Multiple Regression I: Basics

CLPS2908: Multivariate Statistical Techniques

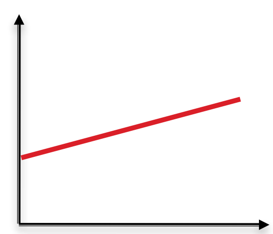
Feb 5, 2019

Beginnings

- **Situation**:
  Y_i = outcome/criterion/dependent variable (DV) $\quad i = 1 \ldots n$
  X_i = predictor/independent variable (IV) $\quad i = 1 \ldots n$
- Find a way to predict Y_i from X_i: $\boxed{Y_i = f(X_i)}$
- Different scales, imperfect measurement, but if at least some degree of linear relationship (correlation), then:

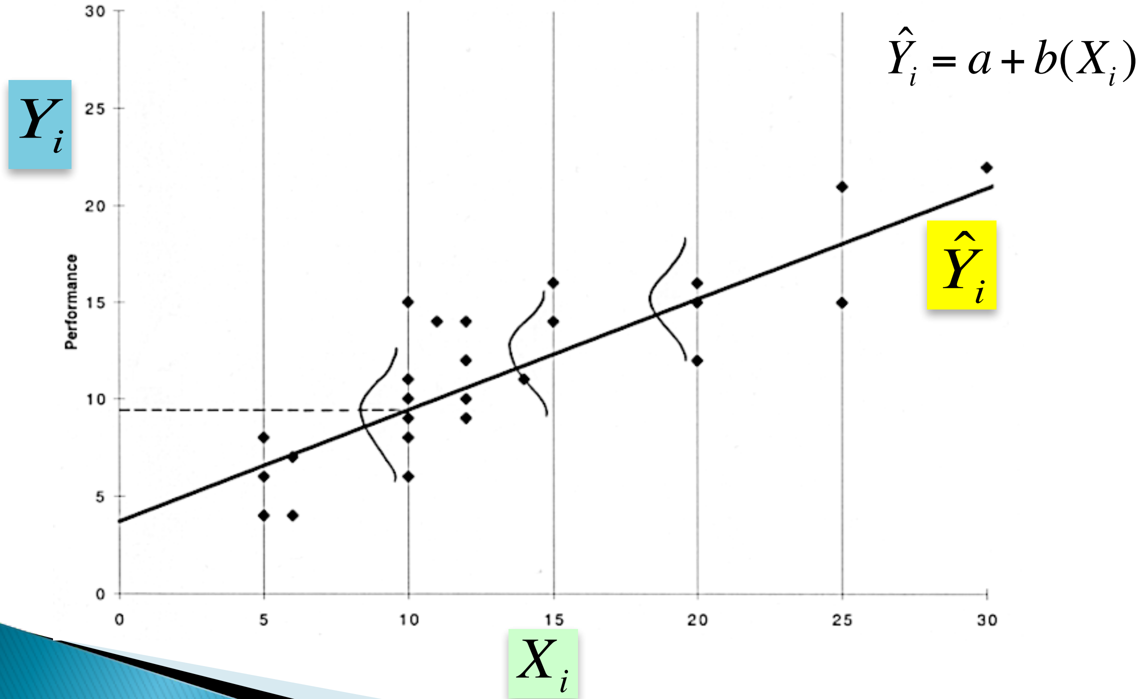
$$\boxed{f(X_i) = a + b(X_i) + \varepsilon_i = \widehat{Y}_i}$$

  - Find best b.
    - **Criterion** = minimize prediction errors
    - **Prediction error** = $\boxed{Y_i - \hat{Y}_i}$



a = intercept
b = slope

Prediction errors = Residuals



Variance Partition of Y_i

- $SS_{\text{total}} = SS_{\text{predicted}} + SS_{\text{residual}}$
- $SS_{\text{total}} = \sum (Y_i - \bar{Y})^2$
- $SS_{\text{predicted}} = \sum (\hat{Y}_i - \bar{Y})^2$
- $SS_{\text{residual}} = \sum (Y_i - \hat{Y}_i)^2$

$$R^2 = \frac{SS_{\text{pred}}}{SS_{\text{total}}}$$

- Maximize prediction portion by minimizing residuals (error)

From Univariate to Multivariate

$Y_i = a + b(X_i) + \varepsilon$ single X (= univariate)

$\mathbf{y} = \mathbf{X} \cdot \beta + \mathbf{e}$ multiple X_{ij} $(j = 1 \ldots p)$, **still single** Y_i

$n \times 1$ $n \times p$ $p \times 1$ $n \times 1$ (p + 1 instead if you have intercept)

- Goal: minimize residual squared error (**e'e**):

$$\boxed{\mathbf{e'e} \rightarrow \min;\ (\mathbf{y} - \mathbf{X}\beta)^2 \rightarrow \min}$$

Solved by : Why?

$$\mathbf{X'X}\beta = \mathbf{X'y}$$

$$\beta = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

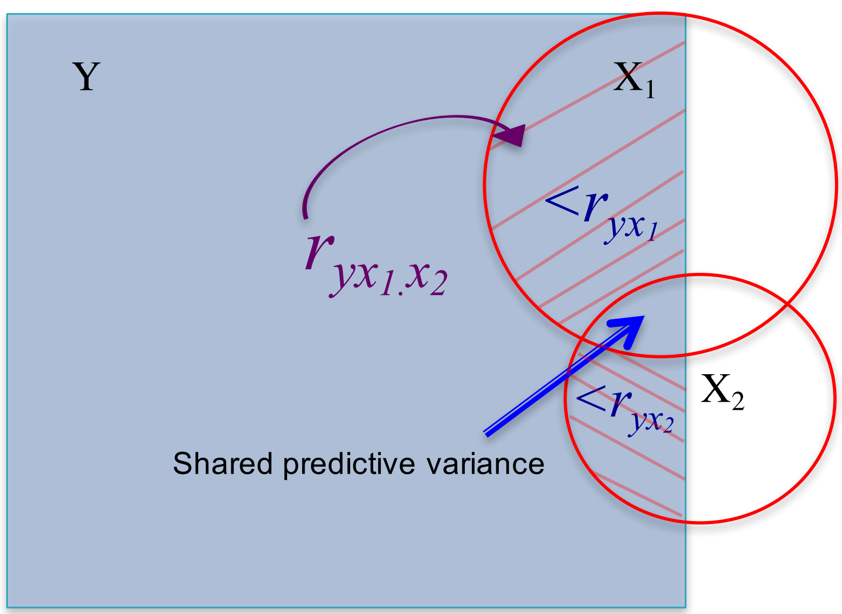
Correlation matrices: Xs among each other, Xs and Y

Deriving Multivariate Equation

$\mathbf{X'e} = 0$	(1) Assumption in the least-squres solution
$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$	(2) The definition of error (residuals)
$\hat{\mathbf{Y}} = \mathbf{Xb}$	(3) The regression equation
$\mathbf{e} = \mathbf{Y} - \mathbf{Xb}$	(4) by substituting $\hat{\mathbf{Y}}$ from (3) in (2)
$\mathbf{X'(Y - Xb) = 0}$	(5) By substituting **e** from (4) in (1)
$\mathbf{X'Y - X'Xb = 0}$	(6) By multiplying out (5)
$\mathbf{X'Y = X'Xb}$	(7) Rearranging (6)
$\mathbf{(X'X)^{-1}X'Y = (X'X)^{-1}X'Xb}$	(8) Dividing both sides by **X'X**
$\mathbf{(X'X)^{-1}X'Y = b}$	(9) Solving for **b**

Major Complication

- **Predictors are correlated**
  - In the equation, each variable's contribution (its ***β* weight**) is <u>adjusted for</u> other variables' contributions (their β weights); β weights are mutually dependent and are "semi-partial."
  - Each predictor's **zero-order correlation** with Y is <u>not adjusted</u> for any other variables; thus it will <u>differ from</u> its regression-based ("semi-partial") correlation.
  - **"Shared variance problem"**
    (= shared predictive variance)
    - Even greater challenge: When there is near collinearity; when there are **suppressor effects** → **Thursday**



~~*regression coefficients will often (although not always) change depending on the variables included in our regression equation.* This development certainly does not argue for the scientific respectability of our findings, however, nor does it bode well for the scientific respectability of multiple regression. If our conclusions change depending on the variables we include in our analyses, then knowledge and conclusions depend on our skill and honesty in selecting variables for analysis. Research findings should be more constant and less ephemeral if they are to form the basis for understanding, knowledge, and theory. Furthermore, this change in findings and conclusions means that, to some extent, we can find what we want by choosing the variables to include in our regression.~~

Keith, T. Z. (2006). *Multiple regression and beyond.* Boston, MA: Pearson.

Multiple Regression II: Expansions

CLPS2908: Multivariate Statistical Techniques

Feb 7, 2019

Expansions

- Methods of Entry
- Interactions
- Understand difference between zero-order and (semi-)partial correlations
- Clarify "suppressor" effects
  - Usual explanation; better approach
  - Weak and strong cases
    - SPSS output examples, Excel calculation file

Methods of Entry

- Backward, Stepwise as Model testing; ~~Forward~~
  - R^2 change (F test)

```
/DEPENDENT=varlist

[/METHOD=]{STEPWISE [varlist]    } [...] [/...]
          {FORWARD [varlist]     }
          {BACKWARD [varlist]    }
          {ENTER [varlist]       }
          {REMOVE varlist        }
          {TEST(varlist)(varlist)...}
```

```
REGRESSION VARIABLES=POP15,POP75,INCOME,GROWTH,SAVINGS
 /DEPENDENT=SAVINGS
 /METHOD=ENTER POP15,POP75,INCOME
 /METHOD=ENTER GROWTH.
```

When more than one METHOD subcommand is specified, each METHOD subcommand is applied to the equation that resulted from the **previous** METHOD subcommands.

Interactions (often “Moderators”)

- CASE 1: Group variable
  - e.g., does an experimental manipulation have an impact on the relationship between two variables?

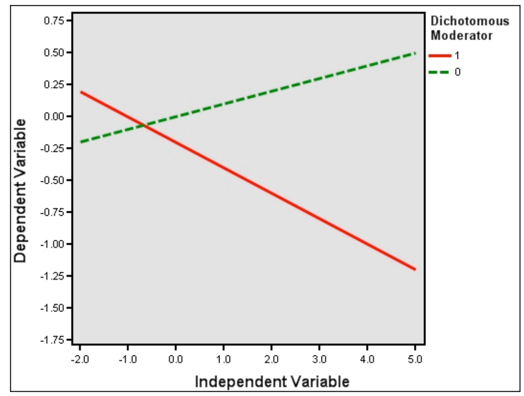
$$Y = b_0 + b_1X_1 + b_2G + b_3GX_1$$

1. $G = 0$

$$Y(0) = \boxed{b_0 + b_1X_1}$$

2. $G = 1$

$$Y(1) = b_0 + b_1X_1 + b_2 + b_3X_1 =$$
$$= \boxed{(b_0 + b_2) + (b_1 + b_3)X_1}$$



If b_2 substantial → different intercepts;
if b_3 substantial → different slopes

1

Interactions

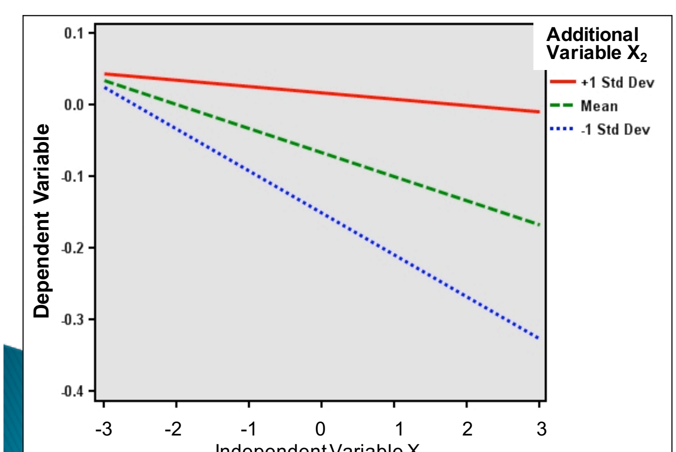
Centering of the continuous IVs allows comparisons of the predictors’ “solo main effects” at the *mean* of the other IV rather than at the 0 point.

- CASE 2: Additional continuous variable X_2

1. Center/standardize the data

2. COMPUTE X12IA $= X_1 \times X_2$

3. **Graph** $Y = f(X_1)$ pattern at +/- 1*SD* of X_2



$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2$$

(1) For all $X_2 \geq 1SD$, set X_2 to 1

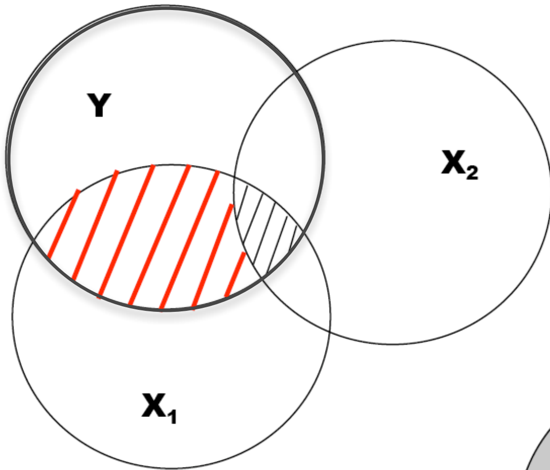
$$Y(\geq 1) = b_0 + b_1X_1 + b_2 + b_3X_1 =$$
$$= (b_0 + b_2) + (b_1 + b_3)X_1$$

(2) For all $X_2 \leq 1SD$, set X_2 to -1

$$Y(\leq 1) = b_0 + b_2X_1 - b_1 - b_3X_1 =$$
$$= (b_0 - b_1) + (b_2 - b_3)X_1$$

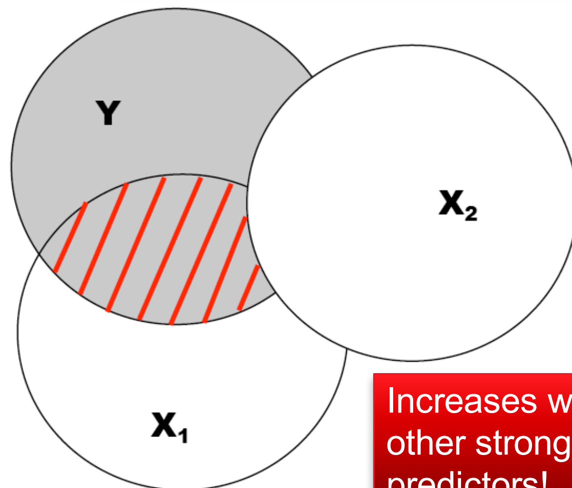
3D graphic software….

2



2. Fully partial (or "partial") variance component (r^2_p) = proportion of red area (unique **X_1** contribution) out of **Y** area not accounted for by other variables.

1. Semi-partial (or "part") variance component (r^2_{sp}) = proportion of red area (unique **X_1** contribution) out of total **Y** area.



Increases with other strong predictors!

3

Shared Variance & Suppression

Separately calculated

Predicting blame from various other ascriptions

Model 1	Semi-partial	squared	total shared
has_choice	0.611	0.373	
has_soul	0.445	0.198	
			0.029
	R^2	0.60	

Model 2	Semi-partial	squared	total shared
has_choice	0.138	**0.019**	
has_soul	0.269	**0.072**	
folk_fw	0.392	0.154	**0.505**
	R^2	0.75	

Suppressor Effects

Suppression of what?
Standard: X2 suppresses residual variance in Y, or nonexplaining variance in X1

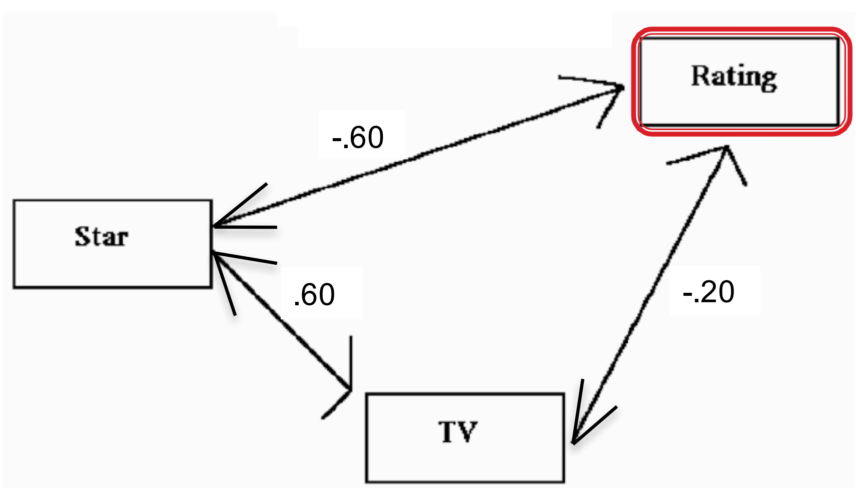
Clinton's Suppression

RATING	Rating of Clinton (1 = very unfavorable towards Clinton, 10 = very favorable)
TV	Did the voter see the 60 minutes story? (1 = yes, 0 = No)
STAR	Did the voter see the Star article? (1 = yes, 0 = no)

Correlation:

		TV	STAR
RATING	1.000	-.200	-.600
TV		1.000	.600
STAR			1.000

Clinton



Clinton

Equation Number 1 Dependent Variable.. RATING

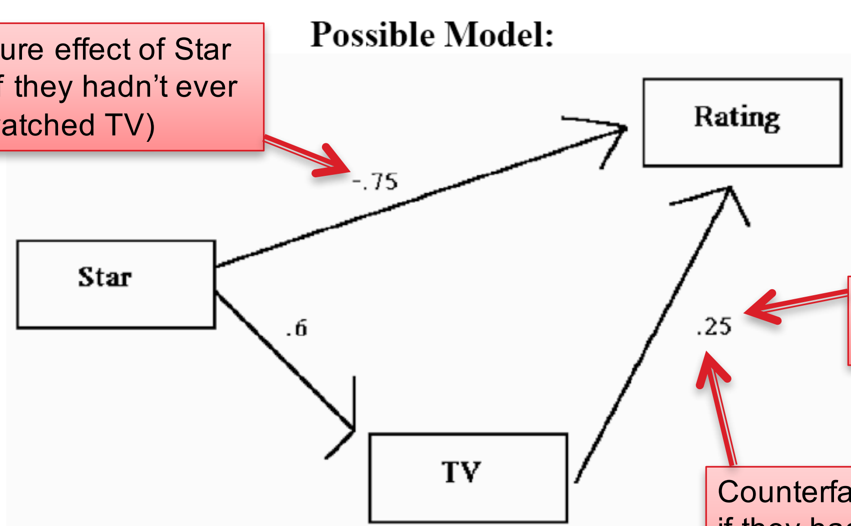
------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
TV	1.091703	.133907	$\boxed{.250000}$	8.153	.0000
STAR	-3.061224	.125162	-.750000 ✔	-24.458	.0000
(Constant)	6.896979	.064122		107.560	.0000

Can be interpreted as a moderator effect: those who read the *Star* article (→ dropping opinion) and then watched the TV announcement had a not-as-bad opinion as the ones who only read the *Star* article.

Clinton

Pure effect of Star (if they hadn't ever watched TV)



Pure effect of TV

Counterfactual world: if they hadn't ever read the Star article

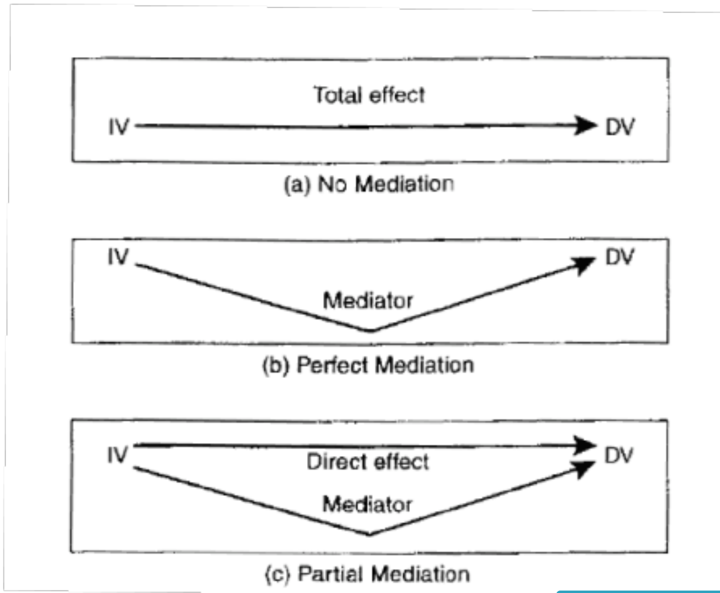
Other Examples

- **How is HEAD Start (see http://eclkc.ohs.acf.hhs.gov/hslc/hs/about) related to scholastic achievement?**
  - (Zero-order correlation is **negative** (as if participating in the program lowered achievement)
  - Add poverty as a predictor →
  - Controlling for poverty, HEAD Start's semi-partial correlation is now ***positive***
  - **Suppressor effect**: HEAD Start's positive correlation with poverty (poor kids are more likely to go into HEAD start) suppresses its predictive power (poor kids have lower achievement) → need to control for poverty.

Value of Comparing r_{xy} and r_{sp}

- r_{xy} = zero-order correlation, "structure coefficient"
- r_{sp} = semi-partial correlation of X_i with Y holding all other X_i constant
- Shows change once other predictors are introduced
- When squared → allows calculation of all shared variance components.
- **Walking through example in SPSS**
- **Excel suppression calculator**

Mediation (Simple version)



Test for Mediation:

1. Show that β(IV) ➛ DV
2. Show that β(Med)➛ DV
3. Show that β(IV) ➛ 0

When β(Med) controlled for

More sophisticated versions: Andrew Hayes…