

Week 4 Math Notes: Multilevel Data

Stat 209

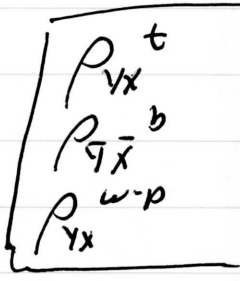
Starting point $Y_{ij} = \bar{Y}_{i..} + (Y_{ij} - \bar{Y}_{i..})$ $X_{ij} = \bar{X}_{i..} + (X_{ij} - \bar{X}_{i..})$ between w/in
 Basic Levels of Analysis Relations (DCD) 2 Levels, indiv vs w/ groups.

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

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

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

Standard Decompositions: ^{Stat 200} between w/in
 conditional variance $\text{Var}(X) = \text{Var}(E(X|u)) + E(\text{Var}(X|u))$
 $\text{Var}(Y) = \text{Var}(E(Y|u)) + E(\text{Var}(Y|u))$

conditional covariance $\text{Cov}(Y, X) = \text{Cov}(E(X|u), E(Y|u)) + E(\text{Cov}(X, Y|u))$

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

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

regress to above score 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

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

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

Levels of Analysis, Aggregation Bias

Stat 209
week 4

Robinson Table 3 2x2 Table review

```
> litnat = matrix(c(1304, 11913, 2614, 81441), nr = 2,
  dimnames = list("Literacy" = c("No", "Yes"),
  "ForB" = c("Y", "N")))
> litnat
```

| Literacy | Y | N |
|----------|-------|-------|
| No | 1304 | 2614 |
| Yes | 11913 | 81441 |

Ecological fallacy ex

48 states
-.53

Literacy, Foreign Born
Correlations

9 regions
-.62

```
> chisq.test(litnat, correct = FALSE)
```

Pearson's Chi-squared test

data: litnat

X-squared = 1348.6, df = 1, p-value < 2.2e-16

sample
 $\phi_k = \sqrt{\frac{\chi^2}{n+1}}$
11 phi_k coeff

```
> phicoeff = sqrt(1348.6/97272)
```

```
> phicoeff
```

vs Robinson reported corr = .118

```
[1] 0.1177464
```

```
> # get conditional probabilities (cond'l on column)
```

```
> prop.table(litnat, 2)
```

relative risk (n/f)

ForBorn

| Literacy | Y | N |
|----------|------------|------------|
| No | 0.09866082 | 0.03109869 |
| Yes | 0.90133918 | 0.96890131 |

or odds ratio

```
> # so relative risk illit > 3
```

vcd package

Multilevel, Contextual effects measured, var's

β_{yx}^t
Indiv level regression
ignoring groups

$\beta_{\bar{y}\bar{x}}^b$
group level regression

Aggregation bias

NEELS data $\beta_{yx \cdot \bar{x}}^t = 3.6$ $\beta_{\bar{y}\bar{x}}^b = 7$
 $\beta_{\bar{y}\bar{x}}^b = 2.1$

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