Class 12: Meta-Analysis for Binary Outcome

R section EPH 705

Min Lu

Division of Biostatistics
University of Miami

Spring 2017

Overview

1. Object:
   - More effect sizes and conversions
   - Fixed effect and random effect models
   - Multivariate meta-analysis
2. R Example
3. Exercise

Effect size and standard error

	Treatment	No treatment	Row total
Case	A_i	B_i	N_{1i}
Controls	C_i	D_i	N_{2i}
Column total	M_{1i}	M_{2i}	T_i

- Odds ratio (OR): $OR = \dfrac{A_i/D_i}{B_i/C_i}$

  $SE_{LogOR_i} = \sqrt{\frac{1}{A_i} + \frac{1}{B_i} + \frac{1}{C_i} + \frac{1}{D_i}}$

- Risk Ratio (RR): $RR = \dfrac{A_i/N_{1i}}{B_i/N_{2i}}$

  $SE_{LogRR_i} = \sqrt{\frac{1}{A_i} + \frac{1}{N_{1i}} + \frac{1}{C_i} + \frac{1}{N_{2i}}}$

- Risk Difference (RD): $RD = A_i/N_{1i} - C_i/N_{2i}$

  $SE_{RD} = \sqrt{\frac{A_i * B_i}{N_{1i}} + \frac{C_i * D_i}{N_{2i}}}$

Contingency Coefficient

Phi Coefficient is a measure of association based on adjusting chi-square significance to factor out sample size. The range of it is between -1 and 1 for 2-by-2 tables, and is between 0 and min(sqrt(rows - 1), sqrt(columns - 1)). Computationally, phi is the square root of chi-square divided by n, the sample size. The phi coefficient is often used as a measure of association in 2-by-2 tables formed by true dichotomies.

Contingency Coefficient is an adjustment to phi coefficient, intended to adapt it to tables larger than 2-by-2. The contingency coefficient is computed as the square root of chi-square divided by chi-square plus n, the sample size. The contingency coefficient will be always less than 1 and will be approaching 1.0 only for large tables. The larger the contingency coefficient the stronger the association. Some researchers recommend it only for 5-by-5 tables or larger. For smaller tables it will underestimated the level of association.

Cramer's V is the most popular of the chi-square-based measures of nominal association because it is designed so that the attainable upper limit is always 1. Cramer's V equals the square root of chi-square divided by sample size, n, times m, which is the smaller of (rows - 1) or (columns - 1).

Phi coefficient

In statistics, the phi coefficient (also referred to as the "mean square contingency coefficient" and denoted by ϕ (or r_ϕ)) is a measure of association for two binary variables. Introduced by Karl Pearson, this measure is similar to the Pearson correlation coefficient in its interpretation. In fact, a Pearson correlation coefficient estimated for two binary variables will return the phi coefficient. The square of the Phi coefficient is related to the chi-squared statistic for a 2 by 2 contingency table (see Pearson's chi-squared test) $\phi^2 = \frac{\chi^2}{n}$, where n is the total number of observations. Two binary variables are considered positively associated if most of the data falls along the diagonal cells. In contrast, two binary variables are considered negatively associated if most of the data falls off the diagonal. If we have a 2 by 2 table for two random variables x and y.

	$y = 1$	$y = 0$	total
$x = 1$	n_{11}	n_{10}	$n_{1\bullet}$
$x = 0$	n_{01}	n_{00}	$n_{0\bullet}$
total	$n_{\bullet 1}$	$n_{\bullet 0}$	n

The phi coefficient that describes the association of x and y is $\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1\bullet}n_{0\bullet}n_{\bullet 0}n_{\bullet 1}}}$.

Phi is related to the point-biserial correlation coefficient and Cohen's d and estimates the extent of the relationship between two variables (2X2).[4] The Phi coefficient can also be expressed using only n , n_{11} , $n_{1\bullet}$, and $n_{\bullet 1}$, as

$$\phi = \frac{nn_{11} - n_{1\bullet}n_{\bullet 1}}{\sqrt{n_{1\bullet}n_{\bullet 1}(n - n_{1\bullet})(n - n_{\bullet 1})}}.$$

Phi coefficient

	$y = 1$	$y = 0$	total
$x = 1$	n_{11}	n_{10}	$n_{1\bullet}$
$x = 0$	n_{01}	n_{00}	$n_{0\bullet}$
total	$n_{\bullet 1}$	$n_{\bullet 0}$	n

The phi coefficient that describes the association of x and y is

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1\bullet}n_{0\bullet}n_{\bullet 0}n_{\bullet 1}}}$$

$$SE_{\varphi} = \frac{1}{\sqrt{n_{..}}}\left(\begin{array}{l} 1-\hat{\varphi}^2+\hat{\varphi}\left(1+\frac{\hat{\varphi}^2}{2}\right)\frac{(n_{1\bullet}-n_{0\bullet})(n_{\bullet 1}-n_{\bullet 0})}{\sqrt{n_{1\bullet}n_{0\bullet}n_{\bullet 1}n_{\bullet 0}}} \\ -\frac{3}{4}\hat{\varphi}^2\left[\frac{(n_{1\bullet}-n_{0\bullet})^2}{n_{1\bullet}n_{0\bullet}}+\frac{(n_{1\bullet}-n_{\bullet 0})^2}{n_{\bullet 1}n_{\bullet 0}}\right] \end{array}\right)^{1/2}$$

Converting the log odds ratio to Cohan's d

$$d = LogOR \times \frac{\sqrt{3}}{\pi}$$

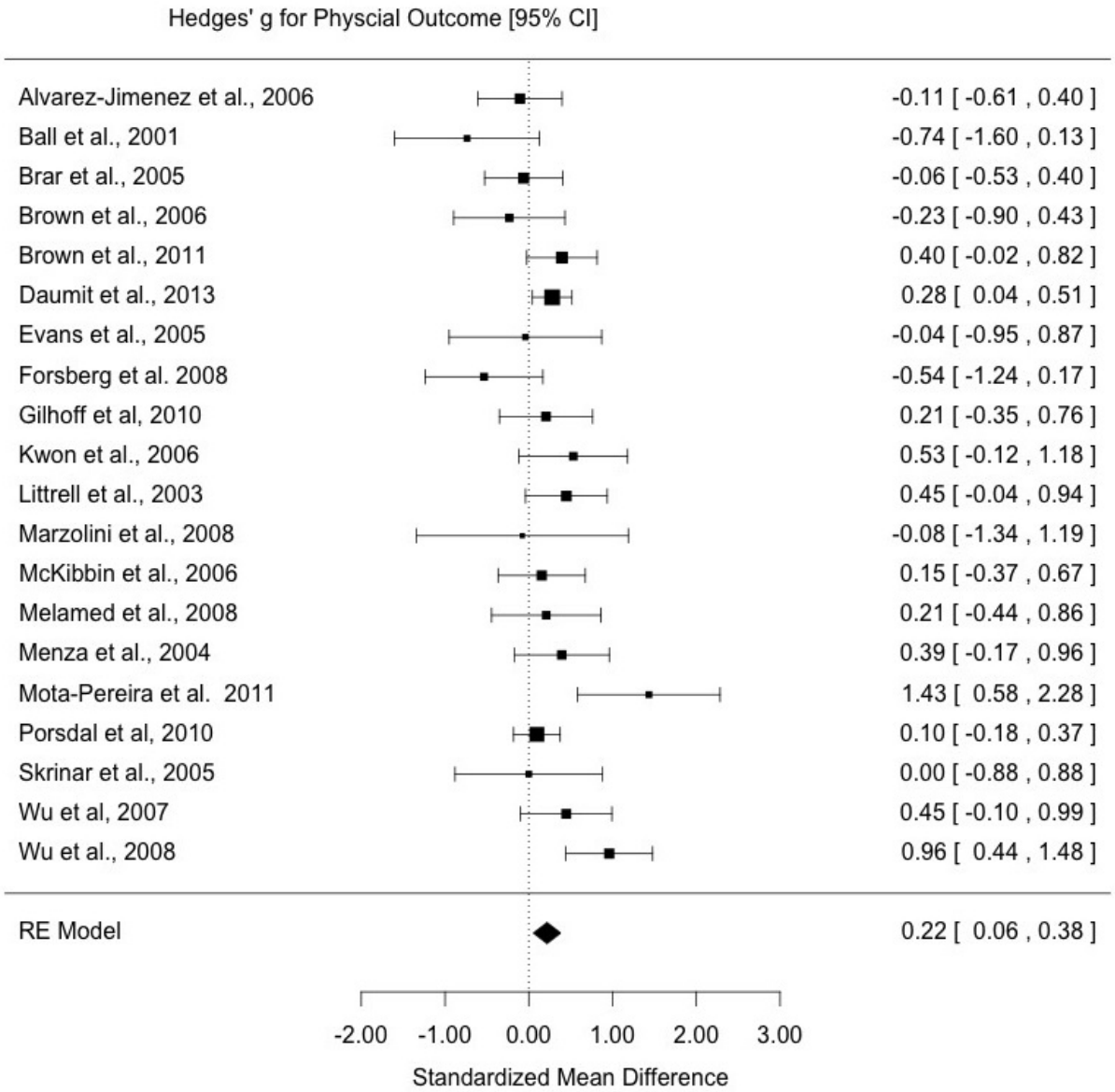
$$SE_d = SE_{LogOR} \times \frac{3}{\pi^2}$$

The corrected d is called Hedge's g:

$$J = 1 - \frac{3}{4(N_{total} - 2) - 1}$$

$g = J * d$ and $SE_g = J * SE_d$ Check out more effect sizes at http://handbook.cochrane.org/chapter_9/9_2_types_of_data_and_effect_measures.htm http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Formulas.php and http://cebcp.org/practical-meta-analysis-effect-size-calculator/

Effect size and standard error



Fixed effect model and random effect model

- A fixed-effect meta-regression model allows for within-study variability but not between-study variability:

$$y_i \sim \mathcal{N}(\mu_i, \sigma_{\eta_i})$$

$$\mu_i = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots$$

Or,

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \eta_j$$

- A random effects meta-regression is called a mixed effects model when moderators are added to the model.

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \eta + \varepsilon_j$$

Here $\sigma^2_{\varepsilon_j}$ is the variance of the effect size in study j , and σ^2_{η} is between study variance

Fixed effect model and random effect model

- Fixed-effect meta-regression assumes that the sampled effect size θ is normally distributed with $\mathcal{N}(\theta, \sigma_\theta)$ where σ_θ^2 is the within-study variance of the effect size. A fixed-effect meta-regression model thus allows for within-study variability but not between-study variability because all studies have an identical expected fixed effect size θ , i.e. $\varepsilon = 0$. $y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \eta_j$
  Here $\sigma_{\eta_j}^2$ is the variance of the effect size in study j . Fixed effect meta-regression ignores between study variation.
- Random effects meta-regression rests on the assumption that θ in $\mathcal{N}(\theta, \sigma_i)$ is a random variable following a (hyper-)distribution $\mathcal{N}(\theta, \sigma_\theta)$. A random effects meta-regression is called a mixed effects model when moderators are added to the model.
  $y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \eta + \varepsilon_j$
  Here $\sigma_{\varepsilon_j}^2$ is the variance of the effect size in study j . Between study variance σ_η^2 is estimated using common estimation procedures for random effects models (restricted maximum likelihood (REML) estimators).

Multivariate meta-analysis

The data for `mvmeta` comprise the point estimate, y_i, and the within-study variance–covariance matrix, S_i, for each study $i = 1$ to n.

We assume the model

$$\begin{aligned} y_i &\sim N(\mu_i, S_i) \\ \mu_i &\sim N(\mu, \Sigma) \\ \Sigma &= \begin{pmatrix} \tau_1^2 & \kappa_{12}\tau_1\tau_2 & . \\ \kappa_{12}\tau_1\tau_2 & \tau_2^2 & . \\ . & . & . \end{pmatrix} \end{aligned}$$

where y_i, μ_i, and μ are $p \times 1$ vectors, and S_i and Σ are $p \times p$ matrices. The within-study variance, S_i, is assumed to be known. Our aim is to estimate μ and Σ.

We set $W_i = (\Sigma + S_i)^{-1}$, noting that this depends on the unknown Σ. If Σ were known (or assumed to be the zero matrix, as in fixed-effects meta-analysis), then we would have

$$\widehat{\mu} = \left(\sum_i W_i\right)^{-1} \left(\sum_i W_i y_i\right)$$

Studies on the Effectiveness of the BCG Vaccine Against Tuberculosis

Results from 13 studies examining the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis.

variable	discreption
trial	trial number
author	author(s)
year	publication year
tpos	# of TB positive cases in the treated (vaccinated) group
tneg	# of TB negative cases in the treated (vaccinated) group
cpos	# of TB positive cases in the control (non-vaccinated) group
cneg	# of TB negative cases in the control (non-vaccinated) group
ablat	absolute latitude of the study location (in degrees)
alloc	method of treatment allocation (random, alternate, or systematic assignment)

Calculating Phi-coefficients as effect sizes

```
library(metafor)
### load BCG vaccine data
dat <- get(data(dat.bcg))
### calculate log relative risks and corresponding sampling variances
dat <- escalc(measure = "PHI", ai = tpos, bi = tneg, ci = cpos, di = cneg, data = dat.bcg)
head(dat)
```

```
##   trial               author year tpos  tneg cpos  cneg ablat     alloc      yi
## 1     1              Aronson 1948    4   119   11   128    44    random -0.1001
## 2     2     Ferguson & Simes 1949    6   300   29   274    55    random -0.1635
## 3     3      Rosenthal et al 1960    3   228   11   209    42    random -0.1067
## 4     4    Hart & Sutherland 1977   62 13536  248 12619    52    random -0.0684
## 5     5 Frimodt-Moller et al 1973   33  5036   47  5761    13 alternate -0.0092
## 6     6      Stein & Aronson 1953  180  1361  372  1079    44 alternate -0.1798
##       vi
## 1 0.0032
## 2 0.0011
## 3 0.0017
## 4 0.0000
## 5 0.0001
## 6 0.0003
```

Display the meta-regression result

```
library(metafor)
```

```
## Warning: package 'metafor' was built under R version 3.6.3
```

```
## Loading required package: Matrix
```

```
## Loading 'metafor' package (version 2.4-0). For an overview
## and introduction to the package please type: help(metafor).
```

```
res <- rma(yi, vi, mods = ~ ablat + year, data = dat)
res
```

```
##
## Mixed-Effects Model (k = 13; tau^2 estimator: REML)
##
## tau^2 (estimated amount of residual heterogeneity):     0.0010 (SE = 0.0006)
## tau (square root of estimated tau^2 value):             0.0316
## I^2 (residual heterogeneity / unaccounted variability): 93.85%
## H^2 (unaccounted variability / sampling variability):   16.27
## R^2 (amount of heterogeneity accounted for):            72.55%
##
## Test for Residual Heterogeneity:
## QE(df = 10) = 98.9908, p-val < .0001
##
## Test of Moderators (coefficients 2:3):
## QM(df = 2) = 25.3327, p-val < .0001
##
## Model Results:
##
##          estimate      se     zval    pval     ci.lb    ci.ub
## intrcpt   -7.7675  2.6213  -2.9633  0.0030  -12.9050  -2.6299  **
## ablat     -0.0014  0.0009  -1.6023  0.1091   -0.0031   0.0003
## year       0.0039  0.0013   2.9794  0.0029    0.0013   0.0065  **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Converts log odds ratio to Hedge's g

```
library("compute.es")
```

```
## Warning: package 'compute.es' was built under R version 3.6.3
```

```
lores(lor = -0.9387, var.lor = 0.3571, n.1 = 123, n.2 = 139, level = 95, cer = 11/(11 + 128), dig = 2, verbose = TRUE, id = NULL, data = NULL)
```

```
## Mean Differences ES:
##
##  d [ 95 %CI] = -0.52 [ -1.16 , 0.13 ]
##   var(d) = 0.11
##   p-value(d) = 0.12
##   U3(d) = 30.24 %
##   CLES(d) = 35.72 %
##   Cliff's Delta = -0.29
##
##  g [ 95 %CI] = -0.52 [ -1.16 , 0.13 ]
##   var(g) = 0.11
##   p-value(g) = 0.12
##   U3(g) = 30.29 %
##   CLES(g) = 35.76 %
##
##  Correlation ES:
##
##  r [ 95 %CI] = -0.25 [ -0.36 , -0.13 ]
##   var(r) = 0.02
##   p-value(r) = 0
##
##  z [ 95 %CI] = -0.26 [ -0.38 , -0.13 ]
##   var(z) = 0
##   p-value(z) = 0
##
##  Odds Ratio ES:
##
##  OR [ 95 %CI] = 0.39 [ 0.12 , 1.26 ]
##   p-value(OR) = 0.12
##
##  Log OR [ 95 %CI] = -0.94 [ -2.11 , 0.23 ]
##   var(lOR) = 0.36
##   p-value(Log OR) = 0.12
##
##  Other:
##
##  NNT = -19.14
##  Total N = 262
```

Multivariate meta-analysis

```
library("metavcov")
data(Geeganage2010)
# D Death
# DD death or disability
#nt_D    Number of people in "1 Drug" Group
#nc_D    Number of people in "control " Group
#st_D    Number of people in "1 Drug" Group reporting D status
#sc_D    Number of people in "control " Group reporting D status

## set the correlation coefficients list r
r12 <- 0.71
r.Gee <- lapply(1:nrow(Geeganage2010), function(i) {
    matrix(c(1, r12, r12, 1), 2, 2)
})


computvocv <- lgOR.vcov(nt = subset(Geeganage2010, select = c(nt_DD, nt_D)), nc = subset(Geeganage2010,
    select = c(nc_DD, nc_D)), st = subset(Geeganage2010, select = c(st_DD, st_D)), sc = subset(Geeganage2010,
    select = c(sc_DD, sc_D)), r = r.Gee)
# name computed log odds ratio as an input
Input <- computvocv$lgOR
colnames(Input) <- c("lgOR.DD", "lgOR.D")
# name variance-covariance matrix of trnasformed z scores as covars
covars <- computvocv$lgOR.cov
library(mvmeta)
mvmeta_RE <- summary(mvmeta(cbind(lgOR.DD, lgOR.D), S = covars, data = as.data.frame(Input),
    method = "reml"))
mvmeta_RE
```

```
## Call:  mvmeta(formula = cbind(lgOR.DD, lgOR.D) ~ 1, S = covars, data = as.data.frame(Input),
##     method = "reml")
##
## Multivariate random-effects meta-analysis
## Dimension: 2
## Estimation method: REML
##
## Fixed-effects coefficients
##          Estimate  Std. Error       z  Pr(>|z|)  95%ci.lb  95%ci.ub
## lgOR.DD    0.0621      0.0853  0.7275    0.4669   -0.1052    0.2294
## lgOR.D     0.0933      0.0993  0.9387    0.3479   -0.1015    0.2880
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Between-study random-effects (co)variance components
## Structure: General positive-definite
##          Std. Dev     Corr
## lgOR.DD    0.0614  lgOR.DD
## lgOR.D     0.0262       -1
##
## Multivariate Cochran Q-test for heterogeneity:
## Q = 31.3827 (df = 32), p-value = 0.4976
## I-square statistic = 1.0%
##
## 17 studies, 34 observations, 2 fixed and 3 random-effects parameters
##   logLik       AIC       BIC
## -19.1964   48.3928   55.7215
```

In class exercise

Using the same data but conduct meta-analysis using correlation coefficients

Class over