CHRISTOPHER PRENER, PH.D.
FALL, 2016

WEEK 15
LECTURE 16

QUANTITATIVE ANALYSIS

MULTIPLE REGRESSION

AGENDA

1. Follow-up
2. Project Checklist
3. Multiple Regression
4. Assessing Model Fit
5. Additional Assumptions

1 FOLLOW-UP

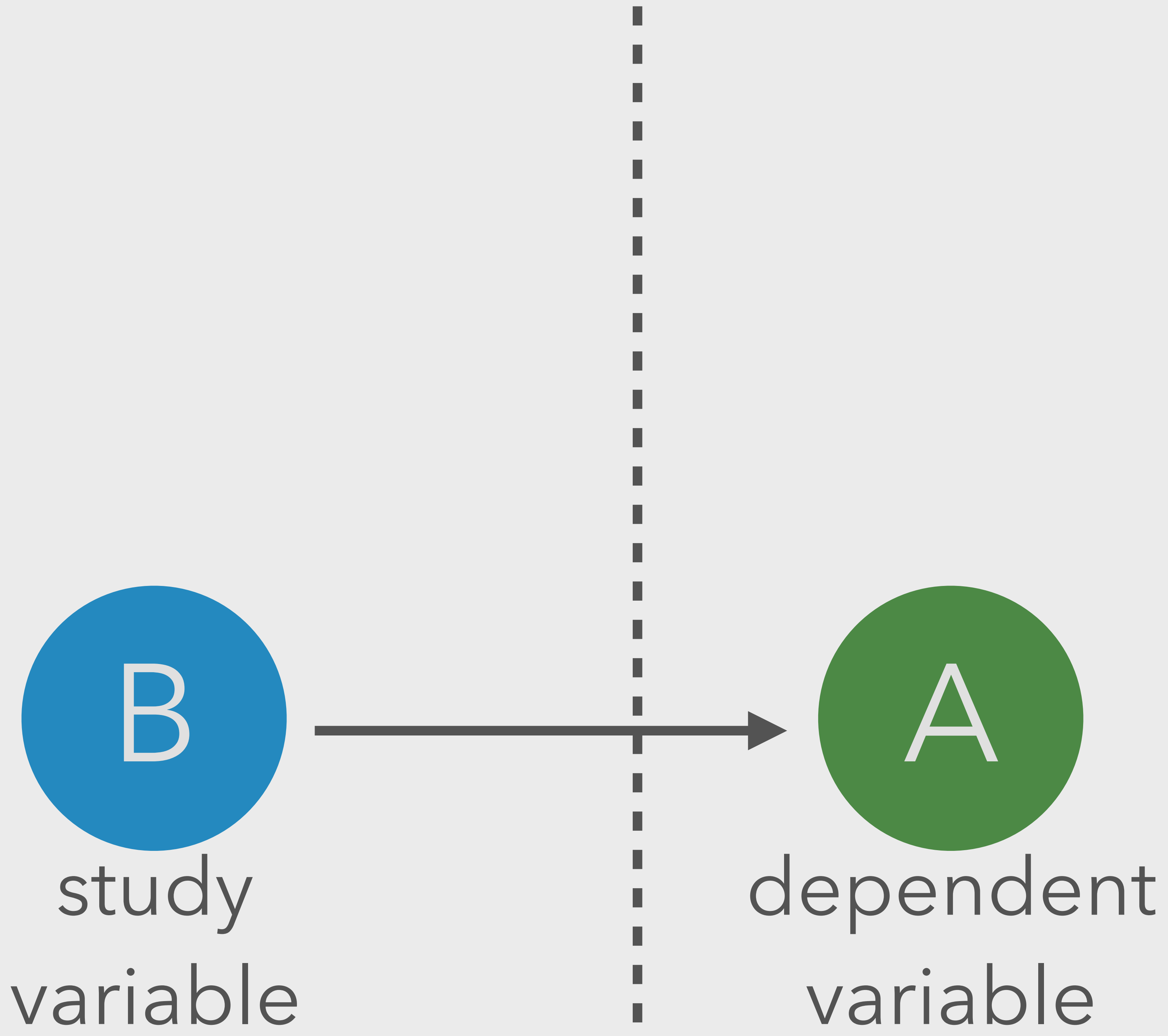
2 PROJECT CHECKLIST

LONG'S CHECKLIST

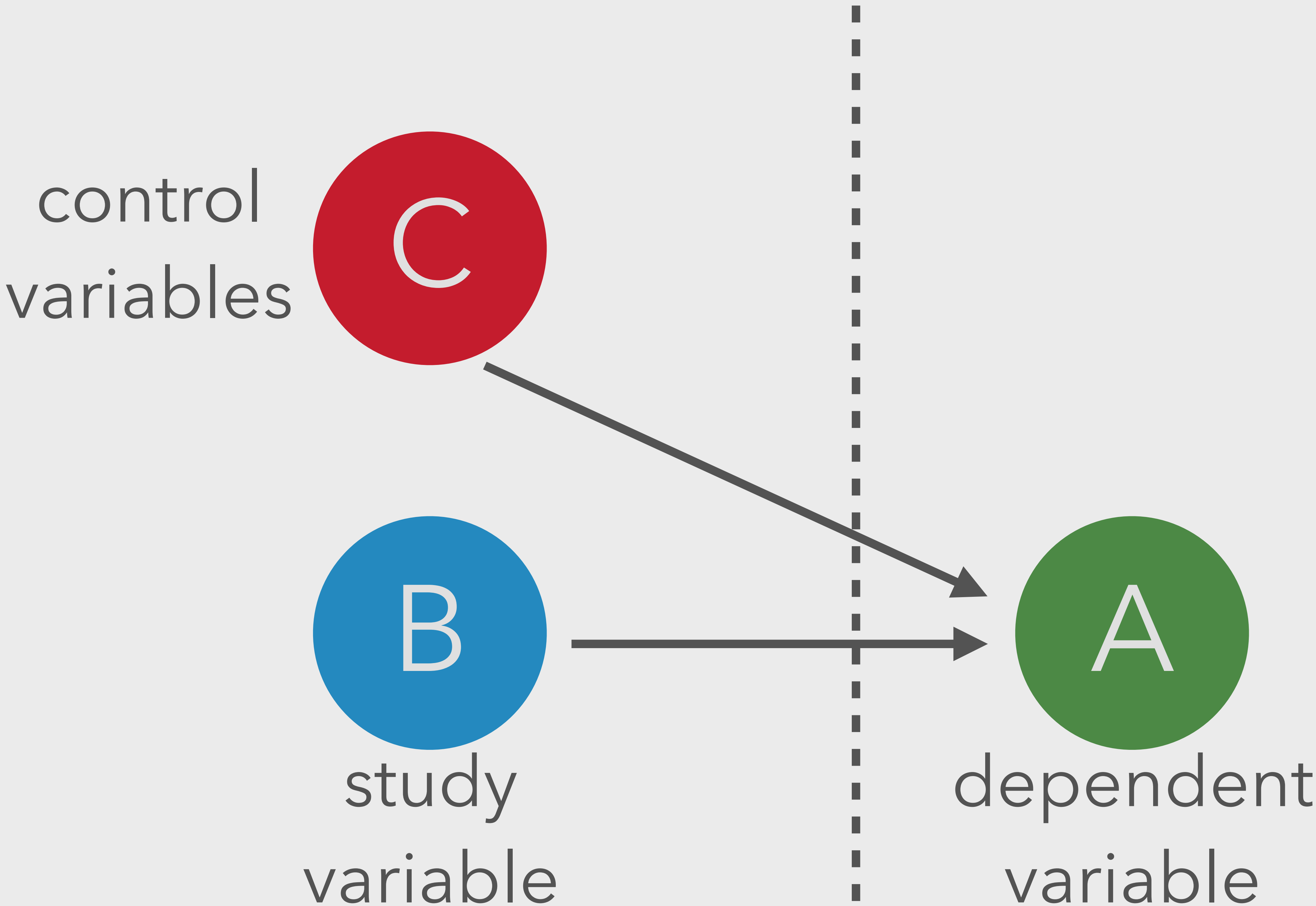
- VERIFY YOUR RESEARCH LOG
- VERIFY THAT YOUR FULL STACK EXECUTES CLEANLY AND CONSISTENTLY
- CLEAN UP STRAY FILES
- LINK RESULTS IN PAPER TO OUTPUT AND CODE
- MAKE YOUR CODE AND ANALYSES AVAILABLE
- ARCHIVE ALL FILES, DOCUMENTATION, AND DRAFTS OF YOUR PAPER

3 MULTIPLE REGRESSION

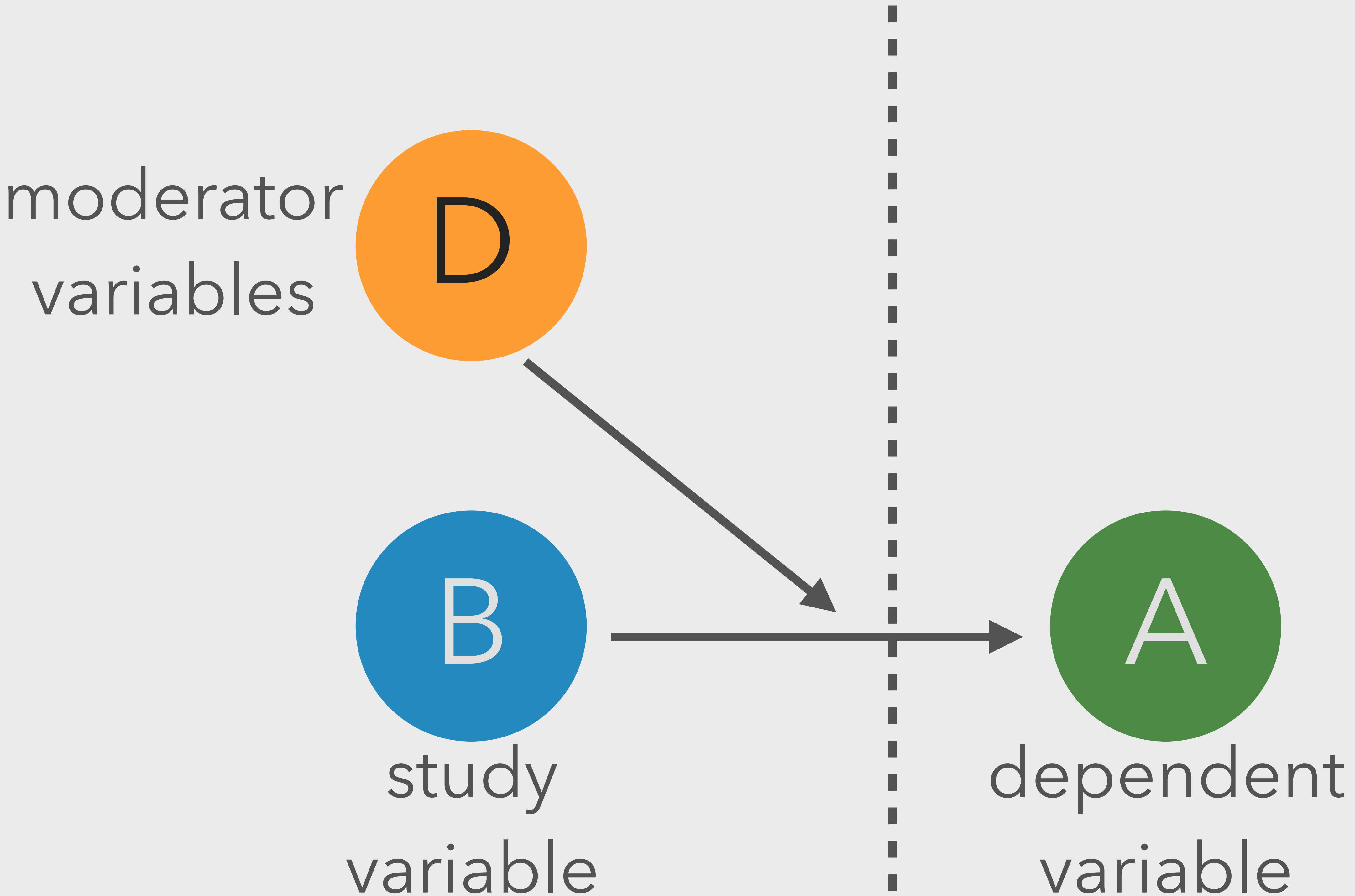
ACCOUNTING FOR VARIATION



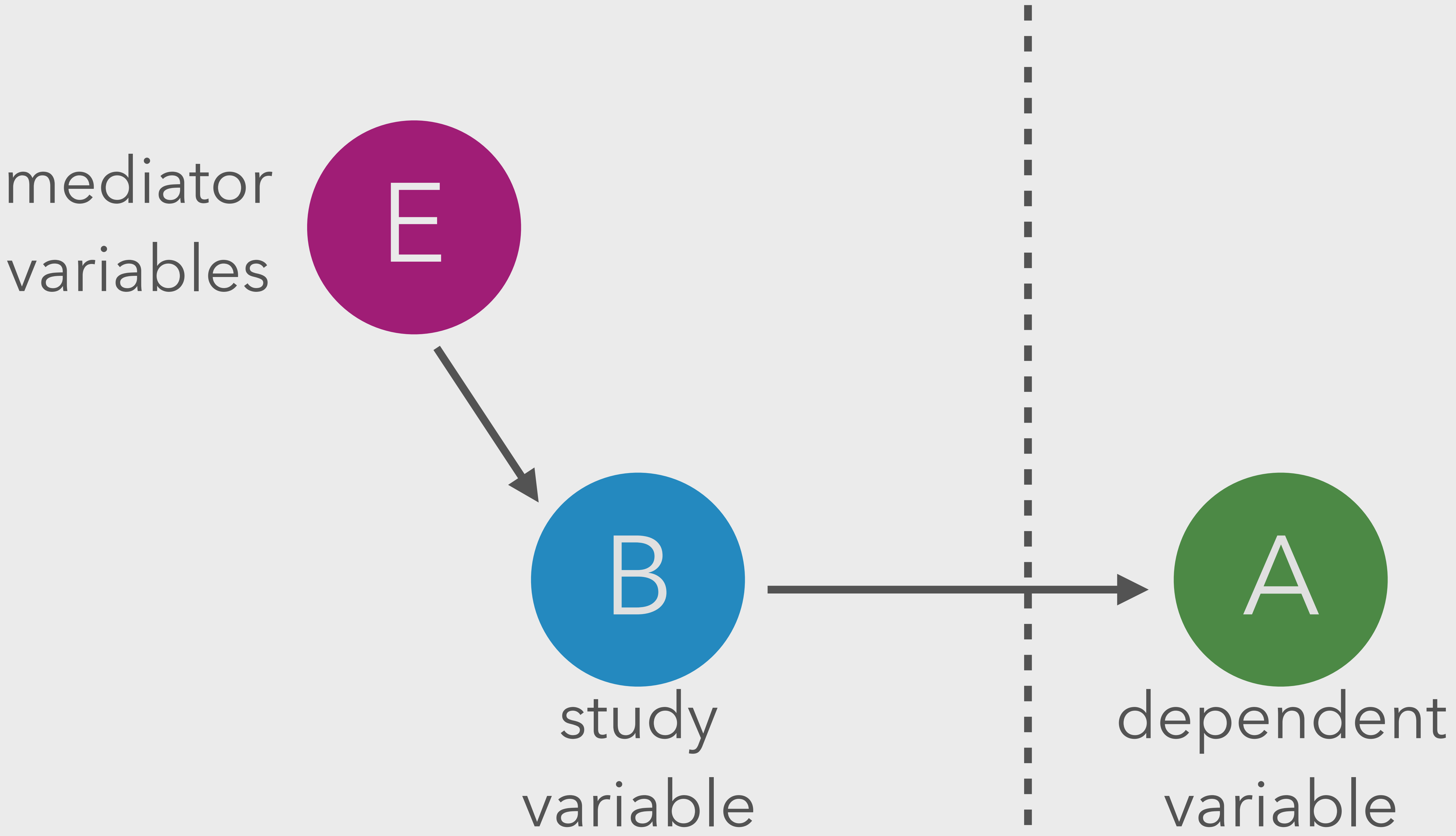
ACCOUNTING FOR VARIATION



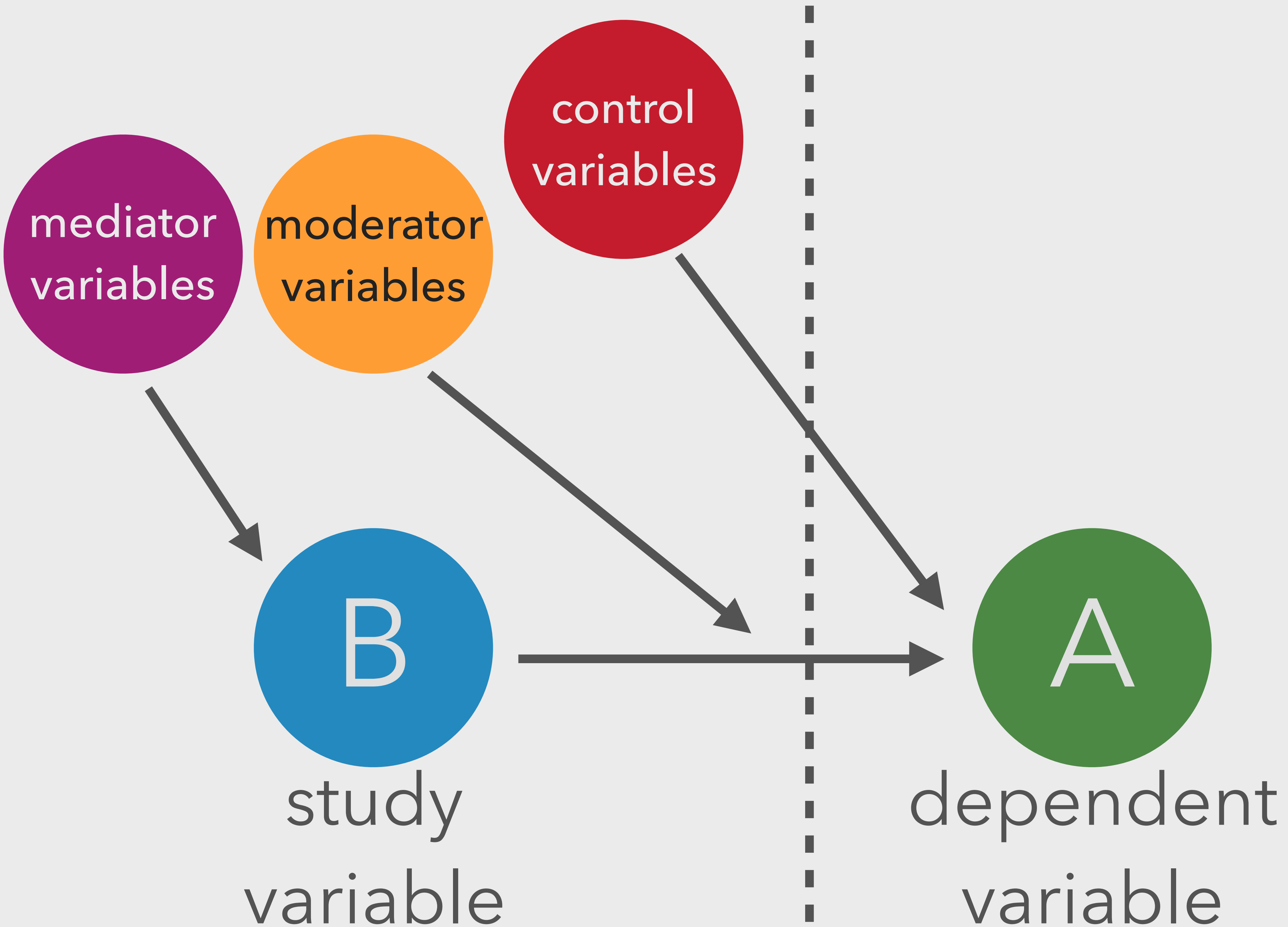
ACCOUNTING FOR VARIATION



ACCOUNTING FOR VARIATION



ACCOUNTING FOR VARIATION



MODEL BUILDING

What is the effect of vehicle weight on engine efficiency, controlling for differences in length, displacement, and gear ratio, and the effect of being produced outside of the U.S.?

MODEL BUILDING

$y = b_0 + b_i x_i + \varepsilon$

y = dependent variable
b_0 = constant
x_i = independent variable i
b_i = beta value of IV i

DV = mpg
IV = weight, length, displacement, gear_ratio, and foreign (0 = domestic, 1 = foreign)

$$y_{mpg} = b_0 + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

MODEL BUILDING

$$y_{mpg} = b_0 + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

MODEL BUILDING

$$\boxed{y_{mpg}} = b_0 + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

DV

MODEL BUILDING

$$y_{mpg} = \boxed{b_0} + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

constant

MODEL BUILDING

$$y_{mpg} = \boxed{b_0} + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

constant

A hypothetical car with a weight, length, displacement, and gear_ratio of 0 that was produced in the United States.

MODEL BUILDING

$$y_{mpg} = b_0 + \boxed{b_1 x_{weight}} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

study variable

MODEL BUILDING

$$y_{mpg} = b_0 + b_1 x_{weight} + \underbrace{b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio}}_{\text{control* variables - size}} + b_5 x_{foreign} + \epsilon$$

MODEL BUILDING

$$y_{mpg} = b_0 + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

control* variable - place of manufacture

MODEL BUILDING

$$y_{mpg} = b_0 + b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign} + \epsilon$$

error

MODEL BUILDING

$$y_{mpg} = b_0 + \boxed{b_1 x_{weight} + b_2 x_{length} + b_3 x_{displacement} + b_4 x_{gear\_ratio} + b_5 x_{foreign}} + \epsilon$$

10-15 obs per predictor

MODEL BUILDING

```
. regress mpg weight

      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(1, 72)        =    134.62
       Model |   1591.9902         1   1591.9902   Prob > F        =    0.0000
    Residual |  851.469256        72  11.8259619   R-squared       =    0.6515
-------------+----------------------------------   Adj R-squared   =    0.6467
       Total |  2443.45946        73  33.4720474   Root MSE        =    3.4389

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0060087   .0005179   -11.60   0.000    -.0070411   -.0049763
       _cons |   39.44028   1.614003    24.44   0.000     36.22283    42.65774
------------------------------------------------------------------------------
```

MODEL BUILDING

```
. regress mpg weight length displacement gear_ratio

      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(4, 69)        =     34.02
       Model |   1621.3296         4    405.3324   Prob > F        =    0.0000
    Residual |   822.12986        69  11.9149255   R-squared       =    0.6635
-------------+----------------------------------   Adj R-squared   =    0.6440
       Total |  2443.45946        73  33.4720474   Root MSE        =    3.4518

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0042442   .0019929    -2.13   0.037    -.0082199   -.0002684
      length |  -.0798266   .0563051    -1.42   0.161    -.1921521     .032499
displacement |   .0071945   .0115007     0.63   0.534    -.0157488    .0301377
  gear_ratio |   .8116129    1.59094     0.51   0.612    -2.362225    3.985451
       _cons |   45.24802   8.476583     5.34   0.000      28.3377    62.15834
------------------------------------------------------------------------------
```

MODEL BUILDING

```
. regress mpg weight length displacement gear_ratio foreign

      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(5, 68)        =     29.17
       Model |  1666.45626         5  333.291252   Prob > F        =    0.0000
    Residual |   777.0032         68  11.4265176   R-squared       =    0.6820
-------------+----------------------------------   Adj R-squared   =    0.6586
       Total |  2443.45946        73  33.4720474   Root MSE        =    3.3803

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0041683    .001952    -2.14   0.036    -.0080634   -.0002731
      length |  -.0907552   .0554125    -1.64   0.106    -.2013292    .0198188
displacement |   .0077379   .0112658     0.69   0.495    -.0147428    .0302185
  gear_ratio |   2.372516   1.744781     1.36   0.178    -1.109141    5.854172
     foreign |  -2.446102   1.230878    -1.99   0.051    -4.902281    .0100763
       _cons |   42.98684   8.378649     5.13   0.000     26.26751    59.70618
------------------------------------------------------------------------------
```

MODEL BUILDING

	model 1	model 2	model 3
weight	-0.006***	-0.004*	-0.004*
length		-0.80	-0.91
displacement		0.007	0.007
gear_ratio		0.812	2.373
foreign			-2.446
constant	39.440***	45.248***	42.987***
adjusted r^2	0.644	0.644	0.657
anova	34.02***	34.02***	29.17***

EFFECT SIZES - ETA SQUARED

```
. estat esize

Effect sizes for linear models

---------------------------------------------------------------------
             Source |   Eta-Squared     df     [95% Conf. Interval]
--------------------+------------------------------------------------
              Model |   .6820069         5     .5242855     .7457006
                    |
             weight |   .0628428         1            .     .1951824
             length |   .0379503         1            .     .1570285
       displacement |   .0068897         1            .      .090549
         gear_ratio |   .0264713         1            .     .1366094
            foreign |   .0548899         1            .     .1836488
---------------------------------------------------------------------
```

ETA-SQUARED CAN BE INTERPRETED AS: 0.01 (SMALL), 0.06 (MODERATE), AND 0.14 (LARGE)

4 ASSESSING MODEL FIT

ROOT MEAN SQUARED ERROR

```
. regress mpg weight length displacement gear_ratio foreign

      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(5, 68)        =     29.17
       Model |  1666.45626         5  333.291252   Prob > F        =    0.0000
    Residual |   777.0032         68  11.4265176   R-squared       =    0.6820
-------------+----------------------------------   Adj R-squared   =    0.6586
       Total |  2443.45946        73  33.4720474   Root MSE        =    3.3803

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0041683    .001952    -2.14   0.036    -.0080634   -.0002731
      length |  -.0907552   .0554125    -1.64   0.106    -.2013292    .0198188
displacement |   .0077379   .0112658     0.69   0.495    -.0147428    .0302185
  gear_ratio |   2.372516   1.744781     1.36   0.178    -1.109141    5.854172
     foreign |  -2.446102   1.230878    -1.99   0.051    -4.902281    .0100763
       _cons |   42.98684   8.378649     5.13   0.000     26.26751    59.70618
------------------------------------------------------------------------------
```

AIC AND BIC

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |           74 -234.3943  -192.0025       6     396.005   409.8294
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

MODEL BUILDING

	model 1	model 2	model 3
weight	-0.006***	-0.004*	-0.004*
length		-0.80	-0.91
displacement		0.007	0.007
gear_ratio		0.812	2.373
foreign			-2.446
constant	39.440***	45.248***	42.987***
adjusted r^2	0.644	0.644	0.657
anova	34.02***	34.02***	29.17***
rmse	3.439	3.452	3.380
aic	394.777	398.183	396.005
bic	399.386	409.703	409.829

5 ADDITIONAL ASSUMPTIONS

BASIC ASSUMPTIONS REVIEW

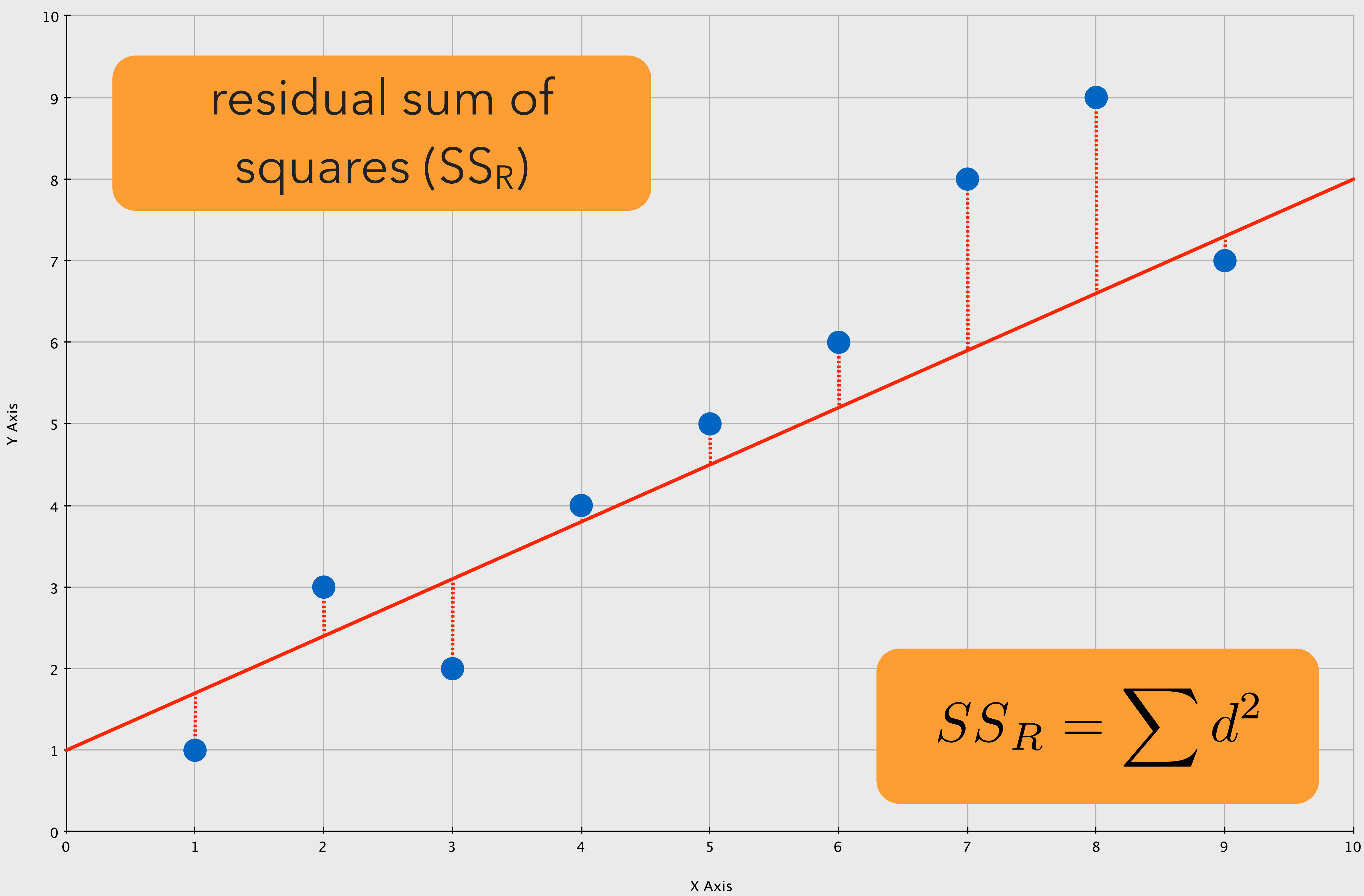
- DV must be continuous
- IV may be:
  - Dummy
  - Ordinal
  - Continuous
- IV's must have a variance > 0
- Relationship is linear
- DV should be normal
- No significant outliers

AUTOCORRELATION

- DV must be continuous
- IV may be:
  - Dummy
  - Ordinal
  - Continuous
- IV's must have a variance > 0
- Relationship is linear
- DV should be normal

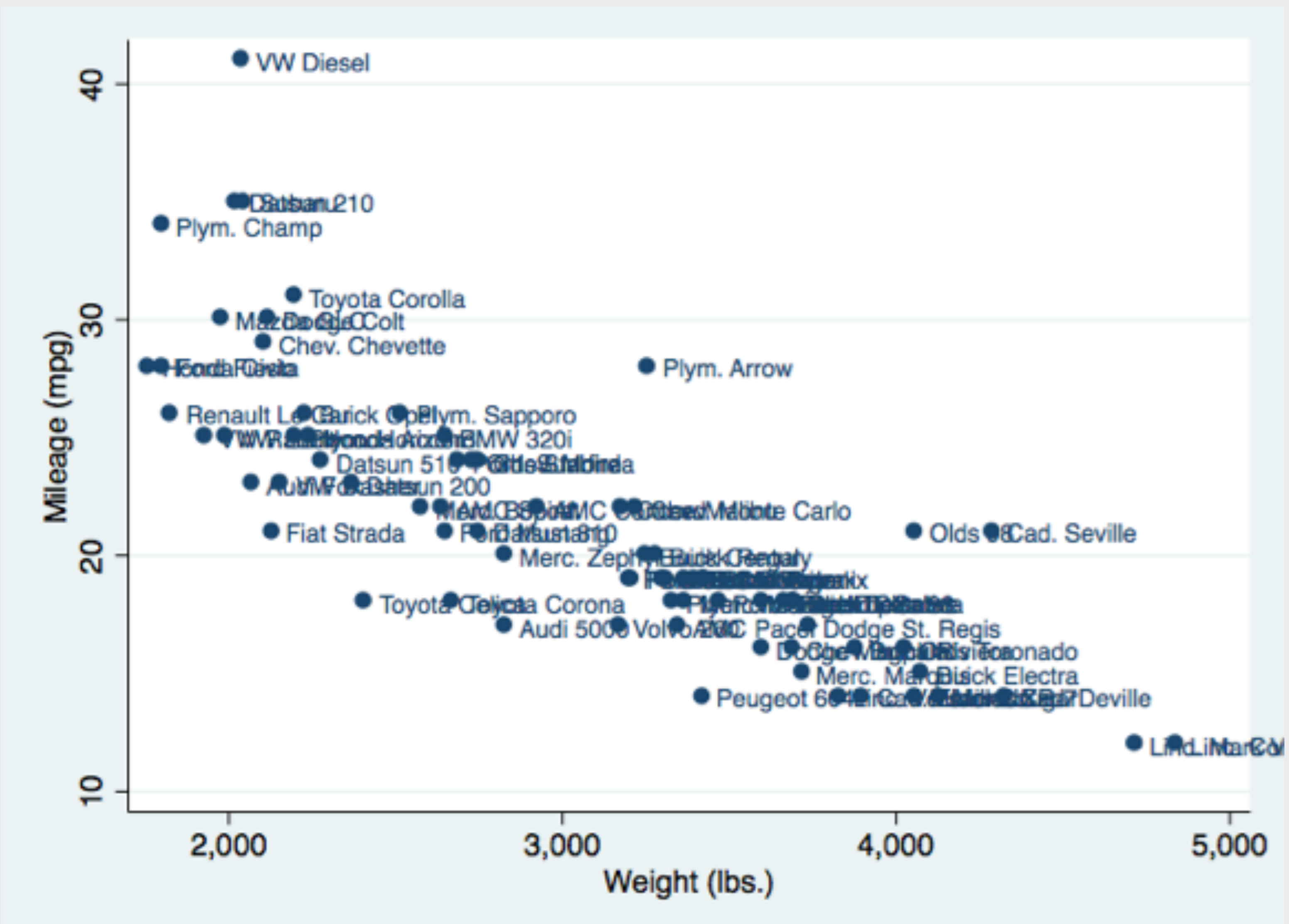
Regression models should be free of outliers and influential observations.

THE MECHANICS OF OLS REGRESSION



DETECTING OUTLIERS

```
. scatter mpg weight, mlabel(make)
```




```
. list make mpg if mpg > 32

     +--------------------+
     | make           mpg |
     |--------------------|
 67. | Plym. Champ     34 |
 72. | Datsun 210      35 |
 73. | Subaru          35 |
 74. | VW Diesel       41 |
     +--------------------+
```

DETECTING OUTLIERS

```
predict varName [, val]

. predict yhat

. predict rstandard, rstandard

. predict residual, residual
```

"STANDARDIZED" RESIDUALS ARE STANDARDIZED USING A Z-SCORE TO MAKE COMPARISONS EASY

DETECTING OUTLIERS

“STANDARDIZED” RESIDUALS ARE STANDARDIZED USING A Z-SCORE TO MAKE COMPARISONS EASY

```
predict varName [, val]

. predict yhat

. predict rstandard, rstandard

. predict residual, residual

. list make mpg yhat residual rstandard if abs(rstandard) > 1.96

     +------------------------------------------------------+
     | make            mpg       yhat   residual   rstand~d |
     |------------------------------------------------------|
 13. | Cad. Seville     21   14.61354   6.386456   2.045138 |
 35. | Olds 98          21   14.52349   6.476507   1.978079 |
 57. | Datsun 210       35   26.58223   8.417767   2.563203 |
 66. | Subaru           35   26.90177   8.098228   2.480626 |
 71. | VW Diesel        41   27.63491   13.36509   4.103542 |
     +------------------------------------------------------+
```

DETECTING OUTLIERS

“STANDARDIZED” RESIDUALS ARE STANDARDIZED USING A Z-SCORE TO MAKE COMPARISONS EASY

```
predict varName [, val]

. predict yhat

. predict rstandard, rstandard

. predict residual, residual

. list make mpg yhat residual rstandard if abs(rstandard) > 2.58

     +---------------------------------------------------+
     | make          mpg       yhat    residual   rstand~d |
     |---------------------------------------------------|
 71. | VW Diesel      41   27.63491    13.36509   4.103542 |
     +---------------------------------------------------+
```

DETECTING INFLUENTIAL OBSERVATIONS

```
predict varName [, val]

. predict leverage, leverage
```

LEVERAGE IS A MEASURE OF HOW INFLUENTIAL A PARTICULAR OBSERVATION IS

DETECTING INFLUENTIAL OBSERVATIONS

```
predict varName [, val]

. predict leverage, leverage

. list make leverage if leverage > (2*5+2)/74

     +---------------------------+
     | make             leverage |
     |---------------------------|
  2. | AMC Pacer        .1756483 |
  7. | Buick Opel       .3513954 |
 14. | Chev. Chevette   .1852391 |
 42. | Plym. Arrow      .2429746 |
     +---------------------------+
```

LEVERAGE IS A MEASURE OF HOW INFLUENTIAL A PARTICULAR OBSERVATION IS

$$leverage > \frac{2k+2}{n}$$

DETECTING INFLUENTIAL OBSERVATIONS

dfbeta

```
. dfbeta
  _dfbeta_1: dfbeta(weight)
  _dfbeta_2: dfbeta(length)
  _dfbeta_3: dfbeta(displacement)
  _dfbeta_4: dfbeta(gear_ratio)
  _dfbeta_5: dfbeta(foreign)
```

DFBETA IS A MEASURE OF DIFFERENCE BETWEEN BETA WHEN OBS IS INCLUDED AND EXCLUDED

DETECTING INFLUENTIAL OBSERVATIONS

DFBETA IS A MEASURE OF DIFFERENCE BETWEEN BETA WHEN OBS IS INCLUDED AND EXCLUDED

$$DFBeta > 2/\sqrt{n}$$

dfbeta

```
. dfbeta
  _dfbeta_1: dfbeta(weight)
  _dfbeta_2: dfbeta(length)
  _dfbeta_3: dfbeta(displacement)
  _dfbeta_4: dfbeta(gear_ratio)
  _dfbeta_5: dfbeta(foreign)

. list make _dfbeta_1 if abs(_dfbeta_1) > 2/sqrt(74)
     +---------------------------+
     | make            _dfbeta_1 |
     |---------------------------|
  2. | AMC Pacer      -.4294335 |
  7. | Buick Opel      .2481316 |
 13. | Cad. Seville    .5760965 |
 42. | Plym. Arrow     .9853229 |
 64. | Peugeot 604    -.4080613 |
     |---------------------------|
 71. | VW Diesel       .3212889 |
     +---------------------------+
```

IDENTIFYING PROBLEMATIC OBSERVATIONS

Outliers

```
     +-------------------------+
     | make           rstand~d |
     |-------------------------|
 13. | Cad. Seville   2.045138 |
 35. | Olds 98        1.978079 |
 57. | Datsun 210     2.563203 |
 66. | Subaru         2.480626 |
 71. | VW Diesel      4.103542 |
     +-------------------------+
```

Leverage

```
     +----------------------------+
     | make              leverage |
     |----------------------------|
  2. | AMC Pacer         .1756483 |
  7. | Buick Opel        .3513954 |
 14. | Chev. Chevette    .1852391 |
 42. | Plym. Arrow       .2429746 |
     +----------------------------+
```

Influence

```
     +---------------------------+
     | make            _dfbeta_1 |
     |---------------------------|
  2. | AMC Pacer       -.4294335 |
  7. | Buick Opel       .2481316 |
 13. | Cad. Seville     .5760965 |
 42. | Plym. Arrow      .9853229 |
 64. | Peugeot 604     -.4080613 |
     |---------------------------|
 71. | VW Diesel        .3212889 |
     +---------------------------+
```

IDENTIFYING PROBLEMATIC OBSERVATIONS

Outliers

	make	rstand~d
13.	Cad. Seville	2.045138
35.	Olds 98	1.978079
57.	Datsun 210	2.563203
66.	Subaru	2.480626
71.	VW Diesel	4.103542

Leverage

	make	leverage
2.	AMC Pacer	.1756483
7.	Buick Opel	.3513954
14.	Chev. Chevette	.1852391
42.	Plym. Arrow	.2429746

Influence

	make	_dfbeta_1
2.	AMC Pacer	-.4294335
7.	Buick Opel	.2481316
13.	Cad. Seville	.5760965
42.	Plym. Arrow	.9853229
64.	Peugeot 604	-.4080613
71.	VW Diesel	.3212889

IDENTIFYING PROBLEMATIC OBSERVATIONS

Outliers

	make	rstand~d
13.	Cad. Seville	2.045138
35.	Olds 98	1.978079
57.	Datsun 210	2.563203
66.	Subaru	2.480626
71.	VW Diesel	4.103542

Leverage

	make	leverage
2.	AMC Pacer	.1756483
7.	Buick Opel	.3513954
14.	Chev. Chevette	.1852391
42.	Plym. Arrow	.2429746

Influence

	make	_dfbeta_1
2.	AMC Pacer	-.4294335
7.	Buick Opel	.2481316
13.	Cad. Seville	.5760965
42.	Plym. Arrow	.9853229
64.	Peugeot 604	-.4080613
71.	VW Diesel	.3212889

```
. generate insample = 1

. replace insample = 0 if make == “AMC Pacer” | ///
    make == “Buick Opel” | make == “Cad. Seville” | ///
    make == “Plym. Arrow” | make == “VW Diesel”
```

REMOVING PROBLEMATIC OBSERVATIONS

```
. regress mpg weight length displacement gear_ratio foreign if insample == 1

      Source |       SS           df       MS      Number of obs   =        69
-------------+----------------------------------   F(5, 63)        =     41.80
       Model |  1505.58963         5  301.117926   Prob > F        =    0.0000
    Residual |  453.888632        63  7.20458147   R-squared       =    0.7684
-------------+----------------------------------   Adj R-squared   =    0.7500
       Total |  1959.47826        68  28.8158568   Root MSE        =    2.6841

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0093139   .0022109    -4.21   0.000     -.013732   -.0048958
      length |   .0462243   .0599751     0.77   0.444    -.0736264    .1660749
displacement |   .0179895   .0111534     1.61   0.112    -.0042988    .0402779
  gear_ratio |   1.771099   1.421635     1.25   0.217    -1.069811     4.61201
     foreign |  -2.544551   .9992889    -2.55   0.013    -4.541471   -.5476319
       _cons |   32.20851    7.94102     4.06   0.000     16.33965    48.07737
------------------------------------------------------------------------------
```

REMOVING PROBLEMATIC OBSERVATIONS

	model 3	model 4
weight	-0.004*	-0.009***
length	-0.91	0.046
displacement	0.007	0.017
gear_ratio	2.373	1.771
foreign	-2.446	-2.545*
constant	42.987***	32.209***
n	74	69
adjusted r^2	0.657	0.750
anova	29.17***	41.80***
rmse	3.380	2.684
aic	396.005	337.792
bic	409.829	351.197

AUTOCORRELATION

- DV must be continuous
- IV may be:
  - Dummy
  - Ordinal
  - Continuous
- IV's must have a variance > 0
- Relationship is linear
- DV should be normal
- No significant outliers

Autocorrelation refers to the normality of the residuals estimated during the regression analysis.

CHECKING FOR NORMALITY

```
predict varName [, val]

. predict residualIS if insample == 1, residual

. swilk residualIS

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
    residual |         69    0.91591      5.116     3.547    0.00019
```

HOMOSKEDASTICITY

- DV must be continuous
- IV may be:
    - Dummy
    - Ordinal
    - Continuous
- IV's must have a variance > 0
- Relationship is linear
- DV should be normal
- No significant outliers
- Residuals are normally distributed

Homoskedasticity refers to the residuals having a constant variance.

CHECKING FOR HOMOSKEDASTICITY

```
. estat hettest


Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of mpg

         chi2(1)      =    11.48
         Prob > chi2  =   0.0007
```

A STATISTICALLY SIGNIFICANT TEST MEANS THAT THE ERRORS ARE HETEROSKEDASTIC (NOT CONSTANT)

CHECKING FOR HOMOSKEDASTICITY

```
. estat hettest


Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of mpg

         chi2(1)      =    11.48
         Prob > chi2  =   0.0007
```

HOWEVER, THIS TEST ASSUMES THAT THE ERRORS ARE NORMALLY DISTRIBUTED

CHECKING FOR HOMOSKEDASTICITY

```
. estat imtest, white


White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(19)     =     28.70
         Prob > chi2  =    0.0709

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |      28.70     19    0.0709
            Skewness |      14.79      5    0.0113
            Kurtosis |       3.86      1    0.0493
---------------------+-----------------------------
               Total |      47.35     25    0.0045
---------------------------------------------------
```

THE WHITE'S TEST RELAXES THE ASSUMPTION OF NORMALLY DISTRIBUTED RESIDUALS BUT HAS LESS POWER, PARTICULARLY AS THE NUMBER OF PREDICTORS INCREASES

FIXING POSSIBLE HETEROSKEDASTIC ERRORS

- Respecify the model - have predictors been left out?
- Transform variables - are some relationships non-linear? Transformation (outside scope of course) can help fix this.
- Use robust standard errors.

REMOVING PROBLEMATIC OBSERVATIONS

```
. regress mpg weight length displacement gear_ratio foreign if insample == 1, robust

Linear regression                               Number of obs     =         69
                                                F(5, 63)          =      44.23
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7684
                                                Root MSE          =     2.6841

------------------------------------------------------------------------------
             |               Robust
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0093139   .0020066    -4.64   0.000    -.0133238    -.005304
      length |   .0462243   .0516243     0.90   0.374    -.0569388    .1493873
displacement |   .0179895   .0090925     1.98   0.052    -.0001804    .0361594
  gear_ratio |   1.771099   1.453085     1.22   0.227     -1.13266    4.674859
     foreign |  -2.544551   1.041033    -2.44   0.017    -4.624889   -.4642131
       _cons |   32.20851   7.474281     4.31   0.000     17.27235    47.14467
------------------------------------------------------------------------------
```

ADJUSTING FOR HETEROSKEDASTICITY

	model 4	model 5
weight	-0.009***	-0.009***
length	0.046	0.046
displacement	0.017	0.017^
gear_ratio	1.771	1.771
foreign	-2.545*	-2.545*
constant	32.209***	32.209***
n	69	69
adjusted r^2	0.750	0.750
anova	41.80***	41.80***
rmse	2.684	2.684
aic	337.792	337.792
bic	351.197	351.197

HOMOSKEDASTICITY

- DV must be continuous
- IV may be:
  - Dummy
  - Ordinal
  - Continuous
- IV's must have a variance > 0
- Relationship is linear
- DV should be normal
- No significant outliers
- Residuals are normally distributed
- Residuals are homoskedastic

Predictors must not have multicollinearity (must not be highly correlated with each other)

DETECTING OUTLIERS

```
. estat vif

    Variable |       VIF       1/VIF  
-------------+----------------------
      weight |     27.63    0.036188
      length |     16.93    0.059053
displacement |      9.73    0.102749
  gear_ratio |      3.86    0.259257
     foreign |      2.02    0.493873
-------------+----------------------
    Mean VIF |     12.04
```

VIF =
VARIANCE INFLATION FACTOR

DETECTING OUTLIERS

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
      weight |     27.63    0.036188
      length |     16.93    0.059053
displacement |      9.73    0.102749
  gear_ratio |      3.86    0.259257
     foreign |      2.02    0.493873
-------------+----------------------
    Mean VIF |     12.04
```

VIF VALUES SHOULD BE < 10 WITH A MEAN VIF < 1

DETECTING OUTLIERS

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
      weight |     27.63    0.036188
      length |     16.93    0.059053
displacement |      9.73    0.102749
  gear_ratio |      3.86    0.259257
     foreign |      2.02    0.493873
-------------+----------------------
    Mean VIF |     12.04
```

IF VIF VALUES ARE PROBLEMATIC, RE-SPECIFY THE MODEL

DOCUMENT DETAILS

Document produced by Christopher Prener, Ph.D for the Saint Louis University course SOC 5050: QUANTITATIVE ANALYSIS - APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository README.md file for additional details.