

Meta-Analysis

The purpose of this TP is to practice using **R** for carrying out meta-analysis for two data sets.

As usual, make sure that you read the **help** for any new functions and data that you use.

Smoking data

Load the **HSAUR3** package and then the **smoking** data. Read the help for this data set so that you understand a little about the scientific background and question of interest as well as the variables included in the data. The main aim of the primary studies is to investigate whether patients receiving nicotine gum helps in quitting smoking compare to patients in the control group. The data are in the form of a 2 x 2 table, so we will be analyzing using the *odds ratio*.

Make exploratory boxplots separately for each group as in the **smoking** help.

The measure of difference (effect size) that we consider here is the *odds ratio*. We will first calculate this 'by hand' in **R**. To get the *pooled odds ratio* (across the studies), type:

```
odds <- function(x) (x[1] * (x[4] - x[3])) / ((x[2] - x[1]) * x[3])
weight <- function(x) ((x[2] - x[1]) * x[3]) / sum(x)
W <- apply(smoking, 1, weight)
Y <- apply(smoking, 1, odds)
sum(W * Y) / sum(W)
```

Try to understand how the 2 functions that you created above (**function(x) ...**), and what the **W** and **Y** variables are for.

There is an **R** package called **rmeta**, and we can use functions in this package to simplify the analysis.

```
library("rmeta")
smokingOR <- meta.MH(smoking[["tt"]], smoking[["tc"]], smoking[["qt"]], smoking[["qc"]], names =
rownames(smoking))
```

You can look at the results using the **summary** command on **smokingOR**. What is the combined estimate of the odds ratio? It should be about the same (rounded) as what you computed above. Is the heterogeneity test significant?

Make a *forest plot* using the **plot** command on **smokingOR**. What can you conclude from this plot?

If interested, you can get more information on making and customizing forest plots by reading the help for **metaplot**. Another useful resource is the vignette for the **forestplot** package.

We now quantify the evidence from the forest plot by getting an *overall* estimate of the odds ratio. We have already done a *fixed effects* meta-analysis; we will now carry out a *random effects*

meta-analysis to compare the results:

```
smokingDSL <- meta.DSL(smoking[["tt"]], smoking[["tc"]], smoking[["qt"]], smoking[["qc"]], names =
rownames(smoking))
```

To get a graphical assessment for the presence of *publication bias*, you can make a *funnel plot*:

```
funnelplot(smokingDSL$logOR, smokingDSL$seLogOR, summ = smokingDSL$logOR, xlim = c(-1.,1.7))
abline(v = 0, lty = 2)
```

Does there appear to be evidence of publication bias?

With a little more work, we can apply Fisher's method of *p*-value combination. The package **metap** computes a number of different methods for combining *p*-values. You don't need to read this now, but [this tutorial](#) gives a lot of detailed information on the various methods.

And here we go!!

```
library(metap) # load metap
arr <- array(unlist(t(smoking)),c(2,2,nrow(smoking))) # put all the tables in a 3-d array
arr # look at the array and compare to original data
smoking.normOR <- smokingOR$logOR/smokingOR$seLogOR # standardize log ORs
smoking.logOR.p <- pnorm(smoking.normOR, lower.tail=FALSE) # get one-sided normal p-values
sumlog(smoking.logOR.p) # Fisher combined p
```

What is your overall conclusion here? What evidence is there for the effect of nicotine gum on quitting smoking?

BCG vaccine data

These data are for a meta-analysis on the efficacy of BCG vaccination against tuberculosis (TB). Carry out both a fixed effects and random effects analysis as above. You will need to change the data set **smoking** to **BCG**, and the components (**tt**, etc.) to the component names for the BCG data - you will have to find these (for examples using **names** on **BCG**) The study names are in **BCG\$Study**.

You can look at the summary and forest plot as above. Find the *p*-value for the heterogeneity test - you will see that it is very highly significant. This indicates that we should use the *random effects* meta-analysis.

If you have looked at the **BCG** data, you will see that there are additional covariates: **Latitude** and **Year**. We will now carry out a weighted *meta-regression* of *log-odds ratio* on these variables. The weights are the inverse sum of between-study variance and within study variance. Assuming that you have named your random effects meta-analysis (the result of **meta.DSL**) **BCG.DSL**, this is given by:

```
studyweights <- 1 / (BCG.DSL$tau2 + BCG.DSL$selogs^2)
y <- BCG.DSL$logS
BCG.lm <- lm(y ~ Latitude + Year, data = BCG, weights = studyweights)
```

Look at the summary of this object. You can also make a plot of the estimated log-OR for each variable like you made in the linear regression lab, e.g.:

```
plot(y ~ Latitude, data = BCG, ylab="Estimated log-OR")
abline <- lm(y ~ Latitude, data = BCG, weights = studyweights)
```

Is there evidence of an effect of either variable on the log-odds ratio?

Since the number of studies is small, we will not make a funnelplot here.

Briefly summarize your results and conclusions. Is the vaccine effect in preventing TB?

Toothpaste data

If you have time and are interested, you can carry out a meta-analysis of the **toothpaste** data. Calculate an appropriate measure of effect size (*hint: it is NOT the odds ratio*) for each study the do the meta-analysis of these results. What conclusions do you draw?