Special topics
Interactions
Interactions

- Interactions test whether the combination of variables affects the outcome differently than the sum of the main (or individual) effects.
- For example, how would we test whether defendants are sentenced to death more frequently for killing white strangers than you would expect from the coefficients on white victim and on victim stranger?

```
. tab wv vs
```

|     vs          |      0 |      1 | Total |
|-----------------+-------+-------+-------|
| wv              |       |       |       |
| 0               |  12   |  14   |  26   |
| 1               |  37   |  37   |  74   |
| Total           |  49   |  51   | 100   |
Interactions

To interpret interactions, substitute the appropriate values for each variable.
E.g., what’s the effect for

- 0.099 wv+.108 vs+.330 wvXvs
- White, non-stranger: 0.099(1)+.108(0)+.330(1)*(0) = 0.099
- White, stranger: 0.099(1)+.108(1)+.330(1)*(1) = 0.537
- Black, non-stranger: 0.099(0)+.108(0)+.330(0)*(0) = comparison
- Black, stranger: 0.099(0)+.108(1)+.330(1)*(0) = 0.108
## Interactions

The code used to produce the table is:

```
.tab wv vs, sum(death)
```

### Means, Standard Deviations and Frequencies of death

<table>
<thead>
<tr>
<th>wv</th>
<th>vs</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.16666667</td>
<td>.28571429</td>
<td>.23076923</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.38924947</td>
<td>.46880723</td>
<td>.42966892</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>14</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.40540541</td>
<td>.75675676</td>
<td>.58108108</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.49774265</td>
<td>.43495884</td>
<td>.4967499</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>37</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.34693878</td>
<td>.62745098</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.48092881</td>
<td>.48829435</td>
<td>.50241839</td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>51</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Importance of a variable
Death penalty example

```
. sum death bd- yv

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>death</td>
<td>100</td>
<td>.49</td>
<td>.5024184</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>bd</td>
<td>100</td>
<td>.53</td>
<td>.5016136</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>wv</td>
<td>100</td>
<td>.74</td>
<td>.440844</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ac</td>
<td>100</td>
<td>.4366667</td>
<td>.225705</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>fv</td>
<td>100</td>
<td>.31</td>
<td>.4648232</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>vs</td>
<td>100</td>
<td>.51</td>
<td>.5024184</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>v2</td>
<td>100</td>
<td>.14</td>
<td>.3487351</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ms</td>
<td>100</td>
<td>.12</td>
<td>.3265986</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>yv</td>
<td>100</td>
<td>.08</td>
<td>.2726599</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```
Death penalty example

```
. reg death bd-yv , beta

| death | Coef.   | Std. Err. | P>|t| | Beta    |
|--------|---------|-----------|-----|---------|
| bd     | -.0869168 | .1102374  | 0.432 | -.0867775 |
| wv     | .3052246   | .1207463  | 0.013 | .2678175  |
| ac     | .4071931   | .2228501  | 0.071 | .1829263  |
| fv     | .0790273   | .1061283  | 0.458 | .0731138  |
| vs     | .3563889   | .101464   | 0.001 | .3563889  |
| v2     | .0499414   | .1394044  | 0.721 | .0346649  |
| ms     | .2836468   | .1517671  | 0.065 | .1843855  |
| yv     | .050356    | .1773002  | 0.777 | .027328   |
| _cons  | -1.189227  | .1782999  | 0.506 |         |
```
Importance of a variable

- Three potential answers
  - Theoretical importance
  - Level importance
  - Dispersion importance
Importance of a variable

- Theoretical importance
  - Theoretical importance = Regression coefficient (b)
  - To compare explanatory variables, put them on the same scale
    - E.g., vary between 0 and 1
Importance of a variable

- Level importance: most important in particular times and places
  - E.g., did the economy or presidential popularity matter more in congressional races in 2006?
  - Level importance $= b_j^* x_j$
Importance of a variable

- Dispersion importance: what explains the variance on the dependent variable
  - E.g., given that the GOP won in this particular election, why did some people vote for them and others against?
  - Dispersion importance =
    - Standardized coefficients, or alternatively
    - Regression coefficient times standard deviation of explanatory variable
    - In bivariate case, correlation
Which to use?

- Depends on the research question
  - Usually theoretical importance
  - Sometimes level importance
  - Dispersion importance not usually relevant
Partial residual scatter plots
Partial residual scatter plots

- Importance of plotting your data
- Importance of controls
- How do you plot your data after you’ve adjusted it for control variables?
- Example: inferences about candidates in Mexico from faces
Greatest competence disparity: pairing 10

- Gubernatorial race
- A more competent
- Who won?
  - A by 65%
Stata: dotplot

- ○ mean of competent
- ● mean of honest
- + mean of sympathetic
### Regression

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0.082003892</td>
<td>3</td>
<td>0.027334631</td>
</tr>
<tr>
<td>Residual</td>
<td>0.190333473</td>
<td>29</td>
<td>0.006563223</td>
</tr>
<tr>
<td>Total</td>
<td>0.272337365</td>
<td>32</td>
<td>0.008510543</td>
</tr>
</tbody>
</table>

Number of obs = 33
F( 3, 29) = 4.16  
Prob > F = 0.0144  
R-squared = 0.3011  
Adj R-squared = 0.2288  
Root MSE = 0.08101

| vote_a | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|------|------|-------------------|
| competent | 0.1669117 | 0.0863812 | 1.93 | 0.063 | -0.0097577 to 0.343581 |
| incumbent | 0.0110896 | 0.0310549 | 0.36 | 0.724 | -0.0524248 to 0.074604 |
| party_a | 0.2116774 | 0.1098925 | 1.93 | 0.064 | -0.013078 to 0.4364327 |
| _cons  | 0.2859541 | 0.0635944 | 4.50 | 0.000 | 0.1558889 to 0.4160194 |

- vote_a is vote share for Candidate A
- incumbent is a dummy variable for whether the party currently holds the office
- party_a is the vote share for the party of Candidate A in the previous election

- We want to create a scatter plot of vote_a by competent controlling for incumbent and party_a
Calculating partial residuals

First run your regression with all the relevant variables

```stata
reg vote_a competent incumbent party_a
```

To calculate the residual for the full model, use

```stata
predict e, res
```
(This creates a new variable “e”, which equals to the residual.)

Here, however, we want to generate the residual controlling only for some of the variables. To do this, we could manually predict vote_a based only on incumbent and party_a:

```stata
g y_hat = 0*.167 + incumbent*.011 + party_a*.212
```

We can then generate the partial residual

```stata
g partial_e = vote_a - y_hat
```

Instead, can use the Stata adjust

```stata
adjust competent = 0, by(incumbent party_a) gen(y_hat)
g partial_e = vote_a - y_hat
```
Calculating partial residuals

- Regression of the partial residual on competent should give you the same coefficient as in the earlier regression. It does.

```
. reg partial_e competent
    Source |       SS  df    MS
-----------------------------+---------------------
     Model |  .027767147  1  .027767147
   Residual |  .190333468  31  .006139789
-----------------------------+---------------------
       Total |  .218100616  32  .006815644

 Number of obs =      33
 F(  1,    31) =    4.52
 Prob > F      =  0.0415
   R-squared    =  0.1273
 Adj R-squared =  0.0992
 Root MSE      =  .07836

------------------------------------------------------------------------------
   e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+----------------------------------------
  competent |   .1669117    .078487     2.13   0.042     .0068364     .326987
     _cons |  -7.25e-09    .0470166   -0.00   1.000    -.0958909     .0958909
------------------------------------------------------------------------------
```
- Compare scatter plot (top) with residual scatter plot (bottom)
- Residual plots especially important if results change when adding controls
You can also use `avplot` to generate residual scatter plots.

- `avplot competent, scheme(lean2)`

Unlike the method above, `avplot` also conditions your explanatory variable.
Imputing missing data
Imputing missing data

- Variables often have missing data
- Sources of missing data
- Missing data reduces estimate precision and may bias estimates
- To rescue data with missing cases: impute using other variables
- Imputing data can
  - Increase sample size and so increase precision of estimates
  - Reduce bias if data is not missing at random
Imputation example

- Car ownership in 1948
- Say that some percentage of sample forgot to answer a question about whether they own a car
- The data set contains variables that predict car ownership: family_income, family_size, rural, urban, employed
Stata imputation command

- `impute depvar varlist [weight] [if exp] [in range], generate(newvar1)`
  - `depvar` is the variable whose missing values are to be imputed.
  - `varlist` is the list of variables on which the imputations are to be based.
  - `newvar1` is the new variable to contain the imputations.

- Example

  - `impute own_car family_income family_size rural suburban employed, g(i_own_car)`
Rules about imputing

- Before you estimate a regression model, use the summary command to check for missing data.
- Before you impute, check that relevant variables actually predict the variable with missing values (use regression or other estimator).
- Don’t use your studies’ dependent variable or key explanatory variable in the imputation (exceptions).
- Don’t impute missing values on your studies’ dependent variable or key explanatory variable (exceptions).
- Always note whether imputation changed results.
- If too much data is missing, imputation won’t help.