Moving Forward with Bivariate and Multivariate Analysis: Dependent and Independent Variables

Instructor: Marshall A. Taylor
844 Flanner Hall
mtaylo15@nd.edu
marshalltaylor.net

What's next?

- Moving forward with bivariate—and especially mutlivariate—inferential tests, it is important to make a distinction between two different types of variables: **dependent variables** and **independent variables**.

- We have been using these with our bivariate inferential tests!

Dependent Variables

- A **dependent variable** is a variable whose **change** you want to model. This change is **dependent** on other variables. What are these other variables?

Independent Variables

- These other variables are the **independent variables**. These variables do not depend (i.e., are **independent**) of the dependent variable and the other independent variables.

Example

- Let’s say job earnings is the dependent variable and sex is the independent variable. This means that job earnings **depend** on sex, but sex **does not depend** on job earnings.

- So, within the context of an independent samples t-test, the continuous variable is your dependent variable and the grouping variable is your independent variable. You are looking at hour job earnings **changes** between males and females—but you are not looking at how being a male or female changes because of job earnings.

Stata Example

- The continuous variable is respondent's income. The grouping variable is whether or not the respondent has a benefit plan. What's the DV? IV?

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     yes |     253    42546.29     1988.02     31621.4    38631.03    46461.54
      no |     403    32621.34    1766.293    35458.09    29149.01    36093.67
---------+--------------------------------------------------------------------
combined |     656     36449.1    1341.024    34346.98    33815.88    39082.33
---------+--------------------------------------------------------------------
    diff |            9924.946    2729.693                4564.927    15284.97
------------------------------------------------------------------------------
    diff = mean(yes) - mean(no)                                   t =   3.6359
Ho: diff = 0                                     degrees of freedom =      654

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9999         Pr(|T| > |t|) = 0.0003          Pr(T > t) = 0.0001
```

Stata Example

- You are modeling change in income. **Income** is your **DV**.

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     yes |     253    42546.29     1988.02     31621.4    38631.03    46461.54
      no |     403    32621.34    1766.293    35458.09    29149.01    36093.67
---------+--------------------------------------------------------------------
combined |     656     36449.1    1341.024    34346.98    33815.88    39082.33
---------+--------------------------------------------------------------------
    diff |            9924.946    2729.693                4564.927    15284.97
------------------------------------------------------------------------------
    diff = mean(yes) - mean(no)                                   t =   3.6359
Ho: diff = 0                                     degrees of freedom =      654

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9999         Pr(|T| > |t|) = 0.0003          Pr(T > t) = 0.0001
```

Stata Example

- You think change in income is explained by sex. **Sex** is your **IV**.

```
Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     yes |     253    42546.29     1988.02     31621.4    38631.03    46461.54
      no |     403    32621.34    1766.293    35458.09    29149.01    36093.67
---------+--------------------------------------------------------------------
combined |     656     36449.1    1341.024    34346.98    33815.88    39082.33
---------+--------------------------------------------------------------------
    diff |            9924.946    2729.693                4564.927    15284.97
------------------------------------------------------------------------------
    diff = mean(yes) - mean(no)                                   t =   3.6359
Ho: diff = 0                                     degrees of freedom =      654

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9999         Pr(|T| > |t|) = 0.0003          Pr(T > t) = 0.0001
```

Another Stata Example

- The proportion you are separating out by group is feelings toward suicide. The grouping variable is whether or not the respondent has been told they are depressed. What's the DV? IV?

```
Two-sample test of proportions                  Dep: Number of obs =        131
                                            Not dep: Number of obs =        690
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Dep |   .6793893   .0407768                      .5994683    .7593103
     Not dep |   .8130435   .0148423                       .783953    .8421339
-------------+----------------------------------------------------------------
        diff |  -.1336542    .043394                     -.2187048   -.0486035
             |  under Ho:   .0387011    -3.45   0.001
------------------------------------------------------------------------------
        diff = prop(Dep) - prop(Not dep)                          z =  -3.4535
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0003         Pr(|Z| > |z|) = 0.0006          Pr(Z > z) = 0.9997
```

Another Stata Example

- You are modeling change in opinions on suicide. **Opinions on suicide** is your **DV**.

```
Two-sample test of proportions                  Dep: Number of obs =      131
                                            Not dep: Number of obs =      690
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Dep |   .6793893   .0407768                      .5994683    .7593103
     Not dep |   .8130435   .0148423                       .783953    .8421339
-------------+----------------------------------------------------------------
        diff |  -.1336542    .043394                     -.2187048   -.0486035
             |  under Ho:   .0387011    -3.45   0.001
------------------------------------------------------------------------------
        diff = prop(Dep) - prop(Not dep)                          z =  -3.4535
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0003         Pr(|Z| > |z|) = 0.0006          Pr(Z > z) = 0.9997
```

Another Stata Example

- You are trying to account for this change by differentiating between those who have been told they are depressed and those who have not. **Whether or not they have been told they have depression** is your **IV**.

```
Two-sample test of proportions                   Dep: Number of obs =      131
                                             Not dep: Number of obs =      690
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Dep |   .6793893   .0407768                       .5994683    .7593103
     Not dep |   .8130435   .0148423                        .783953    .8421339
-------------+----------------------------------------------------------------
        diff |  -.1336542    .043394                      -.2187048   -.0486035
             |  under Ho:   .0387011    -3.45   0.001
------------------------------------------------------------------------------
        diff = prop(Dep) - prop(Not dep)                          z =  -3.4535
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0003         Pr(|Z| > |z|) = 0.0006          Pr(Z > z) = 0.9997
```

Another Stata Example

- If you are curious if type of job explains labor force status, **type of job** is your **IV** and **labor force status** is your **DV**.

r work for whom	labor force status working f	working p	temp not	Total
for-profit company	417	92	5	514
non-profit or not-for	87	24	4	115
government or governm	109	11	6	126
Total	613	127	15	755

Pearson chi2(4) = 16.0845 Pr = 0.003

Another Stata Example

- If you think level of religious fundamentalism explains views on pre-marital sex, then what is the DV and IV?

how fundamentalist is r currently	sex before marriage always wr	almst alw	sometimes	not wrong	Total
fundamentalist	140	34	58	155	387
moderate	138	55	101	381	675
liberal	38	18	86	375	517
Total	316	107	245	911	1,579

gamma = 0.3974 ASE = 0.031

Another Stata Example

- **Religious fundamentalism** is your **IV** and views on **pre-marital sex** is your **DV**.

how fundamentalist is r currently	sex before marriage always wr	almst alw	sometimes	not wrong	Total
fundamentalist	140	34	58	155	387
moderate	138	55	101	381	675
liberal	38	18	86	375	517
Total	316	107	245	911	1,579

gamma = 0.3974 ASE = 0.031

Important Distinction!

- This will become very important as you move into other tests, especially linear regression.
- Any questions?