Analysis of Literature Coding

Mandy Norrbo

29/01/2019

Required packages

```
library(tidyverse)
```

Loading in the data

```
dat <- read_csv("quantdata.csv")
```

Prevalence of internal meta-analyses

```
#calculating how many articles used internal meta-analysis out of all coded empirical articles
nmini <- dat %>%
  filter(EMPIRICAL == "yes") %>%
  count(USES_MINIMETA == "yes") %>%
  arrange(n)
```

14 out of 66 empirical research articles published between May and September 2019 in the Journal of Experimental Social Psychology used an internal meta-analysis — that is 21%.

Cleaning up data

```
tidydat <- dat %>%
  #selecting columns of interest in analysis
  select(-c("TITLE", "DOI", "EMPIRICAL", "SIGTEST", "COMMENT")) %>%
  #selecting the studies that should be included in the control vs minimeta analysis
  filter(INCLUDE == "yes") %>%
  select(-"PVAL8", -"PVAL9", -"PVAL10", -"PVAL11", -"PVAL12", -"N8", -"N9", -"N10", -"N11", -"N12") %>%
  gather(key = "PVALNO", value = "PVALUES", PVAL1,PVAL2,PVAL3,PVAL4,PVAL5,PVAL6, PVAL7)
```

How many multi-study papers?

```
#no of articles with more than 1 study
multidat <- dat %>%
  select(STUDY_NO) %>%
  filter(STUDY_NO > 1) %>%
  nrow()

#no of articles with just 1 study
singledat <- dat %>%
  select(STUDY_NO) %>%
```

```
  filter(STUDY_NO == 1) %>%
  nrow()
```

59 out of 66 empirical articles reported the findings of more than one study, only 7 reported the findings from a single study.

Number of studies combined in meta-analysis

```
#selecting articles that used minimeta
minidat <- tidydat %>%
  filter(USES_MINIMETA == "yes")

#number of individual studies that were aggregated
nstudymini <- minidat  %>%
  summarise(mean = mean(META_STUDY_NO),
            median = median(META_STUDY_NO),
            sd = sd(META_STUDY_NO))
```

The number of studies combined in the mini meta-analyses coded in this overview range from 2 to 7, with a mean of 4.2, median of 4, and with a standard deviation of 1.55.

Plotting the p-values

```
#labels for plot
supp.labs <- c("Used internal meta-analysis", "Did not use internal meta-analysis")

g1 <- ggplot(tidydat, aes(x=PVALNO, y=PVALUES)) +
  geom_point(aes(colour = STUDY_ID)) +
  geom_line(aes(group = STUDY_ID, colour = STUDY_ID)) +
  geom_segment(aes(x = 0, xend = 7, y =.05, yend = 0.05), colour = "red", size = 0.5, linetype = "longda
  facet_wrap(~CONTROL, labeller = labeller(CONTROL = supp.labs)) +
  theme(axis.text.x  = element_text(size = 5)) +
  theme_bw() +
  theme(legend.position = "none") +
  scale_x_discrete(labels=c("PVAL1" = "1", "PVAL2" = "2",
                            "PVAL3" = "3", "PVAL4" = "4",
                            "PVAL5"= "5", "PVAL6" = "6", "PVAL7" = "7"), name = "Study") +
  scale_y_continuous(name = "p-value") +
  labs(title = "p-value distribution across studies by article type") +
  theme(plot.title = element_text(size = 15, face = "bold"), strip.text.x = element_text(size = 17), axi

ggsave("pvalueplot.png", width = 10, height = 6)
```

How many p > 0.05?

```
#how many p-values > 0.05 grouped by minimeta use (yes/no)
nullp <- tidydat %>%
  filter(PVALUES > 0.05) %>%
  group_by(USES_MINIMETA) %>%
  summarise(n = n())

#how many individual studies in total in 20 articles
studytot <- tidydat %>%
```

```
  filter(PVALUES != "") %>%
  nrow()
```

Out of the 20 articles and 84 individual studies, 13 p-values were nonsignificant. Out of these, 11 were in the articles that used internal meta-analyses, while 2 were in the matched control articles that did not end up using an internal meta-analysis.

Sample sizes

```
#sample size dataframe for plotting & descriptives
ndat <- dat %>%
  filter(INCLUDE == "yes", USES_MINIMETA == "yes") %>%
  select(c("N1", "N2", "N3", "N4", "N5", "N6", "N7", "USES_MINIMETA", "STUDY_ID")) %>%
  gather(key = "NSTUDY", value = "NVALUES",
         N1, N2, N3, N4, N5, N6, N7)

#difference between largest and smallest sample sizes within an article across individual studies
bystudy <- ndat %>%
  group_by(STUDY_ID) %>%
  summarise(min = min(NVALUES, na.rm = TRUE), max = max(NVALUES, na.rm = TRUE)) %>%
  mutate(difference = max - min)
```

Sample sizes within single studies ranged from 40 to 988, with a mean size of 325 and median size of 282.

Plotting sample sizes

```
g2 <- ggplot(ndat, aes(x=NSTUDY, y=NVALUES)) +
  geom_point(aes(colour = STUDY_ID)) +
  geom_line(aes(group = STUDY_ID, colour = STUDY_ID)) +
  theme(axis.text.x  = element_text(size = 5)) +
  theme_bw() +
  theme(legend.position = "none") +
  scale_x_discrete(labels=c("N1" = "1", "N2" = "2",
                            "N3" = "3", "N4" = "4",
                            "N5"= "5", "N6" = "6", "N7" = "7"), name = "Study") +
  scale_y_continuous(name = "Total sample size (N)") +
  labs(title = "Sample size distribution across studies using internal meta-analysis") +
  theme(plot.title = element_text(size = 15, face = "bold"), axis.title = element_text(size = 15))


ggsave("nplot.png", width = 10, height = 6)
```

Pre-registrations

```
#dataframe with prereg information
prereg <- dat %>% filter(INCLUDE == "yes", USES_MINIMETA == "yes") %>% select(STUDY_ID, STUDY_NO,PRE_REG

#percentage pre-registered
percreg <- round(sum(prereg$PRE_REG)/sum(prereg$STUDY_NO),2)*100
#total no. pre-registered
totalreg <- sum(prereg$PRE_REG)
#out of how many studies
outof <- sum(prereg$STUDY_NO)
```

Within the internal meta-analysis articles, 8 out of 47, , that is 17%, were reported to have been pre-registered. Only 1 article (82_2) seemed to have pre-registered all studies within the internal meta-analysis. However, the internal meta-analysis itself did not seem to be pre-registered and thus, when to run it and which studies to include was still flexible.

How many do not report p-values?

```
#articles without p-values were coded as "not sure" in "IS_MINI_SIG" variable
nopval <- dat %>%
  filter(IS_MINI_SIG == "not sure") %>%
  nrow()
```

Out of the 14 coded articles using an internal meta-analysis, 3 did not report an overall significance level (i.e. p-value) for the meta-analysis, instead reporting only effect sizes and confidence intervals.