POLS 6481. Research Design and Quantitative Methods II

Lecture 9. Rescaling~~, Partialling, and Mediation~~

Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

Outline:

1. Effects of Rescaling a Regressor (x)
2. Effects of Rescaling the Regressand (y)
3. Beta Coefficients

1½, 2½ and 3½. Example: *The NYC Marathon*

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. Rescaling …
Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

1. Effects of Rescaling a Regressor

Suppose I double the value of x… let $X = 2x$; then:

Effects on $\widehat{\boldsymbol{\beta}_1}$

$cov(X, y) = 2 \times cov(x, y)$ each "rectangle" has double the area

$var(X) = 4 \times var(x)$ each "square" has quadruple the area

The ratio $\frac{cov(X,y)}{var(X)} = \frac{1}{2} \cdot \frac{cov(x,y)}{var(x)}$: twice as much ΔX is needed to yield the same Δy

Effects on $\boldsymbol{se(\widehat{\beta_1})}$

The residual standard error ($\hat{\sigma}$) will be unchanged

The standard deviation of X doubles: $s_x = 2 \times s_x$, and therefore $\sqrt{n} \cdot s_x$ will double

The ratio $\frac{\hat{\sigma}}{\sqrt{n} \cdot s_X} = \frac{1}{2} \cdot \frac{\hat{\sigma}}{\sqrt{n} \cdot s_x}$

Effects on $\boldsymbol{t}$ statistic

The ratio $\frac{\widehat{\beta_1}}{se(\widehat{\beta_1})}$ will be unchanged

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. Rescaling …
Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

1½. Example: *The NYC Marathon*

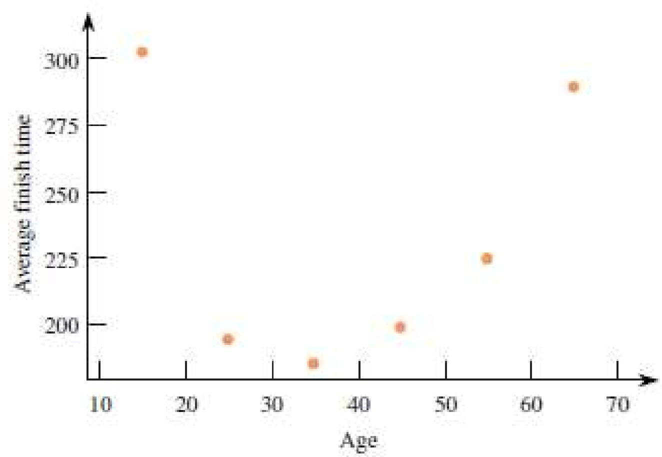
To investigate the effects of age on marathon finish times, the article "Master's Performance in the New York City Marathon" (*British Journal of Sports Medicine* [2004: 408-412]) gave the following data on average finishing time by age group for female participants in 1999:

The dependent variable is *y*, *Average.finish.time* (in minutes) by runners in 10-year age group

The independent variable is *x*, *Representative.age* for generic member of 10-year age group

The dataset is *NYCmarathon.csv*. The R script is *Lecture 9 NYCmarathon.R*.

Age Group	Representative Age	Average Finish Time
10–19	15	302.38
20–29	25	193.63
30–39	35	185.46
40–49	45	198.49
50–59	55	224.30
60–69	65	288.71



POLS 6481. Research Design and Quantitative Methods II
Lecture 9. Rescaling …
Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

1½. Example: *The NYC Marathon*

To investigate the effects of age on marathon finish times, the article "Master's Performance in the New York City Marathon" (*British Journal of Sports Medicine* [2004: 408-412]) gave the following data on average finishing time by age group for female participants in 1999:

The dependent variable is *y*, *Average.finish.time* (in minutes) by runners in 10-year age group

The independent variable is *x*, *Representative.age* for generic member of 10-year age group

The dataset is *NYCmarathon.csv*. The R script is *Lecture 9 scaling variables.R*.

The simple correlation of *x* and *y* is .04 including the 10–19 age group;
the simple correlation of *x* and *y* is .86 excluding the 10–19 age group.

A simple regression of *y* on *x* yields $\widetilde{\beta_1}$ = 2.29 & $se\left(\widetilde{\beta_1}\right)$ = .781 $\rightarrow t$ = 2.933

Suppose we transform the independent variable from years to decades, so that a 1-unit change in *x* refers to being a decade older.

A simple regression of *y* on $\frac{x}{10}$ yields $\widetilde{\beta_1}$ = 22.9 & $se\left(\widetilde{\beta_1}\right)$ = 7.81 $\rightarrow t$ = 2.933

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. Rescaling …
Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

2. Effects of Rescaling the Regressand

Suppose I double the value of y … let $Y = 2y$; then:

Effects on $\widehat{\beta_1}$

$cov(x, Y) = 2 \times cov(x, y)$

$var(x)$ will be unchanged

The ratio $\frac{cov(x,Y)}{var(x)} = 2 \times \frac{cov(x,y)}{var(x)}$, implying the same Δx yields twice as much ΔY

Effects on $se(\widehat{\beta_1})$

The Residual standard error ($\hat{\sigma}$) will double

The standard deviation of x (s_x) will be unchanged, and therefore $\sqrt{n} \cdot s_x$ will be unchanged

The ratio $\frac{\hat{\sigma}}{\sqrt{n} \cdot s_x}$ will be doubled

Effects on t statistic

The ratio $\frac{\widehat{\beta_1}}{se(\widehat{\beta_1})}$ will be unchanged

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. Rescaling …
Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

2½. Example: *The NYC Marathon*

To investigate the effects of age on marathon finish times, the article "Master's Performance in the New York City Marathon" (*British Journal of Sports Medicine* [2004: 408-412]) gave the following data on average finishing time by age group for female participants in 1999:

The dependent variable is *y*, *Average.finish.time* (in minutes) by runners in 10-year age group

The independent variable is *x*, *Representative.age* for generic member of 10-year age group

The dataset is *NYCmarathon.csv*. The R script is *Lecture 9 scaling variables.R*.

The simple correlation of *x* and *y* is .04 including the 10–19 age group;
the simple correlation of *x* and *y* is .86 excluding the 10–19 age group.

A simple regression of *y* on *x* yields $\widetilde{\beta_1} = 2.29$ & $se\left(\widetilde{\beta_1}\right) = .781$ $\rightarrow t = 2.933$

Suppose we transform the dependent variable from minutes to hours, so that a 1-unit change in *y* refers to the race taking an hour longer.

A simple regression of $\frac{y}{60}$ on *x* yields $\widetilde{\beta_1} = .038$ & $se\left(\widetilde{\beta_1}\right) = .013$ $\rightarrow t = 2.933$

POLS 6481. Research Design and Quantitative Methods II
Lecture 9. Rescaling …
Readings: Wooldridge, *Introductory Econometrics 5e,* 2.4a + 6.1

3. Standardized Coefficients

Regression coefficients indicate how *y* changes in response to a 1–unit change in *x*:

- sign indicates whether *y* increases (+) or decreases (–);
- magnitude indicates by how many units *y* changes.

Standardized coefficients indicate how *y* changes in response to a 1–*standard deviation* change in *x*:

- sign indicates whether *y* increases (+) or decreases (–);
- magnitude indicates by how many *standard deviations* *y* changes.

Characteristics of standardized coefficients:

1. in a simple regression, beta equals the correlation coefficient ($\widehat{beta}_j = \hat{\beta}_j \cdot \frac{s_y}{s_{x_j}}$)

2. just like correlation coefficients, betas will be between –1 and +1

Approach A. Standardize all variables prior to running model.
can use **Make.Z** function in **QuantPsyc** package

Approach B. Modify results afterward using **lm.beta*(model)*** function in **QuantPsyc** package