

Week 4 Math Notes: Multilevel Data

Stat 209

Starting point $Y_{ij} = \bar{Y}_{i..} + (Y_{ij} - \bar{Y}_{i..})$ $X_{ij} = \bar{X}_{i..} + (X_{ij} - \bar{X}_{i..})$ between w/in
 Basic Levels of Analysis Relations (DCD) 2 Levels, indivs w/ groups.

Regression $\beta_{YX}^t = n_x^2 \beta_{\bar{Y}\bar{X}}^b + (1 - n_x^2) \beta_{YX}^{w-p}$

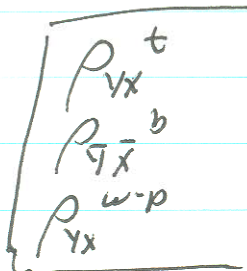
Correlation: $\rho_{YX}^t = n_x n_y \rho_{\bar{Y}\bar{X}}^b + \sqrt{(1 - n_x^2)(1 - n_y^2)} \rho_{YX}^{w-p}$

$\rho_{YX}^t = \rho_{\bar{Y}\bar{X}}^b$ iff $\rho_{YX}^{w-p} = \left(\frac{1 - n_x n_y}{\sqrt{1 - n_x^2} \sqrt{1 - n_y^2}} \right) \rho_{YX}^t$

where β_{YX}^t indiv ignore group

$\beta_{\bar{Y}\bar{X}}^b$ group mean "between"

β_{YX}^{w-p} w/in group pooled, relative standing



grouping measures (0,1) $n_x^2 = \frac{Var(X)}{Var(X)}$

$n_y^2 = \frac{Var(Y)}{Var(Y)}$

Derivation Y, X (unobserved) grouping var U
 means $E(X|U) \cdot E(Y|U)$

Standard Decompositions: Stat 200 between w/in

conditional variance $Var(X) = Var(E(X|U)) + E(Var(X|U))$

$Var(Y) = Var(E(Y|U)) + E(Var(Y|U))$

conditional covariance $Cov(Y, X) = Cov(E(X|U), E(Y|U)) + E(Cov(X, Y|U))$

Define $n_x^2 = \frac{Var(E(X|U))}{Var(X)}$ $\beta_{YX}^t = \frac{Cov(Y, X)}{Var(X)}$ $\beta_{\bar{Y}\bar{X}}^b = \frac{Cov(E(X|U), E(Y|U))}{Var(E(X|U))}$

$\beta_{YX}^{w-p} = E(Cov(Y, X|U)) / E(Var(X|U))$

$\frac{Cov(Y, X)}{Var(X)} = \frac{Cov(E(X|U), E(Y|U))}{Var(X)} \left(\frac{Var(E(X|U))}{Var(E(X|U))} \right)$

$+ \frac{E(Cov(X, Y|U))}{Var(X)} \left(\frac{E(Var(X|U))}{E(Var(X|U))} \right)$

requires to above ~~same~~ repeat for ρ

Deep review: Anova, Nested Designs

Neter et al Ch 26, Training Example

$Y_{ijk} = \mu + \alpha_i + \beta_j(i) + \epsilon_{ijk}$

α effect of school

$\beta_j(i)$ instructors nested w/in school

$Y_{ijk} - \bar{Y}_{...} = (\bar{Y}_{i..} - \bar{Y}_{...}) +$

(A main effect)

$(Y_{ij.} - \bar{Y}_{i..}) + (Y_{ijk} - \bar{Y}_{ij.})_{error}$

Anova Table

Schools	SSA
Instructors w/in school	SSB(A)
Error	SSE

Levels of Analysis, Aggregation Bias: Stat 209

Week 4

Individual level

```

> #####2x2 Tables Review, Robinson (1950) Table 1 data re Stat141
>
> litrace = matrix(c(1512, 7780, 2406, 85574), nr = 2, dimnames =
  list("Literacy" = c("No", "Yes"), "Race" = c("B", "W")))
> litrace
      Race
Literacy B      W
No      1512  2406
Yes     7780  85574

> chisq.test(litrace, correct = FALSE)
Pearson's Chi-squared test
data:  litrace
X-squared = 3984.29, df = 1, p-value < 2.2e-16

> phicoeff = sqrt(3984/97272)
> phicoeff # vs Robinson reported corr = .203
[1] 0.2023791

get conditional probabilities (cond'l on column)
> prop.table(litrace, 2)
      Race
Literacy B      W
No      0.1627206 0.02734712
Yes     0.8372794 0.97265288
          1          1

Relative risk of illiteracy (B/W)
> relrisk = prop.table(litrace, 2)[1,1]/prop.table(litrace, 2)[1,2]
> relrisk
[1] 5.950191
> #ratio of conditional probabilities
> .1627/.02735
[1] 5.948812

Same via Odds Ratio
> library(vcd)
> or = oddsratio(litrace, log = F)
> summary(or)
Odds Ratio
[1,] 6.9122
> confint(or)
      lwr      upr
[1,] 6.455426 7.401367
    
```

Ecological fallacy ex
 Literacy, Race correlations
 State level .773
 Region .946
 compared .203

sample
 $\phi = \sqrt{x^2/n_{++}}$
 "phi" coeff

Multilevel, contextual effects measured, var's

β^t_{yx}
 Indiv level regression ignoring groups

$\beta^b_{\bar{y}\bar{x}}$
 group level regression

Aggregation bias

$$\beta_{y\bar{x} \cdot \bar{x}} = \beta^b_{\bar{y}\bar{x}} - \beta^{w-p}_{yx}$$
 NEELS data $\hat{\beta}^t = 3.6$ $\hat{\beta}^b = 7$
 $\hat{\beta}^{w-p} = 2.1$

Contextual effect group on individual
 mult regr interpretation
 "increase in Y for increase X with X constant"
 As if by experiment.