# Week 4 Math Notes: Multilevel Data

Stat 209 [redacted]

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 2 Levels, indivs w/ groups.

## Basic Levels of Analysis Relations (DCD)

Regression $\beta^{t}_{YX} = \eta^2_X \beta^{b}_{\bar{Y}\bar{X}} + (1-\eta^2_X)\beta^{w-p}_{YX}$

Correlation: $\rho^{t}_{YX} = \eta_X \eta_Y \rho^{b}_{\bar{Y}\bar{X}} + \sqrt{(1-\eta^2_X)(1-\eta^2_Y)}\,\rho^{w-p}_{YX}$

$\rho^{t}_{YX} = \rho^{b}_{\bar{Y}\bar{X}}$ iff $\rho^{w-p}_{YX} = \left(\dfrac{1-\eta_X\eta_Y}{\sqrt{1-\eta^2_X}\sqrt{1-\eta^2_Y}}\right)\rho^{t}_{YX}$

where $\beta^{t}_{YX}$ indiv ignore group $\beta^{b}_{\bar{Y}\bar{X}}$ group mean "between" $\beta^{w-p}_{YX}$ w/in group pooled, relative standing

$\rho^{t}_{YX}$
$\rho^{b}_{\bar{Y}\bar{X}}$
$\rho^{w-p}_{YX}$

grouping measures $(0,1)$ $\eta^2_X = \dfrac{Var(\bar{X})}{Var(X)}$ $\eta^2_Y = \dfrac{Var(\bar{Y})}{Var(Y)}$

---

## 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 $\eta^2_X = \dfrac{Var(E(X|u))}{Var(X)}$ $\beta^{t}_{YX} = \dfrac{Cov(X,Y)}{Var(X)}$

$\beta^{w-p}_{YX} = E(Cov(Y,X|u))/E(Var(X|u)))$ $\beta^{b}_{\bar{Y}\bar{X}} = Cov(E(X|u),E(Y|u))/Var(E(X|u))$

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

regression to above ~~same~~ repeat for $\rho$

---

## Deep review: Anova, Nested Designs

Neter et al Ch 26, Training Example

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

$\alpha$ effect of school

$\beta_{j(i)}$ instructors nested w/in school

$Y_{ijk} - \bar{Y}_{...} = \underbrace{(\bar{Y}_{i..} - \bar{Y}_{...})}_{\text{A main effect}} +$

B w/in A $(Y_{ij.} - \bar{Y}_{i..}) + (Y_{ijk} - \bar{Y}_{ij.})_{\text{error}}$

Anova Table

| | |
|---|---|
| Schools | SSA |
| Instructors w/in School | SSB(A) |
| Error | SSE |

# Levels of Analysis, Aggregation Bias : Stat 209

week 4

```
> ####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

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
```

Individual level

Ecological fallacy ex

Literacy, Race correlations

| | r |
|---|---|
| State level | .773 |
| Region | .946 |

compare w/ .203

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

---

## Multilevel, Contextual effects measured, var's

$\beta^t_{YX} - \beta^b_{\bar{Y}\bar{X}}$ Aggregation bias

$\beta^t_{YX}$: indiv level regression ignoring groups

$\beta^b_{\bar{Y}\bar{X}}$: group level regression

$$\beta_{Y\bar{X}\cdot X} = \beta^b_{\bar{Y}\bar{X}} - \beta^{w-p}_{YX}$$

Contextual effect: effect of group on individ

mult regr interpretation "increase in Y for increase $\bar{X}$ with X constant"

As if by experiment?

NELS data $\hat{\beta}^t = 3.6$ $\hat{\beta}^b = 7$ $\hat{\beta}^{wp} = 2.1$