actually in the interval?

- We do not know, and we will never know
- What we do know is that when the procedure we used is applied to many random samples of data from the same population, then $1-\alpha\%$ of all the interval estimates constructed using this procedure will contain the true parameter
- The interval estimation procedure "works" $1-\alpha\%$ of the time
- What we can say about the interval estimate based on our one sample is that, given the reliability of the procedure, we would be "surprised" if β_2 is not in the interval if the model reflects the true data generating process
- Surprised in the sense that should only happen rarely ($\alpha\%$ of the time!)
- What's the right level of confidence then?

Sample Problems

You have the results from the single variable regression $y = \beta_0 + \beta_1 x + u$ with 51 observations.

a. The estimated error variance is $\widehat{\sigma^2} = 2.04$. What is the residual sum of squares?
b. The variance of $\widehat{\beta_1}$ is 0.00098. What is the standard error? What is $\sum(x_i - \bar{x})$?
c. If y is the mean income in a state and x is the share of 18 year olds who are high school graduates and $\widehat{\beta_1} = 0.18$, what does this mean?
d. If $\bar{x} = 69$ and $\bar{y} = 15.18$ with the coefficient as in part c, what is $\widehat{\beta_0}$?

Sample Problems

- a. $\widehat{\sigma^2} = 2.04$, => RSS/(N-df) = 2.04.
- DF = 2 =>2.04*49 = RSS = 99.96

- b. $v(\widehat{\beta_1}) = 0.0098$. $SE\widehat{(\beta_1)} = \sqrt{0.0098}$= 0.031

- c. if we increase the proportion of high school graduates by 1%, the mean income in the state increases by 0.18 (units) of dollars.

- d. the intercept guarantees the regression line runs through the mean values.
- $15.18 = \widehat{\beta_0} + 0.18 * 69$; $\widehat{\beta_0}$ = 15.18-12.42=2.76

Example Problems

- Sample Data drawn from a normally distributed variable X with unknown mean and variance
- X = [1.3,2.1,0.4,1.3,0.5,0.2,1.8,2.5,1.9,3.2]

1. What is $\bar{x}$?
2. What is s^2?
3. Test the hypothesis that the mean is greater than 2
4. Test the hypothesis that the variance equals 0.5.

Example Problems

- Sample Data drawn from a normally distributed variable X with unknown mean and variance
- X = [1.3,2.1,0.4,1.3,0.5,0.2,1.8,2.5,1.9,3.2]

What is $\bar{x}$?

1/10*(1.3+2.1+0.4+1.3+0.5+0.2+1.8+2.5+1.9+3.2)=1.52

What is s^2? MATLAB CODE:

```
RSS = (x-mean(x)).^2
s2 = sum(RSS)/(length(x)-1)
```

```
s2 =

    0.9418
```

Example Problems

- Sample Data drawn from a normally distributed variable X with unknown mean and variance
- Test that the mean is greater than 2 at a 5% level of confidence:

$$\frac{x - \bar{x}}{\ }\sim t(n-1)$$

```
>> (2-mean(x))/sqrt(s2)

ans =

    0.4946
```

```
>> tcdf((2-mean(x))/sqrt(s2),length(x)-1)

ans =

    0.6836
```

Example Problems

- Sample Data drawn from a normally distributed variable X with unknown mean and variance

Test the hypothesis that the variance equals 0.5.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_9^2$$

```
>> 9*s2/0.5

ans =

   16.9520
```

```
>> chi2cdf(ans,9)

ans =

    0.9505
```

Next time: Violations of GM Assumptions

We saw that we used a subset of the GM assumptions to prove different properties of the OLS estimator.

It's important to think this through when reviewing model's test performance. What if we fail normality tests? What about non-sphericality?

What if there is a relationship between the errors (or functions of the errors) and the X's?