Decomposing, Probing, and Plotting Interactions in Stata

https://stats.idre.ucla.edu/stata/seminars/interactions-stata/

Outline

- Following types of interactions (in linear regression):
  - Continuous by continuous
  - Continuous by categorical
  - Categorical by categorical

- *probe* or *decompose* (defined later) each of these interactions by asking the following research questions:
  - What is the *predicted* Y given a particular X and W? (predicted value)
  - What is *relationship* of X on Y at particular values of W? (simple slopes/effects)
  - Is there a *difference* in the relationship of X on Y for different values of W? (comparing simple slopes)

Requirements

- Basic notions of linear regression
- Stata installed
- Dataset loaded into Stata

```
use https://stats.idre.ucla.edu/wp-content/uploads/2020/06/exercise, clear
```

- Create value labels

```
label define progl 1 "jog" 2 "swim" 3 "read"
label define genderl 1 "male" 2 "female"
label values prog progl
label values gender genderl
```

- Download the complete Stata code here:
  - https://stats.idre.ucla.edu/wp-content/uploads/2020/07/interactions20200724.do

Introduction

- Motivation
- Main vs. Simple effects (slopes)
- Predicted Values vs. Slopes

Motivation

- Different types of questions
  - people who spend more time exercising lose more weight (simple regression)
  - more effort people put into their workouts, less time they need to spend exercising (cont x cont)
  - Females and males differ in the amount of weight they lose for the same amount of time (cat x cont)
  - Certain exercise programs may be more effective for females than males (cat x cat)
- Also, *visualize* the interaction to help us understand these relationships.

Weight Loss Study

- 900 participants in a year-long study
- **loss**: weight loss (continuous), positive = weight loss, negative scores = weight gain
- **hours**: hours spent exercising (continuous)
- **effort**: effort during exercise (continuous), 0 = minimal physical effort and 50 = maximum effort

- 3 different exercise programs, jogging, swimming and reading (control)
- **prog**: exercise program (categorical)
  - jogging=1
  - swimming=2
  - reading=3

- **gender**: participant gender (binary)
  - male=1
  - female=2

Definitions

- **decompose:**
  - break down the interaction into its lower order components (i.e., predicted means or simple slopes)
- **probe**:
  - hypothesis testing to assess the statistical significance of simple slopes and simple slope differences (i.e., interactions)
- **plot:**
  - visually display the interaction in the form of simple slopes such as values of the dependent variable are on the y-axis, values of the predictor is on the x-axis, and the moderator separates the lines or bar graphs
- Elements in the regression model
  - **DV**: dependent variable (Y), the outcome of your study (e.g., weight loss)
  - **IV**: independent variable (X), the predictor of your outcome (e.g., time exercising)
  - **MV**: moderating variable (W) or moderator, a predictor that changes the relationship of the IV on the DV (e.g, effort)
  - **coefficient**: estimate of the direction and magnitude of the relationship between an IV and DV
  - **continuous variable**: a variable that can be measured on a continuous scale, e.g., weight, height
  - **categorical or binary variable**: a variable that takes on discrete values, binary variables take on exactly two values, categorical variables can take on 3 or more values (e.g., gender, ethnicity)
- Elements of an interaction
  - **main effects** or slopes: effects or slopes for models that do not involve interaction terms
  - **simple slope**: when a continuous IV interacts with an MV, its slope at a particular level of an MV
  - **simple effect**: when a categorical IV interacts with an MV, its effect at a particular level of an MV

Regression (Main Effects) Model

- Outcome Y, two IV's X and W

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X + \hat{b}_2 W$$

  - **b0**: the intercept, or the predicted outcome when X=0 and W=0.
  - **b1**: the slope (or **main** effect) of X; for a one-unit change in X the predicted change in Y
  - **b2**: the slope (or **main** effect) of W; for a one-unit change in W the predicted change in Y
- Only intercept is interpreted at zero
- Interactions are formed by the *product* of any two variables.

Regression (Interaction) Model

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X + \hat{b}_2 W + \hat{b}_3 XW$$

- **b0**: the intercept, or the predicted outcome when X=0 and W=0.
- **b1**: the **simple** effect or slope of X, for a one-unit change in X the predicted change in Y at W=0
- **b2**: the **simple** effect or slope of W, for a one-unit change in W the predicted change in Y at X=0
- **b3**: the **interaction** of X and W, the change in the slope of X for a one unit increase in W (or vice versa)

- the intercept fixed at 0 of X and W,
- each coefficient of an IV interacted with an MV is interpreted at zero of the MV.
- effect X varies by levels of W
- identically, effect W varies by levels of X.

Regression (Interaction) Model

- X being the IV and W being the MV, rearrange:

$$\hat{Y} = \hat{b}_0 + \hat{b}_2 W + (\hat{b}_1 + \hat{b}_3 W) X$$

- coefficient for X is now **b1+b3***W
  - X is a function of W
- Ex. if W=0 slope of X is **b1**
- Ex. if W=1 slope of X is **b1+b3**
- **b3** additional increase in the effect or slope of X as W increases by one unit.

Predicted Values vs. Slopes

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 * \text{Hours}$$

```
regress loss hours
```

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hours |   2.469591   .9478805     2.61   0.009      .6092722     4.32991
       _cons |    5.07572   1.955005     2.60   0.010      1.238809    8.912632
------------------------------------------------------------------------------
```

$$\hat{\text{WeightLoss}} = 5.08 + 2.47 * \text{Hours}.$$

We can plug in **Hours=2** to get

$$\hat{\text{WeightLoss}} = 5.08 + 2.47(2) = 10.02.$$

predicted weight loss is 10.02 pounds from 2 hours of exercise

Stata's margins command

- `margins` command (Stata 11)
  - post-estimation command to obtain marginal means, predicted values and simple slopes.
  - run a model before running margins (regress)

```
margins, at(hours=2)


Expression   : Linear prediction, predict()
at           : hours           =           2

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    10.0149   .4685077    21.38   0.000     9.095405     10.9344
------------------------------------------------------------------------------
```

Understanding slopes in regression

$$b = \frac{\text{change in } Y}{\text{change in } X} = \frac{\Delta Y}{\Delta X}$$

$$\Delta Y = y_2 - y_1$$

$$\Delta X = x_2 - x_1$$

If delta X = 1, then $b = \Delta Y$

$$\widehat{\text{WeightLoss}} = 5.08 + 2.47 * \text{Hours}.$$

$$\widehat{\text{WeightLoss}}|_{\text{Hours}=0} = 5.08 + 2.47(0) = 5.08.$$

$$\widehat{\text{WeightLoss}}|_{\text{Hours}=1} = 5.08 + 2.47(1) = 7.55.$$

m = y2 – y1 = 7.55 − 5.08 = 2.47

Slopes in Stata

- instead of using the **`at`** option, we use the option **`dydx`** which stands for the slope

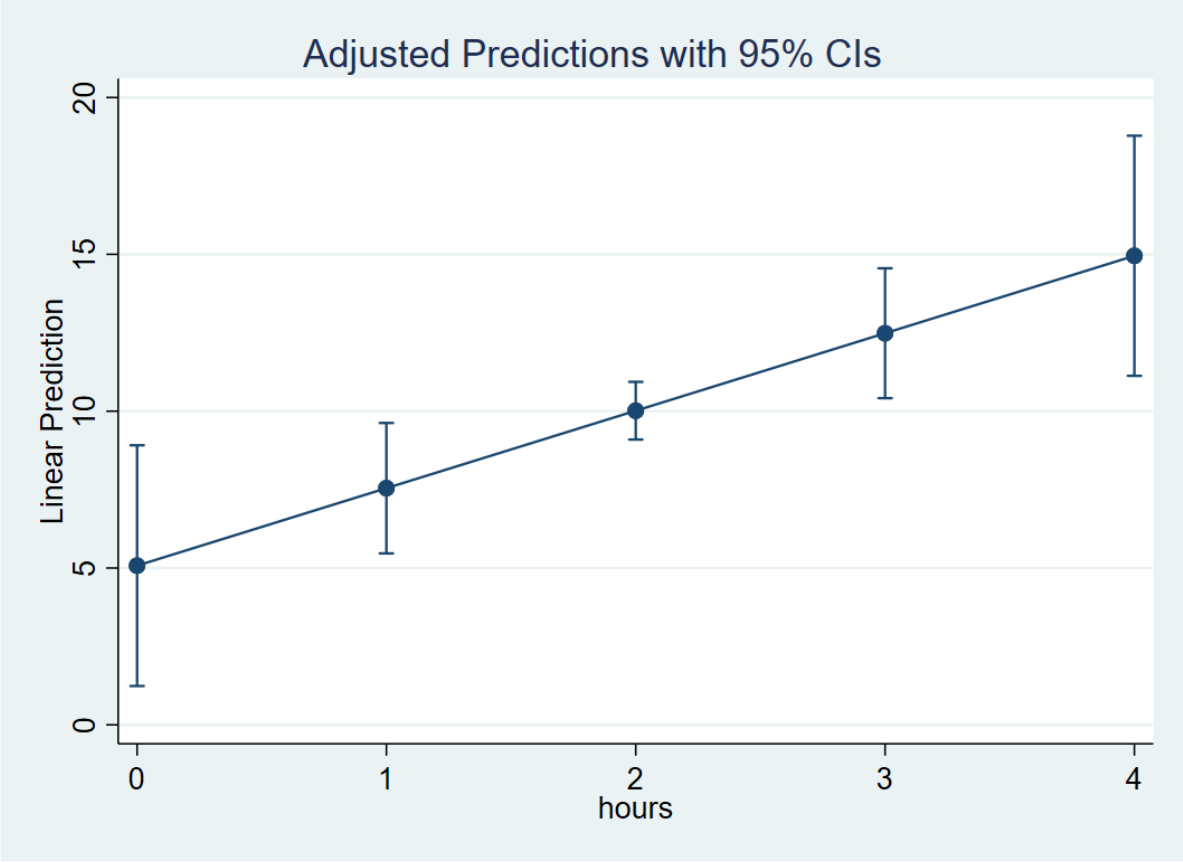
```
margins, dydx(hours)

Expression   : Linear prediction, predict()
dy/dx w.r.t. : hours

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hours |   2.469591   .9478805     2.61   0.009     .6092722     4.32991
------------------------------------------------------------------------------
```

$$b = \frac{\text{change in } Y}{\text{change in } X} = \frac{\Delta Y}{\Delta X}$$

Plotting a regression slope

```
margins, at(hours=(0(1)4))
marginsplot
```

Sequence 0, 1, 2, 3, 4



Look at the x-axis

Quiz #1

True or False?

In the margins command, the option **dydx** is used to estimate predicted values and **at** is used to estimate simple slopes.

Answers are on the last slide.

Exercise 1

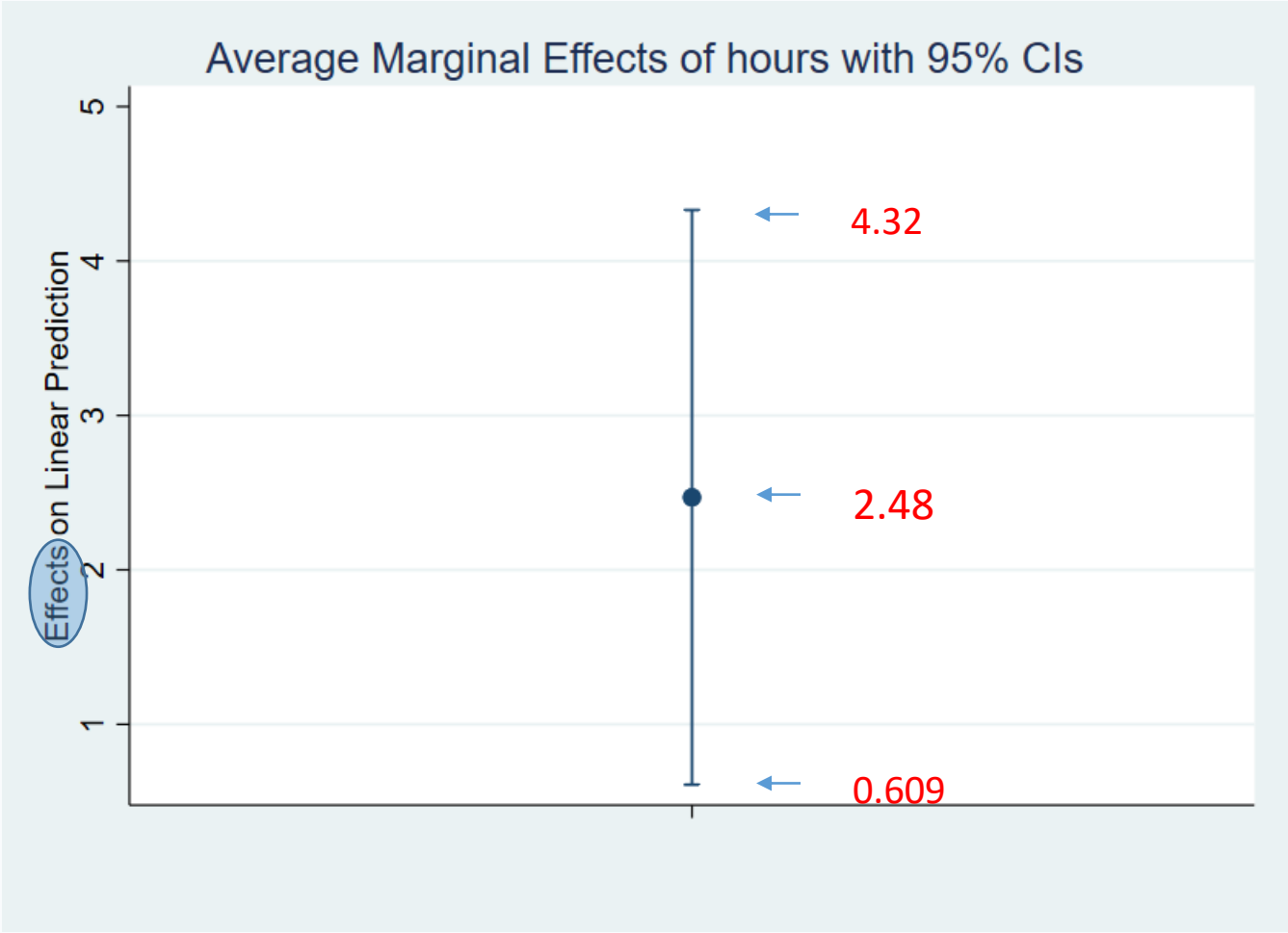
Refer to the following command

```
margins, at(hours=(0(1)4))
marginsplot
```

What would the plot look like if you replaced the first command with **`margins, dydx(hours)?`**

Answer is on the next slide.

Exercise 1 (solution)



	Delta-method dy/dx	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hours	2.469591	.9478805	2.61	0.009	.6092722	4.32991

Exercise 2

Predict two values of weight loss for Hours = 10 and Hours = 20 using **at**, then calculate the slope by hand.

How do the results compare with **dydx**?

Answer is on the next slide.

Exercise 2 (solution)

```
. margins, at(hours=(10 20))

Adjusted predictions                                  Number of obs     =        900
Model VCE    : OLS

Expression   : Linear prediction, predict()

1._at        : hours           =          10

2._at        : hours           =          20
```

	Delta-method Margin	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_at						
1	29.77163	7.595229	3.92	0.000	14.86517	44.6781
2	54.46754	17.066	3.19	0.001	20.97365	87.96144

```
. display (54.5-29.8)/(20-10)
2.47
```

Continuous by Continuous

- Model
- Plotting
- Simple slopes
- Differences in predicted values at fixed moderator values

Cont x Cont Model

- Does effort (W) moderate the relationship of Hours (X) on Weight Loss (Y)?

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1\text{Hours} + \hat{b}_2\text{Effort} + \hat{b}_3\text{Hours*Effort}$$

```
regress loss c.hours##c.effort
```

Equivalent to:

```
regress loss hours effort c.hours#c.effort
```

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hours |  -9.375681   5.663921    -1.66   0.098    -20.49178    1.740415
      effort |  -.0802763   .3846469    -0.21   0.835    -.8351902    .6746375
             |
c.hours#c.effort |   .3933468   .1875044     2.10   0.036     .0253478    .7613458
             |
       _cons |   7.798637   11.60362     0.67   0.502    -14.97479    30.57207
------------------------------------------------------------------------------
```

Model Output

$$WeightLoss = 7.8 - 9.4 * Hours - 0.08 * Effort + 0.39 * Hours * Effort$$

loss	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hours	-9.375681	5.663921	-1.66	0.098	-20.49178	1.740415
effort	-.0802763	.3846469	-0.21	0.835	-.8351902	.6746375
c.hours#c.effort	.3933468	.1875044	2.10	0.036	.0253478	.7613458
_cons	7.798637	11.60362	0.67	0.502	-14.97479	30.57207

- b0 **`_cons`**: intercept, or the predicted outcome when Hours = 0 and Effort = 0.
- b1 **`hours`**: **simple** slope of Hours, for a one unit change in Hours, the predicted change in weight loss at Effort=0.
- b2 **`effort`**: **simple** slope of Effort, for a one unit change in Effort the predicted change in weight loss at Hours=0.
- b3 **`c.hours#c.effort`**: **interaction** of Hours and Effort, change in the slope of Hours for every one unit increase in Effort (or vice versa).

Extrapolation (not good)

- we want to find the predicted weight loss given two hours of exercise and an effort of 30.

```
summarize effort
```

Variable	Obs	Mean	Std. Dev.	Min	Max
effort	900	29.65922	5.142764	12.949	44.07604

```
margins, at(hours=2 effort=30)
```

predicted weight loss is 10.2 pounds if we put in two hours of exercise and an effort level of 30

	Delta-method				
	Margin	Std. Err.	t	P>|t|	[95% Conf. Interval]
_cons	10.23979	.4530791	22.60	0.000	9.350572 11.12901

Extrapolation (not good)

Predicted weight loss is -10.2 pounds (!!) if we put in two hours of exercise and an effort level of 0. We gain weight from exercising if effort is zero! Nobody in the sample had an effort of zero. (Unlikely scenario)

```
margins, at(hours=2 effort=0)
Expression   : Linear prediction, predict()
at           : hours           =           2
               effort          =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -10.95273   2.647602    -4.14   0.000    -16.14895   -5.756502
------------------------------------------------------------------------------
```

Spotlight analysis (cont x cont)

There are an infinite number of (non-extrapolated) simple slopes, use

- **prior research** to guide you
- **spotlight analysis**: high, medium or low

high

medium

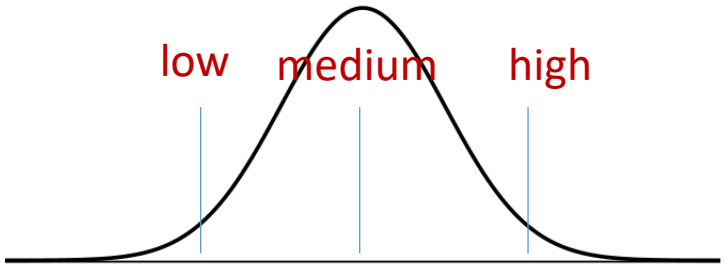
low

$$\text{EffA} = \overline{\text{Effort}} + \sigma(\text{Effort})$$
$$\text{Eff} = \overline{\text{Effort}}$$
$$\text{EffB} = \overline{\text{Effort}} - \sigma(\text{Effort}).$$

```
summarize effort
return list
```

```
r(mean) =  29.65921892801921

r(sd) =  5.142763519103716
```

```
global effa = round(r(mean) + r(sd),0.1)
global eff = round(r(mean),0.1)
global effb = round(r(mean) - r(sd),0.1)
```




```
display $effa
34.8

display $effb
24.5
```

Spotlight analysis output

```
margins, dydx(hours) at(effort=($effa $eff $effb))
```

```
1._at          : effort          =          34.8

2._at          : effort          =          29.7

3._at          : effort          =          24.5
```

Can we marginsplot after this?

```
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
         _at |
          1  |   4.312787   1.308387     3.30   0.001     1.744927    6.880646
          2  |   2.306719   .9148823     2.52   0.012       .511157    4.102281
          3  |   .2613151   1.352052     0.19   0.847    -2.392243    2.914874
------------------------------------------------------------------------------
```

Slope of Hours is 4.31 at Effort = 34.8 (High)

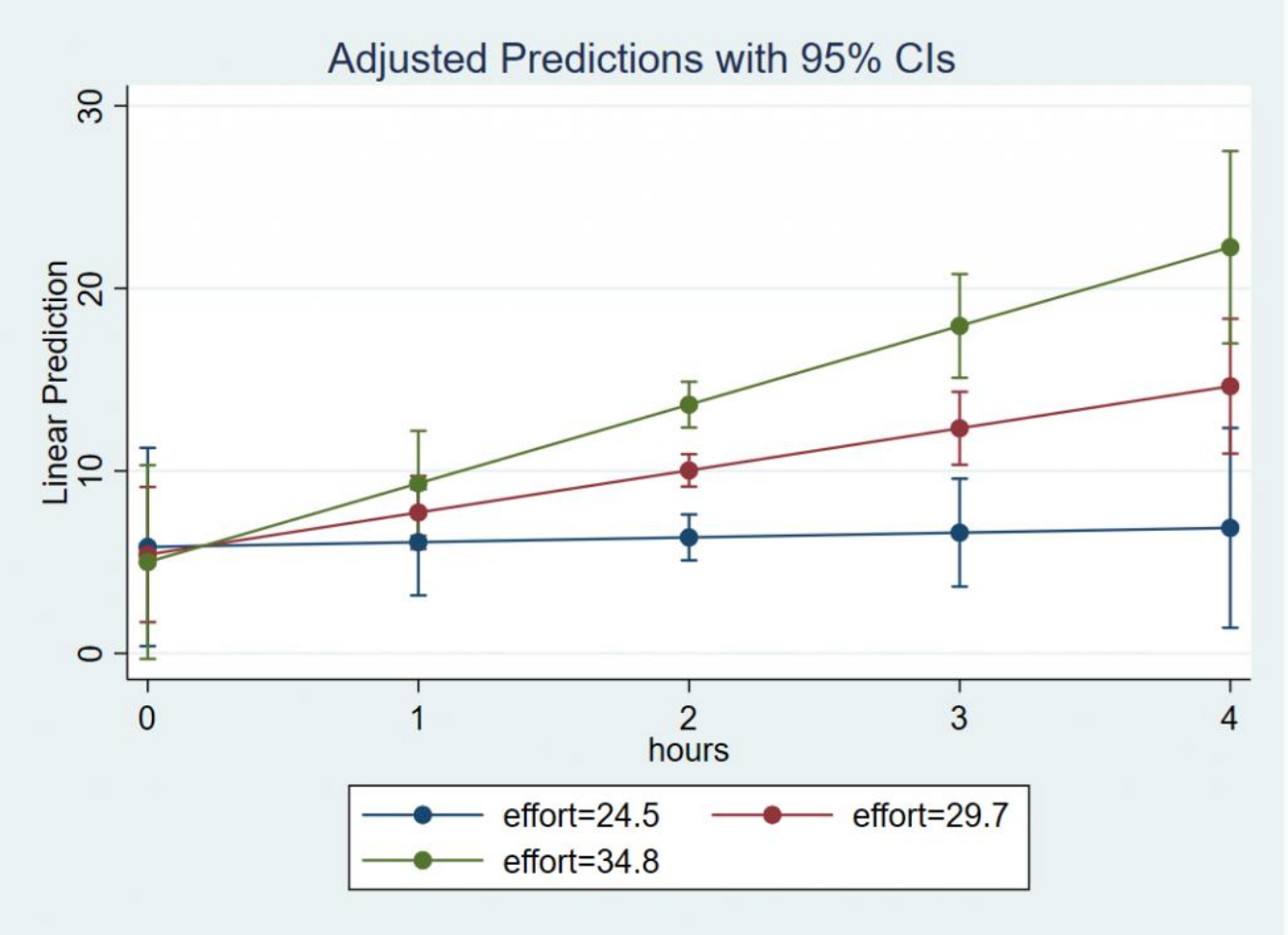
Plotting cont x cont interaction

order matters

x-axis

split lines

```
margins, at(hours=(0(1)4) effort=($effa $eff $effb))
marginsplot
```



hours spent exercising is only effective for weight loss if we put in more effort (HIIT)

Quiz #2

True or False?

The command **`margins, at(hours=(0(1)4) effort=($effa $eff $effb))`** tells Stata to plot Hours as the independent variable and Effort as the moderator.

Testing simple slopes (cont x cont)

```
margins, dydx(hours) at(effort=($effa $eff $effb)) pwcompare(effects)
```

```
Expression   : Linear prediction, predict()
dy/dx w.r.t. : hours

1._at        : effort          =        34.8

2._at        : effort          =        29.7

3._at        : effort          =        24.5
```

Recall simple slopes of hours

1	4.312787
2	2.306719
3	.2613151

2.30 – 4.31 = –2.01

	Contrast	Delta-method Std. Err.	Unadjusted t	Unadjusted P>|t|	Unadjusted [95% Conf.	Interval]
hours						
_at						
2 vs 1	-2.006068	.9562721	-2.10	0.036	-3.882862	-.1292739
3 vs 1	-4.051472	1.931295	-2.10	0.036	-7.841861	-.2610826
3 vs 2	-2.045404	.975023	-2.10	0.036	-3.958999	-.1318087

T- and P- values compared to Interaction

From **regress**

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
                        .... (output omitted) ...
c.hours#c.effort |   .3933468   .1875044     2.10   0.036     .0253478    .7613458
------------------------------------------------------------------------------
```

From **margins**

Contrast	Unadjusted	
dy/dx	t	P>|t|
-2.006068	-2.10	0.036
-4.051472	-2.10	0.036
-2.045404	-2.10	0.036

Notice sign flip of t-statistic

Exercise 3 (Challenge)

Recreate the interaction using **`margins`** and **`pwcompare`**

```
regress loss hours effort c.hours#c.effort
```

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
                          .... (output omitted) ...
c.hours#c.effort |   .3933468   .1875044     2.10   0.036     .0253478    .7613458
------------------------------------------------------------------------------
```

Note: this exercise is exclusive to the slides!
Answer is given on the next slide.

Answer to Exercise 3

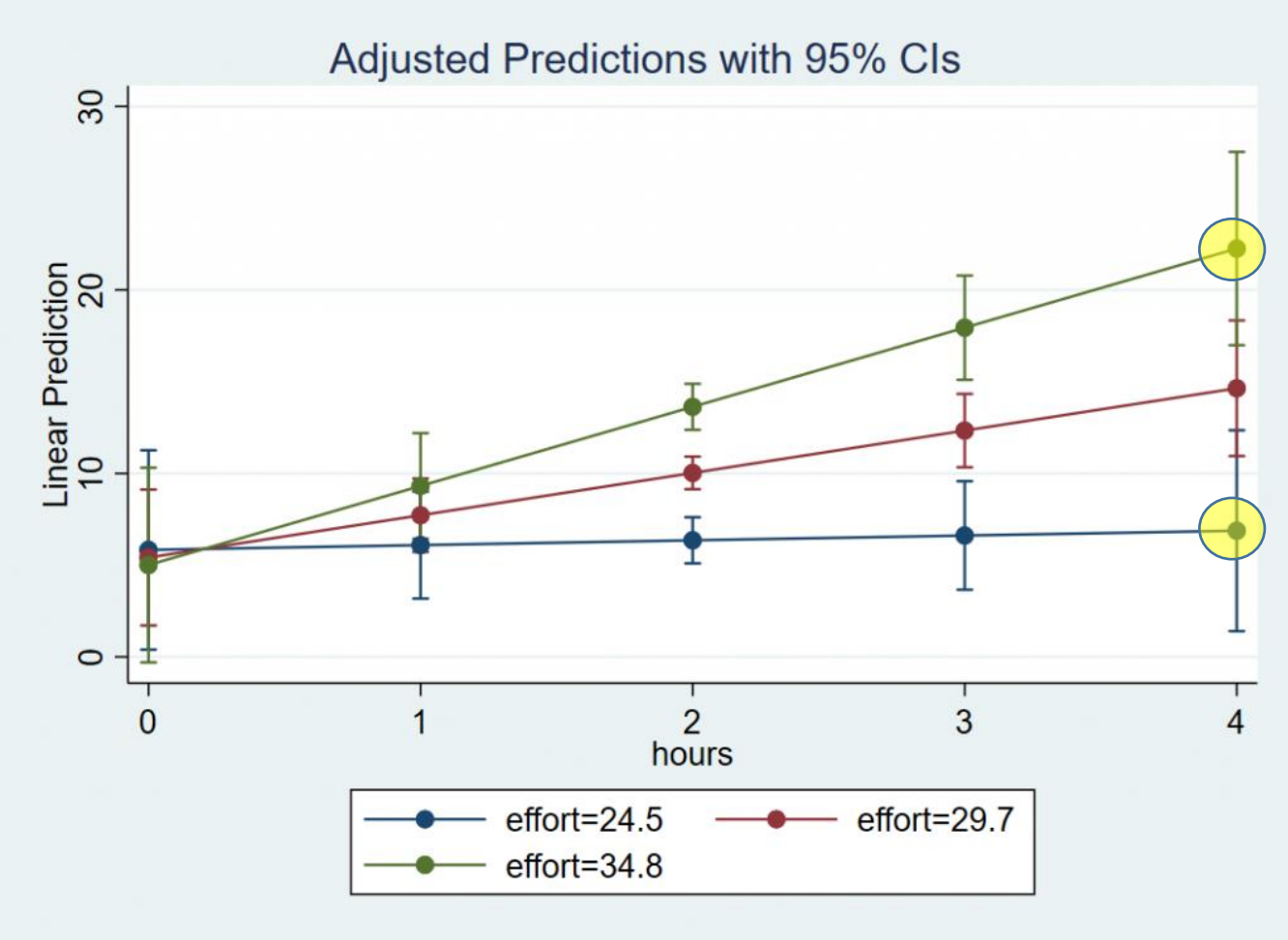
```
. margins, dydx(hours) at(effort=(0 1))

Average marginal effects                        Number of obs     =        900
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : hours

1._at        : effort          =           0

2._at        : effort          =           1
```

	dy/dx	Delta-method Std. Err.	t	P>|t|	[95% Conf.	Interval]
hours						
_at						
1	-9.375681	5.663921	-1.66	0.098	-20.49178	1.740415
2	-8.982335	5.478961	-1.64	0.101	-19.73543	1.770757

-8.982- (-9.376) = 0.394

Testing differences in predicted values

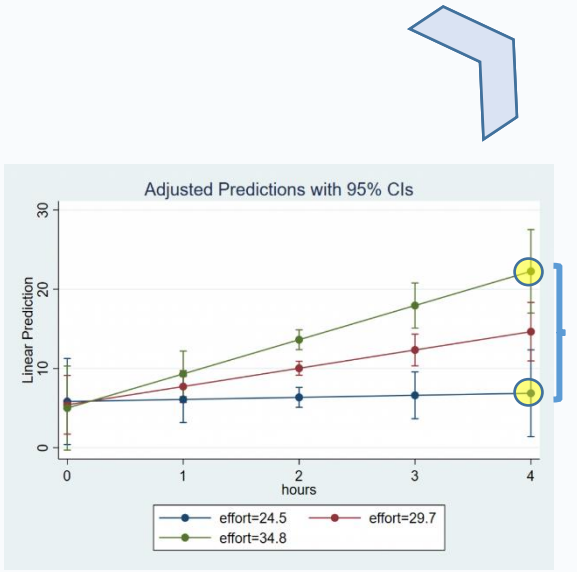
- Instead of testing the difference in slopes (lines), test difference of two predicted values (points)



Testing differences in predicted values

```
margins, at(hours=4 effort=($effa $effb))
```

```
Expression   : Linear prediction, predict()

1._at        : hours           =           4
               effort          =        34.8

2._at        : hours           =           4
               effort          =        24.5

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   22.25617   2.683877     8.29   0.000     16.98875    27.52359
          2  |   6.877127   2.789889     2.47   0.014     1.401648    12.35261
------------------------------------------------------------------------------
```

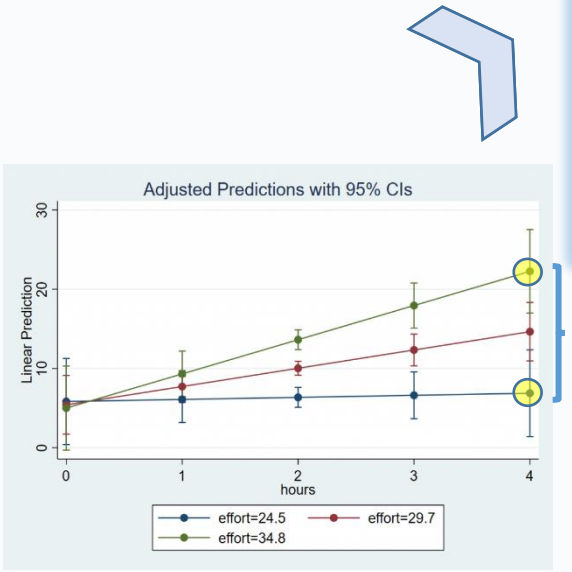


Testing differences in predicted values

```
margins, at(hours=4 effort=($effa $effb)) pwcompare(effects)
```

```
------------------------------------------------------------------------------
             |            Delta-method    Unadjusted           Unadjusted
             |   Contrast   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
    2 vs 1   |  -15.37904   3.972949    -3.87   0.000    -23.17641    -7.58167
------------------------------------------------------------------------------
```

6.88-22.26 = -15.38



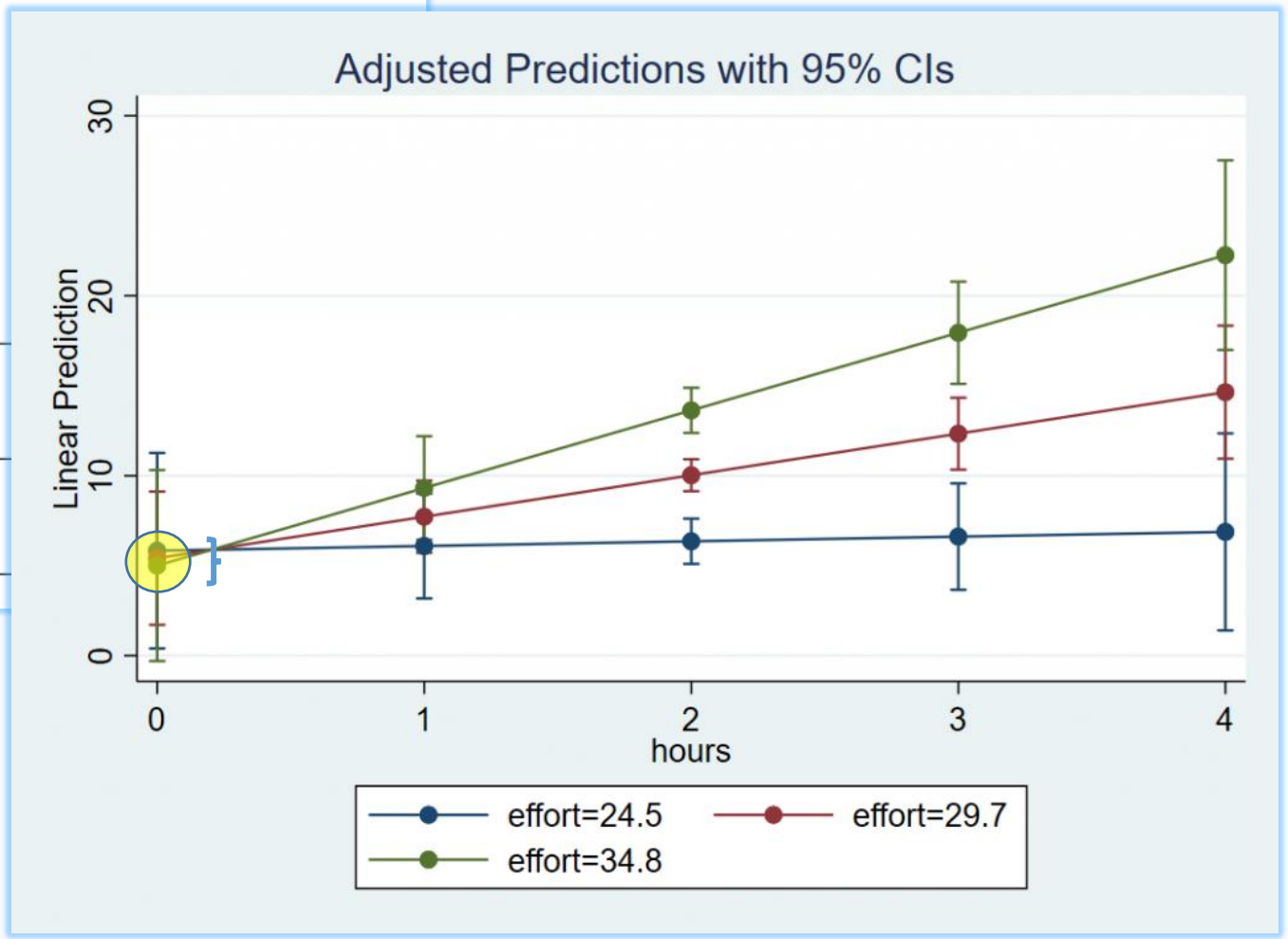
Exercise 4

Estimate the difference in Weight Loss for Low versus High levels of Effort at Hours=0. What is the actual value from Stata? Verify with plot.

Answer to Exercise 4

```
. margins, at(hours=0 effort=($effa $effb)) pwcompare(effects)

Pairwise comparisons of adjusted predictions     Number of obs     =        900
Model VCE    : OLS

Expression   : Linear prediction, predict()

1._at        : hours           =           0
               effort          =        34.8

2._at        : hours           =           0
               effort          =        24.5
```

	Contrast	Delta-method Std. Err.	Unadjusted t	P>|t|
_at 2 vs 1	.8268463	3.961863	0.21	0.835



Continuous by Categorical

- Dummy Coding
- Model
- Simple slopes
- Plotting

Dummy coding

$$D = \begin{cases} 1 & \text{if } X = x \\ 0 & \text{if } X \neq x \end{cases}$$

$$D_{female} = 1 \text{ if } Gender = female$$
$$D_{female} = 0 \text{ if } Gender = male$$

Note: only $k - 1$ dummy codes are required in the regression model
e.g., For gender, $k = 2$ so only 1 dummy code is required

```
tab gender
```

```
tab gender, nolabel
```

gender	gender	Freq.	Percent	Cum.
male	1	450	50.00	50.00
female	2	450	50.00	100.00
Total	Total	900	100.00	

DFEMALE = 0 if Gender = 1

DFEMALE = 1 if Gender = 2

Dummy codes in regression

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1\text{Hours} + \hat{b}_2 D_{male} + \hat{b}_3\text{Hours} * D_{male}$$

`i.` notation makes the lowest value the reference group (Gender = 1 or males)

```
regress i.gender
```

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |
      female |   -.1841939    .940519    -0.20   0.845    -2.030065    1.661677
       _cons |    10.11293   .6650473    15.21   0.000     8.807707    11.41816
------------------------------------------------------------------------------
```

Changing the reference group

$$\widehat{WeightLoss} = \hat{b}_1 + \hat{b}_2 D_{male}$$

ib2. means make the value of 2 the reference group (Gender = 2 or females)

```
regress loss ib2.gender
                          ... (output omitted) ...
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |
        male |   .1841939    .940519     0.20   0.845    -1.661677    2.030065
       _cons |   9.928741   .6650473    14.93   0.000     8.623513    11.23397
------------------------------------------------------------------------------
```

Quiz #3

Multiple Choice

Refer to the equation

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 \text{Hours} + \hat{b}_2 D_{male} + \hat{b}_3 \text{Hours} * D_{male}$$

What would the equation look like if we made males the reference group?

Answer 1: Leave the equation as is.

Answer 2: Change DMALE to DFEMALE.

Answer 3: Add b4*DFEMALE and b5*HOURS*DFEMALE to the equation

Quiz #4

Multiple Choice

Suppose gender = 1 codes for Male and gender = 2 codes for Female. Write the regression equation for the Stata command **`regress i.gender`**

Answer 1: WEIGHTLOSS = b0 + b1*DFEMALE

Answer 2: WEIGHTLOSS = b0 + b1*DMALE

Answer 3: WEIGHTLOSS = b0 + b1*DMALE + b2*DFEMALE

Cont x Cat Model

- Do men and women (MV) differ in the relationship between Hours (IV) and Weight loss?

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 \text{Hours} + \hat{b}_2 D_{male} + \hat{b}_3 \text{Hours} * D_{male}$$

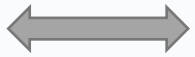
If interacted, the simple slopes are interpreted at 0 of the *other* variable

- b0 **`_cons`**: the intercept, or the predicted weight loss when Hours = 0 in the reference group of Gender, which is Dmale=0 or females.
- b1 **`hours`**: **simple slope** of Hours for the reference group Dmale=0 or females.
- b2 **`male`**: **simple effect** of Gender or the difference in weight loss between males and females at Hours = 0.
- b3 **`gender#c.hours`**: the **interaction** of Hours and Gender, the difference in the *simple slopes* of Hours for males versus females.

Simple slopes by cat moderator (cont x cat)

- simple slopes of Hours by gender

```
margins gender, dydx(hours)
```



```
margins, dydx(hours) over(gender)
```

```
Expression   : Linear prediction, predict()
dy/dx w.r.t. : hours

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
      gender |
        male |   1.591136     1.3523     1.18   0.240    -1.062908    4.245181
      female |   3.315068   1.331649     2.49   0.013     .7015529    5.928582
------------------------------------------------------------------------------
```

Quiz #5, 6, 7

True or False?

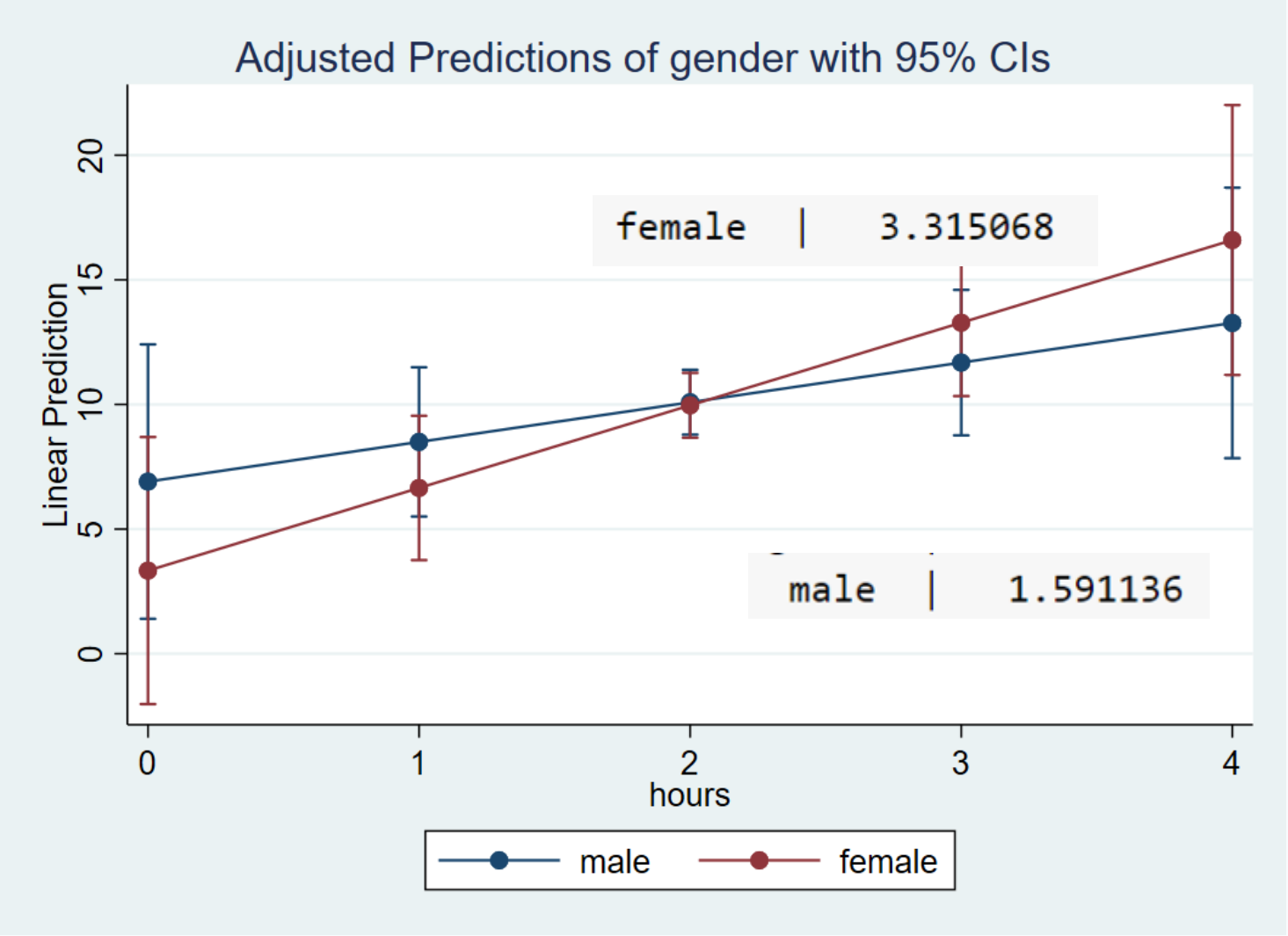
If both simple slopes of Hours for males and females are significantly different from zero, it implies that the interaction of Hours*Gender is not significant.

The command **`margins gender, dydx(hours)`** requests the simple effect of Gender split by levels of Hours.

The command **`margins gender, dydx(hours) pwcompare(effects)`** requests pairwise differences in the predicted values of Hours for females versus males.

Plotting cont x cat interaction

```
margins gender, at(hours=(0 1 2 3 4))
marginsplot
```



Quiz #8, 9

True or False?

Looking at the plot in the previous slide, since Hours is on the x-axis it is the IV and Gender separates the lines so it is the moderator (MV).

Multiple Choice

Refer to the command **`margins gender, at(hours=(0 1 2 3 4))`**. What is an equivalent way to specify the margins command above, so that we are clear that gender is the moderator?

Answer 1: margins, at(hours=(0 1 2 3 4)) over(gender)

Answer 2: margins hours#gender

Answer 3: margins gender#hours

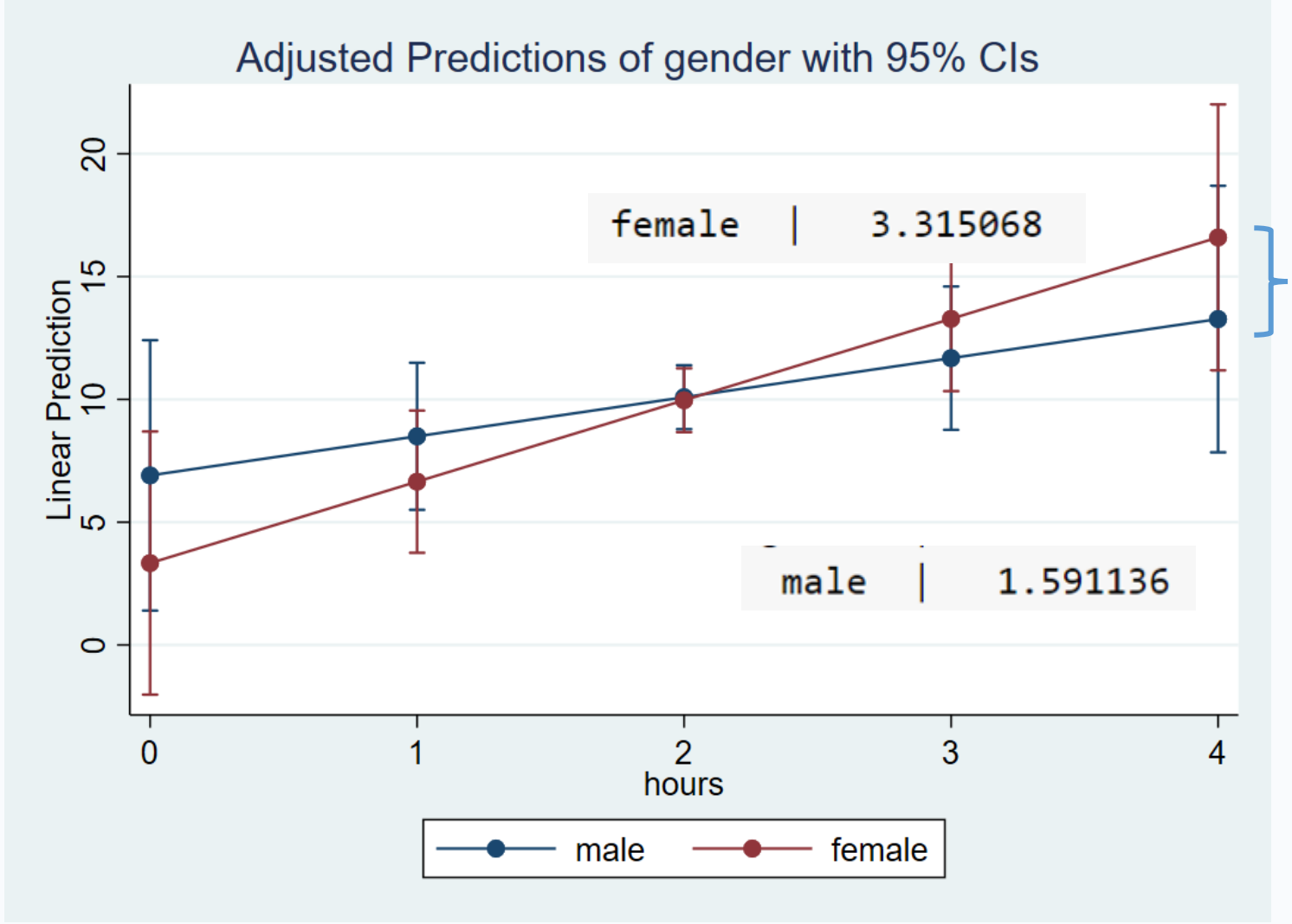
Testing differences in slopes

```
margins gender, dydx(hours) pwcompare(effects)
```

```
Expression   : Linear prediction, predict()
dy/dx w.r.t. : hours

------------------------------------------------------------------------------
             |   Contrast Delta-method    Unadjusted           Unadjusted
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
      gender |
female vs male |   1.723931   1.897895     0.91   0.364    -2.000906    5.448768
------------------------------------------------------------------------------
```

Testing differences in slopes (cont x cat)

```
margins gender, at(hours=(0 1 2 3 4))
marginsplot
```



3.315-1.591 = 1.724

Compare to regression table

```
regress loss c.hours##ib2.gender
```

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
                          ... (output omitted) ...
gender#c.hours |
        male  |  -1.723931   1.897895    -0.91   0.364    -5.448768    2.000906
              |
       _cons  |   3.334652    2.73053     1.22   0.222    -2.024328    8.693632
------------------------------------------------------------------------------
```

3.315-1.591 = 1.724

Why are the signs flipped?

Categorical by Categorical

- Model
- Simple effects
- Plotting

Dummy coding (2 categories)

$$D = \begin{cases} 1 & \text{if } X = x \\ 0 & \text{if } X \neq x \end{cases}$$

$$D_{male} = 1 \text{ if } Gender = male$$
$$D_{male} = 0 \text{ if } Gender = female$$

Note: only $k - 1$ dummy codes are required in the regression model
e.g., For gender, $k = 2$ so only 1 dummy code is required

```
tab gender    tab gender, nolabel

     gender |      gender |      Freq.     Percent        Cum.
------------+-------------+-----------------------------------
       male |           1 |        450       50.00       50.00
     female |           2 |        450       50.00      100.00
------------+-------------+-----------------------------------
      Total |       Total |        900      100.00
```

ib2.gender

DMALE = 0 if Gender = 2

DMALE = 1 if Gender = 1

Dummy Coding (3 categories)

- Does type of exercise (W) moderate the gender effect (X)?
  - do males and females lose weight differently depending on the type of exercise

$$D_{male} = 1 \text{ if } Gender = male$$
$$D_{male} = 0 \text{ if } Gender = female$$

only k-1 needed, k=2

$$D_{jog} = 1, D_{swim} = 0 \text{ if } Prog = jog$$
$$D_{jog} = 0, D_{swim} = 1 \text{ if } Prog = swim$$
$$D_{jog} = 0, D_{swim} = 0 \text{ if } Prog = read$$
$$D_{jog} = 1, D_{swim} = 1 \text{ if } Prog = ?$$

only k-1 needed, k=3

$$\widehat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 D_{male} + \hat{b}_2 D_{jog} + \hat{b}_3 D_{swim} + \hat{b}_4 D_{male} * D_{jog} + \hat{b}_5 D_{male} * D_{swim}$$

Value labels

Recall

```
label define progl 1 "jog" 2 "swim" 3 "read"
label define genderl 1 "male" 2 "female"
label values prog progl
label values gender genderl
```

Verify

```
tab prog
```

```
       prog |      Freq.     Percent        Cum.
------------+-----------------------------------
        jog |        300       33.33       33.33
       swim |        300       33.33       66.67
       read |        300       33.33      100.00
------------+-----------------------------------
      Total |        900      100.00
```

```
tab prog, nolabel
```

```
       prog |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        300       33.33       33.33
          2 |        300       33.33       66.67
          3 |        300       33.33      100.00
------------+-----------------------------------
      Total |        900      100.00
```

Stata i. notation

```
ib2.gender
```

Gender = 2 reference

Female

```
ib3.prog
```

DMALE

Prog = 3 reference

Reading

DJOG, DSWIM

Quiz #10

True or False

When we specify **`ib2.prog`** Stata internally creates two dummy variables for Categories 1 and 3

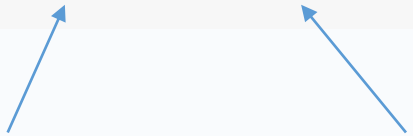
Cat x Cat Model

$$\widehat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 D_{male} + \hat{b}_2 D_{jog} + \hat{b}_3 D_{swim} + \hat{b}_4 D_{male} * D_{jog} + \hat{b}_5 D_{male} * D_{swim}$$

```
regress loss ib2.gender##ib3.prog
```

Equivalent to:

```
regress loss i.gender i.prog ib2.gender#ib3.prog
```



must have i. notation or Stata will think the variable is continuous

Model Interpretation (Cat x Cat)

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 D_{male} + \hat{b}_2 D_{jog} + \hat{b}_3 D_{swim} + \hat{b}_4 D_{male} * D_{jog} + \hat{b}_5 D_{male} * D_{swim}$$

loss	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gender						
male	-.3354569	.7527049	-0.45	0.656	-1.812731	1.141818
prog						
jog	7.908831	.7527049	10.51	0.000	6.431556	9.386106
swim	32.73784	.7527049	43.49	0.000	31.26057	34.21512
gender#prog						
male#jog	7.818803	1.064486	7.35	0.000	5.729621	9.907985
male#swim	-6.259851	1.064486	-5.88	0.000	-8.349033	-4.170669
_cons	-3.62015	.5322428	-6.80	0.000	-4.66474	-2.575559

- **b0 _cons:** intercept or the predicted weight loss when Dmale=0 and Djog=0,Dswim=0 (i.e., reading females)
- **b1 male:** simple effect of males for Djog=0,Dswim=0 (i.e., male – female weight loss in reading)
- **b2 jog:** simple effect of jogging when Dmale=0 (i.e., difference in weight loss between jogging vs reading for females)
- **b3 swim:** simple effect of swimming when Dmale=0 (i.e., difference in weight loss between swimming vs reading for females)

Model Interpretation (Cat x Cat)

$$\hat{\text{WeightLoss}} = \hat{b}_0 + \hat{b}_1 D_{male} + \hat{b}_2 D_{jog} + \hat{b}_3 D_{swim} + \hat{b}_4 D_{male} * D_{jog} + \hat{b}_5 D_{male} * D_{swim}$$

loss	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gender						
male	-.3354569	.7527049	-0.45	0.656	-1.812731	1.141818
prog						
jog	7.908831	.7527049	10.51	0.000	6.431556	9.386106
swim	32.73784	.7527049	43.49	0.000	31.26057	34.21512
gender#prog						
male#jog	7.818803	1.064486	7.35	0.000	5.729621	9.907985
male#swim	-6.259851	1.064486	-5.88	0.000	-8.349033	-4.170669
_cons	-3.62015	.5322428	-6.80	0.000	-4.66474	-2.575559

- **b4 male#jog:** interaction of Dmale and Djog, the male effect (male – female) in jogging vs the male effect in reading. Also, jogging effect (jogging – reading) for males vs the jogging effect for females
- **b5 male#swim:** interaction of Dmale and Dswim, the male effect (male – female) in swimming vs male effect in reading. Also, swimming effect (swimming- reading) for males vs the swimming effect for females

Interaction as the additional effect

$\hat{b}_1 + \hat{b}_4$ **`male+male#jog`** male effect for jogging

- **`b1 male:`** male effect (male – female) weight loss in reading
- **`b4 male#jog:`** male effect (male – female) in jogging vs the male effect in reading, (i.e., *additional* effect of jogging)

$\hat{b}_1 + \hat{b}_5$ **`male+male#swim`** male effect for swimming

- **`b5 male#swim:`** male effect (male – female) in swimming vs male effect in reading, (i.e., *additional* male effect for swimming)

Predicted Values (cat x cat)

```
margins gender#prog
```

categorical predictors come before comma (not an option)

```
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 gender#prog |
   male#jog  |   11.77203   .5322428    22.12   0.000     10.72744    12.81662
  male#swim  |   22.52238   .5322428    42.32   0.000     21.47779    23.56697
  male#read  |  -3.955606   .5322428    -7.43   0.000    -5.000197   -2.911016
 female#jog  |   4.288681   .5322428     8.06   0.000     3.244091    5.333272
female#swim  |   29.11769   .5322428    54.71   0.000      28.0731    30.16228
female#read  |   -3.62015   .5322428    -6.80   0.000     -4.66474   -2.575559
------------------------------------------------------------------------------
```

Simple effects not = interaction (cat x cat)

```
margins prog, dydx(gender)
```

Even though gender is a categorical variable we must specify dydx after comma

```
Expression   : Linear prediction, predict()
dy/dx w.r.t. : 1.gender

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.gender     |
        prog |
        jog  |   7.483346   .7527049     9.94   0.000     6.006072    8.960621
       swim  |  -6.595308   .7527049    -8.76   0.000    -8.072582   -5.118033
       read  |  -.3354569   .7527049    -0.45   0.656    -1.812731    1.141818
-------------+----------------------------------------------------------------
2.gender     |  (base outcome)
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Simple male effects

reference group, **ib2.gender**

Interaction = Difference of Simple Effects (continued)

```
regress loss i.gender i.prog ib2.gender#ib3.prog
```

```
------------------------------------------------------------------------------
        loss |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
                                ... (output omitted) ...
 male#swim   |  -6.259851   1.064486    -5.88   0.000    -8.349033   -4.170669
             |
       _cons |   -3.62015   .5322428    -6.80   0.000     -4.66474   -2.575559
------------------------------------------------------------------------------
```

```
margins prog, dydx(gender)
```

```
1.gender     |
        prog |
        jog  |   7.483346   .7527049     9.94   0.000     6.006072    8.960621
       swim  |  -6.595308   .7527049    -8.76   0.000    -8.072582   -5.118033
       read  |  -.3354569   .7527049    -0.45   0.656    -1.812731    1.141818
```

Male effect swimming

Male effect reading

-6.595 – (-.3354) = - 6.259 Difference of simple effects

$\hat{b}_1 + \hat{b}_5$ **male+male#swim** male effect for swimming Additional effect

Quiz #11, 12

True or False

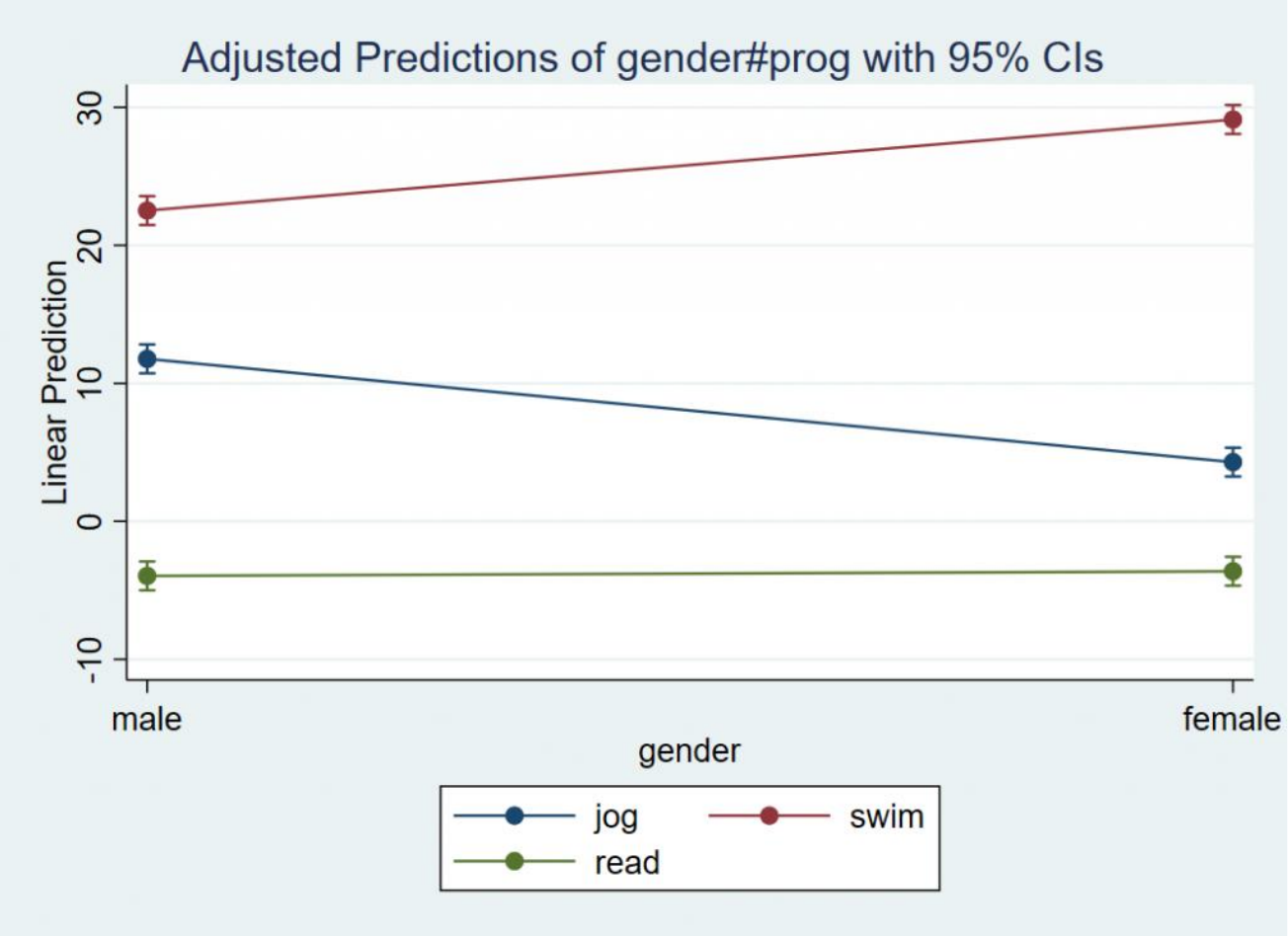
Compare to the Stata command **`regress loss ib2.gender##ib3.prog`**.
Equivalent syntax is **`regress loss gender prog ib2.gender#ib3.prog`**.

The interaction **`male#jog`** is the male effect for the jogging condition.

Plotting cat x cat interaction

```
margins gender#prog
marginsplot
```

both categorical so comes before comma



x-axis

```
margins iv#mv
```

separate lines

Quiz #13,14

True or False

The code **`margins prog#gender`** tells **`marginsplot`** that we want **`prog`** on the x-axis with lines corresponding to levels of **`gender`**.

Multiple Choice

How would we plot exercise type (prog) along the x-axis split by gender?

Answer 1: margins prog#gender followed by marginsplot

Answer 2: margins gender#prog followed by marginsplot

Answer 3: margins gender#prog followed by marginsplot, xdimension(gender)

Answers to Quiz Questions

1. F

2. T

3. Answer 2

4. Answer 1

5. F, The test of simple slopes is not the same as the test of the interaction, which tests the difference of simple slopes.

6. F, We are not obtaining the simple effect of Gender but simple slopes of Hours. The statement dydx(hours) indicates the simple slope we are requesting. Since gender is categorical, it comes before the comma which means we want the simple slope of Hours by Gender.

7. F, This is the pairwise difference in the slope of Hours for females versus males. Recall that dydx(hours) obtains simple slopes and at obtains predicted values.

8. T

9. Answer 1

10. T

11. F, Without the i. prefix for the simple effects, Stata treats gender and prog as continuous variables despite the correct ib#. specification in the interaction term.

12. F, The male jogging effect alone does not capture the interaction. The interaction is the difference of simple effects.

13. T

14. Answer 1