How to: manage and report your results

Violetta Korsunova

Dept. of Sociology, HSE in St. Petersburg

Contrasting your outcomes

By default, the baseline is the first outcome in your variable

For example:

Variable ‘social-economic status’ has levels:

‘low’, ‘medium’, ‘high’

The baseline is…

Contrasting your outcomes

By default, the baseline is the first outcome in your variable

For example:

Variable ‘social-economic status’ has levels:

‘low’, ‘medium’, ‘high’

The baseline is ‘low’

Contrasting your outcomes

That means that ‘medium’ SES is compared to ‘low’, and ‘high’ SES is compared to ‘low’

But what if you need to compare ‘high’ SES to the rest?

You can change the baseline, which is called ‘contrasting’

Let’s look at data ‘hsbde’

```
data = read.dta("https://stats.idre.ucla.edu/stat/data/hsbdemo.dta")
```

The data contains information about programs types choice:

- prog – program type: ‘general’, ‘academic’, ‘vocational’
- ses – socio-economic status: ‘low’, ‘medium’, ‘high’
- math – math test results

```
> table(data$prog)

 general academic vocation
      45      105       50
```

What is the baseline here?

```
> table(data$prog)

 general academic vocation
      45      105       50
```

What is the baseline here?

It is ‘general’

```
> test <- multinom(prog ~ ses + math, data = data)

> stargazer(test, type = 'text')
================================================
                        Dependent variable:
                    ----------------------------
                       academic      vocation
                         (1)            (2)
------------------------------------------------
sesmiddle                0.330         0.927*
                        (0.460)        (0.494)
seshigh                  0.924*         0.354
                        (0.532)        (0.656)
math                    0.082***      -0.071**
                        (0.024)        (0.029)
Constant               -3.929***       2.959**
                        (1.220)        (1.396)
------------------------------------------------
Akaike Inf. Crit.       363.287        363.287
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Model Fit

```
> library(pscl)
> pR2(test)
         llh      llhNull           G2     McFadden
-173.6435783 -204.0966746   60.9061927    0.1492092



> hitmiss(test)

Table of Actual (y) Against Predicted (p)
Classification rule: outcome with highest probability.
           p=general p=academic p=vocation Row PCP
y=general          7          27          11   15.56
y=academic         5          86          14   81.90
y=vocation         4          20          26   52.00

Percent Correctly Predicted, Fitted Model: 59.5%
Percent Correctly Predicted, Null Model  : 52.5%
```

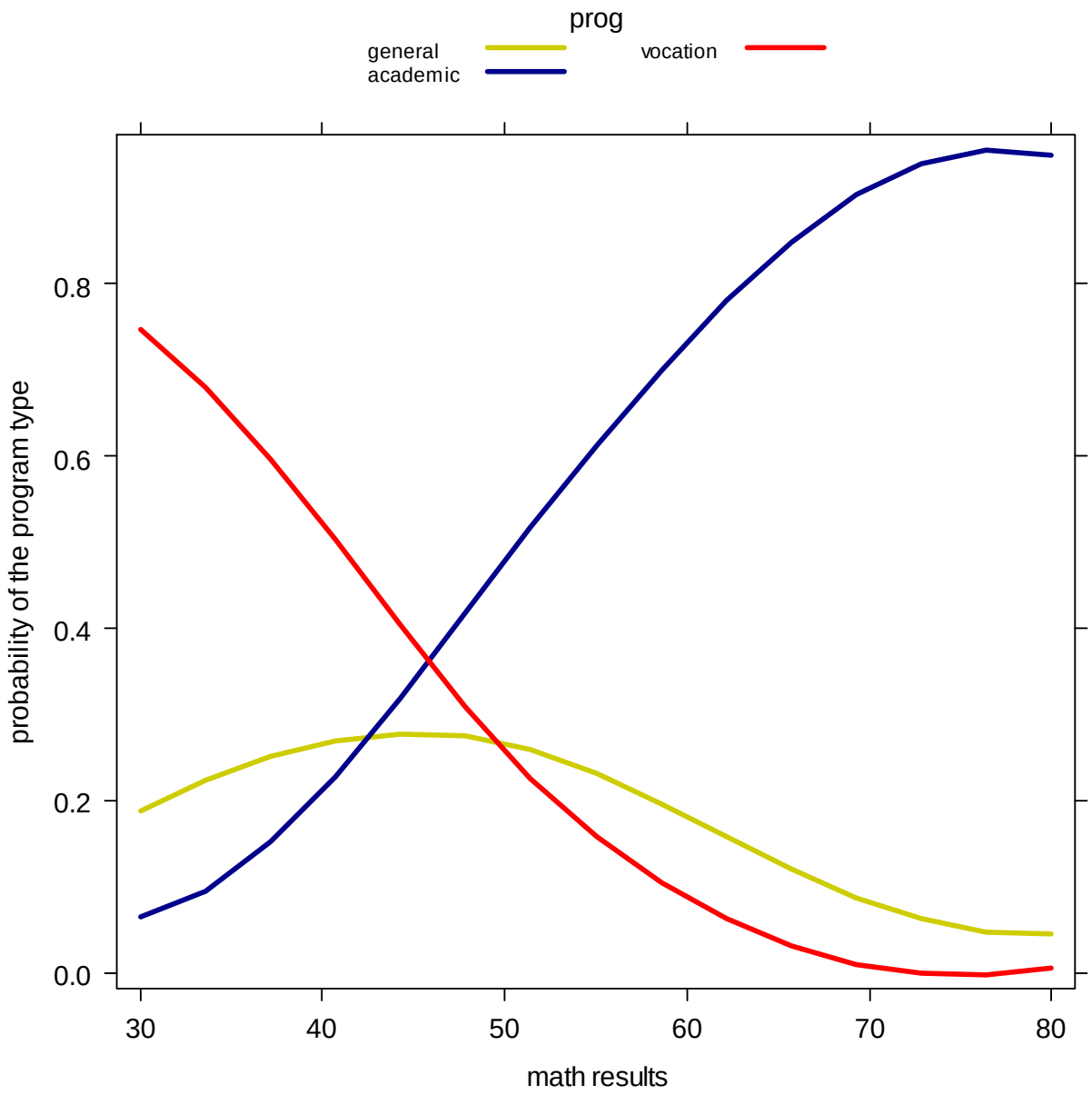
But what if you want to contrast ‘academic’ vs the rest?

You need to relevel your outcomes:

```
test2 <- multinom(relevel(prog, ref = ‘academic’) ~ ses + math, data = data)


plot(Effect(focal.predictors = "math", test),
     lines = list(multiline = T, lwd = 3, col = c("yellow3", "darkblue", "red")), rug = F,
     ylab = "probability of the program type",
     xlab = "math results",
     main = "")
```

```
> test2 <- multinom(relevel(prog, ref = 'academic') ~ ses + math, data = data)

> stargazer(test2, type = 'text')
================================================
                        Dependent variable:
                    ----------------------------
                          general       vocation
                            (1)            (2)
------------------------------------------------
sesmiddle                  -0.330         0.597
                          (0.460)        (0.495)
seshigh                    -0.924*        -0.570
                          (0.532)        (0.612)
math                      -0.082***      -0.153***
                          (0.024)        (0.028)
Constant                   3.930***      6.888***
                          (1.220)        (1.369)
------------------------------------------------
Akaike Inf. Crit.         363.287        363.287
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
> exp(coef(test2))
         (Intercept) sesmiddle   seshigh      math
general     50.88857 0.7192777 0.3968994 0.9214886
vocation   980.46953 1.8175496 0.5656056 0.8585530


> exp(coef(test))
         (Intercept) sesmiddle  seshigh      math
academic  0.01965412  1.390299 2.519593 1.0851969
vocation 19.26990031  2.526794 1.425067 0.9316998
```


prog
general
academic
vocation
probability of the program type
0.0
0.2
0.4
0.6
0.8
30
40
50
60
70
80
math results

How to report the results

2 levels of interpretations:

- *statistical* – what effects you have found
- *substantive* – how your results correspond to the previous research

Statistical interpretation

- Report, which coefficients are significant
  - Remember about the levels of significance and their relativity
- Interpret the direction

- Present visuals that help to understand your results
  - Consider if you should present odds ratios or marginal effects or beta-coefs are enough
  - Think if plots are more helpful

Substantive interpretation

- Do not stop at just reporting the coefficients
  - Remember, you need models to test your suppositions not visa versa
- Go back to your hypotheses and think if your results corroborate them
- Think if any alternative explanations can be found for your results
- Explain how your findings contribute to the current discussion

Validity of your measures

The relevance of your findings depend on the validity of your measures and models

- Consider construct and content validity (do you measure what you need and, if so, do you do that properly?)
- Consider statistical conclusion validity and internal validity (do your results meaningful and can they lead to causal inference?)

Some practice

Load data seminar2.Rdata

Data frame ‘uk’ contains information about 2008 UK elections results

- prvote – party supported: conservative, liberal democrat, labour, other
- age – age in years
- edu – age when completed education
- sex – gender
- income – income group; use this variable as numeric

- Predict which party was supported by age, gender, income group, and education
- Interpret coefficients
- Look at the model fit
- Make a plot to illustrate the relationship between education and party preferences
- Consider theoretical explanations of the results