PS2030
Political Research and Analysis
Unit 1: Fundamentals of Linear Regression
5. Dummy Variable and Mediation Models

Spring 2025, Weeks 4-5

WW Posvar Hall 3600

Professor Steven Finkel

Plan for Session

- Extension of regression to include dichotomous variables, called "dummy" or "indicator" variables
- Estimation and interpretation of coefficients in dummy variable models
- Hypothesis testing
- Dummy variables
- Mediation models
- Examples will come from "ps2030.bank-salaries.data"

Regression with Dummy Independent Variables

- Dummy variables are dichotomous variables coded as 0 or 1. If a nominal dichotomous variable is NOT already coded as 0-1, you need to recode it.

- Multi-category nominal variables can always be recoded into a series of dummy variables, each standing for a different category. E.g., if you have a religious variable where Protestant=1, Catholic=2, Other=3, then you can create three dummy variables:
  - PROT (1=Protestant, 0=all others)
  - CATH (1=Catholic, 0=All others)
  - OTH (1=Other, 0=Protestant, Catholic)

  Note that if you know the value of any two of these three dummy variables, you automatically know the value of the third one. So each pair of dummies is perfectly correlated with the third dummy variable.

  Note also that you should NEVER treat a multi-category nominal variable as anything *but* a series of category-wise dummy variables

- The classic "treatment effects" model (in non-experimental research) starts with an analysis with a dummy variable for treatment exposure, and then adds control variables in multivariate models to take into account pre-existing differences between the treatment and control groups ("selection biases")

- Model:

$$Y_i = \alpha + \beta_1 X_{1i} + \varepsilon_i \quad \text{where}$$

$X_{1i} = 0$ if i is in one (e.g., control) group, $X_{1i} = 1$ if i is in the other (e.g., treatment) group

- So:

$$X_{1i}=0 \qquad Y_i = \alpha + \beta_1(0) + \varepsilon_i$$

$$Y_i = \alpha + \varepsilon_i$$

$$X_{1i}=1 \qquad Y_i = \alpha + \beta_1(1) + \varepsilon_i$$

$$Y_i = (\alpha + \beta_1) + \varepsilon_i$$

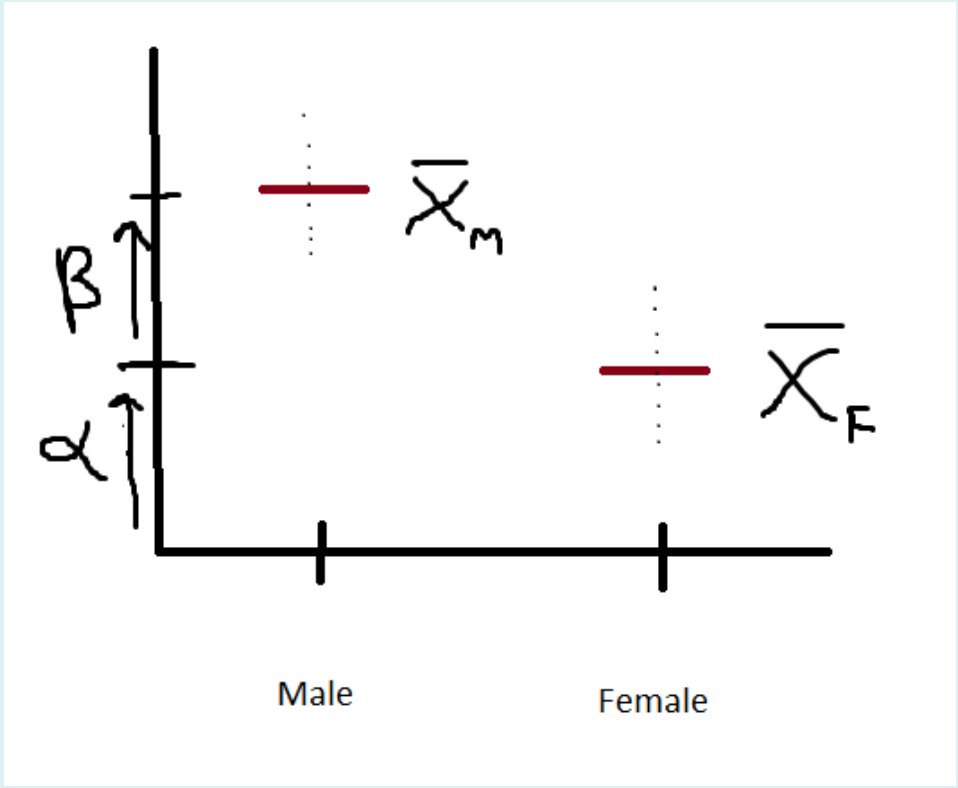
- Implications:
  - α is the intercept for the group where $X_1=0$ (the "control" group)
  - $(\alpha+\beta_1)$ is the intercept for the group where $X_1=1$ (the "treatment" group)
  - β_1 is the difference in intercepts between the group, and hence it is the estimated effect of being in category "1" of the dummy variable versus category "0"
  - Thus the ***slope*** coefficient for the dummy variable represents the ***intercept*** difference between the two groups!
  - This would be the bivariate "treatment effect" in an experimental or quasi-experimental set-up

- Also the case that we can model the expected, or average value of Y for the two groups via the dummy variable model. Say we compare males to females on salary outcomes.

$$E(Y_i) = a + b_1 X_1$$

Female $(X_1 = 0)$ $E(Y_i) = \alpha$

Male $(X_1 = 1)$ $E(Y_i) = (\alpha + \beta_1)$



- So bivariate dummy variable regression is the same thing as testing the significance of the mean difference between the two groups designated by the dummy variable

- The significance of the slope regression coefficient on the dummy variable is the same as the significance of the t-test comparing two groups
- In fact, the t-test for the dummy regression coefficient is equal to the t-test of the difference between the two groups (and F=the t-test value, squared)

```
. regress salnow male

      Source |       SS       df       MS              Number of obs =     474
-------------+------------------------------           F(  1,   472) =  119.80
       Model |  4.4670e+09     1  4.4670e+09           Prob > F      =  0.0000
    Residual |  1.7600e+10   472  37287444.9           R-squared     =  0.2024
-------------+------------------------------           Adj R-squared =  0.2007
       Total |  2.2067e+10   473  46652514.3           Root MSE      =  6106.3

------------------------------------------------------------------------------
      salnow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   6163.945   563.1625    10.95   0.000     5057.329    7270.561
       _cons |   10412.77   415.4841    25.06   0.000     9596.341     11229.2
------------------------------------------------------------------------------
```

```
. ttest salnow, by(male)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     216    10412.77    205.7033    3023.209    10007.32    10818.22
       1 |     258    16576.71    485.5872    7799.685    15620.48    17532.95
---------+--------------------------------------------------------------------
combined |     474    13767.83    313.7244    6830.265    13151.36    14384.29
---------+--------------------------------------------------------------------
    diff |           -6163.945    563.1625               -7270.561   -5057.329
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t = -10.9452
Ho: diff = 0                     degrees of freedom =      472

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

- It is easy to extend this **"dummy variable intercept model"** to a multiple regression context that includes additional explanatory variables:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

Female (X_1=0): $Y_i = \alpha + \beta_1(0) + \beta_2 X_{2i} + \varepsilon_i$

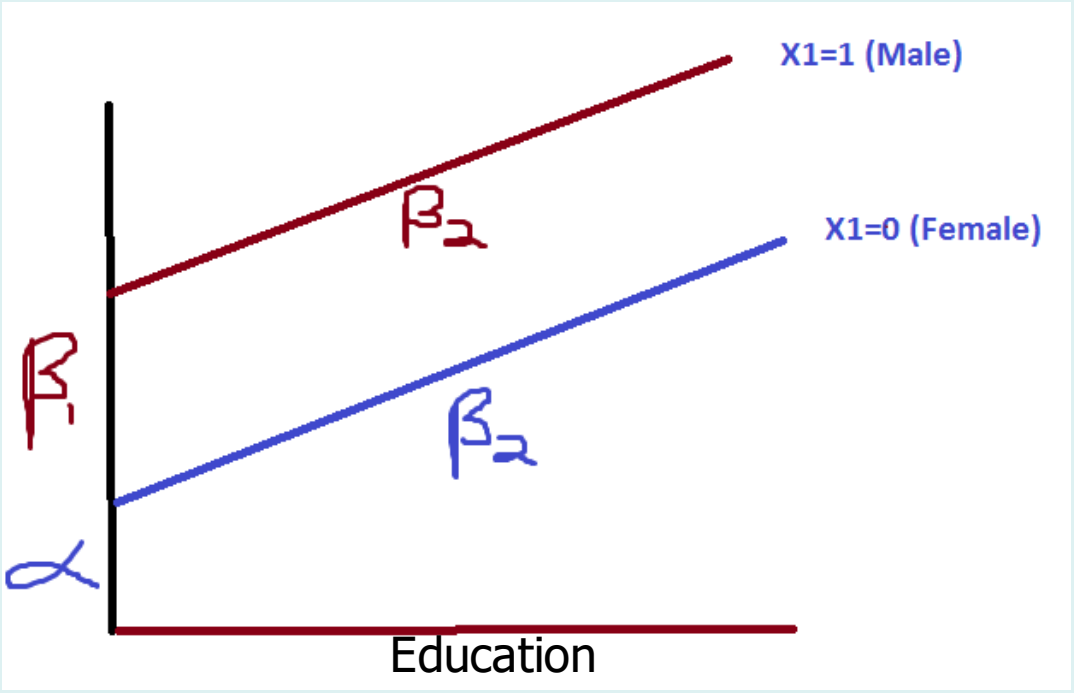
$$Y_i = \alpha + \beta_2 X_{2i} + \varepsilon_i$$

Male (X_1=1): $Y_i = \alpha + \beta_1(1) + \beta_2 X_{2i} + \varepsilon_i$

$$Y_i = (\alpha + \beta_1) + \beta_2 X_{2i} + \varepsilon_i$$

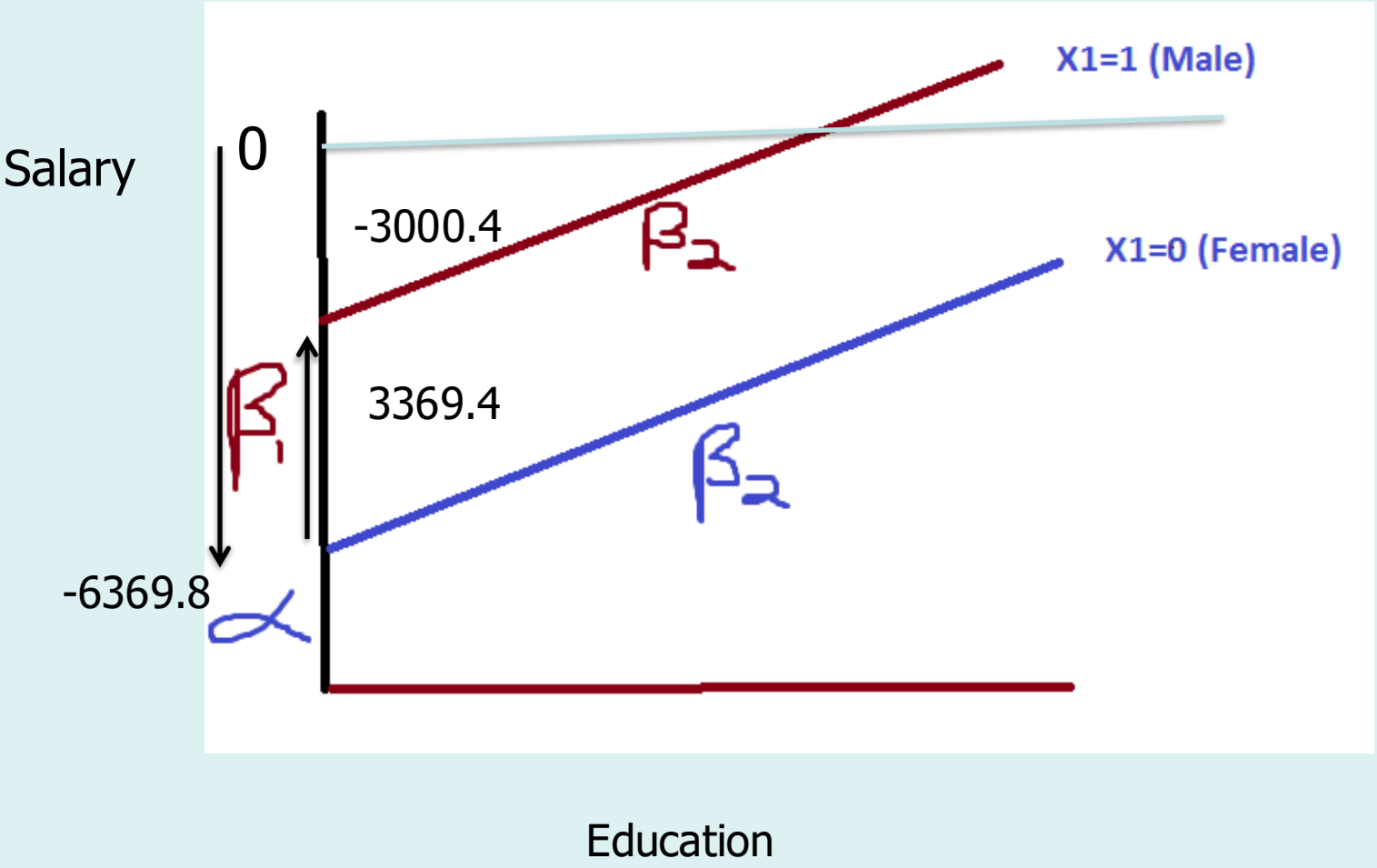
- So:
  - α is the intercept for the X_1 =0 group
  - $(\alpha+\beta_1)$ is the intercept for the X_1 =1 group
  - β_1 is the difference in intercepts between the group, controlling for the effect of X_2
  - Actual models (of course) will include ***all*** observed Xs that are thought to influence Y

Salary




```
. regress salnow male edlevel

      Source |       SS       df       MS              Number of obs =     474
-------------+------------------------------           F(  2,   471) =  225.51
       Model |  1.0794e+10     2  5.3971e+09            Prob > F      =  0.0000
    Residual |  1.1273e+10   471  23933187.2           R-squared     =  0.4892
-------------+------------------------------           Adj R-squared =  0.4870
       Total |  2.2067e+10   473  46652514.3           Root MSE      =  4892.2

------------------------------------------------------------------------------
      salnow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   3369.385   482.8113     6.98   0.000     2420.654    4318.116
     edlevel |   1356.673    83.4395    16.26   0.000     1192.713    1520.633
       _cons |  -6369.781   1084.524    -5.87   0.000    -8500.885   -4238.677
------------------------------------------------------------------------------
```

What This Actually Looks Like!



"Explaining" Group Differences Via Dummy Variable Analysis

- We started out with a $6164 difference between men and women in salary. After controlling for education, we have a $3369 difference left. So education "accounted" for $2795 of the original difference
- This quantity is equal to the regression coefficient for education, multiplied by the mean difference in education between men and women

$$\beta_2 * (\bar{X}_{2M} - \bar{X}_{2F})$$

- We can express this more generally as:

$$\bar{Y} = a + b_1 * \bar{X}_1 + b_2 * \bar{X}_2 + \ldots b_j * \bar{X}_j .. + (\bar{e})$$

$$\bar{Y}_M = a + b_1 + b_2 * \bar{X}_{2M}$$

$$\bar{Y}_F = a + b_2 * \bar{X}_{2F}$$

$$(\bar{Y}_M - \bar{Y}_F) = b_1 + b_2 * (\bar{X}_{2M} - \bar{X}_{2F})$$

$$(16576 - 10412) = 3369 + 1356 * (14.43 - 12.37)$$

- So we can use multiple regression with dummy variables to try to account for differences between groups. Why are the groups different? Start with just the dummy variable difference, then include X2, X3, etc. into the analysis, see if the original bivariate differences diminish, and by how much. This is an *extremely* useful procedure (**if** the slope for X_2 for the two groups is the same – see later slides).

Two (Easy) Dummy Variable Extensions

- Nominal Variables with Multiple Categories
  - Example: 3-Category Religion Variable RELIG, coded as:
    1=Catholic, 2=Other, 3 =Protestant

    Create Three Dummy Variables:
    X_1 (1=Catholic, 0=Non-Catholic)
    X_2 (1=Other, 0=Non-Other)
    X_3 (1=Protestant, 0=Non-Protestant)

- Model is estimated by including (J-1) of the dummy variables associated with the J categories. The omitted variable serves as the "baseline" category, *and all estimated effects are therefore compared with this category.*

- So choose the baseline category purposively – usually the category in which there are the most substantively interesting comparisons to be made. In this case, for example, we would likely *not* choose "X_2" to be the baseline.

- So let's choose Protestant (X_3) to be the baseline category and estimate a model with the two other dummies and another continuous variable X_4.

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{4i} + \varepsilon_i$$

$$Catholic: \quad Y_i = \alpha + \beta_1(1) + \beta_2(0) + \beta_3 X_{4i} + \varepsilon_i$$

$$Other: \quad Y_i = \alpha + \beta_1(0) + \beta_2(1) + \beta_3 X_{4i} + \varepsilon_i$$

$$Protestant: \quad Y_i = \alpha + \beta_1(0) + \beta_2(0) + \beta_3 X_{4i} + \varepsilon_i$$

$So:$

$$Catholic: \quad Y_i = (\alpha + \beta_1) + \beta_3 X_{4i} + \varepsilon_i$$

$$Other: \quad Y_i = (\alpha + \beta_2) + \beta_3 X_{4i} + \varepsilon_i$$

$$Protestant: \quad Y_i = \alpha + \beta_3 X_{4i} + \varepsilon_i$$

Interpretation of Coefficients

- α is the intercept for the baseline group (Protestant)
- $(\alpha+\beta_1)$ is the intercept for Catholics
- $(\alpha+\beta_2)$ is the intercept for Other
- β_1 is the difference in intercepts between the baseline group and Catholics, and hence it is the estimated "effect of being Catholic relative to being a Protestant" (and controlling for X_4)
- β_2 is the difference in intercepts between the baseline group and Other, and hence it is the estimated "effect of being another religion relative to being a Protestant" (and controlling for X_4)
- The significance of these effects is assessed through a t-test for the respective coefficients
- But note that **all effects are interpreted relative to the baseline**, so if you had picked another baseline, you might see a different pattern of effects. That is, Catholics might differ from Protestants, Other might differ from Protestants, but Catholics might not differ from Other.
- You can always re-run the analysis with different baseline categories if you are interested in all of the possible differences between groups, or use the "lincom" procedure in STATA (see subsequent slides)

- Two Dichotomous Independent Variables
  - Example: X_1 (1=Male, 0=Female)
    X_2 (1=Minority, 0=White)

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

Minority Men: $Y_i = \alpha + \beta_1(1) + \beta_2(1) + \varepsilon_i$

White Men: $Y_i = \alpha + \beta_1(1) + \beta_2(0) + \varepsilon_i$

Minority Women: $Y_i = \alpha + \beta_1(0) + \beta_2(1) + \varepsilon_i$

White Women: $Y_i = \alpha + \beta_1(0) + \beta_2(0) + \varepsilon_i$

Interpretation of Coefficients

- α is the intercept for the group that has zeros on both variables (White Women – the baseline category)
- β_1 is the "effect" of being a Man
- β_2 is the "effect" of being a Minority
- The significance of these effects is assessed through a t-test for the respective coefficients
- The effect of being a Man is assumed to be the same for both Minorities and Whites
- The effect of being Minority is assumed to be the same for both men and women
- We can easily model the average values of Y for each of the four groups by adding the intercept and relevant "slope" coefficients (and the means of other Xs weighted by their β)

```
. regress salnow male minority

      Source |       SS       df       MS              Number of obs =     474
-------------+------------------------------           F(  2,   471) =   77.40
       Model |  5.4586e+09     2  2.7293e+09           Prob > F      =  0.0000
    Residual |  1.6608e+10   471  35261163.1           R-squared     =  0.2474
-------------+------------------------------           Adj R-squared =  0.2442
       Total |  2.2067e+10   473  46652514.3           Root MSE      =  5938.1

------------------------------------------------------------------------------
      salnow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   6384.335   549.2216     11.62   0.000     5305.107    7463.563
    minority |  -3505.107   660.9473     -5.30   0.000    -4803.877   -2206.336
       _cons |   11061.86   422.1698     26.20   0.000     10232.29    11891.43
------------------------------------------------------------------------------
```

- E(Salary, Minority Women): 11061.86+-3505.11= 7556.57
- E(Salary, White Women): 11061.86
- E(Salary, Minority Men): 11061.86+6384.34+-3505.11=13941.09
- E(Salary, White Men): 11061.86+6384.34=17446.02

- All of these values are easily obtained via Stata "margins" command

- Note: This procedure will not recover the *exact* means for each of these groups' salaries.
  - E(Salary, Minority Women): 7556.57 Actual Mean: 9225.0
  - E(Salary, White Women): 11061.86 Actual Mean: 10682.7
  - E(Salary, Minority Men): 13940.91 Actual Mean: 12898.4
  - E(Salary, White Men): 17446.02 Actual Mean: 17790.2
- Why not? Assumption of model is that the effect of sex (or race) is the same at all levels of the other variable, i.e., that we have an additive model with no "interaction effect". We will relax this assumption in the next slides. Including an interaction effect **in this case** would "saturate" the model at the group level, i.e., the 4 coefficients estimated would reproduce the 4 group means.
- Final Note: Regression gives you the significance of certain group comparisons but not others, depending on which categories you designated as "baseline" for the model. In this example, we get the effect of "white male" versus "white female" (baseline) directly, but not, e.g., "minority male" versus "white female". Use Stata's "margins" or "lincom" command for this.

"Slope" Dummy Variable Models, or "Interaction Effects"

- What if we think that the **effect** of one variable depends on what group one belongs to, that is, that the slope or the effect of X_1 depends on the level of some dichotomous dummy variable X2?
- We call these kinds of models "interactive" models, "non-additive" models, "conditional effects" models, or **"slope dummy variable models"**, unlike the "intercept dummy variable models" we just considered.
- Example: The effect of Age (X_2) on Salary (Y) depends on Sex (X_1), such that age may *really* boost men's income but not women's by as much

$$Y_i = \alpha + \beta_1 X_{2i} + \beta_2 X_{1i} X_{2i} + \varepsilon_i$$

where $X_{1i} X_{2i}$ is the product of X_1 and X_2

- Conditional effects models are *extremely* common in contemporary political science research (and rightly so, as that is how variables are likely to be related in the "real" world!). Pick out any recent APSR, AJPS, JOP, etc. and you will see examples. You can also look at "heterogeneous treatment effects" as interactions of the treatment with some other characteristic or variable

$$Y_i = \alpha + \beta_1 X_{2i} + \beta_2 X_{1i} X_{2i} + \varepsilon_i$$

where $X_{1i} X_{2i}$ is the product of X_1 and X_2

So:

Women: $Y_i = \alpha + \beta_1 X_{2i} + \beta_2 X_{2i}(0) + \varepsilon_i$

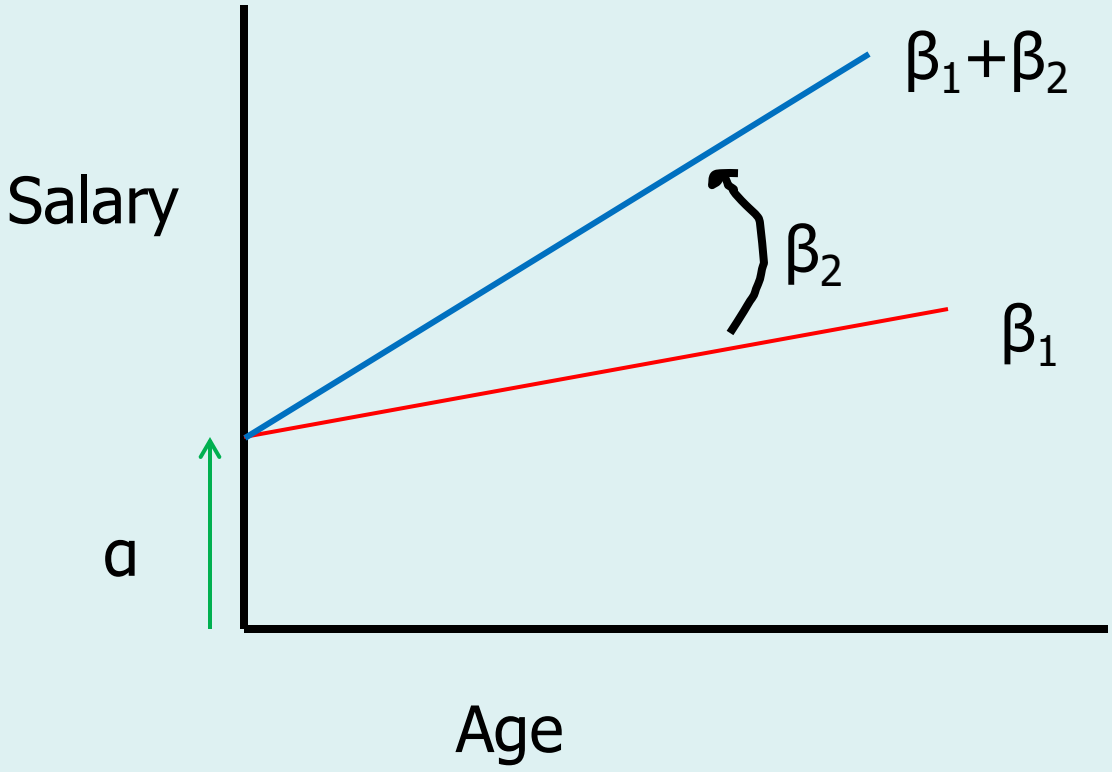
Men : $Y_i = \alpha + \beta_1 X_{2i} + \beta_2 X_{2i}(1) + \varepsilon_i$

or:

Women: $Y_i = \alpha + \beta_1 X_{2i} + \varepsilon_i$

Men: $Y_i = \alpha + (\beta_1 + \beta_2) X_{2i} + \varepsilon_i$

Slope Dummy Variable Relationship

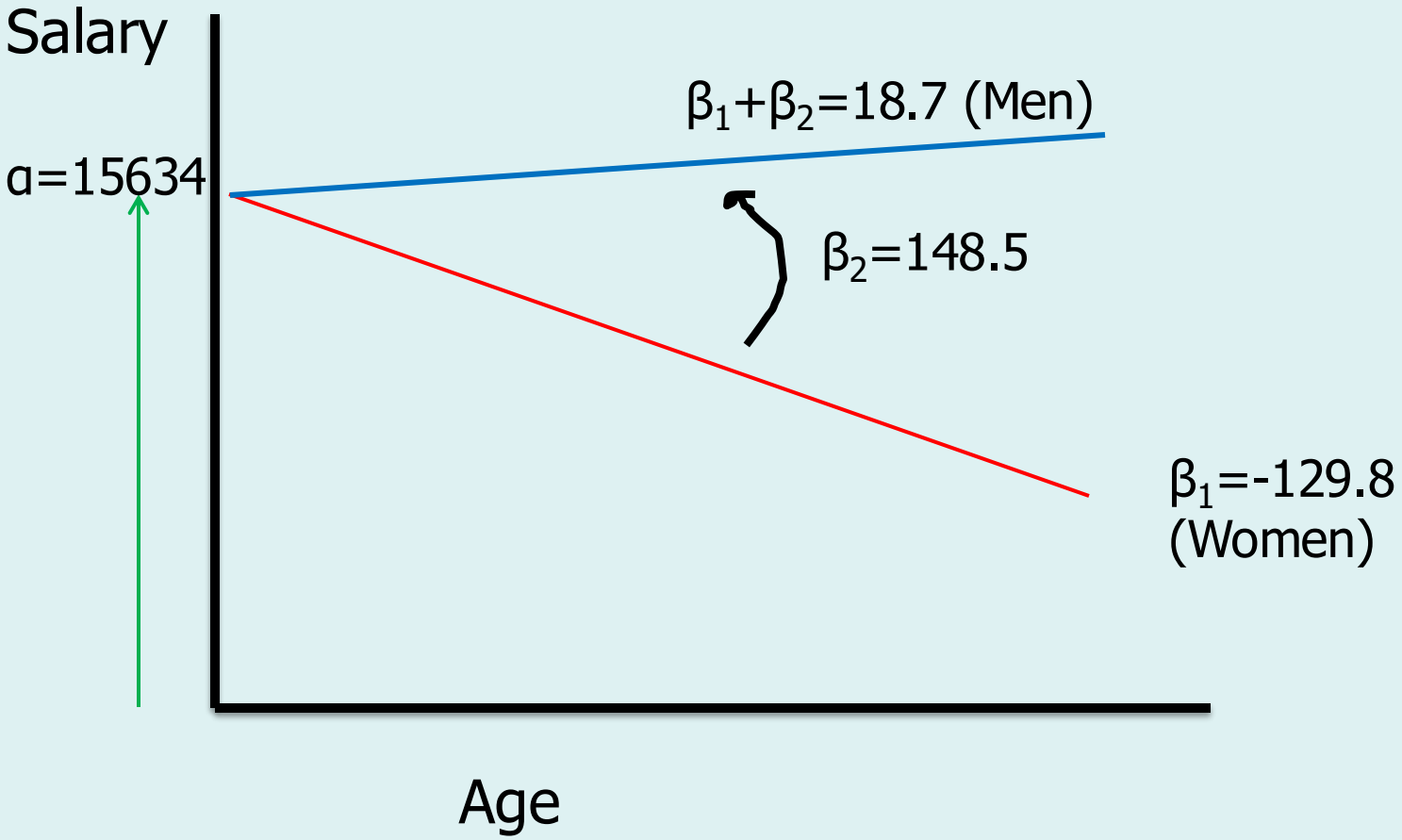


Interpretation of Coefficients

- α is the intercept for both men and women, i.e. the salary they would receive when $X_2=0$ (when age=0)
- β_1 is the effect of age for women (when $X_1=0$)
- $(\beta_1+\beta_2)$ is the effect of age for men (when $X_1=1$)
- β_2 is the difference in the slope of age for men, compared to the "baseline" slope for women
- The significance of the difference in the slopes for men and women is assessed through a t-test for the β_2 coefficient

```
. regress salnow age agemale

      Source |       SS       df       MS              Number of obs =     474
-------------+------------------------------           F(  2,   471) =   58.15
       Model |  4.3695e+09     2  2.1847e+09            Prob > F      =  0.0000
    Residual |  1.7697e+10   471  37573647.3            R-squared     =  0.1980
-------------+------------------------------           Adj R-squared =  0.1946
       Total |  2.2067e+10   473  46652514.3            Root MSE      =  6129.7

------------------------------------------------------------------------------
      salnow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -129.7935      24.32    -5.34   0.000    -177.5826   -82.00439
     agemale |   148.4601   14.57264    10.19   0.000     119.8247    177.0956
       _cons |   15634.62   941.0602    16.61   0.000     13785.42    17483.81
------------------------------------------------------------------------------
```

What This Actually Looks Like!



Extension to More than Two Groups

- It is straightforward to extend the slope dummy variable model to more than two groups. You would just have one baseline group excluded and (J-1) dummy variables that would be multiplied by the "X_2" variable to produce (J-1) interaction terms. The "slope" of each interaction term would represent the difference between the effect of X_2 for the interacted group versus the excluded baseline category.
  - i.e., if you have 4 Religious categories (Protestant, Catholic, Jewish, Other) you would create 3 dummies, multiply each of them by X_2, and then enter X_2 by itself and the 3 interaction/multiplicative terms into the model. The effects of the 3 interaction terms will be the difference in the estimated effects of X_2 on Y between a given group and Protestants (the baseline category, whose effect is represented by the slope on X_2 by itself).
- Same issues as in dummy intercept models for choosing the baseline category for maximum substantive interest
- Same issues for arriving at any comparison of the slopes for any pair of groups by re-running the analysis with difference baseline groups excluded (or using "lincom").
  - So to get the statistical significance of the Jewish-Catholic comparison, re-run the analysis with either one of those groups' dummy interaction variables excluded, and the Protestant interaction variable included instead. Then compare the effect of the Jewish interaction to the baseline effect of Catholic (or the effect of the Catholic interaction to the baseline effect of Jewish)

Are the slopes for each group statistically significant?

- Important: The slope dummy variable analysis tells you two things.
  - 1) Is the slope for X_2 for the baseline category significant? (This is the t-test on β_1); and
  - 2) Is the slope for X_2 for a given other group significantly different from the slope of the baseline group? (This is the t-test on β_2).
- But what about whether the slope for the given other group is significantly different from 0? The analysis so far does not tell us.
- We can test this easily in Stata with the "lincom" command
- Question: Is a "linear combination" of estimates significant or not?. These are based on "conditional standard errors"; that is, the standard error of a slope conditioned on some other variable
  - In this case it is the slope of age, conditioned on a given group being equal to "1". There will be a conditional standard error when "Male=1" and "Male=0", the latter being produced in the normal Stata output
  - We will discuss conditional standard errors in more detail in our later discussion of interaction effects models (next week)

```
. regress salnow age agemale
```

Source	SS	df	MS	Number of obs =	474
Model	4.3695e+09	2	2.1847e+09	F(2, 471) =	58.15
Residual	1.7697e+10	471	37573647.3	Prob > F =	0.0000
				R-squared =	0.1980
				Adj R-squared =	0.1946
Total	2.2067e+10	473	46652514.3	Root MSE =	6129.7

salnow	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-129.7935	24.32	-5.34	0.000	-177.5826	-82.00439
agemale	148.4601	14.57264	10.19	0.000	119.8247	177.0956
_cons	15634.62	941.0602	16.61	0.000	13785.42	17483.81

- The estimated slope for age on salary when Male=0 is: -129.79. This effect is statistically significant at the .05 level
- The estimated *difference* in slopes between Males and Females is 148.46, and this difference is statistically significant at the .05 level

- But what about the estimated slope for age on salary when Male=1; it is (-129.79+148.46)=18.67. Is this value significantly different from 0?
- This is useful information: It says that the effect of age on salary for men is indistinguishable from 0, while the effect of age on salary for women is negative and significant

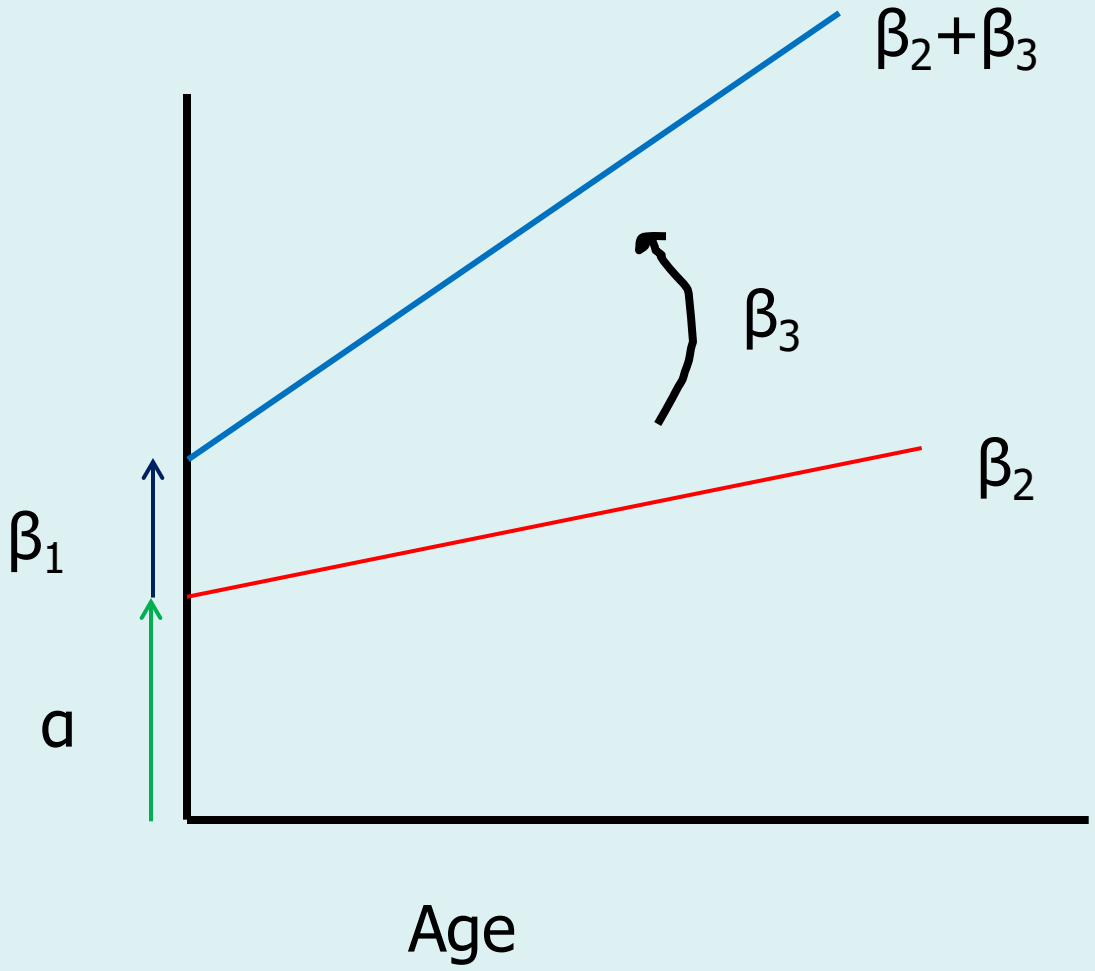
```
. lincom age+agemale*1

 ( 1)  age + agemale = 0
```

salnow	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	18.66663	25.96897	0.72	0.473	-32.36275	69.696

.

Combining Intercept and Slope Dummy Variable Models

- Most often, the slope model will be estimated along with an intercept model. That is, one is usually reluctant to say that *only* the slopes will differ by groups without a corresponding intercept difference as well. (On the other hand, "intercept only" models are very common). So a combined model would suggest that men and women start at different places on salary with "no age", and then increase/decrease at different rates as well. In effect, there are two different regression lines altogether, one for men and one for women



$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$$

where $X_{1i} X_{2i}$ is the product of X_1 and X_2

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \varepsilon_i$$

Female (X_1=0) $\quad Y_i = \alpha + \beta_2 X_{2i} + \varepsilon_i$

Male (X_1=1) $\quad Y_i = (\alpha + \beta_1) + (\beta_2 + \beta_3) X_{2i} + \varepsilon_i$

- So:
  - α is the intercept for the group when X_1=0
  - $(\alpha+\beta_1)$ is the intercept for the group when X_1=1
  - β_2 is the effect of X_2 (Age) on Y for the group when X_1=0
  - $(\beta_2 + \beta_3)$ is the effect of X_2 (Age) on Y for the group when X_1=1

- So, if β_3 is significant, then it means that the effect of X_2 (Age) on Y differs, depending on the level of X_1 (Male)
- It also means that the effect of X_1 (Male) on Y differs, depending on the level of X_2 (Age)! The two statements are equivalent!
- This is the nature of "interaction" or "conditional effects" models
- You can see this by rearranging the model in two ways:

$$Y_i = \alpha + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} + \varepsilon_i$$

$$Y_i = \alpha + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} + \varepsilon_i$$

- Implications:
  - The effect of X_2 on Y depends on X_1 , increasing by an increment of β_3 for every unit change in X_1
  - The effect of X_1 on Y depends on X_2, increasing by an increment of β_3 for every unit change in X_2
  - These are equivalent interpretations of this model

```
. regress salnow male age agemale

      Source |       SS       df       MS              Number of obs =     474
-------------+------------------------------           F(  3,   470) =   43.56
       Model |  4.8011e+09     3  1.6004e+09           Prob > F      =  0.0000
    Residual |  1.7266e+10   470  36735188.6           R-squared     =  0.2176
-------------+------------------------------           Adj R-squared =  0.2126
       Total |  2.2067e+10   473  46652514.3           Root MSE      =    6061

------------------------------------------------------------------------------
      salnow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |    6474.849   1888.884     3.43   0.001     2763.146    10186.55
         age |   -67.06342   30.21845    -2.22   0.027    -126.4434   -7.683441
     agemale |   -10.73527   48.62546    -0.22   0.825    -106.2855    84.81493
       _cons |    12951.39   1215.96     10.65   0.000        10562    15340.78
------------------------------------------------------------------------------

.
```

- Male intercept is $6475 more than female; statistically significant
- Male slope is $11 less than female slope; so for every additional year older, a female makes $67 less, while a man makes $78 less. This difference is not statistically significant, so we cannot reject the hypothesis that the slopes for men and women are equal
- Exercise: Graph what this "Actually Looks Like", i.e., the equivalent of slides 9 and 22

Additional Tests

Is the male slope statistically significant from zero?

```
. lincom age+agemale

 ( 1)  age + agemale = 0
```

salnow	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-77.79869	38.09568	-2.04	0.042	-152.6576	-2.939752

Are the regression lines for men and women statistically different? (Can we reject the null hypothesis that the intercept difference *and* the slope difference are both zero?) This is an F* test for both male (the intercept dummy) and agemale (the slope dummy interaction term)

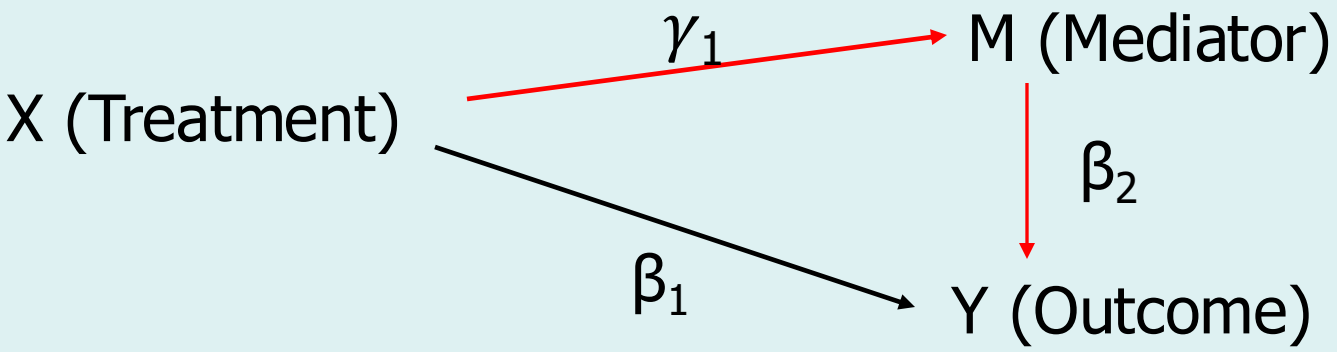
```
. test male agemale

 ( 1)  male = 0
 ( 2)  agemale = 0

       F(  2,   470) =   58.95
            Prob > F =    0.0000
```

Mediation Analysis

- Dummy variable analysis leads nicely into the broader framework of mediation analysis. We may think that a dummy variable X, e.g. treatment/control or male/female, affects an outcome (Y) in two ways: 1) because the treatment affects a mediating variable (M) which then affects Y; and 2), because the treatment affects the outcome directly, over and above the mediation effect. We could call the first process an **"indirect effect"** of X on Y, and the second process a **"direct effect"** of X on Y.
- For example, we may think that there are sex differences in salary, some of which are "mediated" through the differences in education that men and women have obtained.
- How can we determine what the Direct and Indirect Effects are, and by extension, how much of the Total Effect of Sex on Salary is comprised of the Direct and Indirect Effects?
- This is the focus of Mediation Analysis, which, in its modern form of "Causal Mediation Analysis", is now used frequently in political science and elsewhere.

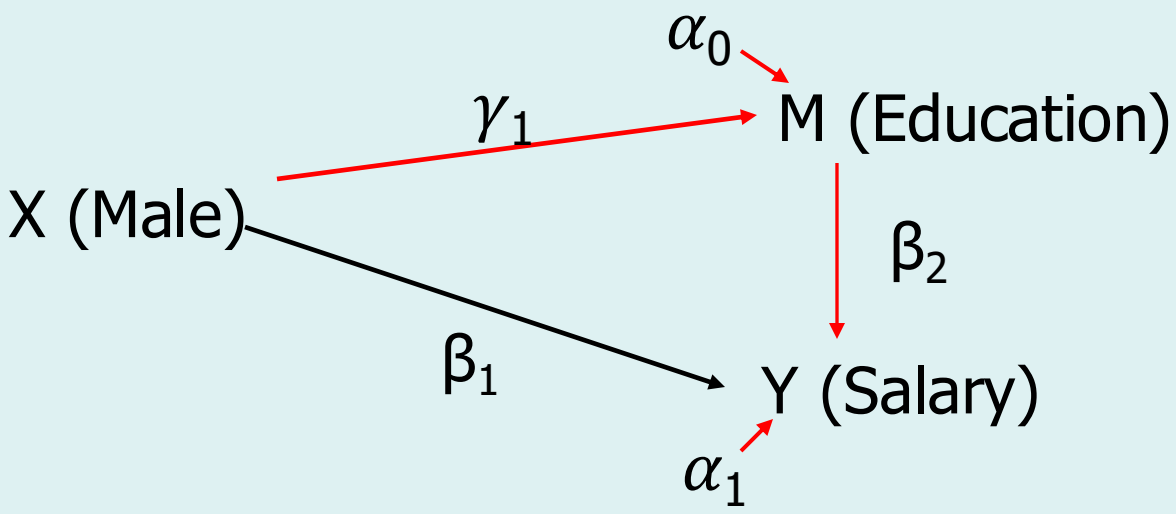
- Classic Treatments in Causal Mediation Analysis
  - Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*.
  - Imai, Kosuke, Luke Keele and Dustin Tingley (2010), "A General Approach to Causal Mediation Analysis", *Psychological Methods*.
  - Judea Pearl and Dana MacKenzie, *The Book of Why* (Basic Books 2018) - an introduction to the Pearl universe of causal inference
  - Tyler VanderWeele, *Explanation in Causal Inference: Methods for Mediation and Interaction* (Oxford University Press 2015).

Causal Mediation Analysis



- We can see the two ways that X may affect Y: 1) The treatment affects a mediating variable (M) which then affects Y; and 2), the treatment affects the outcome directly, over and above the mediation effect.
- You can see the **Indirect Effect** of X on Y in Red (X→M→Y); and the **Direct Effect** in Black (X→Y, controlling for M).

Causal Mediation: Our Sex-Salary Example



- Male is the "Treatment" condition in this example, Female is "Control"
- When there is **no interaction effect** between X and M in predicting Y:

edlevel	Coefficient	Std. err.	t	P>|t|
male	2.059862	.2488915	8.28	0.000
_cons	12.37037	.1836245	67.37	0.000

salnow	Coefficient	Std. err.	t	P>|t|
edlevel	1356.673	83.4395	16.26	0.000
male	3369.385	482.8113	6.98	0.000
_cons	-6369.781	1084.524	-5.87	0.000

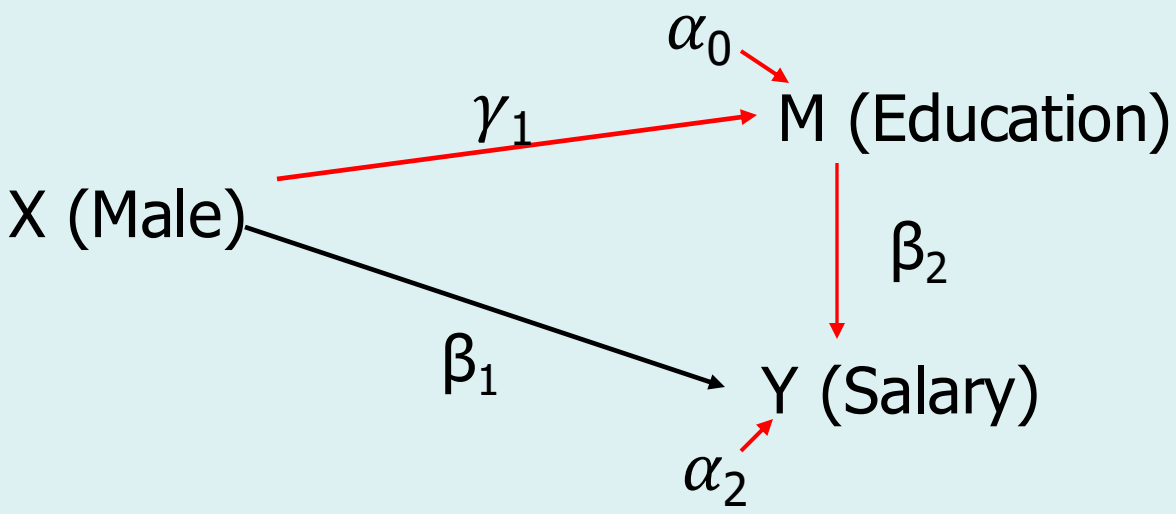
- α_0=12.37 (Female Education Mean)
- γ_1=2.06
- $\alpha_0 + \gamma_1$=14.43 (Male Education Mean)

α_1= -6369.78 (Female Intercept)

β_1 = 3369.39 (Effect of "Male" on Y)

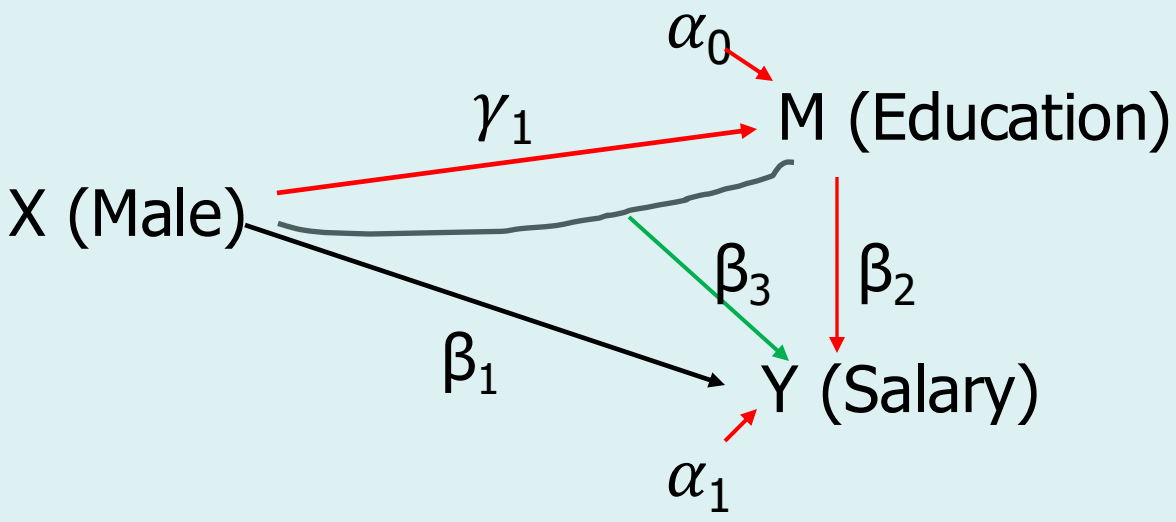
β_2 = 1356.67 (Effect of Education on Y)

Causal Mediation: Our Sex-Salary Example

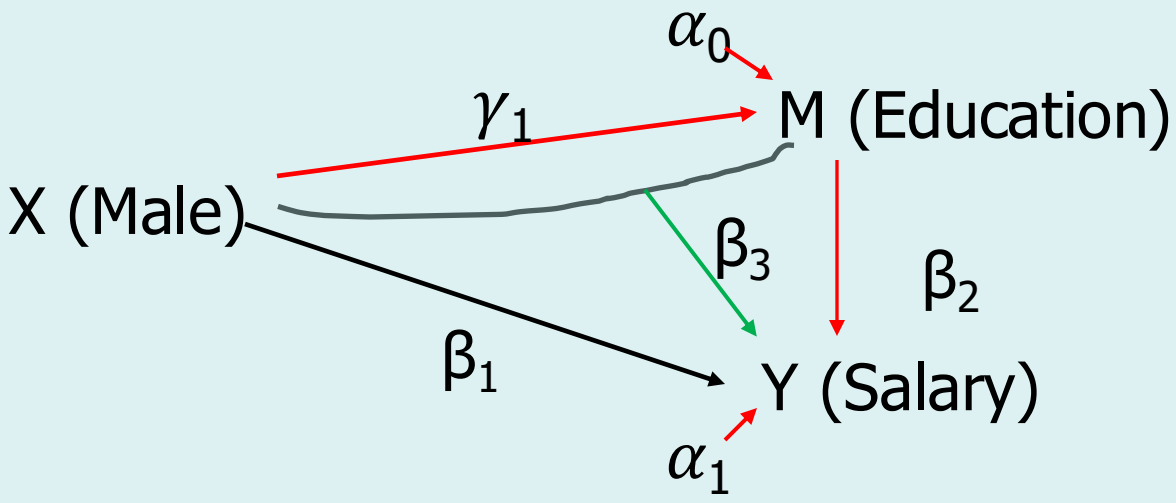


- The Indirect Effect is the product of γ_1 times β_2 (γ_1 β_2), which is the effect of X on M (or the difference in means between X(1) and X(0)), multiplied by the common effect of M on Y.
- The Direct Effect is β_1, the effect of X on Y that does not go through M
- **This is exactly what we calculated earlier on slide 10:**
  - Indirect Effect: (14.43-12.37)=2.06*1356=2795
  - Direct Effect: 3369 = (6164-2795) = Total Effect Minus Indirect Effect
- This is "old style" Mediation Analysis (see Baron and Kenny 1986)

Causal Mediation with an Interaction Effect between X and M



- The general model: $Y=\alpha_1+\beta_1X+\beta_2M+\beta_3XM\ (+\varepsilon)$
- The interaction effect (in green) means that the effect of M on Y differs for groups X=0 and X=1. In this case, that the effect of Education on Salary is different for Males and Females
- Analyzing this kind of model takes us further into the modern world of "Causal Mediation Analysis"



- $Y=\alpha_1+\beta_1X+\beta_2M+\beta_3XM$

  $X=0$: $Y=\alpha_1+\beta_2M$

  $X=1$: $Y=(\alpha_1+\beta_1)+(\beta_2+\beta_3)M$
- The effect of M on Y for $X=0$ is β_2
- The effect of M on Y for $X=1$ is $(\beta_2+\beta_3)$
- The intercept of the Y equation also differs for $X=0$ and $X=1$
  - The intercept of Y for $X=0$ is α_1
  - The intercept of Y for $X=1$ is (α_1+B_1)

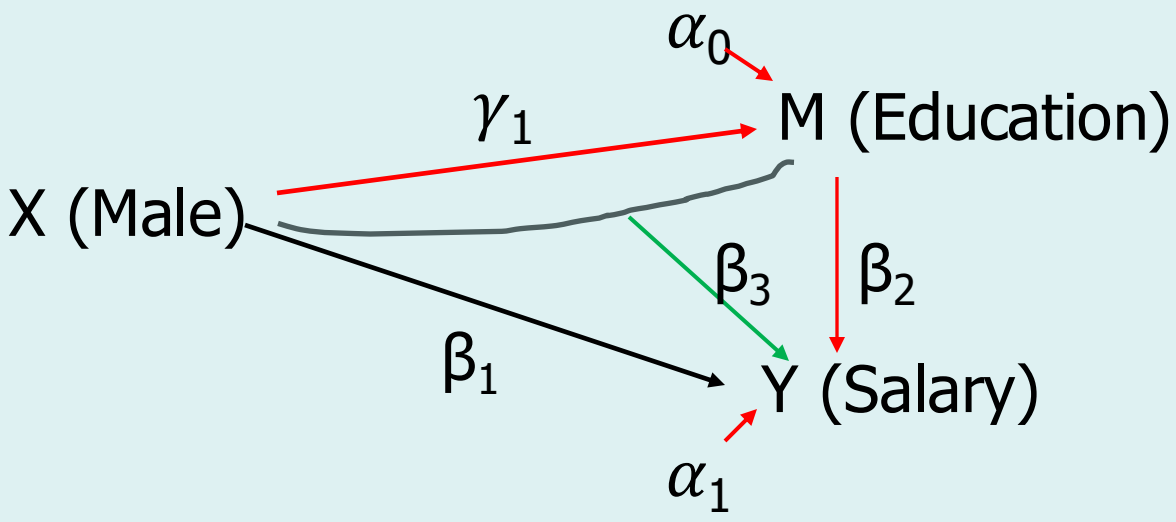
$Y=\alpha_1+\beta 1X+\beta_2M+\beta_3XM$

$X=0:\ Y=\alpha_1+\beta_2M$

$X=1:\ Y=(\alpha_1+\beta_1)+(\beta_2+\beta_3)M$

- Because we have two different equations, one for X=0 and one for X=1, calculating Indirect and Direct Effects is tricky
  - There are different intercepts in the Y equation (α_1 or ($\alpha_1+\beta_1$)) which would be relevant for calculating the "Direct Effect"
  - There are different effects of M on Y for the different levels of X, so calculating the "Indirect Effect" would be different depending on which M→Y effect (β_2 or ($\beta_2+\beta_3$)) was included in the calculation
- It turns out there are two different Direct and two different Indirect Effects, with different combinations of the two producing the Total Effect. This is the conceptual breakthrough provided by causal mediation models.

- To arrive at indirect and direct effects in (modern) causal mediation analysis, we engage in counterfactual reasoning:
  - What would Y look like if you held M at its average level for X=0 or X=1, and then changed X from 0 to 1? That would be the "Direct Effect" of X, since M (the mediator) is held constant and X is counterfactually manipulated.
  - What would Y look like if you held X at either 0 or 1, and then changed M from the average level it has when X=0 to the average level it has when X=1? This would be the "Indirect Effect" of X→M→Y, since X is held constant and M is counterfactually manipulated
  - Note that we don't observe some of these quantities directly, they are counterfactual. For example, when an observation is in the treatment category, we don't observe what its M value would be under the control condition. We'll get more into counterfactual analysis later in the course, but this is the general idea.


α_0
γ_1
M (Education)
X (Male)
β_3
β_2
β_1
Y (Salary)
α_1

```
. gen maleed=male*edlevel

. reg salnow male edlevel maleed
```

Source	SS	df	MS	Number of obs	=	474
				F(3, 470)	=	172.04
Model	1.1549e+10	3	3.8498e+09	Prob > F	=	0.0000
Residual	1.0517e+10	470	22376897.1	R-squared	=	0.5234
				Adj R-squared	=	0.5203
Total	2.2067e+10	473	46652514.3	Root MSE	=	4730.4

salnow	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
male	-9591.421	2279.055	-4.21	0.000	-14069.82	-5113.023
edlevel	698.2664	139.1078	5.02	0.000	424.9162	971.6165
maleed	992.1554	170.7631	5.81	0.000	656.6017	1327.709
_cons	1774.955	1750.657	1.01	0.311	-1665.129	5215.039

Male Equation:

Y = (1774.96+-9591.42)+(698.27+992.16)*Edlevel

Y = -7816.46 + 1690.43 *Edlevel

$(\alpha_1 + \beta_1)$ $(\beta_2 + \beta_3)$

Female Equation:

Y= 1774.96 +. 698.27*Edlevel

(α_1) (β_2)

Difference in Effects of Education on Salary Between Males and Females: 992.16 (β_3)

Difference in Intercepts Between Males and Females: -9591.42 (β_1)

Male Mean Education = 14.43 $(\alpha_0 + \gamma_1)$

Female Mean Education= 12.37 (α_0)

Difference in Education Means: 2.06 (γ_1)

Indirect Effects

To obtain one Indirect Effect of Male→Education→Salary, we take the male regression equation and counterfactually manipulate M from the average female level to the average male level. What would salary be if average education changed from female to male along with male's education effect on salary?

- Male Y(1, 1) = -7816.46+1690.43*14.43=16576.45
  - Male Regression Equation with Male Average Education
- Male Y(1, 0)= -7816.46+1690.43*12.37= 13094.16
  - Male Regression Equation with Female Average Education

Difference =3482.3 (Indirect Effect 1, Effects on Y set to Male)

Calculation Shortcut: $(\beta_2 + \beta_3)*(\gamma_1)$ from the full model. It is the effect of Education on Salary for males (1690.43) multiplied by 2.06, the effect of Male on Education (i.e., the difference in means on Education between males and females). 1690.43*2.06=3482.3

Indirect Effects

We could also obtain another Indirect Effect from the "female perspective" by taking the female equation and counterfactually manipulate M in the same way. What would salary be if average education changed from female to male along with female's education effect on salary?

- Female Y(0,1) = 1774.96 + 698.27*14.43 = 11851
  - Female regression equation with Male Average Education
- Female Y(0,0) = 1774.96 + 698.27*12.37 = 10412.6
  - Female regression equation with Female Average Education

Difference =1438.4 (Indirect Effect 0, Effects on Y set to Female)

Calculation Shortcut: $(\beta_2)*(\gamma_1)$. The effect of Education on Salary for females, multiplied by the effect of Male on Education (i.e., the difference in means on Education between males and females). 698.27*2.06=1438.4

Direct Effects

To obtain one Direct Effect of Male→Salary, controlling for Education, we compare the male equation for salary with the female equation for salary, holding education at its female level. What would salary be for a male (with male regression coefficients) with average female level of education compared to a female (with female coefficients) with female level of education?

- Male Y(1,0) = -7816.46+ (698.27+992.16)*12.37 = 13094.16
  - Male regression equation with Female levels of education
- Female Y(0,0) = 1774.96 + 698.27*12.37 = 10412.6
  - Female regression equation with Female levels of education

Difference =2681.5 (Direct Effect 0, Setting Mediator at Female)

This is the effect of changing X from the control to treatment, holding the mediator M at its control level.

Calculation Shortcut: $(\beta_1 + \beta_3(\alpha_0))$. It is the difference in intercepts between the male and female equations plus the difference in effects of education for males and females multiplied by the female average level of education. -9591.4+992.2*12.37=2682.1

Direct Effects

We could also obtain another Direct Effect by holding education at the male mean. What would salary be for a male (with male regression coefficients) with male average level of education compared to a female (with female regression coefficients) with male level of education? [You could say this is the direct effect of Male from the "male perspective"]

- Male Y(1,1) = -7816.46 + (698.27+992.16)*14.43=16576.4
  - Male regression equation with Male Average Education
- Female Y(0,1) = 1774.96 + 698.27* 14.43 =1774.96+10076.04=11850
  - Female regression equation with Male Average Education

Difference =4725.4 (Direct Effect 1, Setting Mediator at Male)

This is the effect of changing X from control to treatment, holding the mediator M at its treatment level.

Calculation Shortcut: $(\beta_1 + (\beta_3 * (\alpha_0 + \gamma_1)))$. It is the difference in intercepts between males and females multiplied by the difference in effects of education for males and females multiplied by the male average level of education. -9591.4+992.2*14.43=4726

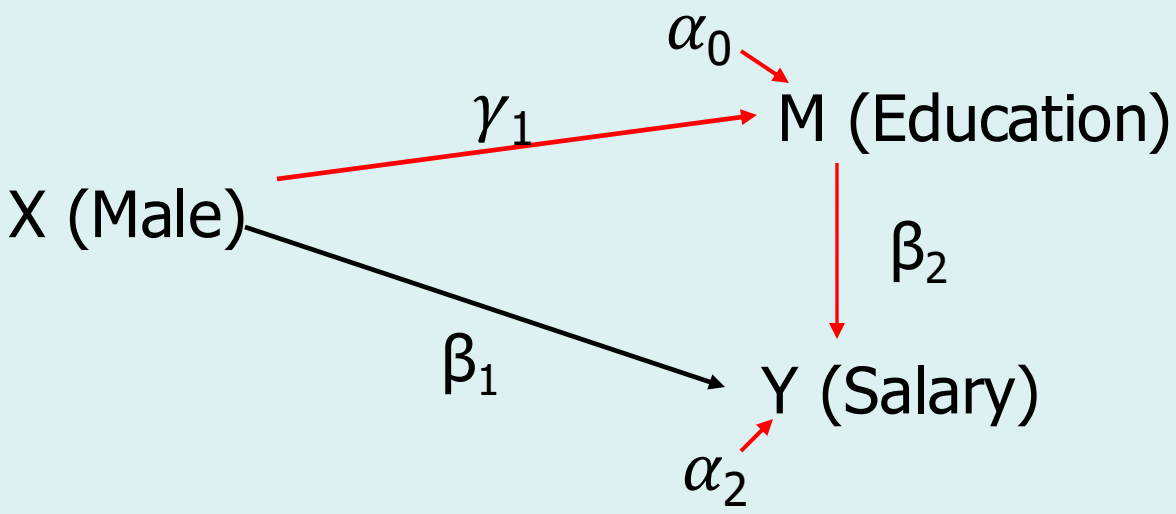
Causal Mediation Quantities and Terminology

- Direct Effects are *usually* expressed using the control group average on M and then comparing Y with X=1 regression coefficients to Y with X=0 regression coefficients. That's what we've called Direct Effect (0, Setting Mediator to Female).
  - Imai et. al. (2010) calls this "Direct Effect (0)"
  - Pearl, VanderWeele) call this the "Pure Natural Direct Effect" or PNDE. It is "natural" to set the mediator to the control condition and then vary X from control to treatment and see the effects on Y
  - Stata just calls it NDE
- Indirect effects are *usually* expressed using the "treatment group" (X=1) equation and changing M from what it would be under (X=0) versus (X=1). We've called this Indirect Effect (1, Effects on Y Set to Male).
  - Imai et. al. call this the "Average Causal Mediation Effect (1)", ACME(1)
  - Pearl,VanderWeele call it the "Total Natural Indirect Effect" or TNIE
  - Stata just calls it NIE

Total Effects

- Total Effect = 6163.8
  - Imai: DE(0)+ACME (1) = 2681.5 + 3482.3 = 6163.8
  - Pearl/VanderWeele: PNDE + TNIE = 2681.5 + 3482.3 = 6163.8
- **This is the Stata default!**
- You can also calculate Total Effects by adding together the two other quantities we calculated earlier: Indirect Effects (0, effects on Y set to Female) and Direct Effects (1, Mediator set to Male).
  - Imai: DE(1) + ACME (0)
  - Pearl/VanderWeele: TNDE+PNIE ("Total Natural Direct Effect" plus "Pure Natural Indirect Effect")
- Imai *et al.*'s exposition also includes the average of the two indirect effects and the average of the two direct effects, sometimes also expressed as proportions of the total effect that is explained by each kind of effect

Causal Mediation Assumptions



1. No omitted variable ("unmeasured confounding") in the X→Y equation
2. No unmeasured confounding in the M→Y equation
3. No unmeasured confounding in the X→M equation

- Randomization ensures that 1 and 3 hold, but not 2 because even if the treatment X is randomized, the mediator may not be
- Solutions: Instrumental variables and/or other methods for causal inference we will consider in subsequent weeks