New methods of interpretation using marginal effects for nonlinear models

Scott Long[1]

[1]Departments of Sociology and Statistics
Indiana University

EUSMEX 2016: Mexican Stata Users Group

Mayo 18, 2016

Version: 2016-05-03b

Road map for talk

Goals

1. Demonstrate new methods for using marginal effects
2. Exploit the power of `margins`, factor syntax, and `gsem`
3. Illustrate the SPost13 `m*` commands

Outline

1. Statistical background
   - Binary logit model
   - Standard definitions of marginal effects
   - Generalizations of marginal effects
2. Stata commands
   - Estimation: factor notation, storing estimates, and `gsem`
   - Post-estimation: `margins` and `lincom`
   - SPost13's `m*` commands
3. Example: explaining the occurrence of diabetes

Logit model

Probability as outcome

1. Nonlinear in probabilities

$$\pi(\mathbf{x}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})} = \Lambda(\mathbf{x}'\boldsymbol{\beta})$$

2. Interpretation with marginal effect: additive change in π for change in x_k holding other variables at *specific values*

Odds as outcome

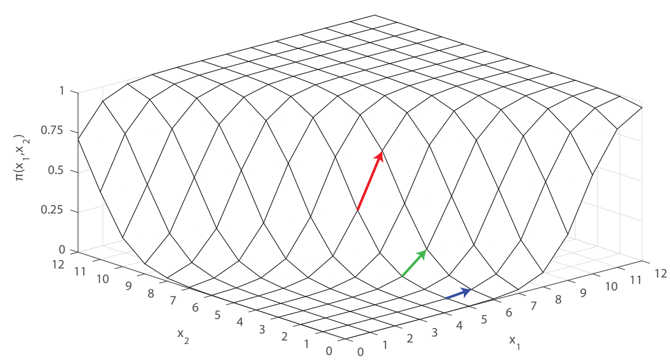
3. Multiplicative in odds

$$\Omega(\mathbf{x}) = \frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})} = exp(\mathbf{x}'\boldsymbol{\beta})$$

4. Interpretation with odds ratio: multiplicative change in $\Omega(\mathbf{x})$ for change in x_k holding other *variables constant*

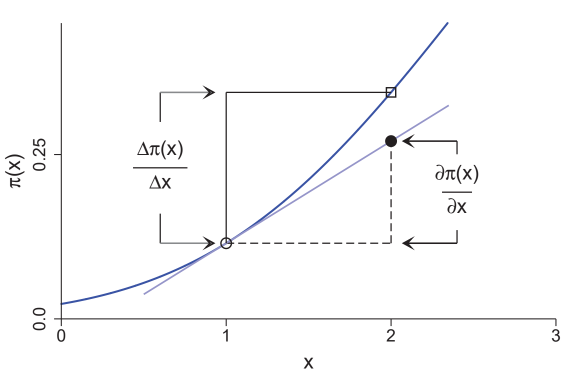
Logit model: nonlinear in probabilities

1. **Odds ratios**: identical at each arrow
2. **Marginal effects**: different at each arrow



Marginal and discrete change

1. **Marginal change**: instantaneous rate of change in $\pi(x)$
2. **Discrete change**: change in $\pi(x)$ for discrete change in x



Definition of discrete change

1. x_k changes from start to end
2. $\mathbf{x} = \mathbf{x}^*$ contains specific values of other variables
3. Discrete change of x_k

$$\text{DC}(x_x) = \frac{\Delta\pi(\mathbf{x})}{\Delta x_k(\text{start} \rightarrow \text{end})} = \pi(x_k = \text{end}, \mathbf{x} = \mathbf{x}^*) - \pi(x_k = \text{start}, \mathbf{x} = \mathbf{x}^*)$$

4. Interpretation

   For a change in x_k from start to end, the probability changes by $\text{DC}(x_k)$*, holding other variables at the specified values.*

Examples of discrete change

1. At observed values for observation i

$$\frac{\Delta\pi(\mathbf{x}_i)}{\Delta x_{ik}(x_{ik} \rightarrow x_{ik}+1)} = \pi(x_k = x_{ik}, \mathbf{x}_i) - \pi(x_k = x_{ik}+1, \mathbf{x}_i)$$

2. At representative values $\mathbf{x}^*$

$$\frac{\Delta\pi(\mathbf{x}^*)}{\Delta x_k(0 \rightarrow 1)} = \pi(x_k = 1, \mathbf{x}^*) - \pi(x_k = 0, \mathbf{x}^*)$$

3. Since $\Delta\pi \,/\, \Delta x_k$ depends on where it is evaluated, how should the effect of x_k be summarized?

Common summary measures of discrete change

Discrete change at the mean (DCM)

$$\mathrm{DCM}(x_k) = \frac{\Delta\pi(\overline{\mathbf{x}})}{\Delta x_k(\text{start} \rightarrow \text{end})} = \pi(x_k = \text{end}, \overline{\mathbf{x}}) - \pi(x_k = \text{start}, \overline{\mathbf{x}})$$

For someone who is average on all variables, increasing x_k from start to end changes the probability by DCM(x_k).

Average discrete change (ADC)

$$\mathrm{ADC}(x_k) = \frac{1}{N}\sum_{i=1}^{N} \frac{\Delta\pi(\mathbf{x} = \mathbf{x}_i)}{\Delta x_{ik}(\text{start} \rightarrow \text{end})}$$

On average, increasing x_k from start to end changes the probability by ADC(x_k).

Variation for computing discrete change

* indicates generalization of standard methods

Conditional and average change
- Conditional effects
  - At observed values
  - At mean values
  - At representative values
- Average effects
  - Average in full sample
  - Average in sub-sample *

Amount of change
- Additive change
- Proportional change *
- Changes as function of predictors *
- Change a component in multiplicative measure *

Number of variables changed
- One variable
- Two or more variables *

Comparing discrete changes

Comparisons within a model
- Effects of different variables
  - H_0: $DC(gender) = DC(age)$
- One variable's effect at different locations
  - H_0: $DC(age \mid age = 50, \mathbf{x}^*) = DC(age \mid age = 65, \mathbf{x}^*)$

Comparisons across models
- Different samples or groups
  - DC(weight) for whites compared to non-whites
- Model specifications
  - DC(weight) in different model specifications

Stata: Overview

1. Requires Stata 12 or later; some examples need Stata 14
2. Assumes `spost13_ado` package is installed
3. Estimation uses factor syntax
   - Logit model used but examples generalize
   - Survey estimation can be used
4. Post-estimation with `margins` and `lincom`
5. In Stata, `search eusmex2016` to download
   - `eusmex2016-effects-scott-long.do` and dataset
   - PDF of slides from talk
   - In the slides, [#xx] points to locations in the do-file

Stata: Estimation

1. Fitting a logit model

   `logit` *`dependent independent`* `[ ,` *`options`* `]`

2. Factor variable syntax

   `i.`*`var`*: categorical predictor (e.g., `i.female`)

   `c.`*`var`*: continuous predictor (e.g., `c.age`)

   `c.`*`var1`*`#c.var2`: product (e.g., `c.age#c.age` ≡ `c.age*c.age`)

3. Regression estimates are stored for later use

   `estimates store` *`ModelName`*

4. To replace current estimates with previously stored estimates

   `estimates restore` *`ModelName`*

Stata: post-estimation

1. `margins` estimates functions of predictions from regressions
2. `margins, post` stores these estimates to `e(b)` and `e(V)`
3. `lincom` estimates linear functions of `e(b)`
4. `mchange`, `mtable`, `mgen` and `mlincom` are SPost13 wrappers to generate complex `margins` commands and improve output

Example

1. Health and Retirement Survey[1]: cross-sectional data on health
2. Outcome is patient's report of having diabetes
3. Begin with standard marginal effects to introduce Stata tools
4. Use these tools to compute more complex marginal effects
5. Demonstrate methods for statistically comparing effects

[1] Steve Heeringa generously provided the data used in *Applied Survey Data Analysis* (Heeringa et al., 2010). Complex sampling is not used in my analyses.

Variables and descriptive statistics

```
. use hrs-gme-analysis2, clear
(hrs-gme-analysis2.dta | Health & Retirement Study GME sample | 2016-04-08)
```

Variable	Mean	Min	Max	Label
diabetes	.205	0	1	Respondent has diabetes?
white	.772	0	1	Is white respondent?
bmi	27.9	10.6	82.7	Body mass index
weight	174.9	73	400	Weight in pounds
height	66.3	48	89	Height in inches
age	69.3	53	101	Age
female	.568	0	1	Is female?
hsdegree	.762	0	1	Has high school degree?

Body mass index: $BMI = \frac{weight_{kg}}{height_m^2} = \frac{703 \times weight_{lb}}{height_{in}^2}$

Models of diabetes: estimate and store

1. Two models are fit [#02]
2. Model Mbmi measures body mass with the BMI index
```
logit diabetes c.bmi i.white c.age##c.age i.female i.hsdegree
estimates store Mbmi
```
3. Model Mwt measures body mass with height and weight
```
logit diabetes c.weight c.height i.white c.age##c.age i.female i.hsdegree
estimates store Mwt
```


Models of diabetes: odds ratios and p-values

Variable	Mbmi	Mwt
bmi	1.1046*	
weight		1.0165*
height		0.9299*
white White	0.5412*	0.5313*
age	1.3091*	1.3093*
c.age#c.age	0.9983*	0.9983*
female Women	0.7848*	0.8743#
hsdegree HS degree	0.7191*	0.7067*
_cons	0.0000*	0.0001*
bic	14991.26	14982.03

Note: # significance at .05 level; * at the .001 level.

Summarizing effects with average discrete change

1. `mchange` from SPost13 is a great first step for assessing effects [#03]
```
. estimates restore Mbmi
. mchange, amount(sd)
logit: Changes in Pr(y) | Number of obs = 16071
```

	Change	p-value
white		
White vs Non-white	-0.099	0.000
bmi		
+SD	0.097	0.000

(*output omitted*)

2. Interpretation

On average the probability of diabetes is .099 less for white respondents than non-white respondents.

Increasing BMI by one standard deviation on average increases the probability of diabetes .097.

3. Where did these numbers come from?

Tool: margins, at(...) and atmeans

1. By default, margins
   1.1 Computes prediction for each observation
   1.2 Then it averages these predictions
2. Average prediction assuming everyone is white

   `margins, at(white=1)`
3. Two average predictions

   `margins, at(white=1) at(white=0)`
4. Prediction if white with means for other variables

   `margins, at(white=1) atmeans`

ADC for binary x_k: ADC(white)

1. ADC for white equals

   $$\text{ADC} = \frac{1}{N}\sum_i \pi(\text{white}=1, \mathbf{x}=\mathbf{x}_i) - \frac{1}{N}\sum_i \pi(\text{white}=0, \mathbf{x}=\mathbf{x}_i)$$
2. margins computes the two average predictions [#04]

```
. margins, at(white=0) at(white=1) post

Expression   : Pr(diabetes), predict()
1._at        : white          =          0
2._at        : white          =          1

             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .2797806   .0073107    38.27   0.000      .265452    .2941092
          2  |   .1805306   .0034215    52.76   0.000     .1738245    .1872367
```

3. 1._at is the average treating everyone as nonwhite

   $$1.\_at = \frac{1}{N}\sum_i \pi(\texttt{white}=0, \mathbf{x}=\mathbf{x}_i)$$
4. 2._at is the average treating everyone as white

ADC for binary x_k: ADC(white)

5. The post option saves the average probabilities

```
. matlist e(b)

             |        1.          2.
             |       _at         _at
-------------+------------------------
          y1 |  .2797806    .1805306
```

6. lincom computes ADC as difference in predictions in e(b)

```
. lincom _b[2._at] - _b[1._at]
 ( 1)  - 1bn._at + 2._at = 0

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |    -.09925   .0082362   -12.05   0.000    -.1153927   -.0831073
------------------------------------------------------------------------------
```

7. Interpretation

   On average, being white decreases the probability of diabetes by .099 ($p < .001$).

TOOL: mlincom simplifies lincom

1. lincom requires column names from e(b) that can be complex

```
lincom (_b[2._at#1.white] - _b[1._at#1.white]) ///
     - (_b[2._at#0.white] - _b[1._at#0.white])
```

2. mlincom uses column numbers which are rows in margins output

```
mlincom (4-2) - (3-1)
```


Tool: margins, at(... = gen(...))

1. at(... = gen(...)) generates new values from observed values
2. Trivially, predictions with observed values of bmi

   `margins, at( bmi = gen(bmi) )`
3. Predictions with observed values of bmi plus 1

   `margins, at( bmi = gen(bmi + 1) )`
4. Both observed and observed plus 1

   `margins, at( bmi = gen(bmi) ) at( bmi = gen(bmi + 1) )`
5. Observed plus a standard deviation

```
1] quietly sum bmi
2] local sd = r(sd)
3] margins, at( bmi = gen(bmi + `sd') )
```


ADC for continuous x_k: ADC(bmi)

1. Compute probabilities at observed bmi and observed+sd [#05]

```
. quietly sum bmi
. local sd = r(sd)

. margins, at(bmi = gen(bmi)) at(bmi = gen(bmi + `sd')) post

Expression   : Pr(diabetes), predict()
1._at        : bmi            = bmi
2._at        : bmi            = bmi + sd

             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .2047166   .0030338    67.48   0.000     .1987704    .2106627
          2  |   .3017056    .005199    58.03   0.000     .2915159    .3118954
```

2. ADC(bmi+sd)

```
. mlincom 2 - 1, stats(all)
             |  lincom        se    zvalue    pvalue        ll        ul
-------------+------------------------------------------------------------
           1 |   0.097     0.004    27.208     0.000     0.090     0.104
```

On average, increasing BMI by one standard deviation, about 6 points, increases the probability of diabetes by .097 ($p < .001$).

Tool: `mtable` wrapper for `margins`

1. `margins` output is complete, not compact
2. `mtable` executes `margins`, then simplifies output (and more)
   - `mtable, commands` lists the `margins` commands used
   - `mtable, detail` shows `margins` output and `mtable` output

DCM for continuous x_k: DCM(bmi)

1. Let `bmi` increase from mean to mean+SD [#06]

```
. qui sum bmi
. local mn = r(mean)
. local mnplus = r(mean) + r(sd)
```

2. Option `atmeans` holds other variables at their means

```
. margins, atmeans at(bmi = `mn´) at(bmi = `mnplus´) post

Expression   : Pr(diabetes), predict()
1._at        : bmi             =    27.89787
               0.white         =    .2284239 (mean)
               1.white         =    .7715761 (mean)
               age             =    69.29276 (mean)
               0.female        =    .4315226 (mean)
               1.female        =    .5684774 (mean)
               0.hsdegree      =    .2375086 (mean)
               1.hsdegree      =    .7624914 (mean)
2._at        : bmi             =     33.6687
               0.white         =    .2284239 (mean)
               1.white         =    .7715761 (mean)
<continued>
```


DCM for continuous x_k: DCM(bmi)

```
               age             =    69.29276 (mean)
               0.female        =    .4315226 (mean)
               1.female        =    .5684774 (mean)
               0.hsdegree      =    .2375086 (mean)
               1.hsdegree      =    .7624914 (mean)

-------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------
         _at |
          1  |   .2097641   .0045531    46.07   0.000     .2008401    .2186881
          2  |   .3202789   .0066246    48.35   0.000      .307295    .3332628
-------------------------------------------------------------------------
```


DCM for continuous x_k: DCM(bmi)

2. Alternatively, `mtable` runs `margins` and reformats the results

```
. mtable, atmeans at(bmi = `mn´) at(bmi = `mnplus´) post
Expression: Pr(diabetes), predict()

          |      bmi     Pr(y)
----------+-------------------
        1 |     27.9     0.210
        2 |     33.7     0.320
Specified values of covariates

          |        1.                  1.          1.
          |     white       age    female    hsdegree
----------+-----------------------------------------
  Current |      .772      69.3      .568        .762
```

3. DCM(bmi+sd)

```
. mlincom 2 - 1

          |   lincom    pvalue        ll        ul
----------+---------------------------------------
        1 |    0.111     0.000     0.102     0.119
```

For an average person, increasing BMI by one standard deviation increases the probability of diabetes by .111 ($p < .001$).

Proportional change in x_k: changing weight

1. Body mass be measured with height and weight

```
logit diabetes c.weight c.height ///
    i.white c.age##c.age i.female i.hsdegree, or
estimates store Mwt
```

2. ADC(weight) increases weight by a constant, say 25 pounds
3. A 25 pound increase in weight means different things
   - A 25% increase from 100 pounds
   - At 14% increase from average weight
   - An 8% increase from 300 pounds
4. The effect of a percentage increase could be more useful than the effect of a 25 pound increase

Proportional change in x_k: ADC(weight+25)

1. Computing ADC(weight + 25) [#07]

```
. estimates restore Mwt

. mtable, at(weight = gen(weight)) at(weight = gen(weight + 25)) post
Expression: Pr(diabetes), predict()

          |    Pr(y)
----------+---------
        1 |    0.205
        2 |    0.271

. quietly mlincom 2 - 1, rowname(ADC add) clear
```

Proportional change in x_k: ADC(weight*1.14)

2. A simple change to gen() computes proportional change

```
. estimates restore Mwt
. mtable, at(weight = gen(weight)) at(weight = gen(weight * 1.14)) post
Expression: Pr(diabetes), predict()

              Pr(y)
  ----------------
     1        0.205
     2        0.273

. mlincom 2 - 1, rowname(ADC pct) add

               lincom     pvalue         ll         ul
  ------------------------------------------------------
   ADC add      0.067      0.000      0.062      0.071
   ADC pct      0.068      0.000      0.063      0.073
```

3. The average effects are close, but is the average a good summary?

Tool: margins, generate()

1. margins, gen(*stub*) creates variables containing predictions for each observation (help margins generate)
2. For example, to save probabilities for 16,071 cases and average them

```
. margins, gen(Prob) at(weight = gen(weight))
Predictive margins                              Number of obs     =     16,071
Expression   : Pr(diabetes), predict()

             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2047166   .0030316    67.53   0.000     .1987747    .2106584
------------------------------------------------------------------------------

. sum Prob1 // matches margins estimate
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       Prob1 |     16,071    .2047166    .1229016   .0123593   .9067207
```

3. Note that gen() is used two ways

Proportional change in x_k: generating variables

1. For ADC(weight*1.14) compute effect and and create variables

```
. mtable, gen(PRpct) at(weight=gen(weight)) at(weight=gen(weight*1.14)) post
Expression: Pr(diabetes), predict()

             Pr(y)
  ---------------
     1       0.205
     2       0.273

. mlincom 2-1, rowname(ADC percent)

                 lincom     pvalue         ll         ul
  -------------------------------------------------------
  ADC percent     0.068      0.000      0.063      0.073
```

2. Compute DC(weight*1.14) for each observation

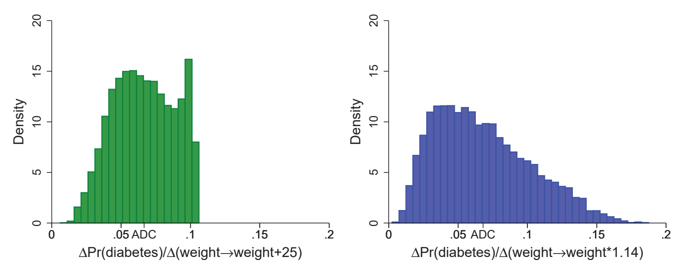
```
. generate DCpct = PRpct2 - PRpct1
. lab var DCpct "DC for 14 percent increase in weight"
```


Proportional change in x_k: generating variables

3. Similarly, ADC(weight + 25)

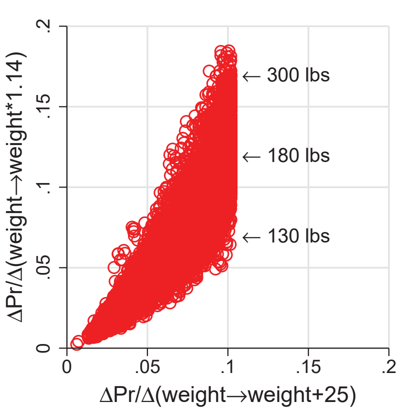
```
. mtable, gen(PRadd) at(weight=gen(weight)) at(weight=gen(weight+25)) post
  (output omitted)
. generate DCadd = PRadd2 - PRadd1
. lab var _DCadd "DC for 25 pound increase"
```

4. DC(weight_i*1.14) and DC(weight_i+25) have quite different distributions



Proportional change in x_k: comparing ADCs

5. Average effects are close, but individual effects can differ greatly



Discrete change with polynomials

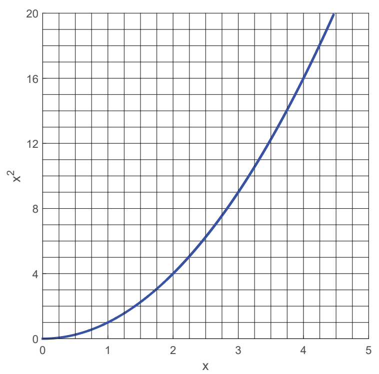
1. A standard discrete change allows only one variable to change
2. With polynomials multiple variables must change together
   - You can't change age, holding age-squared constant
3. For example,

$$\frac{\Delta\pi(\mathbf{x})}{\Delta\text{age}(50 \to 60)} = \pi(\text{age}=60, \text{agesq}=60^2) - \pi(\text{age}=50, \text{agesq}=50^2)$$

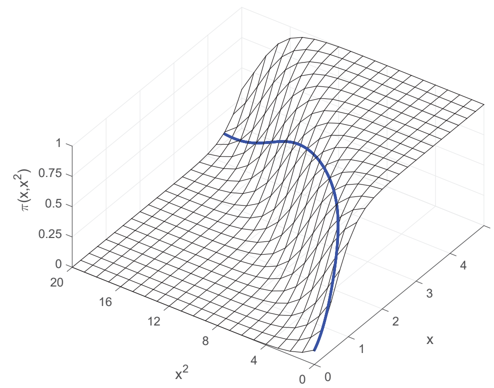
4. This can be computed two ways
   4.1 Automatically with factor syntax
   4.2 Explicitly with at(... = gen(...))

Discrete change with polynomials

1. With x and x^2 only values on the blue curve are mathematically possible

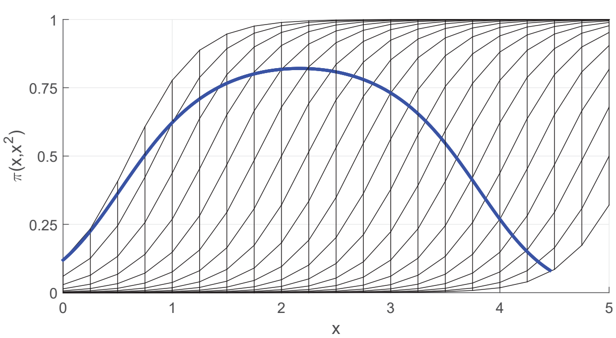


Discrete change with polynomials



2. Changes in the probability reflect linked changes in x and x^2

Discrete change with polynomials



3. The probability can increase and decrease as x and implicity x^2 change

Tool: factor notation for polynomials

Without factor notation

1. Generate age-squared

   `generate agesq = age * age`

2. Model specification

   `logit diabetes c.age c.agesq ...`

With factor notation

1. `c.age##c.age` with two #s does three things (you must include `c.`)
   1.1 Adds `c.age` to the model
   1.2 Create `c.age#c.age` ≡ `c.age*c.age`
   1.3 Adds `c.age#c.age` to the model
2. Model specification

   `logit diabetes c.age##c.age ...`

3. When `c.age` changes, `margins` automatically changes `c.age#c.age`

Discrete change with *age* & *age*^{2}

Correct ADC with factor notation

1. age and age#age automatically change together [#08]

```
. logit diabetes c.age##c.age c.bmi i.white i.female i.hsdegree, or
  (output omitted)

. mtable, at(age = gen(age)) at(age = gen(age+10)) post
Expression: Pr(diabetes), predict()
```

	Pr(y)
1	0.205
2	0.223

```
. mlincom 2 - 1, rowname(FV right)
```

	lincom	pvalue	ll	ul
FV right	0.018	0.000	0.011	0.024

2. Why is the effect of age so small?

Discrete change with *age* & *age*^{2}

<u>Incorrect</u> ADC without factor notation

1. age and agesq are distinct variables

```
. logit diabetes c.age c.agesq c.bmi i.white i.female i.hsdegree, or
  (output omitted)

. mtable, at(age = gen(age)) at(age = gen(age+10)) post
Expression: Pr(diabetes), predict()
```

	Pr(y)
1	0.205
2	0.744

```
. mlincom 2 - 1, rowname(noFV wrong)
```

	lincom	pvalue	ll	ul
noFV wrong	0.540	0.000	0.445	0.634

2. When `margins` changes `age`, variable `agesq` does <u>not</u> change

Discrete change with *age* & *age*^{2}

Correct ADC without factor notation

```
1] . logit diabetes c.age c.agesq c.bmi i.white i.female i.hsdegree, or
   (output omitted)

2]   . mtable, at(age = gen(age)    agesq = gen(agesq) ) ///
3]   >         at(age = gen(age+10) agesq = gen((age+10)^2)) post
     (output omitted)

4]   . mlincom 2 - 1, rowname(noFV right)
     (output omitted)
```

The power of at(gen())

1. With factor syntax you do not need `at(...=gen())` for polynomials
2. However, `at(...=gen())` allows complex links among variables

Discrete change with associated variables

1. Age and age-squared are **mathematically linked**
2. Other variables might be **substantively associated**
3. Example: To examine the effect of cultural capital on health, change all assets together, not just one asset
4. Example: Are "larger people" (taller people with the same body mass) more likely to have diabetes?
   - Use `height` to predict `weight`
   - Use `margins, gen()` to change `height` and `weight` together

   This example illustrates the power of `margins, gen()`

Associated variables: ADC(height, weight)

1. Regress `weight` on `height` and `height` squared [#09]

```
. regress weight c.height##c.height, noci
  (output omitted)
R-squared       =     0.2575
```

weight	Coef.	Std. Err.	t	P>|t|
height	-6.338708	1.61073	-3.94	0.000
c.height#c.height	.0855799	.0120867	7.08	0.000
_cons	217.5991	53.5548	4.06	0.000

2. Save estimates

```
. scalar b0 = _b[_cons]
. scalar b1 = _b[height]
. scalar b2 = _b[c.height#c.height]
```


Associated variables: ADC(height, weight)

3. `margins, gen()` changes weight based on a 6" change in height

```
1]   . mtable, post ///
2]   >     at(height = gen(height)                  /// observed height
3]   >        weight = gen(weight))                 /// observed weight
4]   >     at(height = gen(height+6)                /// +6 inches
5]   >        weight = gen(b0 + b1* (height+6)      ///
6]   >                   + b2*((height+6)^2)))  //  +estimated weight

Expression: Pr(diabetes), predict()
```

	Pr(y)
1	0.205
2	0.208

```
. mlincom 2 - 1
```

	lincom	pvalue	ll	ul
1	0.004	0.601	-0.010	0.017

4. Interpretation

   There is no evidence that being physically larger without greater body mass contributes to the incidence of diabetes.

Summary measures of change: ADC and DCM

1. ADC and DCM are common summaries of a variable's effect
2. Each uses the mean to summarize a distribution
3. ADC: average discrete change

$$\text{ADC}(x_1) = \frac{1}{N}\sum_i \left[\frac{\Delta\pi}{\Delta(x_1|\mathbf{x}=\mathbf{x}_i)}\right]$$

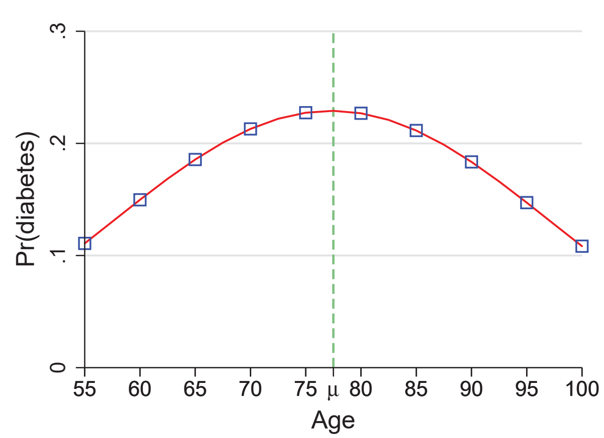
4. DCM: discrete change at the mean

$$\text{DCM}(x_1) = \frac{\Delta\pi}{\Delta(x_1|\mathbf{x}=\overline{\mathbf{x}})} \text{ where } \overline{x}_k = \frac{1}{N}\sum_i x_{ik}$$

5. Hypothetical data shows why means can be misleading

Summary measures of change: ADC and DCM

Hypothetical data



Summary measures of change: distribution of effects

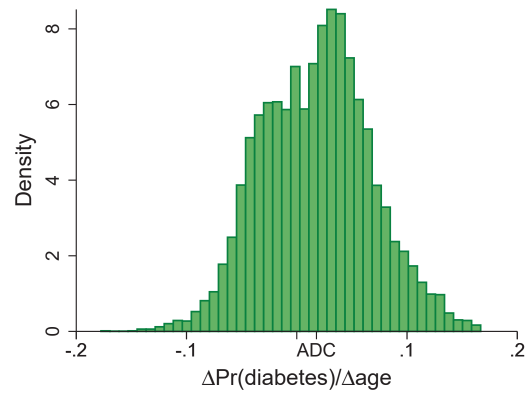
1. To evaluate ADC(age), look at the distribution of DC(age_i)
2. Create a variable with the DC for each observation

```
1]   margins, generate(PRage) ///
2]      at(age = gen(age)) at(age = gen(age+10))

3]   gen DCage10 = PRage2 - PRage1
4]   lab var DCage10 "DC for 10 year increase in age"
```


Summary measures of change: distribution of effects

3. The average effect of age is small, but is large and negative for some people and large and positive for others



Summary measures of change: distribution of effects

1. ADC and DCM are more useful than odds ratios
2. In nonlinear models, summary measure can be very misleading
3. The distribution of effects is valuable for assessing a variable's effect and is simple with `margins, generate()`
   - Long and Freese (2014) do this before the `gen()` option was added
4. The best summary is the one that explains the process being modeled
5. For age, multiple DCRs are more useful than ADC or DCM
   - I use DCR to introduce methods for comparing effects

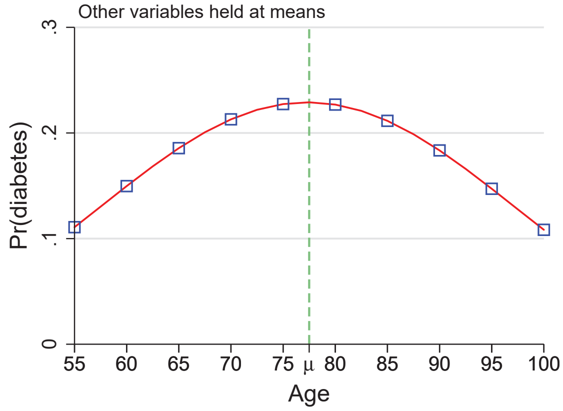
Comparing effects within a model

Examples

1. Compare DCRs for one variable at different values
   - Is the effect of age the same at 60 as at 80?
2. Compare ADCs for two *variables*
   - Does BMI have a larger impact than race?
3. Compare ADCs for two *sub-samples*
   - Does BMI have a larger effect for whites than non-whites?

Comparing DCR(age) at different ages

1. Are the DCR(age) significantly different at different ages?



Comparing DCR(age) at different ages

2. Compute probabilities at 4 ages with other variables at means [#11]

```
. mtable, at(age=(60(10)90)) post atmeans
Expression: Pr(diabetes), predict()

              age     Pr(y)
  
     1         60     0.150
     2         70     0.213
     3         80     0.227
     4         90     0.183
Specified values of covariates
                          1.        1.         1.
              bmi      white    female   hsdegree

  Current     27.9      .772      .568       .762
```

3. DCRs at different ages

```
. mlincom 2-1, clear rowname(DCR60)
. mlincom 3-2, add   rowname(DCR70)
. mlincom 4-3, add   rowname(DCR80)
```

Comparing DCR(age) at different ages

4. Test differences in DCRs

```
. mlincom (2-1) - (3-2), add rowname(DCR60 - DCR70)
. mlincom (2-1) - (4-3), add rowname(DCR60 - DCR80)
. mlincom (3-2) - (4-3), add rowname(DCR70 - DCR80)
```

5. Summarizing

```
. mlincom, twidth(14)
```

	lincom	pvalue	ll	ul
DCR60	0.063	0.000	0.054	0.073
DCR70	0.014	0.004	0.004	0.023
DCR80	-0.043	0.000	-0.061	-0.026
DCR60 - DCR70	0.049	0.000	0.037	0.062
DCR60 - DCR80	0.107	0.000	0.083	0.130
DCR70 - DCR80	0.057	0.000	0.046	0.069

6. Interpretation

The effects of a ten-year increase in age are significantly different at ages 60, 70, and 80 ($p < .001$).

Comparing ADC(white) and ADC(bmi)

1. ADC(race) and ADC(bmi+sd) have similar sizes, but different signs [#12]

```
. est restore Mbmi
(results Mbmi are active now)

. mchange bmi white, amount(sd)
logit: Changes in Pr(y) | Number of obs = 16071
Expression: Pr(diabetes), predict(pr)
```

	Change	p-value
bmi		
+SD	0.097	0.000
white		
White vs Non-white	-0.099	0.000

2. To test if the effects are equal, they must be **estimated simultaneously**

Comparing ADC(white) and ADC(bmi)

3. Simultaneously compute components for ADC(white) and ADC(bmi)

```
. quietly sum bmi
. local sd = r(sd)
. margins, at(white=(0 1)) at(bmi = gen(bmi)) at(bmi = gen(bmi + `sd´)) post
Predictive margins                              Number of obs     =     16,071
Model VCE    : OIM

Expression   : Pr(diabetes), predict()
1._at        : white           =           0
2._at        : white           =           1
3._at        : bmi             = bmi
4._at        : bmi             = bmi + 5.770835041238605
```

_at	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1	.2797806	.0073107	38.27	0.000	.265452	.2941092
2	.1805306	.0034215	52.76	0.000	.1738245	.1872367
3	.2047166	.0030338	67.48	0.000	.1987704	.2106627
4	.3017056	.005199	58.03	0.000	.2915159	.3118954

Comparing ADC(white) and ADC(bmi)

4. Compute effects and test equality

```
. qui mlincom (2-1), rowname(ADC white) clear
. qui mlincom (4-3), rowname(ADC bmi) add
.     mlincom (2-1) + (4-3), rowname(Sum of ADCs) add
```

	lincom	pvalue	ll	ul
ADC female	-0.099	0.000	-0.115	-0.083
ADC bmi	0.097	0.000	0.090	0.104
Sum of ADCs	-0.002	0.809	-0.021	0.016

5. Conclusion

The health cost of being non-white is equivalent to a standard deviation increase in body mass ($p > .80$).

Comparing ADC(bmi) by race

1. An ADC is typically averaged over the estimation sample
2. By averaging within groups, we can examine effects for different groups
   - Is the average effect of BMI the same for whites and non-whites?
3. This requires `margins, over()`

Tool: `margins, over()`

1. By default, `margins` averages over all observations
2. Averages on subsamples are possible with `if` and `over()`
3. Averaging for the non-white subsample

```
margins if white==0, ///
    at(bmi = gen(bmi)) at(bmi = gen(bmi+‘sd’))
```

4. For the white subsample

```
margins if white==1, ///
    at(bmi = gen(bmi)) at(bmi = gen(bmi+‘sd’))
```

5. For both subsamples simultaneously

```
margins, over(white) ///
    at(bmi = gen(bmi)) at(bmi = gen(bmi+‘sd’))
```

Comparing ADC(bmi) by race

1. Use over() to compute components for group specific ADC(bmi) [#13]

```
. margins, over(white) at(bmi = gen(bmi)) at(bmi = gen(bmi + `sd´)) post
Expression   : Pr(diabetes), predict()
over         : white
1._at        : 0.white
                   bmi             = bmi
               1.white
                   bmi             = bmi
2._at        : 0.white
                   bmi             = bmi + 5.770835041238605
               1.white
                   bmi             = bmi + 5.770835041238605
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_at#white						
1#Non-white	.3097249	.0072773	42.56	0.000	.2954616	.3239881
1#White	.173629	.0032892	52.79	0.000	.1671824	.1800757
2#Non-white	.4302294	.009226	46.63	0.000	.4121468	.448312
2#White	.2636564	.0054903	48.02	0.000	.2528955	.2744172

Comparing ADC(bmi) by race

2. Computing ADC(bmi) by group

```
. qui mlincom  4-2, clear rowname(White: ADC bmi)
. mlincom      3-1, add   rowname(Non-white: ADC bmi)
```

	lincom	pvalue	ll	ul
White ADC bmi	0.090	0.000	0.083	0.097
Non-white ADC bmi	0.121	0.000	0.112	0.129

3. A second difference compares effects for the groups

```
. mlincom (4-2) - (3-1), rowname(Difference: ADC bmi)
```

	lincom	pvalue	ll	ul
Difference ADC bmi	-0.030	0.000	-0.034	-0.027

4. Interpretation

   The effect of BMI for non-whites is significantly larger than the effect for whites ($p < .001$).

Comparing DCs across models: examples

Examples of comparing effects from different models

1. Different specifications of predictors
   - Does DC(female) depend on how body mass is measured?
2. Different groups
   - Does DC(bmi) differ for whites and nonwhites

TOOL: joint estimation in Stata

1. gsem simultaneously fits multiple equations
   1.1 Limited to GLM models
   1.2 margins behaves "normally", but is slow
   1.3 Robust standard errors are not required but vce(robust) and vce(cluster *clustvar*) are available
   1.4 Some complex expressions() might not work...
2. suest combines stored estimates
   2.1 Works with most regression models
   2.2 margins computes $\mathbf{x}'\widehat{\beta}$; computing $\widehat{\pi}(\mathbf{x})$ is complicated
   2.3 Average effects for subsamples cannot be computed
   2.4 Robust standard errors must be used
3. Specialized commands like khb (Kohler et al., 2011) are available

Comparing ADC(female) across models

Does the effect of female depend on how body mass is measured?

1. Since female is a factor variables, margins, dydx(female) computes DC(female)
2. Computing ADC(female) for two models

```
. qui logit diabetes c.bmi i.white c.age##c.age i.hsdegree
. qui mtable, dydx(female) rowname(ADC(female) with Mbmi) clear

. qui logit diabetes c.weight c.height i.female i.white c.age##c.age i.hsdegree
. mtable, dydx(female) rowname(ADC(female) with Mwt) below

Expression: Pr(diabetes), predict()
```

	d Pr(y)
ADC(female) with Mbmi	-0.036
ADC(female) with Mwt	-0.020

3. To test if they are equal, we compute the effects simultaneously

Tool: gsem for multiple equations

1. This does not estimate two models

```
gsem ///
  (diabetes <- c.bmi i.female i.white c.age##c.age i.hsdegree, logit) ///
  (diabetes <- c.weight c.height i.female i.white c.age##c.age i.hsdegree, logit)
```

   since it is interpreted as

```
gsem ///
  (diabetes <- c.bmi i.female i.white c.age##c.age i.hsdegree ///
               c.weight c.height, logit)
```

2. The solution is to create clones for each model

```
. clonevar lhsbmi = diabetes // outcome for bmi model
. clonevar lhswt = diabetes // outcome for weight height model
```

Comparing ADC(female) across models

1. Estimating the models simultaneously [#14]

```
. gsem ///
>     (lhsbmi <- c.bmi i.female i.white c.age##c.age i.hsdegree, logit) ///
>     (lhswt  <- c.weight c.height i.female i.white c.age##c.age i.hsdegree ///
>         , logit) ///
> , vce(robust)

Generalized structural equation model           Number of obs     =     16,071

Response       : lhsbmi
Family         : Bernoulli
Link           : logit

Response       : lhswt
Family         : Bernoulli
Link           : logit

Log pseudolikelihood = -14914.007

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lhsbmi <-    |
         bmi |    .099441    .003747    26.54   0.000      .092097    .1067851
             |
      female |
       Women |  -.2423701   .0413006    -5.87   0.000    -.3233177   -.1614225
             |
       white |
       White |   -.614014   .0480926   -12.77   0.000    -.7082738   -.5197543
```


Comparing ADC(female) across models

2. Estimate ADC(female) for both models simultaneously

```
. qui est restore Mgsem

. margins, dydx(female) post

Average marginal effects                        Number of obs     =     16,071
Model VCE    : Robust

dy/dx w.r.t. : 1.female
1._predict   : Predicted mean (Respondent has diabetes?), predict(pr
               outcome(outcome(lhsbmi))
2._predict   : Predicted mean (Respondent has diabetes?), predict(pr
               outcome(lhswt))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.female     |
    _predict |
          1  |  -.0360559   .0061773    -5.84   0.000    -.0481631   -.0239487
          2  |  -.0199213   .0089687    -2.22   0.026    -.0374997   -.0023429
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```


Comparing ADC(female) across models

3. Testing if ADC(female) is the same in both models

```
. mlincom 1-2, stats(all)
             |    lincom        se    zvalue    pvalue        ll        ul
-------------+------------------------------------------------------------
           1 |    -0.016     0.006    -2.526     0.012    -0.029    -0.004
```

4. Interpretation

   The effect of being female is significantly larger when body mass is measured with the BMI index ($p < .02$).

Comparing effects across models

1. Jointly estimating models with `gsem` and computing effects with `margins` is a general approach for comparing effects across models (Mize et al., 2009)
2. `gsem`
   2.1 Fits the GLM class of models, but does not fit non-GLM models
   2.2 `margins` is slow (grumble, grumble)
3. `suest`
   3.1 Fits a much wider class of models
   3.2 `margins` is fast, but hard to use (grumble, grumble)
4. `suest` and `gsem` produce identical results

Comparing groups: outcomes and marginal effects

Linear regression

1. Coefficients differ by group such as $\beta^W_{\texttt{female}}$ and $\beta^N_{\texttt{female}}$
2. Analysis focuses on Chow tests such as $H_0 : \beta^N_{\texttt{female}} = \beta^W_{\texttt{female}}$

Logit and probit

1. Coefficients differ by group such as $\beta^W_{\texttt{female}}$ and $\beta^N_{\texttt{female}}$
2. The coefficients combines
   2.1 The effect of x_k which can differ by group
   2.2 The variance of the error which can differ by group
3. Since regression coefficients are identified to a scale factor, Chow-type tests of $H_0 : \beta^N_k = \beta^W_k$ are invalid (Allison, 1999)
4. Probabilities and marginal effects are identified (Long, 2009)

Comparing groups: outcomes and marginal effects

Group differences can be examined two ways

1. Differences in probabilities

$$H_0\colon\ \pi_W(\mathbf{x} = \mathbf{x}^*) = \pi_N(\mathbf{x} = \mathbf{x}^*)$$

   Is the probability of diabetes the same for white and non-white respondents who have the same characteristics?

2. Differences in marginal effects

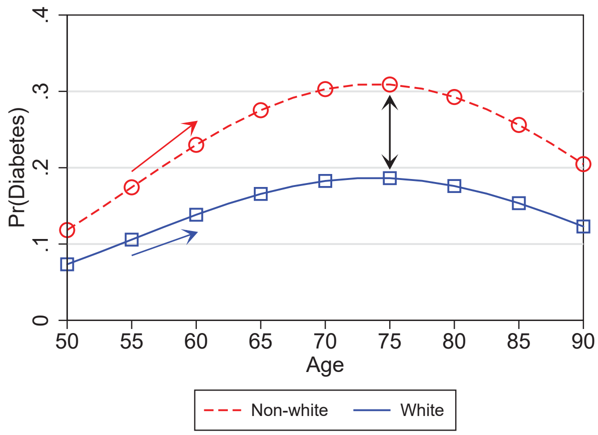
$$H_0\colon\ \frac{\Delta\pi_W}{\Delta x_k} = \frac{\Delta\pi_N}{\Delta x_k}$$

   Is the effect of x_k the same for whites and non-whites?

3. These dimensions of difference are shown in the next graph

Comparing groups: outcome and marginal effects

Hypothetical data



Comparing groups: model estimation

1. Factor syntax allows coefficients to differ by `white`

```
logit diabetes ibn.white ///
    ibn.white#(i.female i.hsdegree c.age##c.age c.bmi), nocon
```

2. This is equivalent to simultaneously estimating

```
logit diabetes i.female i.hsdegree c.age##c.age c.bmi if white==1
logit diabetes i.female i.hsdegree c.age##c.age c.bmi if white==0
```

3. For example [#15]

```
    Variable | Whites     NonWhites
-------------+----------------------
      female |
       Women |  0.713      1.024    <== odds ratios
             |  0.000      0.755    <== p-values
    hsdegree |
   HS degree |  0.706      0.743
             |  0.000      0.000
             |
         age |  1.278      1.369
             |  0.000      0.000
         ::: |  :::::      :::::
```


Group comparison of probabilities by age

1. `dydx(white)` computes DC(white)

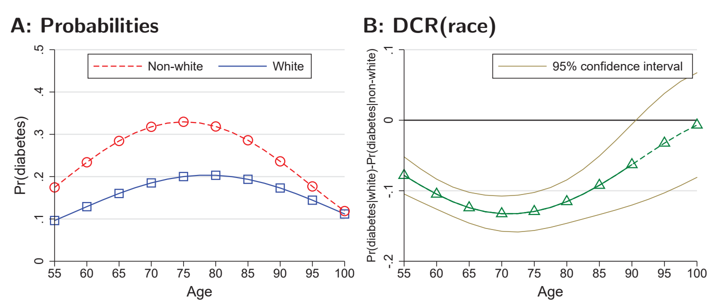
```
. mtable, dydx(white) at(age=(55(10)85)) atmeans stats(est p)
Expression: Pr(diabetes), predict()
          |   age   d Pr(y)      p
----------+------------------------
        1 |    55   -0.078   0.000    <== DCR(white | age=55)
        2 |    65   -0.124   0.000
        3 |    75   -0.129   0.000    <== DCR(white | age=75)
        4 |    85   -0.092   0.000
Specified values of covariates
          |      0.       1.       1.        1.
          |   white    white   female  hsdegree      bmi
----------+----------------------------------------------
  Current |    .228     .772     .568      .762     27.9
```

2. Example of interpretation

   The probability of diabetes is significantly larger for 55 year-old non-whites than whites who are average on other characteristics (p<.01).

3. Graphically we can show effects at multiple ages

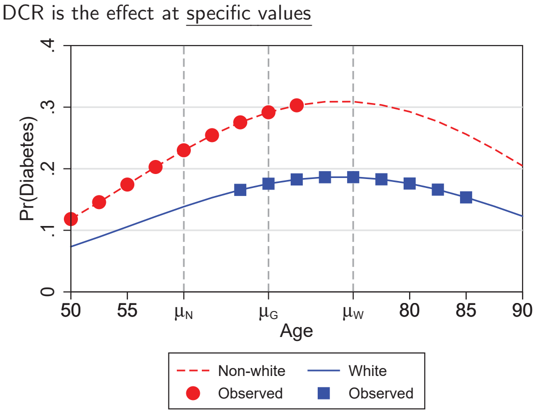
Group comparison of probabilities by age



Group comparison of effects: ADC or DCM?

Hypothetical data

1. ADC reflects the distribution of predictors
2. DCR is the effect at specific values



Group comparison of effects: ADC or DCM?

Comparing ADCs

1. ADCs reflects
   1.1 Differences in the probability curves
   1.2 Differences in distribution of variables
2. Group differences in ADCs reflect both components

Comparing DCRs

1. DCRs show differences in probability curves at a specific location
2. Group differences in DCRs do not depend on the distribution of variables

Which to use?

1. The answer depends on what you want to know?

Group comparison of effects: ADC(bmi+5)

1. To compute ADC(bmi + 5) by race

```
. mtable, over(white) at(bmi = gen(bmi)) at(bmi = gen(bmi+5)) post
Expression: Pr(diabetes), predict()
              |   Pr(y)
--------------+----------
  0.white#c.1 |   0.310
  1.white#c.1 |   0.174
  0.white#c.2 |   0.391
  1.white#c.2 |   0.257
. qui mlincom  3-1,          rowname(ADC(bmi) non) stats(est p) clear
. qui mlincom  4-2,          rowname(ADC(bmi) wht) stats(est p) add
.     mlincom (4-2) - (3-1), rowname(Difference)   stats(est p) add
              |  lincom    pvalue
--------------+--------------------
 ADC(bmi) non |   0.082     0.000
 ADC(bmi) wht |   0.083     0.000
   Difference |   0.002     0.826
```

2. Conclusion

   The average effects of BMI are not significantly different for whites and non-whites (p=.83).

Group comparison of effects: DCR(age+10)

1. Since ADC(age) is not a useful measure, we compare DCR(age+10)
   1.1 Other variables are held at sample means
   1.2 Group specific means could be used (Long and Freese, 2014)
2. For example, DCR(age + 10) at 55

```
mtable, at(age=55 white=(0 1)) at(age=55 white=(0 1)) atmeans post
mlincom  3-1,          rowname(DC nonwhite) stats(est p) clear
mlincom  4-2,          rowname(DC white)    stats(est p) add
mlincom (4-2) - (3-1), rowname(Dif at 55)   stats(est p) add
```

3. And so on, with the following results

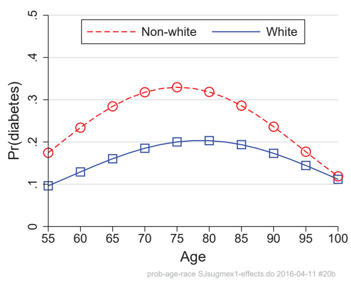
Group comparison of effects: DCR(age+10)

5. DCRs show group differences in

```
                 |   lincom    pvalue
-----------------+--------------------
55:     DC non   |    0.110     0.000
      DC white   |    0.064     0.000
    Difference   |   -0.046     0.001

70:     DC non   |    0.001     0.940
      DC white   |    0.018     0.001
    Difference   |    0.017     0.180

85:     DC non   |   -0.109     0.000
      DC white   |   -0.049     0.000
    Difference   |    0.060     0.003
```



6. These comparisons do not depend on group differences in the distribution of age or other variables

* Decomposing BMI

1. The BMI index measures relative weight or body mass

$$\text{BMI} = \frac{weight_{kg}}{height^2_m} = 703 \times \frac{weight_{lb}}{height^2_{in}}$$

2. Question 1: If BMI is in the model, can we compute the effect of increasing weight?
   - DC(weight) is clearer to patients then DC(bmi)
3. Question 2: Does DC(weight) differ depending on how body mass is included in the model?
4. To do this we create BMI as a product variable

$$\text{BMI} = 703 \times weight \times height^{-1} \times height^{-1}$$

Decomposing BMI: bmi as an interaction

1. Create components of BMI [#16]

```
generate heightinv = 1/height
label var heightinv "1/height"
generate S = 703
label var S "scale factor to convert from metric"
```

2. These models are *identical*

```
logit diabetes c.bmi i.white c.age##c.age i.female i.hsdegree
estimates store Mbmi
logit diabetes c.S#c.weight#c.height_inv#c.height_inv ///
                     i.white c.age##c.age i.female i.hsdegree
estimates store MbmiFV
```

3. The estimates are identical

```
    Variable |    MbmiFV         Mbmi
-------------+--------------------------
c.S#c.weight#|
 c.heightinv#|
 c.heightinv |   1.104553                   <== odds ratio for BMI
             |      0.000
         bmi |                1.1045533     <== odds ratio for BMI
             |                    0.000
       white |
       White |   .5411742      .5411742
             |      0.000         0.000
         ... |        ...           ...
```


Decomposing BMI: ADC(weight)

4. margins with factor syntax makes the rest trivial
5. ADC(weight) in MbmiFV changes only weight

```
. qui estimates restore MbmiFV
. mchange weight, amount(sd) delta(25)
logit: Changes in Pr(y) | Number of obs = 16071
Expression: Pr(diabetes), predict(pr)
             |   Change   p-value
-------------+--------------------
weight       |
         +25 |    0.065     0.000
```

6. ADC(weight) in Mwt is slightly larger

```
. qui estimates restore Mwt
. mchange weight, amount(sd) delta(25)
logit: Changes in Pr(y) | Number of obs = 16071
Expression: Pr(diabetes), predict(pr)
             |   Change   p-value
-------------+--------------------
weight       |
         +25 |    0.067     0.000
```

Decomposing BMI: summary

1. Factor variables and margins make the difficult decompositions trivial
2. Factor syntax understands interactions in model specifications
3. `margins` in turn understands interactions and handles the messy details

* Comparing ADC(weight) in two models

1. To compare ADC(weight) requires joint estimation [#16]

```
. clonevar lhsbmi = diabetes
. clonevar lhswt  = diabetes
. gsem ///
>     (lhsbmi <- c.s#c.weight#c.height_inv#c.height_inv ///
>         i.white c.age##c.age i.female i.hsdegree, logit) ///
>     (lhswt  <- c.weight c.height i.female i.white c.age##c.age i.hsdegree ///
>         , logit) ///
>     , vce(robust)
Generalized structural equation model           Number of obs     =     16,071
Response       : lhsbmi
Family         : Bernoulli
Link           : logit
Response       : lhswt
Family         : Bernoulli
Link           : logit
Log pseudolikelihood = -14914.007
```

(output omitted)

Comparing ADC(weight) in two models

2. Computing the average predictions for both equations

```
. margins, at(weight=gen(weight)) at(weight=gen(weight+25)) post
Predictive margins                              Number of obs     =     16,071
Model VCE    : Robust
1._predict   : Predicted mean (Diabetes?), predict(pr outcome(lhsbmi))
2._predict   : Predicted mean (Diabetes?), predict(pr outcome(lhswt))
1._at        : weight          = weight
2._at        : weight          = weight+25
```

_predict#_at	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1 1	.2047166	.0030419	67.30	0.000	.1987546	.2106786
1 2	.2701404	.0044591	60.58	0.000	.2614007	.27888
2 1	.2047166	.0030394	67.35	0.000	.1987595	.2106737
2 2	.271305	.0044054	61.58	0.000	.2626705	.2799394

Comparing ADC(weight) in two models

3. ADC(weight) for each model and their difference

```
. qui mlincom  2-1, rowname(Mbmi ADC) clear
. qui mlincom  4-3, rowname(Mwt ADC) add
. mlincom (4-3) - (2-1), rowname(Difference) add
```

	lincom	pvalue	ll	ul
Mbmi ADC	0.065	0.000	0.061	0.070
Mwt ADC	0.067	0.000	0.062	0.071
Difference	0.001	0.029	0.000	0.002

4. Conclusion

   The effect of weight on diabetes are nearly identical whether body mass is measured with BMI or with height and weight ($p = .03$).

Conclusions: Stata, margins, and interpretation

Model interpretation and Stata

1. Too often interpretation ends with the estimated coefficients
2. Interpretations using predictions are more informative
3. Without `margins` what I suggested today (and more) would be impractical

Marginal effects is only one method

1. Marginal effects are more useful than odds ratios and should be routinely computed (`mchange` makes this trivial)
2. `margins` allow many extensions to standard marginal effects
3. The best measure is the one that answers your question and might not be a standard measure
4. Marginal effects are one method, not the only or best method. Tables and graphs are often more useful (Long and Freese, 2014)
5. The best interpretation must be motivated by your substantive question

Thanks to many people

Thank *you* for listening

Collaborators Parts of this work were developed with Long Doan, Jeremy Freese, Trent Mize, and Sarah Mustillo. Jeff Pitblado and David Drukker provided valuable help. Mistakes are my own.

Relevant publications There is a large literature on marginal effects and interpreting models. Long and Freese (2014) include many citations. The references directly related to this presentation are given below.

Bibliography

Allison, P. D. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods & Research* 28(2): 186–208.

Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics using Stata*. Revised ed. College Station, Tex.: Stata Press.

Heeringa, S., B. West, and P. Berglund. 2010. *Applied survey data analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Kohler, U., K. B. Karlson, and A. Holm. 2011. Comparing coefficients of nested nonlinear probability models. *Stata Journal* 11(3): 420–438.

Long, J. S. 2009. Group comparisons in logit and probit using predicted probabilities.

Long, J. S., and J. Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata. Third Edition.* College Station, Texas: Stata Press.

Mize, T. D., L. Doan, and J. S. Long. 2009. A General Framework for Comparing Marginal Effects Across Models.