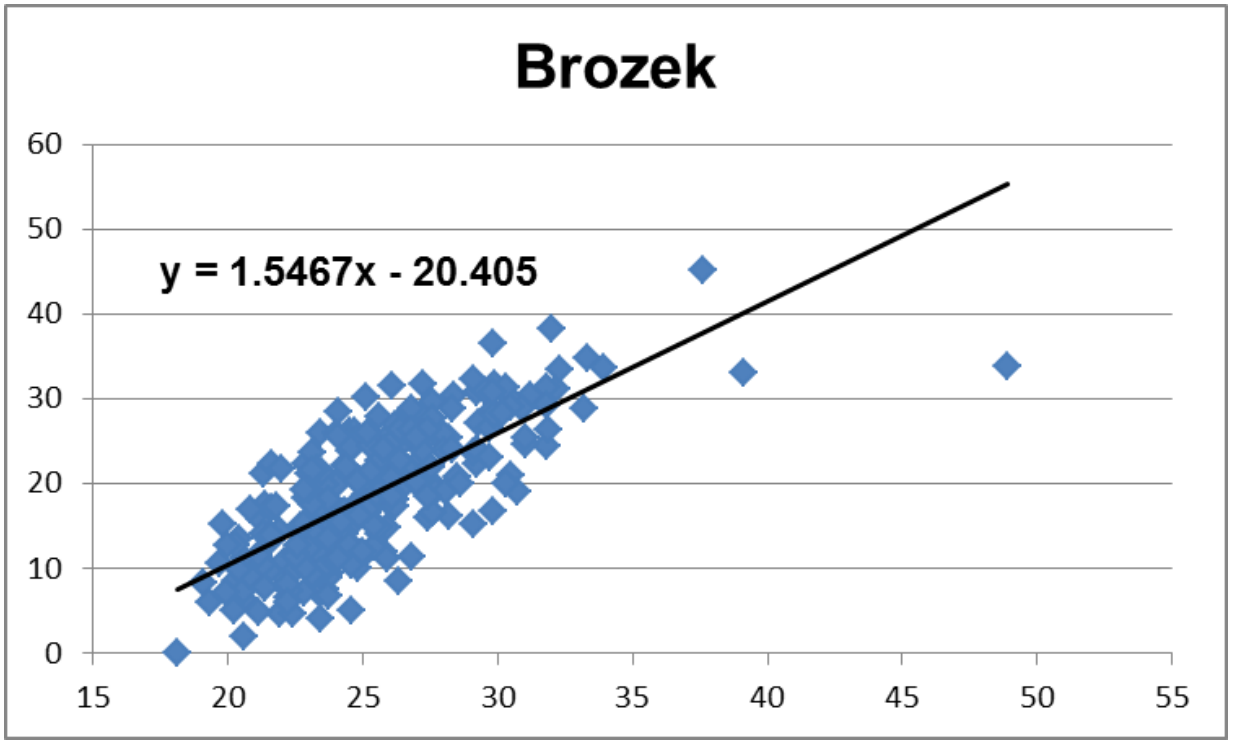
OLS: *Estimation Review & Inference Preview*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =  281.89
       Model |   7,991.510     1  7,991.510             Prob > F      =       0
    Residual |   7,087.507   250     28.350             R-squared     =    0.53
-------------+------------------------------           Adj R-squared =  0.5281
       Total |  15,079.017   251     60.076             Root MSE      =  5.3245

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547    0.092     16.79   0          1.365      1.728
       _cons |     (20.405)   2.367     -8.62   0        (25.067)    (15.743)
------------------------------------------------------------------------------
```

OLS: *Estimation*



OLS: *The Blank Canvas*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F(  1,   250) =
       Model |                                         Prob > F      =
    Residual |                                         R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |                                         Root MSE      =

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
-------------+----------------------------------------------------------------
         BMI |
       _cons |
------------------------------------------------------------------------------
```

OLS: *Estimation*

- OLS: Ordinary Least Squares
  - *Doing*: We are…
    - Fitting lines to data… estimating coefficients… that's all
    - Doing optimization/minimization… nothing more
    - Minimizing SSRs (Sum Squared Residuals)
      - Residuals^2: (Actuals-Predicteds)^2
  - *Not Doing*: We are not…
    - Using any probability or statistics
    - Making any assumptions about the underlying models that might have generated our data
    - Discussing *true* parameters…
      - *What's a true parameter? We're just fitting lines to data!*

OLS Estimation I: *nObs & Coefficients*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =       252
-------------+------------------------------           F(  1,   250) =
       Model |                                         Prob > F      =
    Residual |                                         R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |                                         Root MSE      =
------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547
       _cons |     (20.405)
------------------------------------------------------------------------------
```

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} = \rho_{xy} \frac{S_y}{S_x} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

OLS Estimation II: *Predicteds, SSEs, SSTs & R^2*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =
       Model |  7,991.510                               Prob > F      =
    Residual |                                          R-squared     =    0.53
-------------+------------------------------           Adj R-squared =
       Total |  15,079.017                              Root MSE      =
------------------------------------------------------------------------------

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
------------------------------------------------------------------------------
         BMI |       1.547
       _cons |     (20.405)
------------------------------------------------------------------------------
```

$$SST = \sum (y_i - \bar{y})^2 = (n-1)S_{yy}$$

$$SSE = \sum (\hat{y}_i - \bar{y})^2 = (n-1)S_{\hat{y}\hat{y}}$$

$$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST} = \frac{S_{\hat{y}\hat{y}}}{S_{yy}} = \rho^2_{xy} = \rho^2_{\hat{y}y}$$

OLS *Estimation III: SSRs, MSE & RMSE*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =
       Model |   7,991.510                              Prob > F      =
    Residual |   7,087.507       250   28.350           R-squared     =    0.53
-------------+------------------------------           Adj R-squared =
       Total |  15,079.017                              Root MSE      =  5.3245

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547
       _cons |     (20.405)
------------------------------------------------------------------------------
```

$$SSR = \sum \hat{u}_i^{\,2} = \sum (y_i - \hat{y}_i)^2 = (n-1)S_{\hat{u}\hat{u}}$$

$$MSE = \frac{SSR}{dofs} = \frac{SSR}{n-2}, \qquad RMSE = \sqrt{MSE} = \sqrt{\frac{SSR}{n-2}}$$

OLS/SLR: *Estimation Review*

```
. reg Brozek BMI
```

ANOVA TABLE

Source	SS	df	MS
Model	7,991.510	uninteresting	
Residual	7,087.507	250	28.350
Total	15,079.017	uninteresting	

Goodness of Fit

Number of obs	=	252
F(1, 250)	=	281.89
Prob > F	=	0
R-squared	=	0.53
Adj R-squared	=	MLR
Root MSE	=	5.3245

	Estimation	***Inference***				
Brozek	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
BMI	1.547	0.092	16.79	0	1.365	1.728
_cons	(20.405)	2.367	-8.62	0	(25.067)	(15.743)

OLS: *Inference*

- *Are we close? How close? Are you sure?*
- Tool of Inference:
  - Confidence Intervals
  - Hypothesis Testing
- Cornerstone of Inference: The *t statistic*
- *Doing*: We are …
  - Using probability and statistics… *finally!*
  - Making assumptions about the underlying *stochastic* model that generated our data (*stochastic* = *probabilistic*)
  - Focusing on estimating the *true* parameters of that model
  - Making distributional assumptions (e.g. Normal distribution)

OLS Inference: *Preview*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =
       Model |  7,991.510             uninteresting     Prob > F      =
    Residual |  7,087.507        250  28.350            R-squared     =    0.53
-------------+------------------------------           Adj R-squared =     MLR
       Total |  15,079.017            uninteresting     Root MSE      =  5.3245

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547
       _cons |     (20.405)
------------------------------------------------------------------------------
```

$$B_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_j - \bar{X})^2} = \frac{S_{XY}}{S_{XX}} = \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_j - \bar{X})^2} \qquad B_0 = \bar{Y} - B_1 \bar{X}$$

OLS Inference I: *Standard Errors*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =
       Model |   7,991.510                              Prob > F      =
    Residual |   7,087.507       250     28.350         R-squared     =    0.53
-------------+------------------------------           Adj R-squared =
       Total |  15,079.017                              Root MSE      =  5.3245
------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547     0.092
       _cons |     (20.405)    2.367
------------------------------------------------------------------------------
```

$$se(B_1) = \sqrt{\frac{MSE}{\sum (x_i - \bar{x})^2}} = \frac{RMSE}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{RMSE}{S_x \sqrt{(n-1)}}$$

OLS Inference II: *t Stats & Confidence Intervals*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =
       Model |   7,991.510                              Prob > F      =
    Residual |   7,087.507       250    28.350          R-squared     =    0.53
-------------+------------------------------           Adj R-squared =
       Total |  15,079.017                              Root MSE      =  5.3245
------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547    0.092     16.79   0          1.365    1.728
       _cons |     (20.405)   2.367     -8.62   0        (25.067)  (15.743)
------------------------------------------------------------------------------
```

$$\frac{B_1 - \beta_1}{se(B_1)} \sim t_{n-2} \qquad t\ stat = \frac{B_1 - 0}{se(B_1)} = \frac{B_1}{se(B_1)} \qquad \left[B_1 - c \cdot se(B_1),\ \ B_1 + c \cdot se(B_1)\right]$$

OLS Inference III: *Hypothesis Test I (p values)*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =
       Model |   7,991.510                              Prob > F      =
    Residual |   7,087.507       250     28.350         R-squared     =    0.53
-------------+------------------------------           Adj R-squared =
       Total |  15,079.017                              Root MSE      =  5.3245
------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf.   Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547    0.092     16.79   0.000       1.365     1.728
       _cons |     (20.405)   2.367     -8.62   0.000     (25.067)   (15.743)
------------------------------------------------------------------------------
```

$$p\ value = P\left(\left|t_{n-2}\right| > \left|t\ stat\right|\right)$$

OLS Inference IV: H_0 *Test II (F stats & F probs)*

```
. reg Brozek BMI

      Source |       SS       df       MS              Number of obs =     252
-------------+------------------------------           F(  1,   250) =  281.89
       Model |   7,991.510     1  7,991.510             Prob > F      =       0
    Residual |   7,087.507   250     28.350             R-squared     =  0.5300
-------------+------------------------------           Adj R-squared =  0.5281
       Total |  15,079.017   251     60.076             Root MSE      =  5.3245

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |       1.547    0.092     16.79   0.000     1.365     1.728
       _cons |     (20.405)   2.367     -8.62   0.000    (25.067)  (15.743)
------------------------------------------------------------------------------
```

$$F\ stat = F(1, n-2) = t^2_{\hat{\beta}_1} = (n-2)\frac{R^2}{1-R^2} = (n-2)\frac{SSE}{SSR}$$

$$Prob > F = P\left(F(1, n-2) > Fstat\right)$$

OLS/SLR: *Inference Preview*

```
. reg Brozek BMI
```

ANOVA TABLE

Source	SS	df	MS
Model	7,991.510	uninteresting	
Residual	7,087.507	250	28.350
Total	15,079.017	uninteresting	

Goodness of Fit		
Number of obs	=	252
F(1, 250)	=	281.89
Prob > F	=	0
R-squared	=	0.53
Adj R-squared	=	MLR
Root MSE	=	5.3245

Estimation		*Inference*				
Brozek	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
BMI	1.547	0.092	16.79	0	1.365	1.728
_cons	(20.405)	2.367	-8.62	0	(25.067)	(15.743)

OLS/SLR: *Estimation Preview & Inference Preview*

```
. reg Brozek BMI
```

ANOVA TABLE

Source	SS	df	MS
Model	7,991.510	uninteresting	
Residual	7,087.507	250	28.350
Total	15,079.017	uninteresting	

Goodness of Fit

Number of obs	=	252
F(1, 250)	=	281.89
Prob > F	=	0
R-squared	=	0.53
Adj R-squared	=	MLR
Root MSE	=	5.3245

Estimation / ***Inference***

Brozek	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
BMI	1.547	0.092	16.79	0	1.365	1.728
_cons	(20.405)	2.367	-8.62	0	(25.067)	(15.743)