Handling interactions in Stata, especially with continuous predictors

Patrick Royston & Willi Sauerbrei

*German Stata Users’ meeting, Berlin, 1 June 2012*
Interactions – general concepts

- General idea of a (two-way) interaction in multiple regression is **effect modification**:
  - \( \eta(x_1, x_2) = f_1(x_1) + f_2(x_2) + f_3(x_1, x_2) \)
- Often, \( \eta(x_1, x_2) = E(Y \mid x_1, x_2) \), with obvious extension to GLM, Cox regression, etc.
- Simplest case: \( \eta(x_1, x_2) \) is **linear** in the \( x \)'s and \( f_3(x_1, x_2) \) is the **product** of the \( x \)'s:
  - \( \eta(x_1, x_2) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \)
- Can extend to more general, **non-linear** functions
The simplest type of interaction

- Binary x binary
- E.g. in the MRC RE01 trial in kidney cancer
- 12 month % survival since randomisation
- Substantial treatment effect in patients with low WCC
- Little or no treatment effect in those with high WCC
- But really, WCC is a continuous variable ...

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>White cell count low (&lt;=10)</th>
<th>White cell count high (&gt;10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPA</td>
<td>34% (se 4)</td>
<td>24% (se 4)</td>
</tr>
<tr>
<td>Interferon</td>
<td>49% (se 4)</td>
<td>21% (se 7)</td>
</tr>
</tbody>
</table>
Overview

- Interactions and factor variables (Stata 11/12)
  - Note: I am not an expert on factor variables! I sometimes use them.
- General interactions between continuous covariates in observational studies
  - Focus on continuous covariates ...
  - ... because people don’t appear to know how to handle them!
- Special case: interactions between treatment and continuous covariates in randomized controlled trials
Interactions and factor variables
Scope

- We introduce the topic with a brief introduction to factor variables.
- In this part, we consider only linear interactions:
  - Binary x binary (2 x 2 table)
  - Binary x continuous
  - Continuous x continuous
Factor variables: brief notes

- Implemented via prefixes (unary operators) and binary interaction operators
  - see help fvvarlist
- There are four factor-variable operators:
  
<table>
<thead>
<tr>
<th>Operator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>i.</td>
<td>unary operator to specify indicators (dummies)</td>
</tr>
<tr>
<td>c.</td>
<td>unary operator to treat as continuous</td>
</tr>
<tr>
<td>#</td>
<td>binary operator to specify interactions</td>
</tr>
<tr>
<td>##</td>
<td>binary operator to specify factorial interactions</td>
</tr>
</tbody>
</table>
- Dummy variables are ‘virtual’ – not created *per se*
- Names of regression parameters easily found by inspecting the post-estimation result matrix `e(b)`
Factor variables: i. prefix

- Example from Stata manual [U]11.4.3:

```
. list group i.group in 1/5

<table>
<thead>
<tr>
<th>group</th>
<th>lb.group</th>
<th>2.group</th>
<th>3.group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```
Example dataset

- MRC RE01 trial in advanced kidney cancer
- Of 347 patients, only 7 censored, the rest died
- For simplicity, as a continuous response variable, Y, we use months to death, _t
  - Ignore the small amount of censoring
- There are several *prognostic factors* that may influence time to death
- Some are binary, some categorical, some continuous
Example: factor variable parameters

```
. regress _t i.who

Source |       SS       df       MS              Number of obs =     347
-------------+------------------------------ F(  2,   344) =   15.72
Model |  7780.81413     2  3890.40707           Prob > F      =  0.0000
Residual | 85126.5686   344  247.460955           R-squared =  0.0837
         | 85126 5686 344 247 460955 R
-------------+------------------------------ Adj R-squared =  0.0784
Total |  92907.3828   346  268.518447           Root MSE      =  15.731

------------------------------------------------------------------------------
     _t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      who |     1  |    -4.2783   2.026215    -2.11   0.035    -8.263629   -.2929707
     2  |     2  |  -12.94365  2.354542   -5.50   0.000   -17.57476   -8.312534
     _cons |     1  |   19.17358   1.622518    11.82   0.000     15.98227    22.36488
------------------------------------------------------------------------------
```

```
. matrix list e(b)

```
e(b) [1,4]

```
0b.     1.         2.
     _who     _who     _who     _cons
  y1  0  -4.2782996  -12.943646  19.173577
```
Basic analysis to understand binary x binary: the 2 x 2 table of means

- Example:

```
.table rem sex, contents(mean _t) format(%6.2f)
```

```
x6:           |
  kidney      |
  removed     |
  (Y/N)       |   x2: sex   |
               |   male  female  
---------+---------------
  0 | 13.50  9.34   |
  1 | 13.77  18.19  |
```
Fitting an interaction model

- Consider 3 methods:
  - Method 1: binary operator to specify interactions
    - `regress _t rem#sex`
  - Method 2: binary operator to specify factorial interactions
    - `regress _t rem##sex`
  - Method 3: create multiplicative term(s) yourself
    - `gen byte remsex = rem * sex`
    - `regress _t rem sex remsex`

- Models are identical - all give the same fitted values
- Parameterisation of method 1 is different
Method 1: Binary operator #

```
. regress _t rem#sex

Source |       SS       df       MS              Number of obs =     347
-------------+------------------------------ F(  3,   343) =    2.73
Model | 2168.10384     3  722.701279           Prob > F      =  0.0437
Residual | 90724.0188   343  264.501513           R-squared     =  0.0233
-------------+------------------------------ Adj R-squared =  0.0148
Total | 92892.1227   346  268.474343           Root MSE      =  16.264

------------------------------------------------------------------------------
     _t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    rem#sex  |
     0 1  |  -4.160995   2.888457    -1.44   0.151    -9.842314    1.520323
     1 0  |   .2724983  2.137064     0.13   0.899    -3.930903    4.475899
     1 1  |   4.68417   2.581164     1.81   0.070    -.3927323    9.761072
     _cons |   13.49939   1.610327     8.38   0.000     10.33204    16.66675
------------------------------------------------------------------------------

Notes on parameters:
rem#sex 0 1 (-4.16) is (sex=1) - (sex=0) at rem=0
rem#sex 1 0 (+0.27) is (rem=1) - (rem=0) at sex=0
rem#sex 1 1 (+4.68) is [(rem=1) | sex=1] - [(rem=0) | sex=0]
     _cons (13.50) is intercept (mean of Y at rem=0 & sex=0)
I don’t recommend this parameterisation!
```
Method 2: Binary operator ##

```
. regress _t rem##sex

Source |       SS       df       MS              Number of obs =     347
-------------+------------------------------ F(  3,   343) =    2.73
Model | 2168.10384     3  722.701279           Prob > F      =  0.0437
Residual | 90724.0188   343  264.501513           R-squared     =  0.0233
-------------+------------------------------ Adj R-squared =  0.0148
Total | 92892.1227   346  268.474343           Root MSE      =  16.264

------------------------------------------------------------------------------
|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.rem |   .2724983   2.137064     0.13   0.899    -3.930903    4.475899
1.sex |  -4.160995   2.888457    -1.44   0.151    -9.842314    1.520323
rem#sex |     8.572667   3.792932     2.26   0.024     1.112333      16.033
   _cons |   13.49939   1.610327     8.38   0.000     10.33204    16.66675
------------------------------------------------------------------------------
Notes on parameters:
1.rem (0.27) is (rem=1) - (rem=0) at sex=0
1.sex (-4.16) is (sex=1) - (sex=0) at rem=0
rem#sex (+8.57) is [(rem=1) - (rem=0) at sex=1] - [(rem=1) - (rem=0) at sex=0]
_cons (13.50) is intercept (mean of Y at rem=0 & sex=0)
This is a ‘standard’ parameterisation with P-value as given above
```
Method 3: DIY multiplicative term

```
. generate byte remsex = rem * sex

. regress _t rem sex remsex

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 347</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2168.10384</td>
<td>3</td>
<td>722.701279</td>
<td>F(3, 343) = 2.73</td>
</tr>
<tr>
<td>Residual</td>
<td>90724.0188</td>
<td>343</td>
<td>264.501513</td>
<td>Prob &gt; F = 0.0437</td>
</tr>
<tr>
<td>Total</td>
<td>92892.1227</td>
<td>346</td>
<td>268.474343</td>
<td>R-squared = 0.0233</td>
</tr>
</tbody>
</table>

|     | Coef.   | Std. Err. | t    | P>|t| | 95% Conf. Interval |
|-----|---------|-----------|------|-----|-------------------|
| _t  | .2724983 | 2.137064  | 0.13 | 0.899 | -3.930903    4.475899 |
| rem | -4.160995 | 2.888457  | -1.44 | 0.151 | -9.842314    1.520323 |
| sex | 8.572667  | 3.792932  | 2.26  | 0.024 | 1.112333    16.033  |
| remsex | 13.49939 | 1.610327  | 8.38  | 0.000 | 10.33204    16.66675 |

Notes on parameters:
rem is 1.rem in Method 2
sex is 1.sex in Method 2
remsex is rem#sex in Method 2
This is the same parameterisation as Method 2
```
Interactions in non-normal errors models

Key points:

1. In a 2 x 2 table, an interaction is a ‘difference of differences’
2. Tabulate the 2 x 2 table of mean values of the linear predictor
3. May back-transform values via the inverse link function
   • e.g. exponentiation in hazards models
Binary x continuous interactions

- Use `c.` prefix to indicate continuous variable
- Use the `##` operator

```
. regress _t trt##c.wcc

Source |       SS       df       MS              Number of obs =     347
--------+------------------------------ F(  3,   343) =    7.44
Model   | 5678.62935     3 1892.87645           Prob > F      =  0.0001
Residual| 87228.7534   343  254.311234           R-squared     =  0.0611
         |                |                | Adj R-squared =  0.0529
Total   | 92907.3828   346  268.518447           Root MSE      =  15.947

        _t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+---------------------------------------------------------------
      1  .trt | 12.81405   4.124167     3.11   0.002     4.702208    20.92589
      wcc | -.2867831   .2741174    -1.05   0.296    -.8259457    .2523796
     trt#c.wcc | -1.034239   .4327233    -2.39   0.017    -1.885365    -.1831142
      1  _cons | 14.45292   2.712383     5.33   0.000     9.117919    19.78791
```
Binary x continuous interactions (cont.)

- The main effect of $w_{cc}$ is the slope in group 0.
- The interaction parameter is the difference between the slopes in groups 1 & 0.
- Test of $\text{trt}\#c.w_{cc}$ provides the interaction parameter and test.
- Results are nicely presented graphically:
  - Predict linear predictor $xb$
  - Plot $xb$ by levels of the factor variable
  - Also, ‘treatment effect plot’ (coming later)
Plotting a binary x continuous interaction

```
. regress _t trt##c.wcc
. predict fit
. twoway (line fit wcc if trt==0, sort) (line fit wcc if trt==1, sort lp(-)), legend(lab(1 "trt 0 (MPA)") lab(2 "trt 1 (IFN)") ring(0) pos(1))
```

![Graph showing the interaction between trt 0 (MPA) and trt 1 (IFN) with fitted values for white cell count (per liter x 10^-9).]
Continuous x continuous interaction

- Just use `c.` prefix on each variable

```plaintext
.regress _t c.age##c.t_mt

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 347</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7714.26052</td>
<td>3</td>
<td>2571.42017</td>
<td>F( 3, 343) = 10.35</td>
</tr>
<tr>
<td>Residual</td>
<td>85193.1223</td>
<td>343</td>
<td>248.37645</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>92907.3828</td>
<td>346</td>
<td>268.518447</td>
<td>R-squared = 0.0830</td>
</tr>
</tbody>
</table>

|                | Coef.          | Std. Err. | t    | P>|t|   | [95% Conf. Interval] |
|----------------|----------------|-----------|------|-------|---------------------|
| _t             |                |           |      |       |                     |
| age            | 0.0719063      | 0.0876542 | 0.82 | 0.413 | -0.1005011 0.2443137|
| t_mt           | 0.0659781      | 0.0128802 | 5.12 | 0.000 | 0.040644 0.0913122 |
| c.age#c.t_mt   | -0.0008783     | 0.001861  | -4.72| 0.000 | -0.0012443 -0.0005124|
| _cons          | 8.055213       | 5.256114  | 1.53 | 0.126 | -2.28306 18.39349  |
```
Continuous x continuous interaction

- Results are best explored graphically
- Consider in more detail next
Continuous x continuous interactions
Motivation: continuous x continuous intn.

- Many people only consider linear by linear interactions
- Not sensible if main effect of either variable is non-linear
- Mismodelling the main effect may introduce spurious interactions
  - E.g. false assumption of linearity can create a spurious linear x linear interaction
- Or they categorise the continuous variables
  - Many problems, including loss of power
The MFPIgen approach (1)

- MFP = multivariable fractional polynomials
- I = interaction
- gen = general
- Fractional polynomials (FPs) can be used to model relationships that may be non-linear
- In Stata, FPs are implemented through the standard `fracpoly` and `mfp` commands
- MFPIgen is implemented through a user-written command, `mfpigen`
The MFPIgen approach (2)

- MFPIgen aims to identify non-linear main effects and their two-way interactions
- Assume $x_1$, $x_2$ continuous and $z$ confounders
- Apply MFP to $x_1$ and $x_2$ and $z$
  - Force $x_1$ and $x_2$ into the model
  - FP functions $FP_1(x_1)$ and $FP_2(x_2)$ are selected for $x_1$ and $x_2$
  - Linear functions could be selected
- Add term $FP_1(x_1) \times FP_2(x_2)$ to the model chosen
- Apply likelihood ratio test of interaction
The MFPIgen approach in practice

- Start with a list of covariates
- Check all pairs of variables for an interaction
- Simultaneously, apply MFP to adjust for confounders
- Use a low significance level to detect interactions, e.g. 1%
- Present interactions graphically
- Check interactions for artefacts graphically
- Use forward stepwise if more than one interaction remains
Example: Whitehall 1

- Prospective cohort study of 17,260 Civil Servants in London
- Studied various standard risk factors for common causes of death
- Also studied social factors, particularly job grade
- We consider 10-year all-cause mortality as the outcome
- Logistic regression analysis
Example: Whitehall 1 (2)

- Consider weight and age

```
. mfpigen: logit all10 age wt
```

MFPIGEN - interaction analysis for dependent variable all10

<table>
<thead>
<tr>
<th>variable 1</th>
<th>function 1</th>
<th>variable 2</th>
<th>function 2</th>
<th>dev. diff.</th>
<th>d.f.</th>
<th>P</th>
<th>Sel</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>Linear</td>
<td>wt</td>
<td>FP2(-1 3)</td>
<td>5.2686</td>
<td>2</td>
<td>0.0718</td>
<td>0</td>
</tr>
</tbody>
</table>

Sel = number of variables selected in MFP adjustment model

- Age function is linear, weight is FP2(-1, 3)
- No strong interaction (P = 0.07)
. mfpigen, fplot(40 50 60): logit all10 age wt
Mis-specifying the main effects function(s)

- Assume age and weight are linear
- The \texttt{dfdefault(1)} option imposes linearity

\begin{verbatim}
.mfpigen, dfdefault(1): logit all10 age wt
\end{verbatim}

\texttt{MFPGEN - interaction analysis for dependent variable all10}

\begin{verbatim}
variable 1   function 1   variable 2   function 2  dev. diff.  d.f.    P   Sel
------------------------------------------------------------------------------
age Linear     wt Linear    8.7375    1  0.0031  0
------------------------------------------------------------------------------
Sel = number of variables selected in MFP adjustment model
\end{verbatim}

- There appears to be a highly significant interaction (P = 0.003)
Checking the interaction model

- Linear age x weight interaction seems important
- Check if it’s real, or the result of mismodelling
- Categorize age into (equal sized) groups
  - for example, 4 groups
- Compute running line smooth of the binary outcome on weight in each age group, transform to logits
- Plot results for each group
- Compare with the functions predicted by the interaction model
Whitehall 1: Check of age x weight linear interaction
Interpreting the plot

- Running line smooths are roughly parallel across age groups ⇒ no (strong) interactions
- Erroneously assuming that the effect of weight is linear ⇒ estimated slopes of weight in age-groups indicate strong interaction between age and weight
- We should have been more careful when modelling the main effect of weight
Whitehall 1: 7 variables, any interactions?

.mfpigen, select(0.05): logit all10 cigs sysbp age ht wt chol i.jobgrade

MFPIGEN - interaction analysis for dependent variable all10

<table>
<thead>
<tr>
<th>variable 1</th>
<th>function 1</th>
<th>variable 2</th>
<th>function 2</th>
<th>dev. diff.</th>
<th>d.f.</th>
<th>P</th>
<th>Sel</th>
</tr>
</thead>
<tbody>
<tr>
<td>cigs</td>
<td>FP1(.5)</td>
<td>sysbp</td>
<td>FP2(-2 -2)</td>
<td>0.7961</td>
<td>2</td>
<td>0.6716</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>FP1(.5)</td>
<td>age</td>
<td>Linear</td>
<td>0.0028</td>
<td>1</td>
<td>0.9576</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>FP1(.5)</td>
<td>ht</td>
<td>Linear</td>
<td>2.1029</td>
<td>1</td>
<td>0.1470</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>FP1(.5)</td>
<td>wt</td>
<td>FP2(-2 3)</td>
<td>0.1560</td>
<td>2</td>
<td>0.9249</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>FP1(.5)</td>
<td>chol</td>
<td>Linear</td>
<td>1.7712</td>
<td>1</td>
<td>0.1832</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>FP1(.5)</td>
<td>i.jobgrade</td>
<td>Factor</td>
<td>4.3061</td>
<td>3</td>
<td>0.2303</td>
<td>5</td>
</tr>
<tr>
<td>sysbp</td>
<td>FP2(-2 -2)</td>
<td>age</td>
<td>Linear</td>
<td>3.1169</td>
<td>2</td>
<td>0.2105</td>
<td>5</td>
</tr>
</tbody>
</table>

(remaining output omitted)
What `mfpigen` is doing

- FP functions for each pair of continuous variables are selected
  - Functions are simplified if possible
  - Closed test procedure in `mfp`
  - Controlled by the `alpha()` option
- The `select(0.05)` option tests confounders for inclusion in each interaction model at the 5% significance level
- The `Sel` column in the output shows how many variables are actually included in each confounder model
Results: P-values for interactions

<table>
<thead>
<tr>
<th>Variable</th>
<th>cigs*</th>
<th>sysbp*</th>
<th>age</th>
<th>height</th>
<th>weight*</th>
<th>chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>cigs*</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sysbp*</td>
<td>0.7</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.9</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>height</td>
<td>0.1</td>
<td>0.5</td>
<td>1.0</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight*</td>
<td>0.9</td>
<td>0.5</td>
<td>0.1</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chol</td>
<td>0.2</td>
<td>0.07</td>
<td>0.001</td>
<td>0.8</td>
<td>0.2</td>
<td>−</td>
</tr>
<tr>
<td>grade</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.04</td>
<td>0.4</td>
</tr>
</tbody>
</table>

*FP transformations were selected; otherwise, linear
Graphical presentation of age x chol interaction

. fracgen cigs .5, center(mean)
. fracgen sysbp -2 -2, center(mean)
. fracgen wt -2 3, center(mean)

. mfpigen, linadj(cigs_1 sysbp_1 sysbp_2
> wt_1 wt_2 ht i.jobgrade) df(1)
> fplot(%10 35 65 90): logit all110 age chol

Alternatively:

. logit all110 c.age##c.chol cigs_1 sysbp_1 sysbp_2
> wt_1 wt_2 ht i.jobgrade
. sliceplot age chol, sliceat(10 35 65 90) percent

- sliceplot is a new user-written command
Graphical presentation of age x chol intn.
Check of chol x age interaction

Q1: slope 0.17 (SE 0.06)

Q2: slope 0.22 (SE 0.05)

Q3: slope 0.14 (SE 0.04)

Q4: slope -0.01 (SE 0.04)
Interactions with continuous covariates in randomized trials
MFPI method (Royston & Sauerbrei 2004)

- Continuous covariate $x$ of interest, binary treatment variable $t$ and other covariates $z$
- Independent of $x$ and $t$, use MFP to select an ‘adjustment’ (confounder) model $z^*$ from $z$
- Find best FP2 function of $x$ (in all patients) adjusting for $z^*$ and $t$
- Test $\text{FP2}(x) \times t$ interaction (2 d.f.)
  - Estimate $\beta$’s in each treatment group
  - Standard test for equality of $\beta$’s
- May also consider simpler FP1 or linear functions – choose e.g. by min AIC
MFPI in Stata

- MFPI is implemented as a user command, `mfpi`
- `mfpi` is available on SSC
- Program was updated in 2012 to support factor variables
Treatment effect function

- Have estimated two FP2 functions – one per treatment group
- Plot the difference between functions against $x$ to show the interaction
  - i.e. the treatment effect at different $x$
- Pointwise 95% CI shows how strongly the interaction is supported at different values of $x$
  - i.e. variation in the treatment effect with $x$
Example: MRC RE01 trial in kidney cancer

- Main analysis: Interferon improves survival
- HR: 0.76 (0.62 - 0.95), P = 0.015
- Is the treatment effect similar in all patients?
- Nine possible covariates available for the investigation of treatment-covariate interactions – only one is significant (WCC)
Kaplan-Meier showing treatment effect

Follow-up (months)

Proportion alive

At risk 1: 175 55 22 11 3 2 1
At risk 2: 172 73 36 20 8 5 1

(1) MPA
(2) Interferon
The *mfpi* command

```
.mfpi, select(0.05) fp2(wcc) with(trt) gendiff(d): stcox (whod1 whod2) t_dt t_mt rem mets haem
```

Interactions with *trt* (347 observations). Flex-1 model (least flexible)

<table>
<thead>
<tr>
<th>Var</th>
<th>Main</th>
<th>Interact</th>
<th>idf</th>
<th>Chi2</th>
<th>P</th>
<th>Deviance</th>
<th>tdf</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>wcc</td>
<td>FP2(-1 -.5)</td>
<td>FP2(-1 -.5)</td>
<td>2</td>
<td>6.91</td>
<td>0.0316</td>
<td>3180.194</td>
<td>7</td>
<td>3194.194</td>
</tr>
</tbody>
</table>

idf = interaction degrees of freedom; tdf = total model degrees of freedom

```
.mfpi_plot wcc
[using variables created by gendiff(d)]
```
About 25% of patients, those with WCC > 10 seem not to benefit from interferon
Concluding remarks

- MFPIgen and MFPI should help researchers detect, model and visualize interactions with continuous covariates
- Usually, we are searching for interactions, so small P-values are required
- Other methods not considered
  - STEPP – mainly graphical
  - ...

Thank you.

- Cox identifies 3 types of variable that might appear in interactions:
- **Treatment variables**
  - Can be modified or imposed
  - Treatments, e.g. chemotherapy, surgery
  - Behaviours, e.g. smoking, drinking
- **Intrinsic variables**
  - Cannot be modified
  - Often demographic, e.g. sex, age
- **Unspecific variables**
  - e.g. structural blocks, `random’ factors