



help for umeta and umeta_postestimation

Title

umeta - U-statistics-based random-effects meta-analyses

Description

The **umeta** command performs U-statistics-based random-effects meta-analysis on a dataset of univariate, bivariate or trivariate point estimates, sampling variances, and for bivariate or trivariate data, within-study correlations or covariances. The methodology is described in [Ma and Mazumdar \(2011\)](#).

For each outcome, **umeta** calculates the overall effect and a confidence interval for the effect. The command also displays the between-study variance (or alternatively between-study standard deviation), between-study correlation(s) for bivariate or trivariate data and inconsistency (I-squared) statistics.

umeta Syntax

```
umeta yvar* svar* [wsvar*] [if] [in] [, covvar(string) level(#) predint
tscale(logit|log|asin) noestimates bssd zci i2]
```

where the data are arranged with one line per study: the point estimates are held in variables *yvar**, the sampling variances are held in *svar**, and within-study correlations (or covariances) for 2 or 3 outcomes are held in variable *wsvar**.

For univariate data, *yvar** is *yvar* and *svar** is *svar*

For bivariate data, *yvar** is *yvar1 yvar2*, *svar** is *svar1 svar2* and *wsvar** is *wsvar12*

For trivariate data, *yvar** represents *yvar1 yvar2 yvar3*, *svar** is *svar1 svar2 svar3* and *wsvar** is *wsvar12, wsvar13 wsvar23*

For any unreported outcomes, **umeta** sets the outcome and its variance at 0 and 1E12, respectively.

Options for umeta

covvar(string) For bivariate or trivariate data analysis, you **must** specify **covvar**(rho) or **covvar**(cov) depending on whether you are using within-study correlation(s) or covariance(s).

level(#) specifies the significance level for probability intervals.

predint displays outcome-specific mean estimates with the probability interval of the approximate predictive distribution of a future trial, based on the extent of heterogeneity. No method has been developed as yet for multivariate predictive distribution.

tscale(logit|log|asin) transformation of estimates to original scale, if data was transformed prior to analysis.

bssd reports the between-study standard deviations with confidence intervals (calculated as a function of inconsistency statistic and typical within-study variance as by [White\(2009\)](#)) instead of the default between-study variances.

noestimates prevents display of mean estimates, between-study variances (or standard deviations) and correlation(s)

zci uses z-statistics instead of default t-statistics for confidence interval calculation. This is overridden if option **predint** specified.

i2 reports I-squared statistic for each outcome, together with confidence intervals as is described in [White\(2009\)](#).

umeta, typed without specifying varlist, redisplayes the latest estimation results. All the output options listed above may be used

by...: or **statsby...:** may be used with **umeta** to perform subgroup analyses; see help [by](#) or [statsby](#).

Remarks

Multivariate meta-analysis is used to synthesize multiple outcomes simultaneously taking into account the correlation between the outcomes ([Riley\(2009\)](#)). Likelihood based approaches, in particular, Restricted Maximum Likelihood (REML) method is commonly utilized in this context. REML assumes a multivariate normal distribution for the random-effects model. This assumption is difficult to verify, especially for meta-analysis with small number of component studies. Use of REML also requires iterative estimation between parameters, needing moderately high computation time, especially when the dimension of outcomes is large ([White\(2009\)](#)). [Jackson, White and Thompson\(2010\)](#) have developed a multivariate method of moments (MMM) which has been shown to perform equally well to REML.

Ma and Mazumdar recently proposed a new method for multivariate meta-analysis based on the theory of U-statistic. The motivation for using U-statistic stems from the fact that it provides a a robust, nonparametric and noniterative approach. Additionally, the asymptotic behavior of the related statistics and their estimates are easy to derive being based on theorems already available for U-statistics.

Since the between-study variance matrix for the random-effects meta-analysis model involves second order moments, U-statistic formulation is especially beneficial. It is easily applied to estimate the variance matrix components and to develop their joint asymptotic distribution for related inference. Because the U-statistic-based method does not depend on parametric distributional assumptions for both random effects and sampling errors, it provides robust inference irrespective of the data distribution

For a detailed description of the u-statistic methodology, see [Ma and Mazumdar \(2011\)](#).

By convention, the within-study variances are assumed known and replaced by their sample estimates. Thus imprecision in within-study variance estimates may affect the estimation of pooled effect size especially when the size of within-study variation is relatively large.

This program does not assume that variables need log, logit or arcsin or other transformation(s). However, if study-level outcome data are available as odds ratios, risk ratios or proportions, the user may choose to log-, logit-or arcsin-transform them first. Then **tscale** option may be used to change back to the original scale for reporting if so desired.

The probability interval of the approximate predictive distribution of a future trial, is based on the extent of heterogeneity. This incorporates uncertainty in the location and spread of the random effects distribution using the formula $t(df) \times \sqrt{se2 + \tau2}$ where t is the t-distribution with n-2 degrees of freedom, se2 is the squared standard error and tau2 the heterogeneity statistic and n is the number of observations(studies). This is applied to each outcome separately. For further information see [Higgins, Thompson and Spiegelhalter\(2009\)](#)

I-squared formulated by [Higgins and Thompson \(2002\)](#), describes the percentage of total variation across studies that is attributable to heterogeneity rather than chance and measures impact of heterogeneity. Negative values of I-squared are made equal to 0 so that I-squared lies between 0% and 100%. A value of 0% indicates no observed heterogeneity, and values greater than 50% may be considered substantial heterogeneity. The main advantage of I-squared is that it does not inherently depend on the number of the studies in the meta-analysis

Examples

Example 1: Univariate Data

```
. use umeta_example1, clear
. list yvar svar, clean noobs
. umeta yvar svar
```

Example 2: Bivariate logit-transformed Data, No within-study correlation

```
. use umeta_example2, clear
. list yvar* svar* rho*, clean noobs
. umeta yvar* svar* rho*, p
. umeta yvar* svar* rho*, z bssd p tscale(logit)
```

Example 3: Bivariate Outcomes with missing Data

```
. use umeta_example3, clear
. list yvar* svar* rho*, clean noobs
. umeta yvar* svar* rho*
. umeta yvar* svar* rho*, pred
. umeta, noest i2 z q
```

Example 4: Trivariate Outcomes with Zero within-study covariance matrix

```
. use umeta_example4, clear
. list yvar* svar* rho*, clean noobs
. umeta yvar* svar* rho*
. umeta, noest i2 z q
```

Example 5: Trivariate Outcomes with within-study correlations

```
. use umeta_example5, clear
. list yvar* svar* rho*, clean noobs
. umeta yvar* svar* rho*, pred
```

Saved results

`umeta` saves the following in `e()`:

Scalars

<code>e(N)</code>	number of observations
<code>e(dims)</code>	number of outcomes for meta-analysis
<code>e(df_r)</code>	degrees of freedom for meta-analysis estimation
<code>e(Qdf)</code>	degrees of freedom for homogeneity testing

Macros

<code>e(cmd)</code>	umeta
<code>e(cmdline)</code>	command as typed
<code>e(properties)</code>	b v
<code>e(yvars)</code>	names of study-specific outcome variables (point estimates)
<code>e(svars)</code>	names of study-specific sampling variances
<code>e(predict)</code>	program used to implement predict

Matrices

<code>e(b)</code>	coefficient vector
<code>e(V)</code>	variance-covariance matrix of the estimators
<code>e(Isqmat)</code>	matrix of outcome-specific I^2 values
<code>e(Qmat)</code>	matrix of outcome-specific heterogeneity statistic
<code>e(Vtyp)</code>	typical within-study variance
<code>e(Sigma)</code>	between-study variance-covariance matrix
<code>e(svars)</code>	matrix of study-specific sampling variances
<code>e(rho)</code>	matrix of between-study correlation
<code>e(yvars)</code>	matrix of study-specific point estimates

Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

Authors

Ben A. Dwamena, Department of Radiology, Division of Nuclear Medicine, University of Michigan Medical School, Ann Arbor, Michigan

Yan Ma, Hospital for Special Surgery, Weill Medical College of Cornell University, New York, New York

programming problems:

bdwamena@umich.edu.

u-statistic-based questions:

yam2007@med.cornell.edu.

Title

umeta postestimation — Postestimation tools for umeta

Description

`umeta` is programmed as an Stata estimation command and so supports many of the commands listed under help `estcom` and `postest`. The following standard postestimation commands may be particularly useful:

Command	Description
<u>estat</u>	VCE and estimation sample summary. See help estat
<u>estimates</u>	Cataloging estimation results. See help estimates
<u>lincom</u>	Point estimates, standard errors, testing, and inference for linear combinations of coefficients. See lincom
<u>nlcom</u>	Point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients. See nlcom
<u>predict</u>	predictions, residuals, influence statistics, and other diagnostic measures
<u>test</u>	Wald tests of linear hypotheses. See help test
<u>testnl</u>	Wald tests of non-linear hypotheses. See help testnl

predict Syntax

The syntax of [predict](#) following [umeta](#) is

syntax 1:

```
predict [type] newvarname [if exp] [in range] [, statistic ]
```

syntax 2:

```
predict newvarname [if exp] [in range] [, statistic show(string)]
```

<i>statistic</i>	Description
<u>fixed</u>	prediction of fixed-effects; the default
<u>stfixed</u>	standard error of the fixed-effects prediction
<u>fitted</u>	prediction including random effects
<u>stfit</u>	standard error of <u>fitted</u>
<u>stdf</u>	standard error of the forecast
<u>reffects</u>	predicted random effects
<u>reses</u>	standard error of predicted random effects
<u>rstandard</u>	standardized predicted random effects
<u>lev</u>	leverage (diagonal elements of projection matrix)
<u>cooks</u>	Cook's influence measure

These statistics are available both in and out of sample; type "predict ... if e(sample) ..." if wanted only for the estimation sample.

<i>show</i>	Description
<u>clean</u>	force table format with no divider or separator lines
<u>table</u>	force table format
<u>abbreviate(#)</u>	abbreviate variable names to # characters; default is <u>ab(8)</u>
<u>noobs</u>	do not list observation numbers
<u>divider</u>	draw divider lines between columns
<u>separator(#)</u>	draw a separator line every # lines; default is <u>separator(5)</u>

Options for predict

fixed calculates the linear prediction for the fixed portion of the model.

stfixed calculates the outcome-specific standard error of the fixed-portion linear prediction

stfitted calculates the outcome-specific standard error of the prediction including random effects.

fitted calculates the outcome-specific prediction including random effects, $Xb[i] + u[i]$, also known as the empirical Bayes estimates of the effects in each study.

stdf calculates the outcome-specific standard error of the forecast. This gives the standard deviation of the predicted distribution of the **true** value of *depvar* in a future study $stdf^2 = stdp^2 + tau^2$.

reffects calculates the outcome-specific best linear unbiased predictions (BLUPs) of the random effects, also known as the posterior mean or empirical Bayes estimates of the random effects, or as shrunken residuals.

reses calculates the outcome-specific standard error of predicted random effects.

rstandard calculates the outcome-specific standardized predicted random effects, i.e. the predicted random effects $u[i]$ divided by their (unconditional) standard errors. These may be useful for diagnostics and model checking.

lev calculates the study-specific leverages

cooks calculates the study-specific Cook's influence statistic.

Remarks

Similar to other types of data, it is not uncommon to observe extreme effect size values when conducting a meta-analysis. As the main objective of a meta-analysis is to provide a reasonable summary of the effect sizes of a body of empirical studies, the presence of such outliers may distort the conclusions of a meta-analysis. Moreover, if the conclusions of a meta-analysis hinge on the data of only one or two influential studies, then the robustness of the conclusions are called into question. Researchers, therefore, generally agree that the effect sizes should be examined for potential outliers and influential cases when conducting a meta-analysis.

The most thorough treatment of outlier diagnostics in the context of meta-analysis to date can be found in the classic book by Hedges and Olkin, who devoted a whole chapter to diagnostic procedures for effect size data. However, the methods developed by Hedges and Olkin(1985) are only applicable to fixed-effects models. Given that random- and mixed-effects models are gaining popularity in the meta-analytic context, corresponding methods for outlier and influential case diagnostics need to be developed.

Viechtbauer and Cheung(2010) have introduced several outlier and influence diagnostic procedures for the random- and mixed-effects model in meta-analysis. These procedures are logical extensions of the standard outlier and case-deletion influence diagnostics for regular regression models as in Demidenko and Stukel(2005) and take both sampling variability and between-study heterogeneity into account. The proposed measures provide a simple framework for evaluating the potential impact of outliers or influential cases in meta-analysis.

Examples

```
. use umeta_example5, clear
. umeta yvar* svar* rho*
. predict lev, lev show(clean)
. predict cook, cooks show(clean)
. predict fit, fit
```

```
. predict fix
. predict reff, reff show(clean noobs)
. predict res, res
. predict rst, rst
. predict stpred, stfit

. predict double stdf, stdf
```

Author

Ben A. Dwamena, Department of Radiology, Division of Nuclear Medicine,
University of Michigan Medical School, Ann Arbor, Michigan.
bdwamena@umich.edu.

References

- Demidenko, E., T. A. Stukel. 2005 Influence analysis for linear mixed-effects models *Statistics in Medicine* 24: 893~909
- DerSimonian, R., and N. Laird. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177-188.
- Hedges LV, I. Olkin. 1985. *Statistical Methods for Meta-Analysis* Academic Press: New York.
- Higgins, J. P. T., and S. G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21: 1539-1558.
- Higgins, J. P. T., S. G. Thompson, and D. J. Spiegelhalter. 2009. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A* 172: 137-159.
- Jackson, D., I. R. White, and S. G. Thompson. 2010. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine* 29: 1282-1297.
- Ma, Y., and M. Mazumdar. 2011. Multivariate meta-analysis: a robust approach based on the theory of U-Statistic. *Statistics in Medicine* 30: 2911-2929.
- Riley, R. D. 2009. Multivariate meta-analysis: The effect of ignoring within-study correlation. *Journal of the Royal Statistical Society, Series A* 172: 789-811.
- Viechtbauer, W., M. W.-L. Cheung. 2010. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods* 1: 112-125.
- White, I. R. 2009. Multivariate random-effects meta-analysis. *Stata Journal* 9: 40-56.

Also see

Help: [mvmeta](#) (if installed)