14th International Research/Expert Conference
"Trends in the Development of Machinery and Associated Technology"
TMT 2010, Mediterranean Cruise, 11-18 September 2010

IT SUPPORTED STATISTICAL INFERENCE: A COMBINED META-ANALYSIS OF TERRITORIALLY STRATIFIED DATA

Nina Bijedic
Faculty of Information Technologies, University "Dzemal Bijedic" of Mostar
M. Tita bb, Mostar
Bosnia and Herzegovina

Dzevad Zecic
Faculty of Economy, University of Zenica
Travnicka 1, Zenica
Bosnia and Herzegovina

ABSTRACT
There are various designs for inference about stratified data, based on stratified sampling schemes. The approach in this paper proposes to deploy the methodology of meta-analysis, in particular the combined frequentist and Bayesian model for territorially stratified data.
The research presented in this paper uses meta-regression of statistical data collected in Federation of Bosnia and Herzegovina, territorially stratified in administrative units – 10 cantons. In order to show methodologically richer example, we use odds ratio of new versus old passenger versus transport vehicle purchase, so infringing on one aspect of the consumer power in BiH. The proposed estimate is the REM Bayesian regression (log odds), which is impossible to derive without the use of IT. We use the results of FEM and REM frequentist meta-analysis to frame and test the results.
Keywords: inference, bayesian, meta-regression, frequentist

1. INTRODUCTION

Applied mathematics and the inference it supports is tightly relaying on the development of information technologies (IT). Nowadays, it is difficult to conduct any mathematical analysis without the IT support. In this paper, we present the inference based on bayesian approach, and controlled by the frequentist methodology.
The methodology of meta-analysis serves to analyze the results of already published research on the same topic. It can be either solely frequentist or bayesian, or combined, as presented in [1]. In this paper, we deploy the combined methodology to analyze results published at the same time but from the different regions. We focus on the IT-related details that justify the use of the method.

2. METHODS

From the point of view of data analysis, the combined model is relaying on bayesian meta-regression for the estimate, and is deploying frequentist meta-analysis methods for analysis of heterogeneity, input for informative prior and sensitivity analysis. Therefore, inference is bayesian, and consequently in the terms of probability theory, but subjectivity is avoided with the use of the frequentist "frame".
For the purpose of the research presented in this paper, we chose the topic to be suitable for meta-analysis with the data available at the official web site of statistical agency of Bosnia and Herzegovina [2]. The intent was to analyze data from Cantonal clinical hospitals, but since there are no such data available online, we chose another topic. After the analysis of the available data, we chose to analyze

the odds ratio (OR) of the purchase of new passenger versus new transport vehicle, and therefore conclude about the consumer power of physical individuals (buying passenger vehicles) vs. legal bodies (buying transport vehicles).
The research question is which part of the population is more likely to invest into a new vehicle.

```
model {
 for( i in 1 : N ) {
    kT[i] ~ dbin(pT[i], nT[i])
    kP[i] ~ dbin(pP[i], nP[i])
    logit(pT[i]) <- mi[i]
    logit(pP[i]) <- mi[i] + delta[i]
    mi[i] ~ dnorm(0.0,1.0E-5)
    delta[i] ~ dnorm(d, tau) }
  d ~ dnorm(0.0,1.0E-6)
# REM variance prior choice
# tau ~ dgamma(0.001,0.001)
# sigma <- 1 / sqrt(tau)
# Prior 2
tau<-1/(sigma*sigma)
sigma~dunif(0,10)
delta.new ~ dnorm(d, tau)  }
```

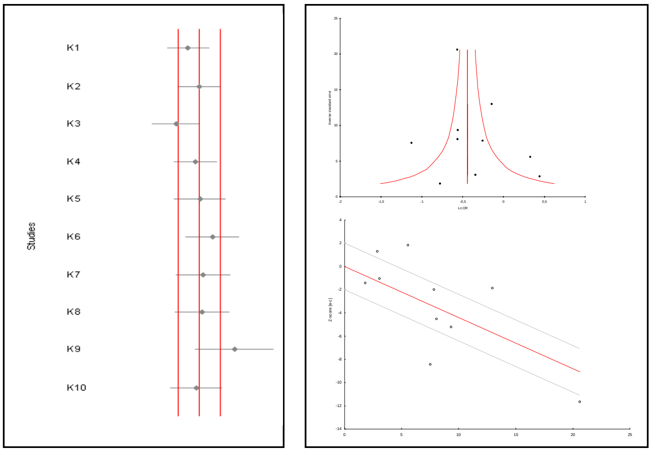

Figure 1. The log-normal model. WinBUGS 1.4.3

Figure 2. Heterogeneity analysis - exclusion sensitivity plot (left) funnel plot 1/se (right-up) and Galbraith plot. MiX 1.7.

After the above steps, there followed the heterogeneity analysis in MiX 1.7 [3]. We used heterogeneity statistics for frequentist fixed effect model (FEM) such as:

- Cohran's Q [4],
- Residual standard deviation H,
- Percent of total variation not attributable neither to sampling error, nor to randomness, I^2 [5], and
- Variability among researches independent of number of statistical units, t^2.

For diagnostics, we show a variety of graphical methods, such as funnel, Gailbright and trim-and-fill plot [6].
For the bayesian informative prior [7], we used binomial priors for the odds of new passenger/transport (k^T/k^P) vs. used vehicles, resulting in log-normal distribution for odds ratio (OR) and uniform prior on standard deviation in WinBUGS 1.4.3 [8] (see Figure 1)
The sensitivity analysis consists of alternative non-informative bayesian Gamma prior on precision, and frequentist random effect model (fREM) Der Simonian and Laird inverse effect estimate of OR (DerSimonian i Laird 1986).

Table 1. fFEM Heterogeneity analysis

Heterogeneity	
Q	77,4912
p-value (two-tailed)	< 0,0001
H	2,9343
95% CI lower limit	2,2775
95% CI upper limit	3,7806
I^2	88,39%
95% CI lower limit	80,72%
95% CI upper limit	93%
t^2	0,1029

Table 2. fFEM meta-analysis as input for bREM

General data	
Number of studies	10
Total number of participants	71977
OR (MH) - FEM	
Meta-analysis outcome	0.6435
95% CI lower limit	0.6035
95% CI upper limit	0.6860
z	13,4912
p-value (two-tailed)	< 0,0001

3. RESULTS

We presented the results of the heterogeneity analysis, (source MiX 1.7), in Table 1 and Figure 2, followed by the results of the frequentist FEM (fFEM) meta-analysis, as input for bayesian REM (bREM) meta-analysis in Table 2.
Furthermore, we present the results of bayesian REM (see Figure 1) after initial 1000 followed by 10000 iterations, in Table 3 and Table 4 for informative prior and the alternative in sensitivity analysis, respectively, source WinBUGS 1.4.3.

Table 3. bREM informative prior

node	mean	sd	MC error
d	-0.352	0.178	0.003
delta.new	-0.357	0.564	0.006
sigma	0.506	0.173	0.005

Table 4. bREM non-informative prior

node	mean	sd	MC error
d	-0.357	0.161	0.002
delta.new	-0.348	0.496	0.005
sigma	0.449	0.146	0.003

The graphical representation in Figure 3 presents the characteristics of Markov chains for the informative prior and the alternative from sensitivity analysis (see Figure 1), source WinBUGS 1.4.3. The results of all analysis are summarized in Table 5, where the results of bayesian analysis are obtained as exponents of delta.new.

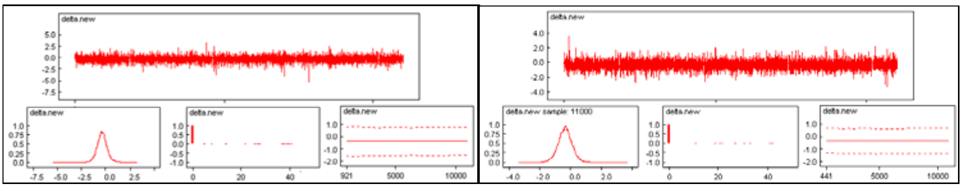

Figure 3. Markov Chain properties for variable delta.new for both informative (left) and non-informative (right) prior imply convergence. Mixing is good (upper), probability density is close to normal (lower-left), autocorrelation is very low (lower-middle) and quartile are stable (lower-right). Source WinBUGS 1.4.3.

4. DISCUSSION

The results of the fFEM heterogeneity analysis (Table 1) show evidence of heterogeneity. From those results we can conclude that situation in Federation of BiH is not the same across the administrative units (Cantons). In order to further analyze heterogeneity, we deploy graphical tests (presented in Figure 2). Both funnel and Galbraith plot imply evidence of heterogeneity. The exclusion sensitivity plot implies that consumer power is lower than expected in Canton 3, and higher in Canton 9. Looking into the background of the problem, one can argue that unemployment rate is very high in Canton 3, and that Canton 9 has the biggest budget in the Federation. Nevertheless, since both confidence intervals (CI) cut the CI of OR, the results should not be excluded from the analysis.
Therefore, we justify deployment of fFEM meta-analysis for informative prior (Table 2).
The graphical tests of Markov chain convergence (Figure 3) validate the results of bayesian meta-analysis presented in Table 3 and Table 4. The results are presented for log odds modeled by variable delta.new.
The left figure is for informative prior and we can see that mixing of the chain is good, even if the space of outcomes is a bit narrow, the distribution is close to Normal, though a bit peaked, autocorrelation is very low, and quartile are stable.
In the right figure we present properties for alternative prior: mixing of the chain is even better than for the informative prior since the space is broader, probability density is close to normal, autocorrelation is very low, and quartile are stable.

Table 5. Results for odds ratio of buying new passenger versus new transport vehicle for all methods

Method	OR: new passenger vs. new transport vehicle	Confidence / credibility interval
fFEM	0.644	(0.604, 0.686)
bREM (informative prior)	0.700	(0.227, 2.162)
fREM	0.686	(0.545, 0.865)
bREM (non- informative prior)	0.706	(0.262, 1.927)

Finally, all of the results presented in Table 5 are justified, and we can compare the methods.
The value of OR used for informative prior is the lowest (0.644) but since it was the input for bayesian meta-analysis we can exclude it from further discussion. The informative prior OR (0.7) is nicely nested into the other values obtained in sensitivity analysis (0.686 and 0.709), so we can argue that the model for informative prior is acceptable.
As expected, frequentist results have narrower CI, while credibility intervals (CI) of bayesian estimates are quite wide. Authors believe that bayesian CI better explain the evidence of heterogeneity.
Taking all of the above discussion in consideration, we conclude that the suggested informative bayesian prior is acceptable estimate of the consumer power in the research.
The answer to the research question “Which part of the population is more likely to invest into a new vehicle?” is: it is more likely that legal entities will invest in new transport vehicle than that physical bodies will invest in new passenger vehicles.

5. CONCLUSSIONS

The research presented in this paper was aiming to deploy well-known methodology in a new manner. That is, as the purpose of meta-analysis is to analyze results from various usually chronologically different researches, we can call it vertical analysis, we deployed it for horizontal analysis in the sense of time. The deployed methods strongly rely on IT support. We focused mostly on the IT component of applied mathematics, namely meta-analysis. Accordingly, we presented the consumer power combined meta-analysis. The variety of methods deployed and two packages, MiX 1.7 and WinBUGS 1.4.3 show the necessity of IT in modern applied mathematics. Presented results favor the use of more complicated and more IT-dependent bayesian method, for the output is in terms of probability theory.

6. REFERENCES

[1] Zecic Dz., Bijedic N., Vukobrat-Bijedic Z.: Frequentist and Bayesian Meta-Analysis: An Example of a Combined Model. TMT 2009 Conference proceedings, Tunisia, 2009.,

[2] Statistical yearbook 2008. Federal Statistical Office of Federation of Bosnia and Herzegovina. http://www.fzs.ba/god2008/GODISNJAK%202008.pdf.,

[3] Bax L, Yu LM, Ikeda N, Tsuruta N, Moons KGM. MIX: Comprehensive free software for meta-analysis of causal research data - version 1.7. http://www.mix-for-meta-analysis.info; 2008.

[4] Cochran WG.: The combination of estimates from different experiments. Biometrics, 10:101-129, 1954.,

[5] Higgins JPT, Thompson SG: Quantifying heterogeneity in a meta-analysis. Statistics in Medicine, 21:1539-1558, 2002.,

[6] Pettiti DB. Meta-analysis, Decision-analysis, and Cost-effectiveness Analysis. New York, NY. Oxford University Press, 1994.,

[7] Berry DA, Stangl DK, editors: Bayesian biostatistics. New York: Marcel Dekker; 1996.,

[8] Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10:325--337. WinBUGS Version 1.4.3 http://www.mrc-bsu.cam.ac.uk/bugs ..