Lecture 24: Partial correlation, multiple regression, and correlation

Ernesto F. L. Amaral

November 21, 2017
Advanced Methods of Social Research (SOCI 420)

Source: Healey, Joseph F. 2015. ”Statistics: A Tool for Social Research.” Stamford: Cengage Learning. 10th edition. Chapter 15 (pp. 405–441).

TEXAS A&M UNIVERSITY.

Chapter learning objectives

- Compute and interpret partial correlation coefficients
- Find and interpret the least-squares multiple regression equation with partial slopes
- Find and interpret standardized partial slopes or beta-weights (b^*)
- Calculate and interpret the coefficient of multiple determination (R^2)
- Explain the limitations of partial and regression analysis

Multiple regression

- Discuss ordinary least squares (OLS) multiple regressions
  - OLS: linear regression
  - Multiple: at least two independent variables
- Disentangle and examine the separate effects of the independent variables
- Use all of the independent variables to predict Y
- Assess the combined effects of the independent variables on Y

Partial correlation

- Partial correlation measures the correlation between *X* and *Y*, controlling for *Z*

- Comparing the bivariate (zero-order) correlation to the partial (first-order) correlation
  - Allows us to determine if the relationship between *X* and *Y* is direct, spurious, or intervening
  - Interaction cannot be determined with partial correlations

Formula for partial correlation

- Formula for partial correlation coefficient for *X* and *Y*, controlling for *Z*

$$r_{yx.z} = \frac{r_{yx} - (r_{yz})(r_{xz})}{\sqrt{1 - r_{yz}^2}\sqrt{1 - r_{xz}^2}}$$

- We must first calculate the zero-order coefficients between all possible pairs of variables (*Y* and *X*, *Y* and *Z*, *X* and *Z*) before solving this formula

Example

- Husbands' hours of housework per week (*Y*)
- Number of children (*X*)
- Husbands' years of education (*Z*)

Scores on Three Variables for 12 Dual–Wage-Earner Families

Family	Husband's Housework (*Y*)	Number of Children (*X*)	Husband's Years of Education (*Z*)
A	1	1	12
B	2	1	14
C	3	1	16
D	5	1	16
E	3	2	18
F	1	2	16
G	5	3	12
H	0	3	12
I	6	4	10
J	3	4	12
K	7	5	10
L	4	5	16

Source: Healey 2015, p.409.

Correlation matrix

- The bivariate (zero-order) correlation between husbands' housework and number of children is +0.50
  - This indicates a positive relationship

Zero-Order Correlations

↓	Husband's Housework (Y)	Number of Children (X)	Husband's Years of Education (Z)
Husband's Housework (Y)	1.00	0.50	−0.30
Number of Children (X)		1.00	−0.47
Husband's Years of Education (Z)			1.00

Source: Healey 2015, p.410.

First-order correlation

- Calculate the partial (first-order) correlation between husbands' housework (*Y*) and number of children (*X*), controlling for husbands' years of education (*Z*)

$$r_{yx.z} = \frac{r_{yx} - (r_{yz})(r_{xz})}{\sqrt{1 - r_{yz}^2}\sqrt{1 - r_{xz}^2}}$$

$$r_{yx.z} = \frac{(0.50) - (-0.30)(-0.47)}{\sqrt{1 - (-0.30)^2}\sqrt{1 - (-0.47)^2}}$$

$$r_{yx.z} = 0.43$$

Interpretation

- Comparing the bivariate correlation (+0.50) to the partial correlation (+0.43) finds little change
- The relationship between number of children and husbands' housework has not changed, controlling for husbands' education
- Therefore, we have evidence of a direct relationship

Bivariate & multiple regressions

- Bivariate regression equation

$$Y = a + bX = \beta_0 + \beta_1 X$$

  - $a = \beta_0 = Y$ intercept
  - $b = \beta_1 =$ slope

- Multivariate regression equation

$$Y = a + b_1 X_1 + b_2 X_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

  - $b_1 = \beta_1 =$ partial slope of the linear relationship between the first independent variable and Y
  - $b_2 = \beta_1 =$ partial slope of the linear relationship between the second independent variable and Y

Multiple regression

$$Y = a + b_1X_1 + b_2X_2 = \beta_0 + \beta_1X_1 + \beta_2X_2$$

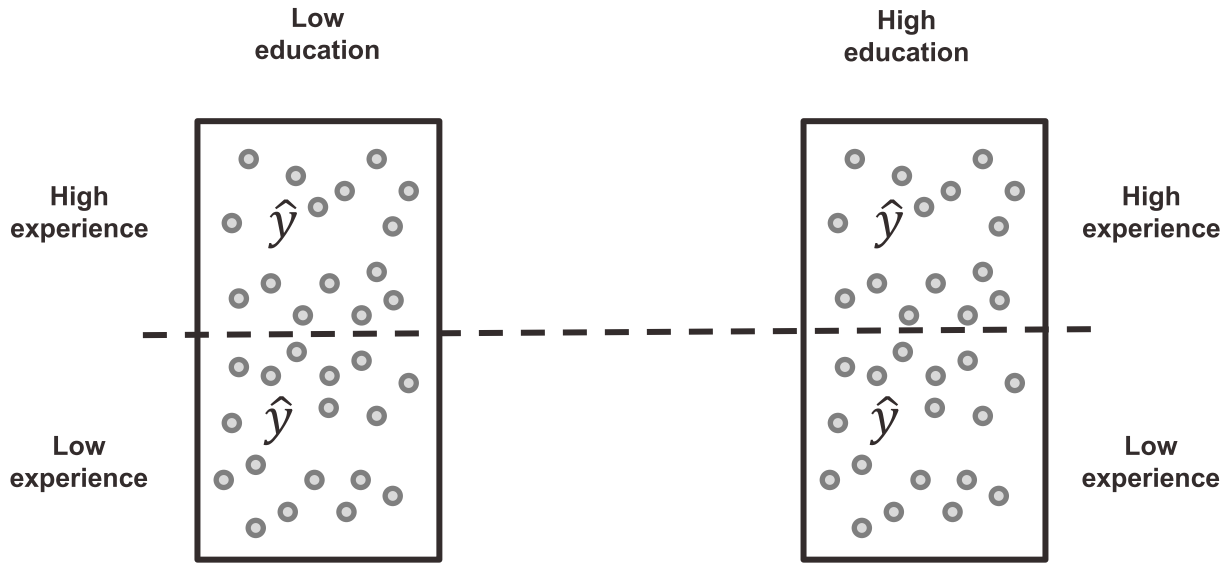
- $a = \beta_0$ = the Y intercept, where the regression line crosses the Y axis
- $b_1 = \beta_1$ = partial slope for X_1 on Y
  - β_1 indicates the change in Y for one unit change in X_1, controlling for X_2
- $b_2 = \beta_2$ = partial slope for X_2 on Y
  - β_2 indicates the change in Y for one unit change in X_2, controlling for X_1

Partial slopes

- The partial slopes indicate the effect of each independent variable on Y
- While controlling for the effect of the other independent variables
- This control is called *ceteris paribus*
  - Other things equal
  - Other things held constant
  - All other things being equal

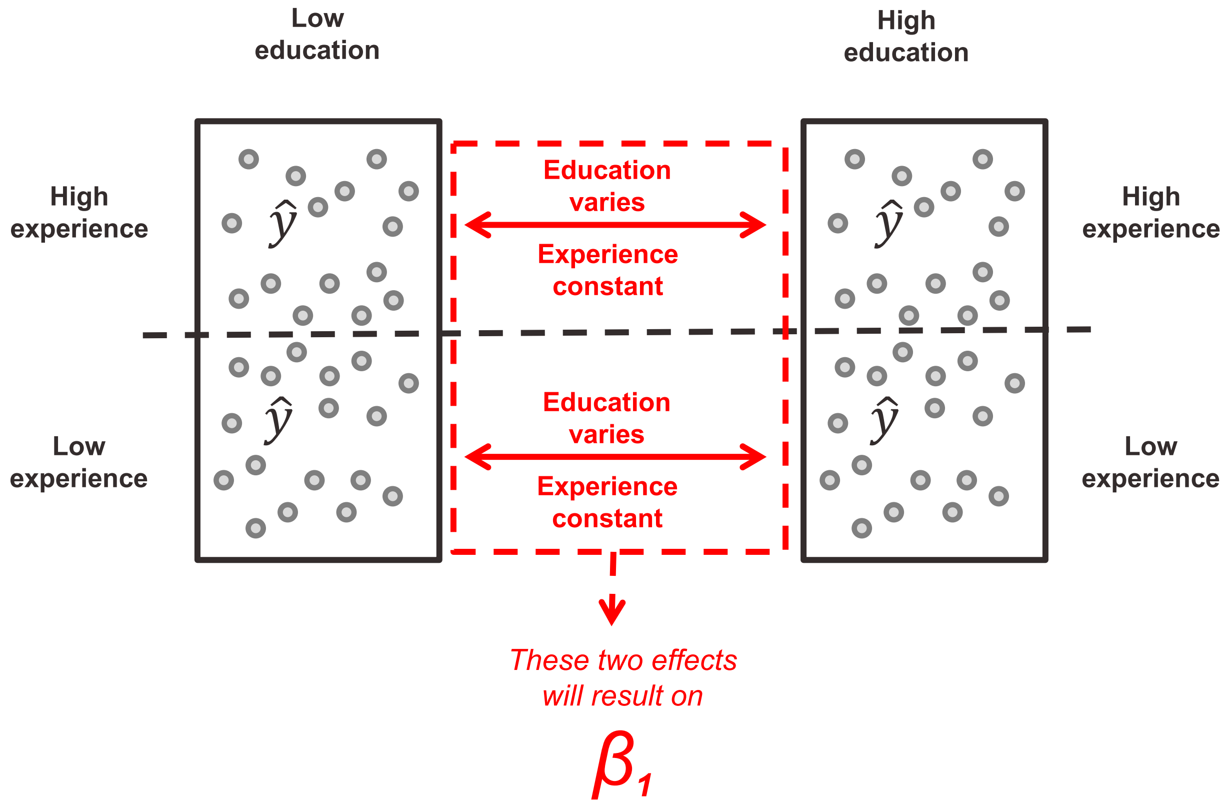
Ceteris paribus

$$Income = \beta_0 + \beta_1 education + \beta_2 experience + u$$



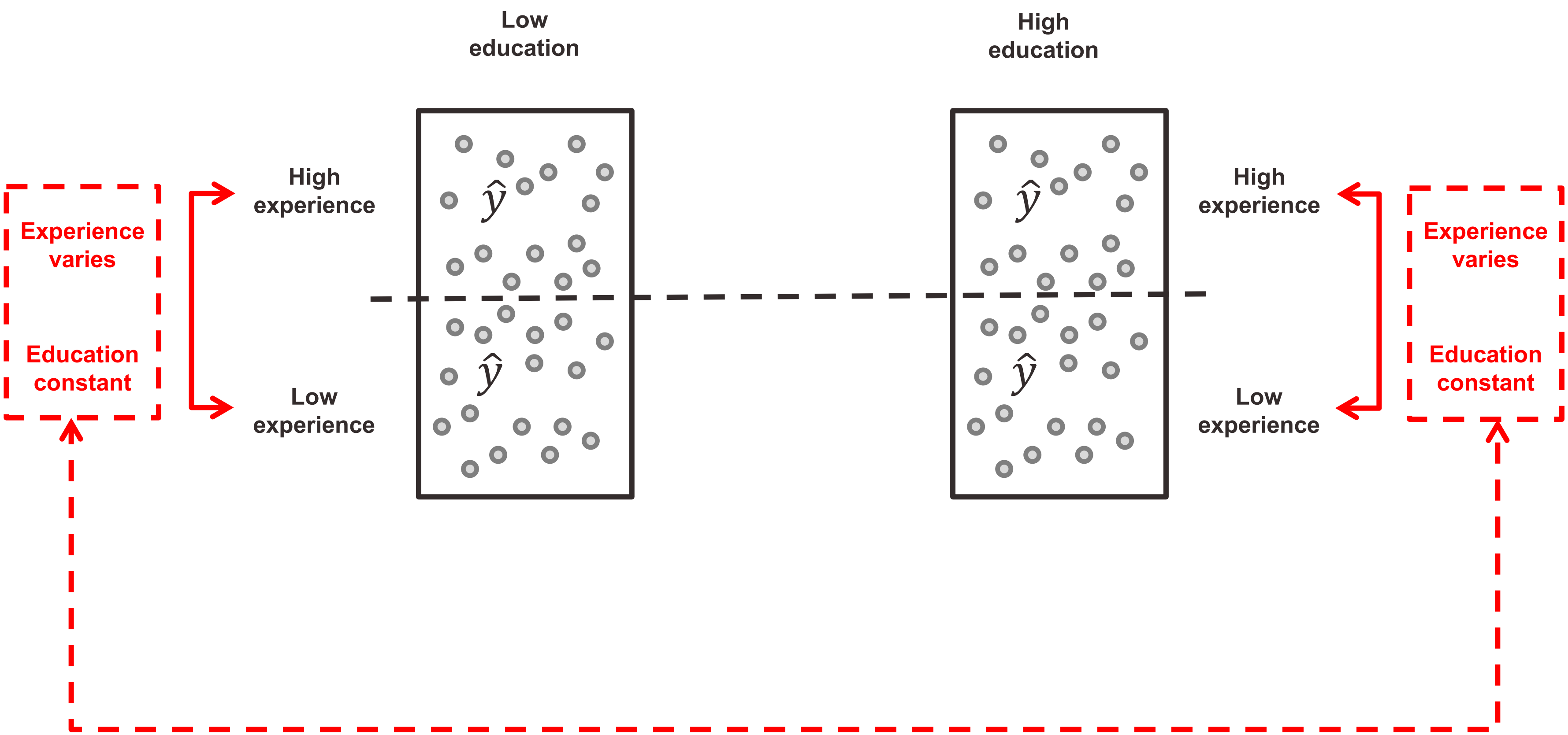
Ceteris paribus

$$Income = \beta_0 + \beta_1 education + \beta_2 experience + u$$



Ceteris paribus

$$Income = \beta_0 + \beta_1 education + \beta_2 experience + u$$

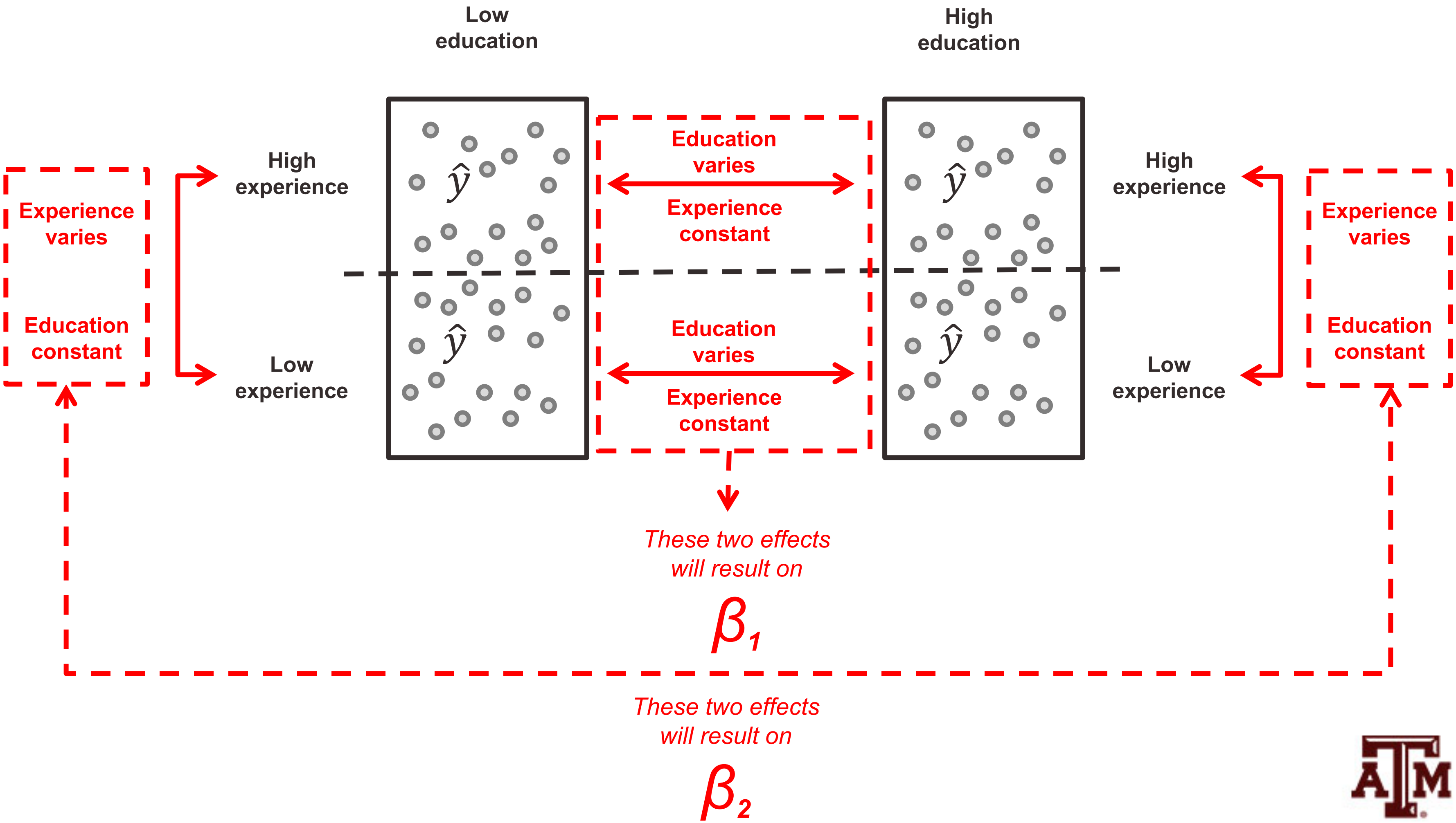


These two effects will result on

β_2

Ceteris paribus

$$Income = \beta_0 + \beta_1 education + \beta_2 experience + u$$



Interpretation of partial slopes

- The partial slopes show the effects of the *X*'s in their original units
- These values can be used to predict scores on *Y*
- Partial slopes must be computed before computing the *Y* intercept (β_0)

Formulas of partial slopes

$$b_1 = \beta_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

$$b_2 = \beta_2 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}\right)$$

$b_1 = \beta_1$ = partial slope of X_1 on Y

$b_2 = \beta_2$ = partial slope of X_2 on Y

s_y = standard deviation of Y

s_1 = standard deviation of the first independent variable (X_1)

s_2 = standard deviation of the second independent variable (X_2)

r_{y1} = bivariate correlation between Y and X_1

r_{y2} = bivariate correlation between Y and X_2

r_{12} = bivariate correlation between X_1 and X_2

Formula of constant

- Once b_1 (β_1) and b_2 (β_2) have been calculated, use those values to calculate the Y intercept

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2$$

Example

- Using information below, calculate the slopes

Husband's Housework	Number of Children	Husband's Education
$\bar{Y} = 3.3$	$\bar{X}_1 = 2.7$	$\bar{X}_2 = 13.7$
$s_y = 2.1$	$s_1 = 1.5$	$s_2 = 2.6$

Zero-Order Correlations

$$r_{y1} = 0.50$$
$$r_{y2} = -0.30$$
$$r_{12} = -0.47$$

Source: Healey 2015, p.414.

Result and interpretation of b_1

$$b_1 = \beta_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

$$b_1 = \beta_1 = \left(\frac{2.1}{1.5}\right)\left(\frac{0.50 - (-0.30)(-0.47)}{1 - (-0.47)^2}\right) = 0.65$$

- As the number of children in a dual-career household increases by one, the husband's hours of housework per week increases on average by 0.65 hours (about 39 minutes), controlling for husband's education

Result and interpretation of b_2

$$b_2 = \beta_2 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}\right)$$

$$b_2 = \beta_2 = \left(\frac{2.1}{2.6}\right)\left(\frac{-0.30 - (0.50)(-0.47)}{1 - (-0.47)^2}\right) = -0.07$$

- As the husband’s years of education increases by one year, the number of hours of housework per week decreases on average by 0.07 (about 4 minutes), controlling for the number of children

Result and interpretation of *a*

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2$$

$$a = \beta_0 = 3.3 - (0.65)(2.7) - (-0.07)13.7$$

$$a = \beta_0 = 2.5$$

- With zero children in the family and a husband with zero years of education, that husband is predicted to complete 2.5 hours of housework per week on average

Final regression equation

- In this example, this is the final regression equation

$$Y = a + b_1X_1 + b_2X_2$$

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2$$

$$Y = 2.5 + (0.65)X_1 + (-0.07)X_2$$

$$Y = 2.5 + 0.65X_1 - 0.07X_2$$

Prediction

- Use the regression equation to predict a husband's hours of housework per week when he has 11 years of schooling and the family has 4 children

$$Y' = 2.5 + 0.65X_1 - 0.07X_2$$

$$Y' = 2.5 + (0.65)(4) + (-0.07)(11)$$

$$Y' = 4.3$$

- Under these conditions, we would predict 4.3 hours of housework per week

Standardized coefficients (b^*)

- Partial slopes ($b_1=\beta_1$; $b_2=\beta_2$) are in the original units of the independent variables
  - This makes assessing relative effects of independent variables difficult when they have different units
  - It is easier to compare if we standardize to a common unit by converting to Z scores
- Compute beta-weights (b^*) to compare relative effects of the independent variables
  - Amount of change in the standardized scores of Y for a one-unit change in the standardized scores of each independent variable
    - While controlling for the effects of all other independent variables
  - They show the amount of change in standard deviations in Y for a change of one standard deviation in each X

Formulas

- Formulas for standardized coefficients

$$b_1^* = b_1\left(\frac{s_1}{s_y}\right) = \beta_1^* = \beta_1\left(\frac{s_1}{s_y}\right)$$

$$b_2^* = b_2\left(\frac{s_2}{s_y}\right) = \beta_2^* = \beta_2\left(\frac{s_2}{s_y}\right)$$

Example

- Which independent variable, number of children (X_1) or husband's education (X_2), has the stronger effect on husband's housework in dual-career families?

$$b_1^* = b_1 \left(\frac{s_1}{s_y}\right) = (0.65)\left(\frac{1.5}{2.1}\right) = 0.46$$

$$b_2^* = b_2 \left(\frac{s_2}{s_y}\right) = (-0.07)\left(\frac{2.6}{2.1}\right) = -0.09$$

  – The standardized coefficient for number of children (0.46) is greater in absolute value than the standardized coefficient for husband's education (–0.09)
  – Therefore, number of children has a stronger effect on husband's housework

Standardized coefficients

- Standardized regression equation

$$Z_y = a_z + b_1^* Z_1 + b_2^* Z_2$$

  – where *Z* indicates that all scores have been standardized to the normal curve

- The Y intercept will always equal zero once the equation is standardized

$$Z_y = b_1^* Z_1 + b_2^* Z_2$$

- For the previous example

$$Z_y = (0.46)Z_1 + (-0.09)Z_2$$

Multiple correlation

- The coefficient of multiple determination (R^2) measures how much of Y is explained by all of the X's combined

- R^2 measures the percentage of the variation in Y that is explained by all of the independent variables combined

- The coefficient of multiple determination is an indicator of the strength of the entire regression equation

$$R^2 = r_{y1}^2 + r_{y2.1}^2\left(1 - r_{y1}^2\right)$$

  - R^2 = coefficient of multiple determination
  - r_{y1}^2 = zero-order correlation between Y and X_1
  - $r_{y2.1}^2$ = partial correlation of Y and X_2, while controlling for X_1

Partial correlation of Y and X_2

- Before estimating R^2, we need to estimate the partial correlation of Y and X_2 ($r_{y2.1}$)

$$r_{y2.1} = \frac{r_{y2} - (r_{y1})(r_{12})}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{12}^2}}$$

- We need three correlations
  - Between X_1 and Y: 0.50
  - Between X_2 and Y: –0.30
  - Between X_1 and X_2: –0.47

$$r_{y2.1} = \frac{(-0.30) - (0.50)(-0.47)}{\sqrt{1 - (0.50)^2}\sqrt{1 - (-0.47)^2}}$$

$$r_{y2.1} = -0.08$$

Result and interpretation

- For this example, R^2 will tell us how much of husband's housework is explained by the combined effects of the number of children (X_1) and husband's education (X_2)

$$R^2 = r_{y1}^2 + r_{y2.1}^2\left(1 - r_{y1}^2\right)$$

$$R^2 = (0.50)^2 + (-0.08)^2(1 - 0.50^2)$$

$$R^2 = 0.255$$

- Number of children and husband's education explain 25.5% of the variation in husband's housework