OLS/SLR Analytics: *Estimating Parameters*

- ***The SLR (Simple Linear Regression) Model Setup***
- ***OLS (Ordinary Least Squares) Estimation: FOCs and SOCs***
- ***OLS and Sample Statistics: Interpreting the OLS Coefficients***
- ***… $\hat{\beta}_1$: A Weighted Average of Slopes***
- ***OLS Predictions, Residuals and SRFs***
- ***… Properties of OLS/SLR Residuals***
- ***Units Of Measurement and Estimated Coefficients***
- ***Economic Significance (Meaningfulness): Beta Regressions and Elasticities***
- ***Examples in Excel and Stata***

The SLR (Simple Linear Regression) Model Setup

1. You have a dataset consisting of n observations of two variables (x, y): $\{(x_i, y_i)\} \quad i = 1, 2, \ldots n$.

2. You believe that except for random noise in the data, there is a linear relationship between the x's and the y's: $y_i \sim \beta_0 + \beta_1 x_i \ldots$ and are interested in estimating the unknown parameters β_0 (the *y intercept*) and β_1 (the *slope*).

3. If there was no noise in the data, then since $y_i = \beta_0 + \beta_1 x_i$ for all observations, we could easily determine β_0 and β_1.[1] But typically, the relationship is not exactly linear in the observed data.

4. Call your parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, and your predicted *y* values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

   a. I will try to be consistent and always use $\beta\,'s$ for true parameter values… and $\hat{\beta}\,'s$ for estimates of the $\beta\,'s$.

5. We call the difference between the observed y_i and the predicted value $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ the residual, $\hat{u}_i$: $\hat{u}_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$.

   residuals = actuals - predicteds

6. One measure of how well the predicteds fit the actuals will be the SSRs, the Sum of the Squared Residuals: $SSR = \sum \hat{u}_i^2 = \sum \left(y_i - \hat{y}_i\right)^2$.

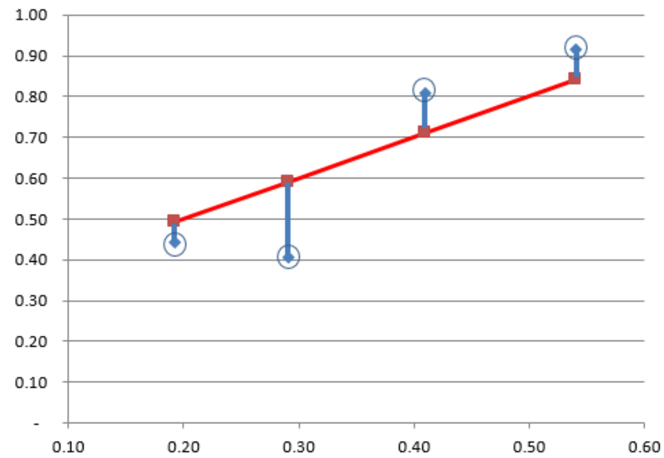
   a. We square the residuals so that positive and negative residuals won't offset one another when we add them up.

[1] β_1 is just the slope of the line connecting any two datapoints, and $\beta_0 = y_i - \beta_1 x_i$, for any datapoint.

7. Here's an example (negative residuals for #1 and #2; positive residuals for #3 and #4):

		intercept	0.3	
		slope	1	
				SSRs
				0.0520
id	x	y	pred	residual
1	0.19	0.44	0.49	(0.05)
2	0.29	0.41	0.59	(0.18)
3	0.41	0.81	0.71	0.10
4	0.54	0.92	0.84	0.08

id	resid^2
1	0.0025
2	0.0338
3	0.0100
4	0.0056
SSRs	0.0520



8. In this example, we have:

   a. pred(icteds) computed using an intercept = .03 and slope = 1… and SSR=.0520

   b. In *Ordinary Least Squares (OLS)* regressions, the goal is to find the coefficient values that minimize the sum of the squared residuals, or SSRs… which is why we call the estimated coefficients *least squares* estimates.

OLS = min SSRs

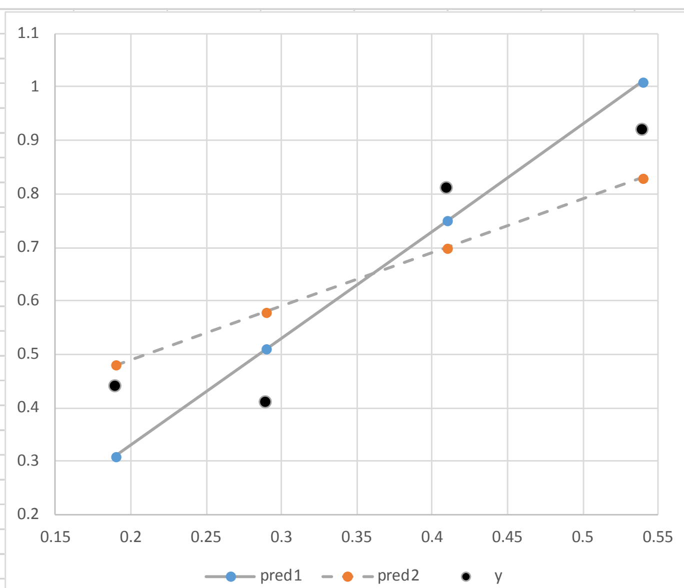
9. Here are two more examples:

   a. pred1: intercept = -.07, slope = 2, and SSR=.0386 (solid line below)

   b. pred2: intercept = .29, slope = 1, and SSR=.0507 (dashed line below)

   c. Note that both predicted values are above data point #2 and below data point #3… and on opposite sides of data points #1 and #3.

		intercept	-0.07	0.29
		slope	2	1
			predicteds	
id	x	y	pred1	pred2
1	0.19	0.44	0.31	0.48
2	0.29	0.41	0.51	0.58
3	0.41	0.81	0.75	0.7
4	0.54	0.92	1.01	0.83
			residuals	
1			0.13	(0.04)
2			(0.10)	(0.17)
3			0.06	0.11
4			(0.09)	0.09
			residuals^2	
1			0.0169	0.0016
2			0.0100	0.0289
3			0.0036	0.0121
4			0.0081	0.0081
			SSRs	
			0.0386	***0.0507***



10. Perhaps we can do better in terms of minimizing SSRs, but at the moment, the pred1 coefficients do the best job of fitting the data, with SSR=.0386; pred2 is second best with SSR = .0507, and pred in the first chart provides the poorest fit to the data, with SSR = .052.

11. ***Take Away***: The fit of the *predicteds* to the *actuals* will vary as we change the intercept and slope coefficients. The goal is to find the coefficient values that provide in some sense the best fit. One way of measuring the fit for each set of coefficients is to look at SSRs the sum of the squared residuals. The OLS coefficients will provide the best fit of predicteds to actuals, in the sense of having the smallest possible SSR. And that's why we call the estimation technique *least squares*… or more formally, *Ordinary Least Squares*.

OLS (Ordinary Least Squares) Estimation: *FOCs and SOCs*

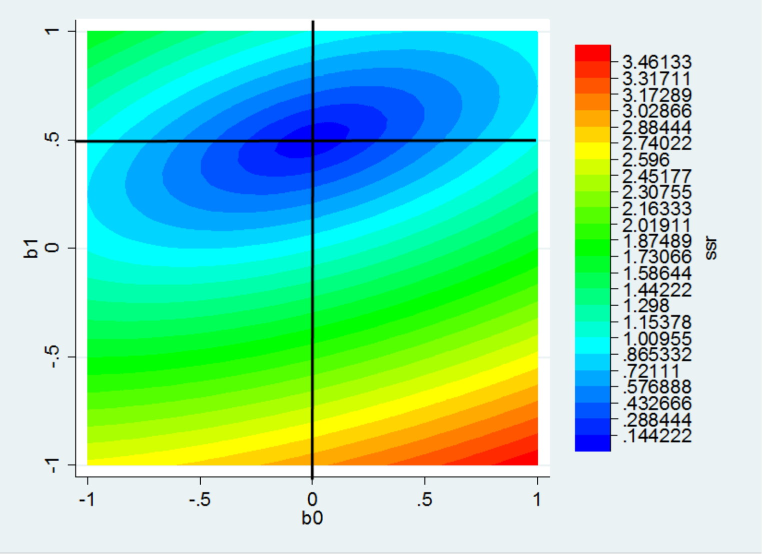
12. OLS: Minimize Sum (of the) Squared Residuals (SSRs)

    a. The challenge in Ordinary Least Squares is to find the slope coefficient (b_1) and intercept coefficient (b_0) that together minimize Sum Squared Residuals (SSR), defined by:

    $$SSR = \sum (u_i)^2 = \sum \left(y_i - (b_0 + b_1 x_i) \right)^2 .$$

    b. To do this, and as you saw in *Getting Started II*, we'll use *First Order Conditions (FOCs)* to identify *least squares* coefficient candidates, and *Second Order Conditions (SOCs)* to ensure that we have indeed minimized SSRs.

    c. Before turning to the math, here's an example of SSR contours for different values of *b0* and *b1*. In the Figure, SSRs are minimized when $b0 = 0$ and $b1 = .5$:



13. OLS I: Working with standardized variables

a. Assume that the x's and y's have been standardized to have mean zero and variance one, so that $\overline{x} = \overline{y} = 0$, $S_x = S_{xx} = S_y = S_{yy} = 1$, and $S_{xy} = \rho_{xy}$.

b. FOCs: Focus on the FOCs for our minimization problem:

minimize $SSR = \sum\left(y_i - \left(b_0 + b_1 x_i\right)\right)^2$ with respect to (wrt) b_0 *and* b_1.

i. FOC 1: Differentiating wrt b_0:

$$\frac{\partial SSR}{\partial b_0} = -2\sum(y_i - b_0 - b_1 x_i) \Rightarrow -2n\overline{y} + 2nb_0 + 2b_1 n\overline{x} = 0\text{, and so } /(n-1)$$

$$\frac{\partial SSR}{\partial b_0} = 0 \Leftrightarrow b_0 = \overline{y} - b_1\overline{x}$$

1. Checking a SOC: $\frac{\partial^2 SSR}{\partial b_0^2} = 2n > 0$, so we have a minimum at b_0.

ii. Since $\overline{x} = \overline{y} = 0$, $b_0^* = \overline{y} - b_1\overline{x} = 0$ is our best estimate for the intercept parameter.[2]

iii. And so our minimization problem becomes:

minimize $SSR = \sum\left(y_i - b_1 x_i\right)^2$ wrt b_1.

iv. FOC 2: Differentiating $SSR = \sum\left(y_i - b_1 x_i\right)^2$ wrt b_1:

$$\frac{dSSR}{db_1} = -2\sum x_i\left(y_i - b_1 x_i\right) = 0\text{. So } \sum\left(x_i y_i\right) = b_1\sum x_i^2\text{, and } b_1 = \frac{\sum\left(x_i y_i\right)}{\sum x_i^2}.$$

1. Checking a SOC: $\frac{d^2 SSR}{db_0^2} = 2\sum x_i^2 > 0$, so we do indeed have a minimum at b_1.

v. Since x and y are standardized, we have several equivalent expressions for the estimated slope coefficient:

$$b_1^* = \frac{\sum\left(x_i y_i\right)}{\sum x_i^2} = \frac{\left[\sum\left(x_i y_i\right)\right]/(n-1)}{\left[\sum x_i^2\right]/(n-1)} = \frac{S_{xy}}{S_{xx}} = \rho_{xy}.$$

c. Accordingly, the predicted values generated by OLS with standardized variables are defined by: $\hat{y}_i = \rho_{xy} x_i$. Now you know why it is sometimes said that OLS parameter estimates capture the correlations between variables. And now you perhaps better understand the results in qFlip01!

[2] The * indicates that the particular coefficient value minimizes SSRs.

14. OLS II: … more generally …

a. Now, turn to the more general case in which the x's and y's have not been standardized.

b. Focusing on the FOCs for our minimization problem:

minimize $SSR = \sum\left(y_i - (b_0 + b_1 x_i)\right)^2$ with respect to (wrt) b_0 *and* b_1.

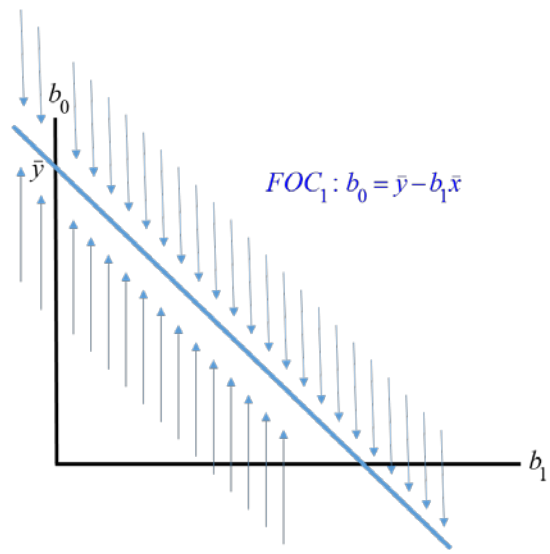
i. FOC 1: Differentiating wrt b_0:

$\frac{\partial SSR}{\partial b_0} = -2\sum(y_i - b_0 - b_1 x_i) \Rightarrow -2n\bar{y} + 2nb_0 + 2b_1 n\bar{x} = 0$, and so

$$\frac{\partial SSR}{\partial b_0} = 0 \Leftrightarrow b_0 = \bar{y} - b_1\bar{x}$$

1. So as before, the intercept estimate will be equal to the mean of the *y's* less the slope estimate times the mean of the *x's*. You don't yet know what the intercept and slope estimates are… but you know that for FOC 1 to be satisfied, they $(b_0 \text{ and } b_1)$ have to satisfy this relationship.

2. The following Figure is illustrative… and assumes $\bar{y} > 0$ *and* $\bar{x} > 0$. FOC 1 implies that the b_0 *and* b_1 that minimize SSRs must lie on the straight line defined by $b_0 = \bar{y} - b_1\bar{x}$. To find the exact SSR minimizing values of b_0 *and* b_1, we turn to FOC 2.



ii. Since $SSR = \sum\left[y_i - (b_0 + b_1 x_i)\right]^2$ and $b_0 = \bar{y} - b_1\bar{x}$, we now want to minimize

$SSR = \sum\left[y_i - (\bar{y} - b_1\bar{x} + b_1 x_i)\right]^2 = \sum\left[(y_i - \bar{y}) - b_1(x_i - \bar{x})\right]^2$, wrt b_1.

iii. FOC 2: Differentiating wrt b_1:

$\frac{dSSR}{db_1} = -2\sum(x_i - \bar{x})\left[(y_i - \bar{y}) - b_1(x_i - \bar{x})\right] = 0$. So

$\sum(x_i - \bar{x})(y_i - \bar{y}) = b_1\sum(x_i - \bar{x})^2$, and $b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$.

c. The SOCs are more complicated and will be skipped, but rest assured that they are satisfied and the FOCs identify the global minimum of the SSRs. For some intuition: the two second derivatives (wrt b_0 and wrt b_1) are both positive, suggesting that we may indeed have identified a minimum with the FOCs:

i. Differentiating FOC 1: $\frac{\partial^2 SSR}{\partial b_0^2} = 2n > 0$, and

ii. Differentiating FOC 2: $\frac{d^2 SSR}{db_1^2} = -2(-1)\sum (x_i - \overline{x})^2 > 0$.

OLS and Sample Statistics : ***Interpreting the OLS coefficients***

15. The OLS estimated coefficients

a. For the given sample, the OLS estimates of the unknown intercept and slope parameters are:

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}, \text{ and } \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

As previously mentioned, we use "hats" to denote estimates.

b. Since $\sum (x_i - \overline{x}) = 0$, $\sum (x_i - \overline{x})(y_i - \overline{y}) = \sum (x_i - \overline{x}) y_i - \overline{y} \sum (x_i - \overline{x}) = \sum (x_i - \overline{x}) y_i \ldots$ as discussed in the Sample Statistics section of *Getting Started II*. Accordingly, we have an alternative expression for the estimated slope coefficient which will prove useful later:

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}.$$

c. $\hat{\beta}_0$ and the sample means

i. Since $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$, the estimated intercept is the sample mean of the *y's* minus $\hat{\beta}_1$ times the sample mean of the *x's*.

ii. The estimate of the intercept assures that the average predicted value, $\hat{\beta}_0 + \hat{\beta}_1 \overline{x}$, is the same as the average observed value $\overline{y}$, since $\hat{\beta}_0 + \hat{\beta}_1 \overline{x} = \left(\overline{y} - \hat{\beta}_1 \overline{x}\right) + \hat{\beta}_1 \overline{x} = \overline{y}$.

d. $\hat{\beta}_1$ and the sample variances, covariance and correlation

i. If we divide the numerator and denominator of the $\hat{\beta}_1$ equation by *(n-1)*, then using the sample statistics notation from *Getting Started II*, we have:

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{\left[\sum (x_i - \overline{x})(y_i - \overline{y})\right] / (n-1)}{\left[\sum (x_i - \overline{x})^2\right] / (n-1)} = \frac{S_{xy}}{S_{xx}}.$$

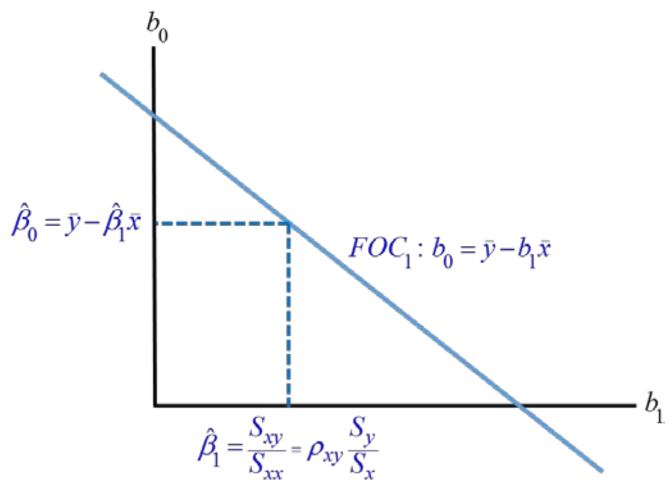
ii. Accordingly, the OLS slope estimator is just the ratio of the sample covariance of *x's* and *y's* and the sample variance of the *x's*:

$$\hat{\beta}_1 = \frac{Sample\ Covariance(x, y)}{Sample\ Variance(x)}$$

iii. Recall that the sample correlation is defined by: $\rho_{xy} = \frac{S_{xy}}{S_x S_y}$, where S_x and S_y are the square roots of the respective sample variances.

iv. Since $\rho_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{S_{xy}}{S_{xx}} \frac{S_x}{S_y}$, we have: $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \rho_{xy} \frac{S_y}{S_x}$.

e. And so once the slope coefficient is determined, the SSR minimizing intercept coefficient follows from FOC 1:



f. ***OLS/SLR Slope Estimate ~ Correlation***: So the regression slope coefficient is the product of the sample correlation between the *x's* and *y's* and the ratio of the two estimated standard deviations:

$$\hat{\beta}_1 = Sample\ Correlation(x, y) \frac{Sample\ StdDev(y)}{Sample\ StdDev(x)}$$

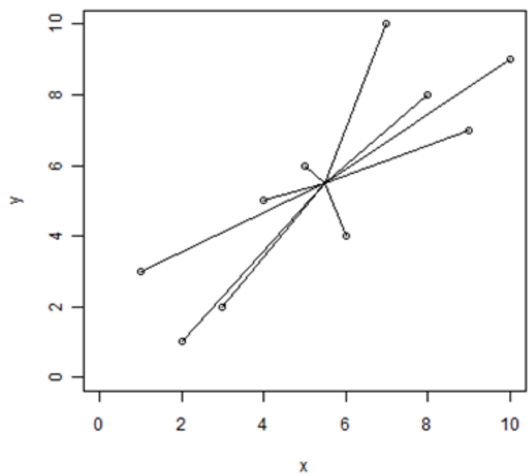
i. If the two sample standard deviations are the same then the estimated slope coefficient will be the estimated correlation between the *x's* and *y's*. You saw this in the first instance when we considered SLR models with standardized variables, with $S_x = S_y = 1$.

ii. Indeed it is not unusual to think of the OLS slope estimate $\hat{\beta}_1$ as reflecting the correlation between the x's and y's. Since $\hat{\beta}_1 = \rho_{xy} \frac{S_y}{S_x}$ the sign of the estimated slope coefficient, $\hat{\beta}_1$, is the same as the sign of the correlation between x and y, ρ_{xy}

(assuming that the ratio of standard deviations positive, which it always is unless one of the standard deviations is zero).

$\hat{\beta}_1$: A Weighted Average of Slopes

16. The estimated slope coefficient is a weighted average of slopes of lines joining the various data points to the sample means $(\bar{x}, \bar{y})$:



$$\hat{\beta}_1 = \sum_i w_i \left[\frac{(y_i - \bar{y})}{(x_i - \bar{x})} \right] = \sum_i w_i \ slope_i \,.$$

a. This result holds because

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_{xx}}$$

$$= \sum_i \left[\frac{(x_i - \bar{x})^2}{(n-1)S_{xx}} \right] \left[\frac{(y_i - \bar{y})}{(x_i - \bar{x})} \right]$$

$$= \sum_i w_i \left[\frac{(y_i - \bar{y})}{(x_i - \bar{x})} \right] = \sum_i w_i \ slope_i \text{, where}$$

$slope_i = \dfrac{(y_i - \bar{y})}{(x_i - \bar{x})}$ is the slope of the line connecting (x_i, y_i) to $(\bar{x}, \bar{y})$, and

$$w_i = \frac{(x_i - \bar{x})^2}{(n-1)S_{xx}} = \frac{(x_i - \bar{x})^2}{\sum_j (x_j - \bar{x})^2} \,.$$

b. By construction, the w_i's are non-negative weights, which sum to 1.

c. Accordingly, in the equation for $\hat{\beta}_1$, the slopes are weighted proportionally to $(x_i - \bar{x})^2$, the square of the various x-distances from the x mean.

d. In this interpretation, note that the data points are not weighted equally (that would be another estimator… but not OLS). Those that are farther away from $\bar{x}$ (in the x dimension) get greater weight, and that weight increases with the square of the x-distance from $\bar{x}$.

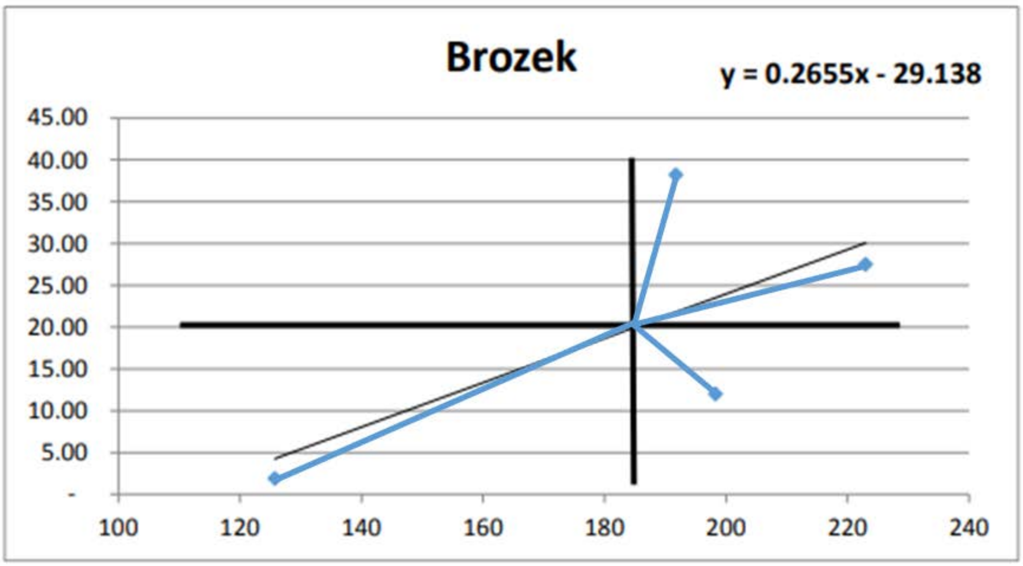
e. ***Here's an example:*** The blue dots are the data points; the horizontal and vertical black lines are at the sample means; the blue lines are the lines connecting the data points to the sample means; and the think black line shows the predicted Brozek values given the slope and intercept estimates.

   i. Note that in the weighted averaging of slopes, one data point, (125.75, 1.90), gets two thirds of the weight, and when combined with (223, 27.50), those two data points get 95% of the weight. So even though there are four data points, the slope estimate is being largely driven by just two of the data points.

OLS Slope Estimate: Weighted Average of slopes

Means	184.688	19.900

Case	wgt	Brozek	x-dist	y-dist	slope	(x-dist)^2	wgt	wgt*slope
172	125.75	1.90	58.94	18.00	**0.31**	3,473.63	**67%**	0.2050
36	191.75	38.20	(7.06)	(18.30)	**2.59**	49.88	**1%**	0.0250
10	198.25	12.00	(13.56)	7.90	**(0.58)**	183.94	**4%**	(0.0207)
205	223	27.50	(38.31)	(7.60)	**0.20**	1,467.85	**28%**	0.0563
						5,175.30	***100%***	***0.2655***



OLS Predictions, Residuals and *SRF*s

17. OLS coefficient estimates will generate predicted values, $\hat{y}'s$, and residuals, $\hat{u}'s$:

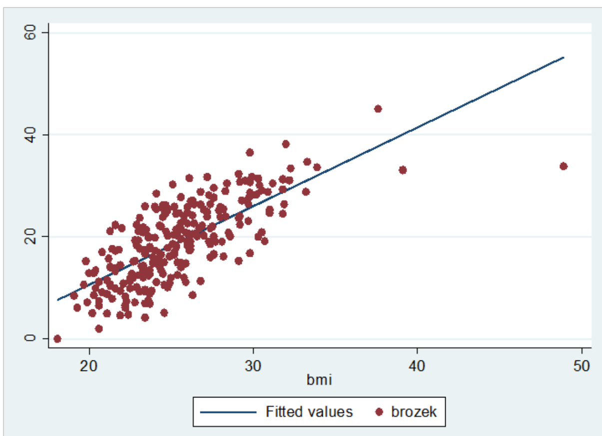
    a. ***Predicted values***: For given x_i, the predicted y_i value given the estimated coefficients is: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ (recall again that we use "hats" for predicted or estimated values).

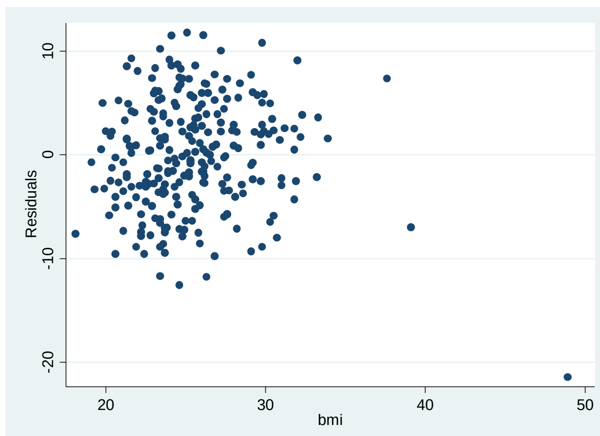
    b. ***Sample Regression Function (SRF)***: The predicted values from the estimated equation, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, comprise the ***Sample Regression Function***.

    c. ***Residuals***: And for the given predicted y_i value, the residual, $\hat{u}_i$, is as above the difference between the actual and predicted values: $\hat{u}_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$.

18. To illustrate *predicteds* (from the SRF) and *residuals*, we turn to the bodyfat dataset and a regression of Brozek on BMI, with an SRF defined by: $\widehat{brozek} = -20.41 + 1.55\, bmi$.

    a. The following chart on the left illustrates the relationship between the SRF (predicteds) and the actuals (the actual $\left(x_i, y_i\right)$ data points).





    b. And on the right you see a graph of the residuals from the analysis.

        i. You shouldn't be surprised to see the SRF slice through the dataset… since we estimated the coefficients by minimizing SSRs.

        ii. And you should not be surprised to see the residuals evenly dispersed above and below 0, since by construction, and as you'll see below, the residuals will have sample mean 0.

        iii. But what about that rogue residual in the lower right corner of the Figure? Need to check on that!

19. SRFs will depend on the actual sample used to estimate the slope and intercept parameters… different samples will typically lead to different parameter estimates and accordingly, different SRFs. But there are some consistent outcomes with SRFs:

a. The SRF always passes through the sample means, $(\overline{x}, \overline{y})$.

   i. $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$ assures that the SRF passes through $(\overline{x}, \overline{y})$ since as previously discussed, the value of the SRF at $\overline{x}$ is $\hat{\beta}_0 + \hat{\beta}_1\overline{x} = \left(\overline{y} - \hat{\beta}_1\overline{x}\right) + \hat{\beta}_1\overline{x} = \overline{y}$

b. $\rho_{y\hat{y}} = \rho_{yx}$: ***SampleCorrelation(predicteds, actuals) = SampleCorrelation(x's, y's)***

   i. The sample correlation between the actuals and predicted values is the same as the sample correlation between the actuals and the *x's*: $\rho_{y\hat{y}} = \rho_{yx}$. Proof below.[3]

… Properties of OLS/SLR Residuals

20. Recall that the residuals are defined by: $\hat{u}_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$. We have the following properties of residuals:

21. ***Average residuals:*** The average residual is zero: $\frac{1}{n}\sum \hat{u}_i = \overline{y} - \left(\hat{\beta}_0 + \hat{\beta}_1\overline{x}\right) = 0$

22. ***Correlation I:*** The sample correlation between the x_i's and the $\hat{u}_i$'s is zero, since $\sum \hat{u}_i\left(x_i - \overline{x}\right) = 0$. Proof below.[4]

23. ***Correlation II:*** The sample correlation between the predicted values ($\hat{y}_i$ s) and the residuals ($\hat{u}_i$ s) is zero. Proof below.[5]

predicteds and residuals are uncorrelated

24. ***Decomposition***: And so OLS essentially decomposes actual y_i's s into two uncorrelated parts, *predicteds* and *residuals*: $y_i = \hat{y}_i + \hat{u}_i$ and $\hat{\rho}_{\hat{y}\hat{u}} = 0$. This result will prove useful later.

$$\boxed{y_i = \hat{y}_i + \hat{u}_i} \quad \boxed{\hat{\rho}_{\hat{y}\hat{u}} = 0}$$

---

[3] Since $\rho_{y\hat{y}} = \frac{S_{y\hat{y}}}{S_y S_{\hat{y}}}$, and since $S_{y\hat{y}} = S_{y(\hat{\beta}_0 + \hat{\beta}_1 x)} = S_{y\hat{\beta}_0} + \hat{\beta}_1 S_{yx} = \hat{\beta}_1 S_{yx}$, and since $S_{\hat{y}\hat{y}} = \hat{\beta}_1^2 S_{xx}$, we have:

$$\rho_{y\hat{y}} = \frac{S_{y\hat{y}}}{S_y S_{\hat{y}}} = \frac{\hat{\beta}_1 S_{yx}}{S_y \hat{\beta}_1 S_x} = \frac{S_{yx}}{S_y S_x} = \rho_{yx}.$$

[4] $\sum \hat{u}_i\left(x_i - \overline{x}\right) = \sum\left(y_i - \hat{y}_i\right)\left(x_i - \overline{x}\right) = \sum\left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)\left(x_i - \overline{x}\right)$. But since $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$, the last expression is $\sum\left(y_i - (\overline{y} - \hat{\beta}_1\overline{x} + \hat{\beta}_1 x_i)\right)\left(x_i - \overline{x}\right) = \sum\left((y_i - \overline{y}) - \hat{\beta}_1(x_i - \overline{x})\right)\left(x_i - \overline{x}\right)$

$= \sum (y_i - \overline{y})\left(x_i - \overline{x}\right) - \hat{\beta}_1 \sum (x_i - \overline{x})^2 = 0$ given the definition of $\hat{\beta}_1$.

[5] $\sum\left(\hat{u}_i - \overline{\hat{u}}\right)\left(\hat{y}_i - \overline{\hat{y}}\right) = \sum \hat{u}_i\left(\hat{y}_i - \overline{y}\right) = \hat{\beta}_1 \sum\left(y_i - \hat{y}_i\right)\left(x_i - \overline{x}\right)$, which is zero (see previous proof).

25. Since the predicted and residuals have zero covariance, the variance of their sum is the sum of their variance: $S_{yy} = S_{\hat{y}\hat{y}} + S_{\hat{u}\hat{u}}$.

Var(actuals) = Var(predicteds) + Var(residuals)

26. Multiplying through by (n-1), we also have:

$$\sum\left(y_i - \overline{y}\right)^2 = \sum\left(\hat{y}_i - \overline{y}\right)^2 + \sum \hat{u}_i^2 ,$$

since the mean of the predicted y's is the mean of the actuals, $\frac{1}{n}\sum \hat{y}_i = \overline{y}$, and since the residuals have mean 0, $\frac{1}{n}\sum \hat{u}_i = 0$. This result will prove especially useful later.

Units of Measurement and Estimated Coefficients

27. It is of so tempting to see large estimated coefficients and to rejoice in thinking that you've found a large effect… or to see a small estimated coefficient and to fall into the depths of depression thinking that you've found no real effect at all. And in both cases, you would be seriously in error… falling into the the trap of thinking that the magnitudes of the estimated coefficients tell you something meaningful. They do not… as they are sensitive to units of measurement.

28. If you change units of measurement, you will change OLS estimated coefficients… or put differently, you can make the magnitudes of those coefficients as large or small as you want just by changing units of measurement.

Customary Units of Measurement - Chart

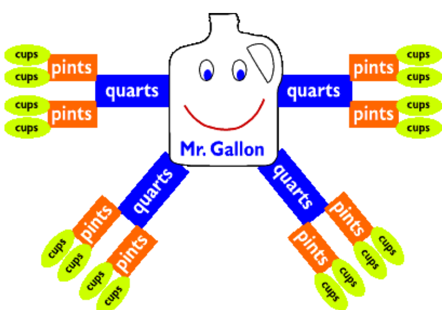
Length	Weight	Capacity	Time
12 in = 1 ft	16 oz = 1lb	128 fl oz = 1 gal	60 sec = 1 min
3 ft = 1 yrd	2000 lb = 1 ton	2 pt = 1 qt	60 min = 1 hr
5,280 ft = 1 mi		8 pt = 1 gl	24 hr = 1 day
1,760 yrd = 1 mi		4 qt = 1 gal	7 days = 1 wk
			52 wk = 1 yr
			12 mon = 1 yr
			365 days = 1 yr

29. Here's why:

    a. Consider the standard SLR model in which you've regressed *y* on *x*. You know from above that the estimated OLS slope and intercept coefficients will be defined by:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \rho_{xy}\frac{S_y}{S_x} \text{ and } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

    b. Now suppose that you rescale x to create a new variable v, $v = \lambda_x x$, where $\lambda_x > 0$ so that you preserve the sign of the variable. For example:



i. Perhaps *x* was originally height measured in feet… and now *v* is height measured in inches. Then $\lambda_x = 12$, and $v = 12x$.

ii. Or maybe *x* was originally volume measured in gallons… and now *v* is volume measured in quarts, so that $\lambda_x = 4$, and $v = 4x$.

You get the idea.

    c. This rescaling will impact the sample mean and standard deviation:

$$\bar{v} = \lambda_x \bar{x} \text{ and } S_v = \lambda_x S_x.$$

    d. Suppose you also rescale the *y*'s as well: $w = \lambda_y y$. This will similarly impact the mean and standard deviation: $\bar{w} = \lambda_y \bar{y}$ and $S_w = \lambda_y S_y$.

    e. But recall that rescaling both the *x*'s and the *y*'s will not impact the sample correlations, so that: $\rho_{vw} = \rho_{xy}$. This will proves to be an important feature.

f. Now if you regress the rescaled variables on one another (you regress *w* on *v*), the new OLS coefficient estimates will be defined by:

$$slope = \rho_{vw}\frac{S_w}{S_v} = \rho_{xy}\frac{\lambda_y S_y}{\lambda_x S_x} = \frac{\lambda_y}{\lambda_x}\rho_{xy}\frac{S_y}{S_x} = \frac{\lambda_y}{\lambda_x}\hat{\beta}_1 \text{ and}$$

$$\_cons = \overline{w} - (slope)\,\overline{v} = \lambda_y\overline{y} - \left(\frac{\lambda_y}{\lambda_x}\hat{\beta}_1\right)\lambda_x\overline{x} = \lambda_y\left(\overline{y} - \hat{\beta}_1\overline{x}\right) = \lambda_y\hat{\beta}_0$$

30. Accordingly:

a. Changes in the units of measurement of the RHS *x* variable will proportionately impact the estimated slope coefficient… and have no impact on the estimated intercept. In a sense, the estimated coefficient will unwind the rescaling of the variable. In the previous examples:

i. feet to inches: x is 12 times larger, slope is 1/12th the size…. and slope*x is unchanged

ii. gallons to quarts: x is 4 times larger, slope is 1/4th the size…. and slope*x is unchanged

b. Changes in the units of measurement of the LHS *y* variable will proportionately impact both the estimated slope and intercept coefficients.

c. And if you rescale both variables, the impacts on the estimated slope and intercept coefficients will be some combination of the above..

31. To repeat: The magnitudes of the estimated coefficients will be dependent on the units of measurement, and in that sense, will typically tell you little about the meaningfulness of the estimated effect. There are exceptions, of course… but they are not the norm.

32. Or put differently: Don't fall into the trap of thinking that the <u>sizes/magnitudes</u> of estimated coefficients tell you anything useful, as they are driven in part by the units of measurement. In contrast: the <u>sign</u> of slope coefficient (which does not change with rescaling) does tell you the direction of the estimated effect. So pay attention to signs… but not so much to magnitudes… unless you have specific reasons for thinking that the magnitudes are meaningful.

33. But then, how do we assess meaningfulness?

Economic Significance (*Meaningfulness*): *Beta Regressions and Elasticities*

34. Once the unknown parameters have been estimated using OLS, the obvious question is: What do those estimates tell you? Do they suggest that there is a meaningful relationship between changes in the *x*'s and predicted changes in the *y*'s? Or maybe not? How do you tell?

35. Later, we will address this question from a statistical perspective, using the tools of statistical inference and the concept of ***statistical significance***. But for now, we focus on a more commonsensical approach to answering the question: How (economically) ***meaningful*** is the estimated relationship? Do you want to brag about it to the world? Or will everyone just laugh at you, and tell you that what you've estimated is trivial, and of little consequence?

36. Meaningfulness is definitely in the eye of the beholder. Nonetheless, there are some systematic ways in which researchers tackle the question: ***Beta Regressions* and *Elasticities***

Meaningfulness I: Beta Regressions

37. One way around the issue of sensitivity to units of measurement is to first standardize your variables before you run your regression. By subtracting means and dividing by standard deviations, you will transform your variables into variables with mean zero and unit variances and standard deviations. More importantly, your standardized variables will be insensitive to units of measurement… or put differently: changes in units of measurement will have no impact on the standardized variables.

38. More formally:

    a. Create the z's by standardizing the x's: $z_i = \dfrac{x_i - \overline{x}}{S_x}$.

    b. Suppose you rescale the x's as above: $v = \lambda_x x$ … so $\overline{v} = \lambda_x \overline{x}$ and $S_v = \lambda_x S_x$.

    c. Then the standardize v's will be defined by: $\dfrac{v_i - \overline{v}}{S_v} = \dfrac{\lambda_x x_i - \lambda_x \overline{x}}{\lambda_x S_x} = \dfrac{x_i - \overline{x}}{S_x} = z_i$.

    d. Changes in units of measurement will have no impact on the standardized variable… standardization negates the impact of any change in units of measurement.

    e. And so regressions run with standardized variables will be unaffected by changes in units of measurement.

39. Beta regressions: With beta regressions, we just regress the standardized y on the standardized x. As you saw earlier in the semester, the OLS estimated intercept will be zero, since both standardized variables have mean zero, and the estimated slope coefficient will just be the sample correlation between the x's and y's (which is unaffected by rescaling).

40. To run these in Stata, just add **, beta** to your reg command. Here's an example, working with the bodyfat dataset:

```
. corr Brozek BMI

             |   Brozek      BMI
-------------+------------------
      Brozek |   1.0000
         BMI |   0.7280   1.0000

. reg Brozek BMI, beta

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(1, 250)       =    281.89
       Model |  7991.50988         1  7991.50988   Prob > F        =    0.0000
    Residual |  7087.50675       250   28.350027   R-squared       =    0.5300
-------------+----------------------------------   Adj R-squared   =    0.5281
       Total |  15079.0166       251  60.0757635   Root MSE        =    5.3245

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
         BMI |   1.546712   .0921238    16.79   0.000                 .7279942
       _cons |  -20.40508   2.367227    -8.62   0.000                        .
------------------------------------------------------------------------------
```

41. The reported Coef.'s are the usual OLS coefficients from regressing y on x. The coefficients for the Beta Regression are on the far right of the results table…. and as expected, the estimated intercept is 0 and the estimated slope is just the sample correlation between Brozek and BMI.

42. There are two interpretations of the estimated beta regression slope coefficient:

    a. As mentioned above, the beta regression slope coefficient will be the sample correlation between x and y, which is invariant with respect to changes in units of measurement.

    b. The results above say that a one standard deviation increase in BMI is on average associated with a .73 standard deviation increase in Brozek. So the beta regression slope coefficient captures effects measured in standard deviation units. And those effects will not vary with units of measurement.

43. Since the slope estimates in Beta regressions are correlations, they are bounded between -1 and +1… and we have a sense of their magnitude: closer to zero, not so meaningful… and closer to -1 or +1, and we'd say that there was a meaningful relationship. For the .73 magnitude above, most would say that the estimated relationship was meaningful… and surely no one would laugh at that claim. Though never forget that meaningfulness is in the eye of the beholder.

Meaningfulness II: Elasticities

44. In economics and mathematics, we typically use derivatives to assess relationships between changes in one variable, say, x, and changes in another, say y. But derivatives are sensitive to units of measurement… and so to circumvent this problem, economists often turn to elasticities, which provide a unit free measure of responsiveness (of the predicted $\hat{y}$'s to changes in the x's):

    $elasticity = \frac{\%\Delta\hat{y}}{\%\Delta x}$ … the elasticity captures the estimated relationship between percentage changes in x and percentage changes in the predicted values.

45. Using the SRF to estimate relationships: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

    a. ***Derivative***: The estimated average marginal relationship between x and $\hat{y}$: $\frac{d}{dx}\hat{y} = \hat{\beta}_1$.

       Note that you can read the derivative right off the regression output (it's the estimated slope coefficient, $\hat{\beta}_1$).

    b. ***(Point) Elasticity***: $\frac{x}{\hat{y}}\frac{d}{dx}\hat{y} = \hat{\beta}_1\frac{x}{\hat{y}}$ evaluated at $(x,\hat{y}) = (x, \hat{\beta}_0 + \hat{\beta}_1 x)$, somewhere along the SRF…. ***Evaluate where? Your call!***

       i. Where you evaluate the elasticity on the SRF is often arbitrary… but be sure to evaluate the elasticity at some point on the SRF. You will typically get different elasticities depending on where along the SRF you estimate the elasticity… but maybe they don't change much as you move along the SRF.

       ii. We often evaluate the elasticity ***at the means*** (which are by definition in the middle of the dataset): $\hat{\beta}_1\frac{\overline{x}}{\overline{y}}$

          1. Recall that the mean of the predicted values will be $\overline{y}$ … and that the SRF passes through $(\overline{x},\overline{y}) = (\overline{x}, \hat{\beta}_0 + \hat{\beta}_1\overline{x})$.

46. You can use the *margins* command in Stata to generate the elasticities. Let's return to the previous example and regress Brozek on BMI. To generate the elasticity, we first run the regression, and then follow that with the **margins** command:

```
. reg Brozek BMI

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(1, 250)       =    281.89
       Model |  7991.50988         1  7991.50988   Prob > F        =    0.0000
    Residual |  7087.50675       250   28.350027   R-squared       =    0.5300
-------------+----------------------------------   Adj R-squared   =    0.5281
       Total |  15079.0166       251  60.0757635   Root MSE        =    5.3245

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |   1.546712   .0921238    16.79   0.000     1.365275     1.72815
       _cons |  -20.40508   2.367227    -8.62   0.000    -25.06733   -15.74283
------------------------------------------------------------------------------
```

```
. margins, eyex(_all) atmeans

Conditional marginal effects                    Number of obs     =        252
Model VCE    : OLS

Expression   : Linear prediction, predict()
ey/ex w.r.t. : BMI
at           : BMI             =     25.4369 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      ey/ex   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         BMI |    2.07744   .1290888    16.09   0.000     1.823199     2.33168
------------------------------------------------------------------------------
```

47. While you can evaluate the elasticity (**eyex** in the syntax) at different points, the **atmeans** options will generate the elasticity at the means… which is 2.08. Most would say that an elasticity of that magnitude (suggesting that a 10% increase in BMI s associated with a 21% increase in predicted Brozek) is highly meaningful.

48. While there's no official border separating elasticities for meaningful effects from those that are not so meaningful, I think it's fair to say that everyone agrees that elasticities above 1 (and even .5) in magnitude suggest a meaningful effect, and those below .05 might suggest a not so meaningful estimated relationship. If I had to pick a zone of indifference, I'd say that it might be in the neighborhood of .1 … but this is clearly a judgement call.

49. Often elasticities are so small or so large, no one needs to worry about picking a dividing line for meaningfulness. But unfortunately, that is not always the case… in which case reasonable people may disagree.

Beta Regressions v. Elasticities:

50. In our bodyfat example, the beta regression and elasticities approaches both suggest that there is a highly meaningful and positive relationship between BMI and Brozek. I should warn you though that while these two approaches almost always lead to consistent interpretations, that won't always happen.

51. In that case, you can throw your hands in the air… or maybe just fall back on the *eyeball* and *laughability* tests. Do your critics laugh at you when you claim to have found a meaningful effect? Or maybe they agree with you, even though no one agrees on exactly how to define meaningfulness.

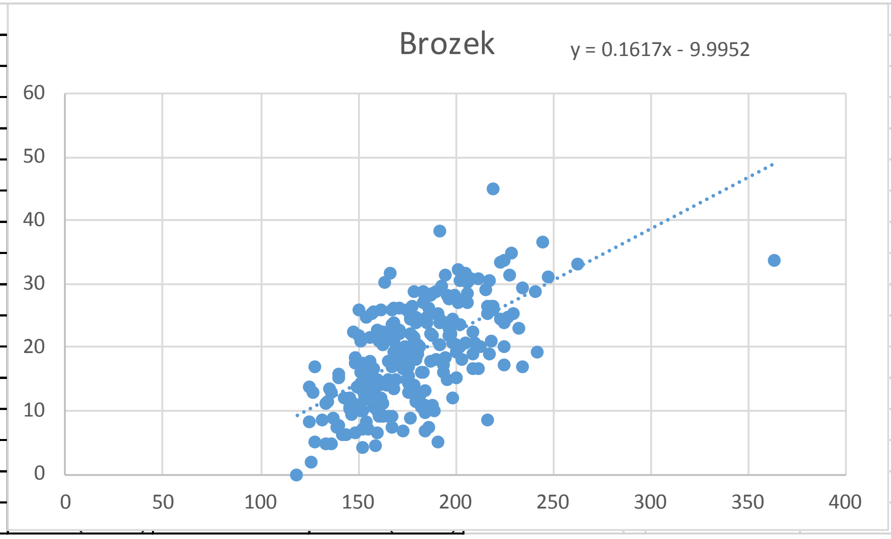
OLS/SLR Analytics: ***Estimating Parameters***

Examples in Excel and Stata

Let's first do this in Excel.

Open the bodyfat dataset in Excel. Generate the x-y scatterplot of Brozek v. wgt, and "add trendline". You should see something like:

Case	wgt	Brozek
1	154.25	12.6
2	173.25	6.9
3	154	24.6
4	184.75	10.9
5	184.25	27.8
6	210.25	20.6
7	181	19
8	176	12.8
9	191	5.1
10	198.25	12
11	186.25	7.5
12	216	8.5
13	180.5	20.5
14	205.25	20.8
15	187.75	21.7
16	162.75	20.5



Trendline fits a straight line to the data… and that straight line is in fact generated by the OLS intercept and slope coefficients!

Excel Trendline = OLS/SLR

For Brozek and wgt, compute sample means, variances, standard deviations, as well as the covariance and correlation, and apply the various formulae for the OLS slope and intercept estimates. You should get something like:

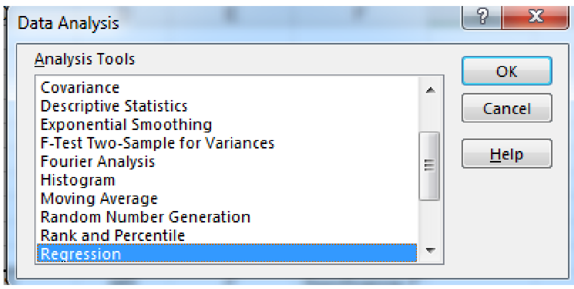
			Sample Variances		**Sample Cov**	**Sample Corr**	**Slope estimates**	
			863.72	60.08	139.67	0.6132	Sxy/Sxx	0.1617
		StDevs	29.39	7.75			corr*(Sy/Sy)	0.1617
			Sum Squares		**Sum**		**Intercept estimate**	
Means	178.924	18.938	216,794.40	15,079.02	35,057.55		Bbar-b1*wbar	(9.9952)
Case	**wgt**	**Brozek**	**wgt-wbar**	**Brozek-Bbar**	**product**			
1	154.25	12.6	(24.67)	(6.34)	156.40			
2	173.25	6.9	(5.67)	(12.04)	68.31			
3	154	24.6	(24.92)	5.66	(141.11)			
4	184.75	10.9	5.83	(8.04)	(46.83)			
5	184.25	27.8	5.33	8.86	47.19			

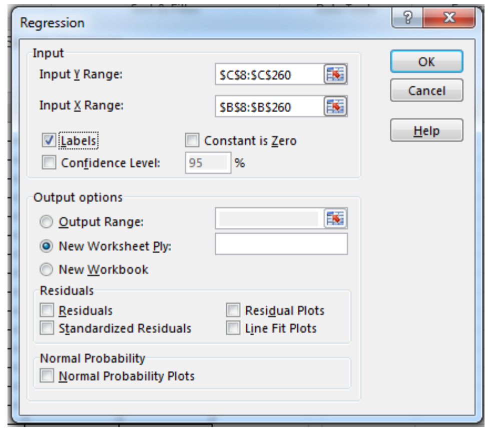
So who knew? The Excel Trendline is generated by OLS!

OLS/SLR Analytics: *Estimating Parameters*

Running regressions in Excel

You can also run the OLS regression in Excel using Data/Data Analysis/Regression (you may have to load the Data Analysis Tool-Pak (go to Options/Add-Ins):



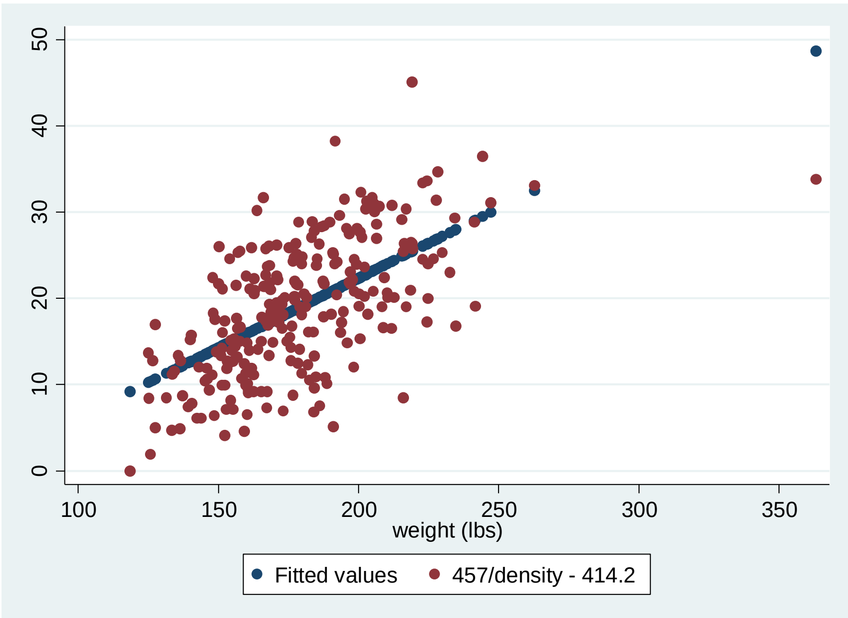


SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.61316					
R Square	0.37596					
Adjusted R Square	0.37346					
Standard Error	6.13511					
Observations	252					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	5,669.11	5,669.11	150.62	2.05905E-27	
Residual	250	9,409.90	37.64			
Total	251	15,079.02				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	(9.9952)	2.3891	(4.18)	3.97276E-05	(14.7004)	(5.2899)
wgt	0.1617	0.0132	12.27	2.05905E-27	0.1358	0.1877

Same OLS slope and intercept!

Now for Stata.

```
. bcuse bodyfat

Contains data from http://fmwww.bc.edu/ec-p/data/wooldridge/bodyfat.dta
  obs:           252
 vars:            24
 size:        40,068
------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
Case            int     %10.0g                Case
Brozek          double  %10.0g                457/density - 414.2
wgt             double  %10.0g                weight (lbs)
------------------------------------------------------------------------------
Sorted by:

. reg Brozek wgt

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(1, 250)       =    150.62
       Model |  5669.11335         1  5669.11335   Prob > F        =    0.0000
    Residual |  9409.90327       250  37.6396131   R-squared       =    0.3760
-------------+----------------------------------   Adj R-squared   =    0.3735
       Total |  15079.0166       251  60.0757635   Root MSE        =    6.1351

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   .1617088   .0131765    12.27   0.000     .1357578    .1876598
       _cons |  -9.995151   2.389056    -4.18   0.000    -14.70039   -5.289908
------------------------------------------------------------------------------

. predict bhat
. scatter bhat Brozek wgt
```



OLS/SLR Analytics: ***Estimating Parameters***

Use the **summarize**, **correlation** and **display** commands to generate the OLS slope and intercept estimates:

```
. summ Brozek wgt

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      Brozek |        252    18.93849    7.750856          0       45.1
         wgt |        252    178.9244    29.38916      118.5     363.15

. corr Brozek wgt, covar

             |   Brozek      wgt
-------------+------------------
      Brozek |  60.0758
         wgt |  139.672  863.723
```

slope coefficient 1: ratio of sample covariance to sample variance

```
. di 139.672 / 863.723
.16170925
```

intercept estimate:

```
. di 18.93849 - .16170925 *  178.9244
-9.9952405

. corr Brozek wgt

             |   Brozek      wgt
-------------+------------------
      Brozek |   1.0000
         wgt |   0.6132   1.0000
```

slope coefficient 2: **(**sample corr) (ratio of sample standard deviations)

```
. di 0.6132 * 7.750856 / 29.38916
.16172034
```

Verify that the correlation of *Brozek* with *wgt* is the same as the correlation of *Brozek* with *bhat*:

```
. corr Brozek bhat wgt

             |   Brozek     bhat      wgt
-------------+---------------------------
      Brozek |   1.0000
        bhat |   0.6132   1.0000
         wgt |   0.6132   1.0000   1.0000
```

OLS/SLR Analytics: ***Estimating Parameters***

Capture the *residuals* and verify that they are uncorrelated with the *predicteds* (*bhats*) and as well with the explanatory variable (*wgt*)

```
. predict resids, res

. corr bhat wgt resids

             |     bhat      wgt   resids
-------------+---------------------------
        bhat |   1.0000
         wgt |   1.0000   1.0000
      resids |  -0.0000  -0.0000   1.0000
```

Elasticity at the means.

Evaluate the elasticity associated with the estimated OLS coefficients:

```
. di .1617088*178.9244/ 18.93849
1.5277696
```

Or just run the ***margins*** command right after the ***reg*** command

```
. reg Brozek wgt
. margins, eyex(_all) atmeans

Conditional marginal effects                      Number of obs     =        252
Model VCE    : OLS

Expression   : Linear prediction, predict()
ey/ex w.r.t. : wgt
at           : wgt             =    178.9244 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      ey/ex   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   1.527769   .1283313    11.90   0.000     1.275021    1.780517
------------------------------------------------------------------------------
```

And how about a Beta Regression?

```
.  reg Brozek wgt, beta

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(1, 250)       =    150.62
       Model |  5669.11335         1  5669.11335   Prob > F        =    0.0000
    Residual |  9409.90327       250  37.6396131   R-squared       =    0.3760
-------------+----------------------------------   Adj R-squared   =    0.3735
       Total |  15079.0166       251  60.0757635   Root MSE        =    6.1351

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
         wgt |   .1617088   .0131765    12.27   0.000                 .6131561
       _cons |  -9.995151   2.389056    -4.18   0.000                        .
------------------------------------------------------------------------------
```

OLS/SLR Analytics: *Estimating Parameters*

As you saw before, the beta regression slope coefficient is just the correlation between Brozek and wgt. The following shows that the Beta regression is as advertised… it's what you get when you first standardize your variables before running OLS. You'll see that you can use use **egen** and the **std(.)** function to easily standardize your variables:

```
. egen zBrozek=std(Brozek)
. egen zwgt=std(wgt)
. reg zBrozek zwgt

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(1, 250)       =    150.62
       Model |  94.3660642         1  94.3660642   Prob > F        =    0.0000
    Residual |  156.633936       250  .626535743   R-squared       =    0.3760
-------------+----------------------------------   Adj R-squared   =    0.3735
       Total |         251       251  .999999999   Root MSE        =    .79154

------------------------------------------------------------------------------
     zBrozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        zwgt |   .6131561   .0499616    12.27   0.000     .5147569    .7115553
       _cons |  -1.39e-09   .0498623    -0.00   1.000    -.0982038    .0982038
------------------------------------------------------------------------------
```

Finally: About that sensitivity to scale… Here are some regression results, with height in feet, inches, meters and centimeters (notice the use of **eststo** and **esttab** to compile the results).

```
. gen hgt_ft = hgt/12
. gen hgt_cm = hgt_m*100
. reg Brozek hgt_ft
. eststo
. reg Brozek hgt
. eststo
. reg Brozek hgt_m
. eststo
. reg Brozek hgt_cm
. eststo
. esttab

                      (1)             (2)             (3)             (4)
                   Brozek          Brozek          Brozek          Brozek
                   (feet)        (inches)        (meters)   (centimeters)
----------------------------------------------------------------------------
hgt_ft             -2.263
                  (-1.41)

hgt                                -0.189
                                  (-1.41)

hgt_m                                              -7.423
                                                  (-1.41)

hgt_cm                                                            -0.0742
                                                                  (-1.41)

_cons               32.17***        32.17***        32.17***        32.17***
                   (3.44)          (3.44)          (3.44)          (3.44)
----------------------------------------------------------------------------
N                     252             252             252             252
----------------------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

As expected, the slope coefficients in (1) and (2) differ by a factor of 12, and those in Models (3) and (4) differ by a factor of 100. And the intercepts are unaffected by the changes in scale of the RHS variable.

And if we put the different height variables on the LHS, the slope and intercept coefficients will reflect the differing units.

```
. qui: reg hgt_ft wgt
. eststo
. qui: reg hgt wgt
. eststo
. qui: reg hgt_m wgt
. eststo
. qui: reg hgt_cm  wgt
. eststo
. esttab

----------------------------------------------------------------------------
                      (1)             (2)             (3)             (4)   
                   hgt_ft             hgt           hgt_m          hgt_cm   
                   (feet)        (inches)        (meters)   (centimeters)
----------------------------------------------------------------------------
wgt               0.00320***       0.0384***     0.000976***       0.0976***
                   (5.12)          (5.12)          (5.12)          (5.12)   

_cons               5.273***        63.27***        1.607***        160.7***
                  (46.54)         (46.54)         (46.54)         (46.54)   
----------------------------------------------------------------------------
N                     252             252             252             252   
----------------------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

As expected, the slope and intercept coefficients in (1) and (2) each differ by a factor of 12, and further, those in Models (3) and (4) each also differ by a factor of 100.

Appendix: A Simple Derivation of those OLS Coefficients

We want to min $SSR = \sum\left[y_i - \left(b_0 + b_1 x_i\right)\right]^2$ wrt b_0 *and* b_1.

Define $\delta = \bar{y} - b_0 - b_1\bar{x}$.

Then if we add and subtract δ inside the square brackets in the SSR expression we have:

$$SSR = \sum\left[y_i - \delta + \delta - \left(b_0 + b_1 x_i\right)\right]^2 = \sum\left[\left(y_i - \bar{y}\right) + \delta - b_1\left(x_i - \bar{x}\right)\right]^2,$$

which can be simplified to

$$\begin{aligned} SSR &= n\delta^2 + \sum\left(y_i - \bar{y}\right)^2 - 2b_1\sum\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right) + b_1^2\sum\left(x_i - \bar{x}\right)^2 \\ &= (n-1)\left[\frac{n}{n-1}\delta^2 + S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx}\right]\ldots \end{aligned}$$

which we want to minimize wrt δ *and* b_1.

Since $\delta^2 \geq 0$ for any b_0 *and* b_1, we minimize SSRs with $\delta = \bar{y} - b_0 - b_1\bar{x} = 0$. So

$$b_0^* = \bar{y} - b_1^*\bar{x} \ldots \text{ (which should look very familiar by now!)}$$

And to minimize the rest of the expression that varies with b_1, $\left[-2b_1 S_{xy} + b_1^2 S_{xx}\right]$, just use a FOC:

$$\frac{dSSR}{db_1} = (n-1)\left[-2S_{xy} + 2b_1 S_{xx}\right] = 0, \text{ or } b_1^* = S_{xy} / S_{xx}.$$

OLS coefficients!