Epidemiology/Biostats VHM 812/802 Course
Winter 2016, Atlantic Veterinary College, PEI
Henrik Stryhn

Lecture 2a — Stata Factor-Notation Basics

For predictors x, z and outcome y, we have the notation

- `i.x` $\sim$ categorical effect of x (x must be integer)
- `c.x` $\sim$ continuous (slope) effect of x (x must be numerical)
- default depends on command:
  - `reg y x` $\sim$ `reg y c.x` (i.e., default is `c.x`)
  - `anova y x` $\sim$ `anova y i.x` (i.e., default is `i.x`)

Combined effects:

- `x#z` $\sim$ interaction $x \times z$; in all commands, the default is `x#z` $\sim$ `i.x#i.z`
- `c.x#c.z` $\sim$ multiplication $x \cdot z$
- always: `x##z` $\sim$ `x z x#z`
- quadratic regression: `c.x##c.x` $\sim$ continuous terms x and x^2

Factor terms can be used in tests, e.g. the Stata commands

- `testparm i.x`
- `testparm c.x##c.x`