Exercises I: Statistical Modeling in R

Henrik Singmann

January 2019 (updated: 2019-05-19)

Formula Interface for Statistical Models: ~

- Allows symbolic specification of statistical model, e.g. linear models: `lm(ACT ~ SATQ, sat.act)`
- Everything to the left of `~` is the dependent variable.
- Independent variables are to the right of the `~`:

Formula	Interpretation
`~ x` or `~1+x`	Intercept and main effect of `x`
`~ x-1` or `~0 + x`	Only main effect of `x` and no intercept (questionable)
`~ x+y`	Main effects of `x` and `y`
`~ x:y`	Interaction between `x` and `y` (and no main effect)
`~ x*y` or `~ x+y+x:y`	Main effects and interaction between `x` and `y`

Continuous Variables: How many Parameters in each Model?

```
lm(ACT ~ SATQ_c + SATV_c, sat.act)   # a
lm(ACT ~ SATQ_c : SATV_c, sat.act)   # b
lm(ACT ~ 0 + SATQ_c:SATV_c, sat.act) # c
lm(ACT ~ SATQ_c*SATV_c, sat.act)     # d
lm(ACT ~ 0+SATQ_c*SATV_c, sat.act)   # e
```

Categorical Variables: How many Parameters in each Model?

```
lm(ACT ~ gender, sat.act)                  # a
lm(ACT ~ 0+gender, sat.act)                # b
lm(ACT ~ gender+education, sat.act)        # c
lm(ACT ~ 0+gender+education, sat.act)      # d
lm(ACT ~ gender:education, sat.act)        # e
lm(ACT ~ 0+gender:education, sat.act)      # f
lm(ACT ~ gender*education, sat.act)        # g
lm(ACT ~ 0+gender*education, sat.act)      # h
lm(ACT ~ gender+gender:education, sat.act) # i
```

```
levels(sat.act$gender) ## 2
```

```
## [1] "male"   "female"
```

```
levels(sat.act$education) ## 6
```

```
## [1] "0" "1" "2" "3" "4" "5"
```