p-values Simpson index

Clémence Métayer metayer.clemence@yahoo.com (main contributor)
Marta Avalos Fernandez marta.avalos-fernandez@u-bordeaux.fr

2022-07-19

Data import

```
library(metafor)
library(forestplot)
library(lme4)
library(lmerTest)

data_study <- read.csv("data_pvalue_simpson.csv", sep = ";",
    dec = ".")
rownames(data_study) <- c("Den16", "Li17", "Hua20", "Kim17",
    "Pra12", "Fil12", "Sor20", "Buy17")
```

Computation of differences and standard deviations with escalc

```
diff_healthy_stable <- escalc(measure = "SMDH", m1i = Mean_healthy,
    sd1i = SD_healthy, n1i = Size_healthy, m2i = Mean_stable,
    sd2i = SD_stable, n2i = Size_stable, data = data_study)
Sum_HS <- summary(diff_healthy_stable)

diff_healthy_exacerbated <- escalc(measure = "SMDH", m1i = Mean_healthy,
    sd1i = SD_healthy, n1i = Size_healthy, m2i = Mean_exacerbated,
    sd2i = SD_exacerbated, n2i = Size_exacerbated, data = data_study)
Sum_HE <- summary(diff_healthy_exacerbated)

diff_healthy_diseased <- escalc(measure = "SMDH", m1i = Mean_healthy,
    sd1i = SD_healthy, n1i = Size_healthy, m2i = Mean_diseased,
    sd2i = SD_diseased, n2i = Size_diseased, data = data_study)
Sum_HD <- summary(diff_healthy_diseased)

diff_stable_exacerbated <- escalc(measure = "SMDH", m1i = Mean_stable,
    sd1i = SD_stable, n1i = Size_stable, m2i = Mean_exacerbated,
    sd2i = SD_exacerbated, n2i = Size_exacerbated, data = data_study)
Sum_SE <- summary(diff_stable_exacerbated)
```

Data preparation

```
Simpson = c()
Variance = c()
Group = c()
Study = c()
Disease = c()
Sample = c()
Continent = c()
Sequencing = c()
Rarefaction = c()
Taxonomy = c()
Control_size = c()
Case_size = c()

coeff = which(is.na(diff_healthy_stable$yi) == FALSE)
name = rownames(diff_healthy_stable)[coeff]
Simpson = c(Simpson, diff_healthy_stable$yi[coeff])
Variance = c(Variance, diff_healthy_stable$vi[coeff])
Group = c(Group, rep("Healthy/Stable", length(diff_healthy_stable$yi[coeff])))
Study = c(Study, diff_healthy_stable$author[coeff])
Disease = c(Disease, diff_healthy_stable$Disease[coeff])
Sample = c(Sample, diff_healthy_stable$Sample[coeff])
Continent = c(Continent, diff_healthy_stable$Continent[coeff])
Sequencing = c(Sequencing, diff_healthy_stable$Sequencing[coeff])
Rarefaction = c(Rarefaction, diff_healthy_stable$Rarefaction[coeff])
Taxonomy = c(Taxonomy, diff_healthy_stable$Taxonomy[coeff])
Control_size = c(Control_size, diff_healthy_stable$Control_size[coeff])
Case_size = c(Case_size, diff_healthy_stable$Case_size[coeff])

coeff = which(is.na(diff_healthy_exacerbated$yi) == FALSE)
name = rownames(diff_healthy_exacerbated)[coeff]
Simpson = c(Simpson, diff_healthy_exacerbated$yi[coeff])
Variance = c(Variance, diff_healthy_exacerbated$vi[coeff])
Group = c(Group, rep("Healthy/Exacerbated",
↪  length(diff_healthy_exacerbated$yi[coeff])))
Study = c(Study, diff_healthy_exacerbated$author[coeff])
Disease = c(Disease, diff_healthy_exacerbated$Disease[coeff])
Sample = c(Sample, diff_healthy_exacerbated$Sample[coeff])
Continent = c(Continent, diff_healthy_exacerbated$Continent[coeff])
Sequencing = c(Sequencing, diff_healthy_exacerbated$Sequencing[coeff])
Rarefaction = c(Rarefaction, diff_healthy_exacerbated$Rarefaction[coeff])
Taxonomy = c(Taxonomy, diff_healthy_exacerbated$Taxonomy[coeff])
Control_size = c(Control_size, diff_healthy_exacerbated$Control_size[coeff])
Case_size = c(Case_size, diff_healthy_exacerbated$Case_size[coeff])

coeff = which(is.na(diff_healthy_diseased$yi) == FALSE)
name = rownames(diff_healthy_diseased)[coeff]
Simpson = c(Simpson, diff_healthy_diseased$yi[coeff])
```

```
Variance = c(Variance, diff_healthy_diseased$vi[coeff])
Group = c(Group, rep("Healthy/Diseased", length(diff_healthy_diseased$yi[coeff])))
Study = c(Study, diff_healthy_diseased$author[coeff])
Disease = c(Disease, diff_healthy_diseased$Disease[coeff])
Sample = c(Sample, diff_healthy_diseased$Sample[coeff])
Continent = c(Continent, diff_healthy_diseased$Continent[coeff])
Sequencing = c(Sequencing, diff_healthy_diseased$Sequencing[coeff])
Rarefaction = c(Rarefaction, diff_healthy_diseased$Rarefaction[coeff])
Taxonomy = c(Taxonomy, diff_healthy_diseased$Taxonomy[coeff])
Control_size = c(Control_size, diff_healthy_diseased$Control_size[coeff])
Case_size = c(Case_size, diff_healthy_diseased$Case_size[coeff])

coeff = which(is.na(diff_stable_exacerbated$yi) == FALSE)
name = rownames(diff_stable_exacerbated)[coeff]
Simpson = c(Simpson, diff_stable_exacerbated$yi[coeff])
Variance = c(Variance, diff_stable_exacerbated$vi[coeff])
Group = c(Group, rep("Stable/Exacerbated", length(diff_stable_exacerbated$yi[coeff])))
Study = c(Study, diff_stable_exacerbated$author[coeff])
Disease = c(Disease, diff_stable_exacerbated$Disease[coeff])
Sample = c(Sample, diff_stable_exacerbated$Sample[coeff])
Continent = c(Continent, diff_stable_exacerbated$Continent[coeff])
Sequencing = c(Sequencing, diff_stable_exacerbated$Sequencing[coeff])
Rarefaction = c(Rarefaction, diff_stable_exacerbated$Rarefaction[coeff])
Taxonomy = c(Taxonomy, diff_stable_exacerbated$Taxonomy[coeff])
Control_size = c(Control_size, diff_stable_exacerbated$Control_size[coeff])
Case_size = c(Case_size, diff_stable_exacerbated$Case_size[coeff])

data_pvalue <- data.frame(Simpson, Variance, Group, Study, Disease,
    Sample, Continent, Sequencing, Rarefaction, Taxonomy, Control_size,
    Case_size)

data_pvalue$Simpson <- as.numeric(data_pvalue$Simpson)
data_pvalue$Variance <- as.numeric(data_pvalue$Variance)
data_pvalue$Study <- as.factor(data_pvalue$Study)
data_pvalue$Group <- as.factor(data_pvalue$Group)
data_pvalue$Group <- relevel(data_pvalue$Group, ref = "Healthy/Stable")
data_pvalue$Disease <- as.factor(data_pvalue$Disease)
data_pvalue$Sample <- as.factor(data_pvalue$Sample)
data_pvalue$Continent <- as.factor(data_pvalue$Continent)
data_pvalue$Sequencing <- as.factor(data_pvalue$Sequencing)
data_pvalue$Rarefaction <- as.factor(data_pvalue$Rarefaction)
data_pvalue$Taxonomy <- as.factor(data_pvalue$Taxonomy)
data_pvalue$Control_size <- as.numeric(data_pvalue$Control_size)
data_pvalue$Case_size <- as.numeric(data_pvalue$Case_size)

summary(data_pvalue)
```

Anova p-values

```
rmaresults <- matrix(NA, ncol = 3, nrow = 18)
colnames(rmaresults) <- c("Disease", "Variable", "ANOVA p-value")

# Asthma
coef_asthma <- which(data_pvalue$Disease == "Asthma")
data_study_asthma <- data_pvalue[coef_asthma, ]

data_study_asthma$Group <- relevel(data_study_asthma$Group, ref = "Healthy/Stable")
asthma_disease_difference <- rma(yi = Simpson, vi = Variance,
    mods = ~Group, method = "REML", data = data_study_asthma)
rmaresults[1, 1] <- "Asthma"
rmaresults[1, 2] <- "Difference"
rmaresults[1, 3] <- asthma_disease_difference$QMp

data_study_asthma$Sample <- relevel(data_study_asthma$Sample,
    ref = "BAL")
asthma_disease_sample <- rma(yi = Simpson, vi = Variance, mods = ~Sample,
    method = "REML", data = data_study_asthma)
rmaresults[2, 1] <- "Asthma"
rmaresults[2, 2] <- "Sample"
rmaresults[2, 3] <- asthma_disease_sample$QMp

data_study_asthma$Continent <- relevel(data_study_asthma$Continent,
    ref = "Asia")
asthma_disease_continent <- rma(yi = Simpson, vi = Variance,
    mods = ~Continent, method = "REML", data = data_study_asthma)
rmaresults[3, 1] <- "Asthma"
rmaresults[3, 2] <- "Continent"
rmaresults[3, 3] <- asthma_disease_continent$QMp

data_study_asthma$Sequencing <- relevel(data_study_asthma$Sequencing,
    ref = "454 pyrosequencing")
asthma_disease_sequencing <- rma(yi = Simpson, vi = Variance,
    mods = ~Sequencing, method = "REML", data = data_study_asthma)
rmaresults[4, 1] <- "Asthma"
rmaresults[4, 2] <- "Sequencing"
rmaresults[4, 3] <- asthma_disease_sequencing$QMp

data_study_asthma$Rarefaction <- relevel(data_study_asthma$Rarefaction,
    ref = "YES")
asthma_disease_rarefaction <- rma(yi = Simpson, vi = Variance,
    mods = ~Rarefaction, method = "REML", data = data_study_asthma)
rmaresults[5, 1] <- "Asthma"
rmaresults[5, 2] <- "Rarefaction"
rmaresults[5, 3] <- asthma_disease_rarefaction$QMp

data_study_asthma$Taxonomy <- relevel(data_study_asthma$Taxonomy,
```

```
    ref = "OTU")
asthma_disease_taxonomy <- rma(yi = Simpson, vi = Variance, mods = ~Taxonomy,
    method = "REML", data = data_study_asthma)
rmaresults[6, 1] <- "Asthma"
rmaresults[6, 2] <- "Taxonomy"
rmaresults[6, 3] <- asthma_disease_taxonomy$QMp

# COPD
coef_copd <- which(data_pvalue$Disease == "COPD")
data_study_copd <- data_pvalue[coef_copd, ]

data_study_copd$Group <- relevel(data_study_copd$Group, ref = "Healthy/Stable")
# copd_disease_difference <-rma(yi=Simpson, vi=Variance,
# mods = ~ Group, method='REML', data=data_study_copd)
rmaresults[7, 1] <- "COPD"
rmaresults[7, 2] <- "Difference"
rmaresults[7, 3] <- "-"

data_study_copd$Sample <- relevel(data_study_copd$Sample, ref = "BAL")
# copd_disease_sample <-rma(yi=Simpson, vi=Variance, mods =
# ~ Sample, method='REML', data=data_study_copd)
rmaresults[8, 1] <- "COPD"
rmaresults[8, 2] <- "Sample"
rmaresults[8, 3] <- "-"

data_study_copd$Continent <- relevel(data_study_copd$Continent,
    ref = "Asia")
# copd_disease_continent <-rma(yi=Simpson, vi=Variance,
# mods = ~ Continent, method='REML', data=data_study_copd)
rmaresults[9, 1] <- "COPD"
rmaresults[9, 2] <- "Continent"
rmaresults[9, 3] <- "-"

data_study_copd$Sequencing <- relevel(data_study_copd$Sequencing,
    ref = "454 pyrosequencing")
copd_disease_sequencing <- rma(yi = Simpson, vi = Variance, mods = ~Sequencing,
    method = "REML", data = data_study_copd)
rmaresults[10, 1] <- "COPD"
rmaresults[10, 2] <- "Sequencing"
rmaresults[10, 3] <- copd_disease_sequencing$QMp

data_study_copd$Rarefaction <- relevel(data_study_copd$Rarefaction,
    ref = "YES")
copd_disease_rarefaction <- rma(yi = Simpson, vi = Variance,
    mods = ~Rarefaction, method = "REML", data = data_study_copd)
rmaresults[11, 1] <- "COPD"
rmaresults[11, 2] <- "Rarefaction"
rmaresults[11, 3] <- copd_disease_rarefaction$QMp
```

```
data_study_copd$Taxonomy <- relevel(data_study_copd$Taxonomy,
    ref = "OTU")
copd_disease_taxonomy <- rma(yi = Simpson, vi = Variance, mods = ~Taxonomy,
    method = "REML", data = data_study_copd)
rmaresults[12, 1] <- "COPD"
rmaresults[12, 2] <- "Taxonomy"
rmaresults[12, 3] <- copd_disease_taxonomy$QMp

# Cystic fibrosis
coef_cf <- which(data_pvalue$Disease == "Cystic fibrosis")
data_study_cf <- data_pvalue[coef_cf, ]

data_study_cf$Group <- relevel(data_study_cf$Group, ref = "Healthy/Exacerbated")
# cf_disease_difference <-rma(yi=Simpson, vi=Variance, mods
# = ~ Group, method='REML', data=data_study_cf)
rmaresults[13, 1] <- "CF"
rmaresults[13, 2] <- "Difference"
rmaresults[13, 3] <- "-"

data_study_cf$Sample <- relevel(data_study_cf$Sample, ref = "Sputum")
cf_disease_sample <- rma(yi = Simpson, vi = Variance, mods = ~Sample,
    method = "REML", data = data_study_cf)
rmaresults[14, 1] <- "CF"
rmaresults[14, 2] <- "Sample"
rmaresults[14, 3] <- cf_disease_sample$QMp

data_study_cf$Continent <- relevel(data_study_cf$Continent, ref = "America")
# cf_disease_continent <-rma(yi=Simpson, vi=Variance, mods
# = ~ Continent, method='REML', data=data_study_cf)
rmaresults[15, 1] <- "CF"
rmaresults[15, 2] <- "Continent"
rmaresults[15, 3] <- "-"

data_study_cf$Sequencing <- relevel(data_study_cf$Sequencing,
    ref = "454 pyrosequencing")
cf_disease_sequencing <- rma(yi = Simpson, vi = Variance, mods = ~Sequencing,
    method = "REML", data = data_study_cf)
rmaresults[16, 1] <- "CF"
rmaresults[16, 2] <- "Sequencing"
rmaresults[16, 3] <- cf_disease_sequencing$QMp

data_study_cf$Rarefaction <- relevel(data_study_cf$Rarefaction,
    ref = "YES")
# cf_disease_rarefaction <-rma(yi=Simpson, vi=Variance,
# mods = ~ Rarefaction, method='REML', data=data_study_cf)
rmaresults[17, 1] <- "CF"
rmaresults[17, 2] <- "Rarefaction"
rmaresults[17, 3] <- "-"
```

```r
data_study_cf$Taxonomy <- relevel(data_study_cf$Taxonomy, ref = "OTU")
# cf_disease_taxonomy <-rma(yi=Simpson, vi=Variance, mods =
# ~ Taxonomy, method='REML', data=data_study_cf)
rmaresults[18, 1] <- "CF"
rmaresults[18, 2] <- "Taxonomy"
rmaresults[18, 3] <- "-"
```