HW_10

Grebeniuk_Alena

1/22/2024

R Markdown

```
df_chineseRCs<-read.table("chinese_meta_analysis.txt",header=TRUE)
head(df_chineseRCs)
```

```
##   study.id              study        y    se
## 1        1     Hsiao et al 03   50.000 25.00
## 2        2    Gibson et al 13 -120.000 48.00
## 3        3  Vas. et al 13, E1  148.500 50.90
## 4        4  Vas. et al 13, E2   82.600 41.20
## 5        5  Vas. et al 13, E3 -109.400 54.80
## 6        6 Jaeg. et al 15, E1   55.617 65.14
```

```
#install.packages("metafor")
library(metafor)
```

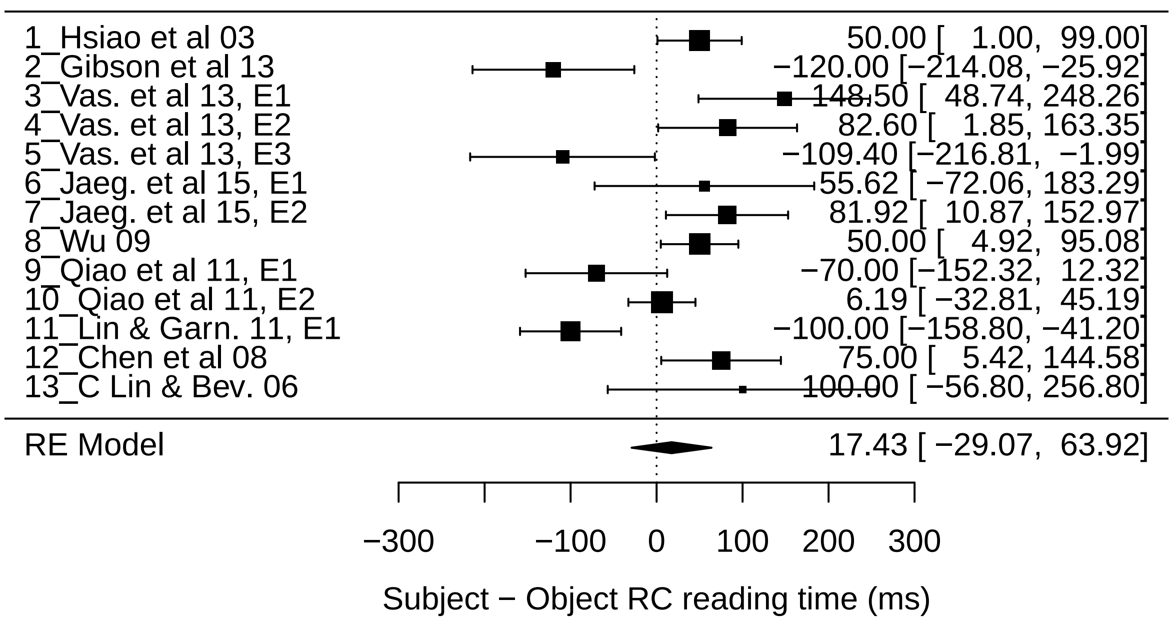
```
## Loading required package: Matrix

## Loading required package: metadat

## Warning: package 'metadat' was built under R version 4.1.2

## Loading required package: numDeriv

##
## Loading the 'metafor' package (version 4.4-0). For an
## introduction to the package please type: help(metafor)
```

```
## run the meta-analysis:
m <- rma(yi=y,
         vi=se^2,
         method = "REML",
         slab = paste(study.id, study, sep = "_"),
         data = df_chineseRCs)
summary(m)
```

```
##
## Random-Effects Model (k = 13; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
## -70.7258  141.4517  145.4517  146.4215  146.7850
##
## tau^2 (estimated amount of total heterogeneity): 5520.9801 (SE = 2950.7285)
## tau (square root of estimated tau^2 value):      74.3033
## I^2 (total heterogeneity / total variability):   82.09%
## H^2 (total variability / sampling variability):  5.58
```

```
##
## Test for Heterogeneity:
## Q(df = 12) = 53.5603, p-val < .0001
##
## Model Results:
##
## estimate       se     zval     pval      ci.lb     ci.ub
##  17.4296  23.7222  0.7347  0.4625  -29.0651  63.9243
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
forest(m,xlab="Subject - Object RC reading time (ms)")
```




```
SEs<-c(48, 54.8, 30, 42, 44.63,19.9, 25, 40.74,23, 65.14,
       35.5,36.25,41.2, 80, 50.9)
n_subj<- c(37,40,48,32,40,24,35,48,40,49,39,49,61,48,60)
chinese_SE_n<-data.frame(SEs,n_subj)
chinese_SE_n
```

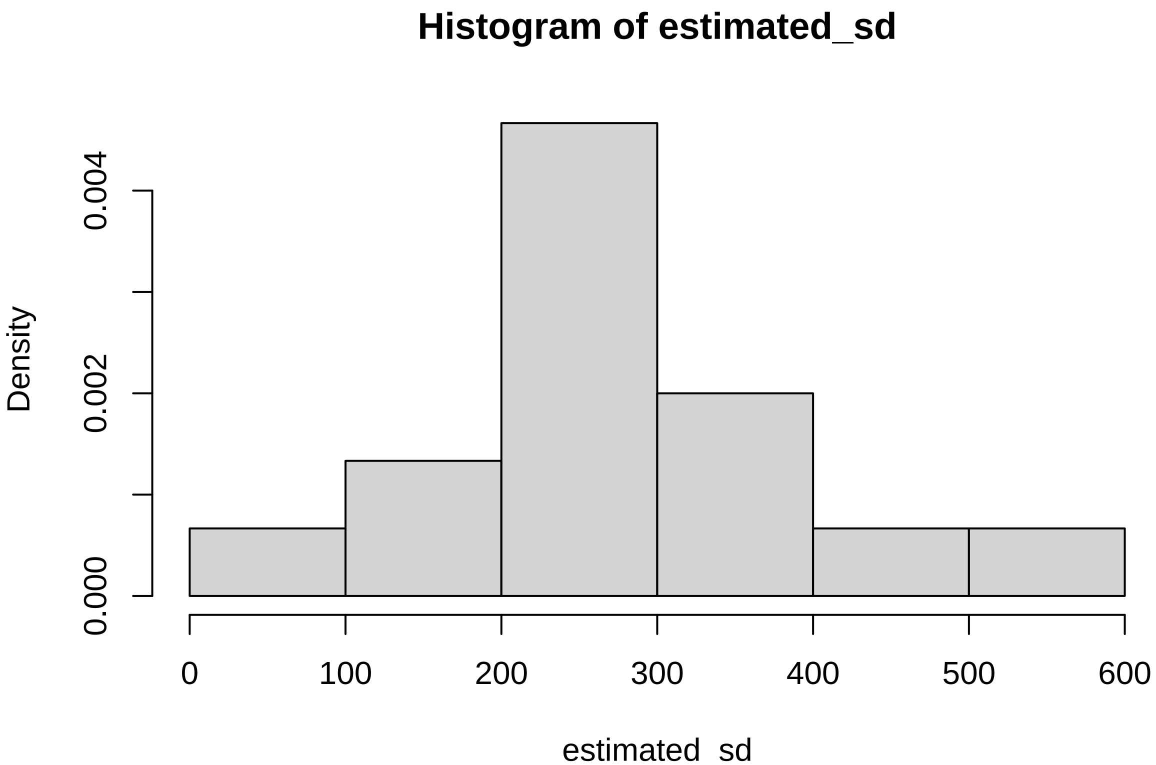
```
##      SEs n_subj
## 1  48.00     37
## 2  54.80     40
## 3  30.00     48
## 4  42.00     32
## 5  44.63     40
## 6  19.90     24
## 7  25.00     35
## 8  40.74     48
## 9  23.00     40
## 10 65.14     49
## 11 35.50     39
## 12 36.25     49
## 13 41.20     61
## 14 80.00     48
```

```
## 15 50.90     60
```

```
estimated_sd<-SEs*sqrt(n_subj)
summary(estimated_sd)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   97.49  214.77  282.25  282.74  334.18  554.26
```

```
hist(estimated_sd,freq=FALSE)
```



Exercise 1

Given that the effect size based on the published data is 17 ms, and assuming that the standard deviation can range from 100 to 560 ms, find out what the sample size (the number of subjects) you would need to have 80% statistical power. Show power calculations for a one-sample t-test assuming a range of standard deviations. For example, do a series of power analyses with sd=100,170, 180,. . . ,560:

```
SDs<-seq(100,560,by=10)
```

```
# Load the pwr package
#install.packages("pwr")
library(pwr)

# Set the effect size (Cohen's d)
effect_size <- 17 / sd(c(100, 560))

# Set the desired power
desired_power <- 0.8

# Perform the power analysis
sample_size <- pwr.t.test(d = effect_size, sig.level = 0.05, power = desired_power, type = "two.sample")
```

```r
# Print the result
cat("The required sample size is:", round(sample_size))
```

```
## The required sample size is: 5748
```

```r
library(pwr)

# Create a sequence of standard deviations
SDs <- seq(100, 560, by = 10)

# Calculate the number of standard deviations
n_SD <- length(SDs)

# Create an empty data frame to store the results
results <- data.frame(SDs, n_subjects_needed = rep(NA, n_SD))

# Perform power analysis for each standard deviation
for (i in 1:n_SD) {
  sd_value <- SDs[i]
  effect_size <- 17 / sd_value

  # Perform the power analysis
  power_result <- pwr.t.test(d = effect_size, sig.level = 0.05, power = 0.8, type = "one.sample")

  # Store the sample size needed for 80% power
  results$n_subjects_needed[i] <- power_result$n
}

# Print the results
print(results)
```

```
##    SDs n_subjects_needed
## 1  100          273.5138
## 2  110          330.5460
## 3  120          393.0101
## 4  130          460.9062
## 5  140          534.2342
## 6  150          612.9940
## 7  160          697.1856
## 8  170          786.8089
## 9  180          881.8641
## 10 190          982.3511
## 11 200         1088.2698
## 12 210         1199.6202
## 13 220         1316.4025
## 14 230         1438.6165
## 15 240         1566.2622
## 16 250         1699.3397
## 17 260         1837.8489
## 18 270         1981.7899
## 19 280         2131.1626
## 20 290         2285.9670
## 21 300         2446.2032
## 22 310         2611.8712
```

```
## 23 320         2782.9709
## 24 330         2959.5023
## 25 340         3141.4655
## 26 350         3328.8604
## 27 360         3521.6870
## 28 370         3719.9454
## 29 380         3923.6356
## 30 390         4132.7574
## 31 400         4347.3110
## 32 410         4567.2964
## 33 420         4792.7135
## 34 430         5023.5623
## 35 440         5259.8428
## 36 450         5501.5551
## 37 460         5748.6992
## 38 470         6001.2749
## 39 480         6259.2824
## 40 490         6522.7217
## 41 500         6791.5927
## 42 510         7065.8954
## 43 520         7345.6299
## 44 530         7630.7961
## 45 540         7921.3940
## 46 550         8217.4237
## 47 560         8518.8851
```

Exercise 2

The meta-analysis estimate actually comes with some uncertainty about the effect size. The estimate is 17 ms with an estimated standard error of 24 ms (look at Model Results below):

```
summary(m)
```

```
##
## Random-Effects Model (k = 13; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
## -70.7258  141.4517  145.4517  146.4215  146.7850
##
## tau^2 (estimated amount of total heterogeneity): 5520.9801 (SE = 2950.7285)
## tau (square root of estimated tau^2 value):      74.3033
## I^2 (total heterogeneity / total variability):   82.09%
## H^2 (total variability / sampling variability):  5.58
##
## Test for Heterogeneity:
## Q(df = 12) = 53.5603, p-val < .0001
##
## Model Results:
##
## estimate       se    zval    pval     ci.lb    ci.ub
##  17.4296  23.7222  0.7347  0.4625  -29.0651  63.9243
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(pwr)

# Create a sequence of standard deviations
SDs <- seq(100, 560, by = 10)

# Calculate the number of standard deviations
n_SD <- length(SDs)

# Create an empty data frame to store the results
results <- data.frame(SDs, n_subjects_needed = rep(NA, n_SD))

# Perform power analysis for each standard deviation
for (i in 1:n_SD) {
  sd_value <- SDs[i]
  effect_size <- 35 / sd_value

  # Perform the power analysis
  power_result <- pwr.t.test(d = effect_size, sig.level = 0.05, power = 0.8, type = "one.sample")

  # Store the sample size needed for 80% power
  results$n_subjects_needed[i] <- power_result$n
}

# Print the results
print(results)
```

```
##    SDs n_subjects_needed
## 1  100          66.01882
## 2  110          79.46965
## 3  120          94.20294
## 4  130         110.21839
## 5  140         127.51583
## 6  150         146.09511
## 7  160         165.95614
## 8  170         187.09885
## 9  180         209.52319
## 10 190         233.22913
## 11 200         258.21663
## 12 210         284.48568
## 13 220         312.03626
## 14 230         340.86835
## 15 240         370.98194
## 16 250         402.37703
## 17 260         435.05359
## 18 270         469.01164
## 19 280         504.25118
## 20 290         540.77217
## 21 300         578.57464
## 22 310         617.65858
## 23 320         658.02399
## 24 330         699.67084
## 25 340         742.59916
## 26 350         786.80894
## 27 360         832.30018
```

```
## 28 370         879.07287
## 29 380         927.12702
## 30 390         976.46262
## 31 400        1027.07967
## 32 410        1078.97818
## 33 420        1132.15814
## 34 430        1186.61956
## 35 440        1242.36242
## 36 450        1299.38671
## 37 460        1357.69250
## 38 470        1417.27971
## 39 480        1478.14838
## 40 490        1540.29849
## 41 500        1603.73004
## 42 510        1668.44307
## 43 520        1734.43753
## 44 530        1801.71344
## 45 540        1870.27080
## 46 550        1940.10962
## 47 560        2011.22988
```

Exercise 3

A researcher carries out an experiment on Chinese relative clauses. Here are the data:

```
df_gibsonwu2<-read.table("chineseRCstudy.txt",header=TRUE)
head(df_gibsonwu2)
```

```
##    subj item condition pos   rt    region
## 9   1m1   15   obj-ext   8  832 head noun
## 20  1m1    8  subj-ext   8 2131 head noun
## 33  1m1   11   obj-ext   8  553 head noun
## 46  1m1   10  subj-ext   8 1091 head noun
## 62  1m1   16  subj-ext   8  598 head noun
## 75  1m1   14  subj-ext   8  645 head noun
```

```
means<-round(with(df_gibsonwu2, tapply(rt, IND=condition, mean)))
bysubj<-aggregate(rt~subj+condition,mean,data=df_gibsonwu2)
means
```

```
##  obj-ext subj-ext
##      447      555
```

```
bysubj
```

```
##    subj condition       rt
## 1   1m1   obj-ext 547.0000
## 2  1m10   obj-ext 310.1667
## 3  1m11   obj-ext 309.5000
## 4  1m12   obj-ext 556.5714
## 5  1m13   obj-ext 524.7500
## 6  1m14   obj-ext 534.5714
## 7  1m15   obj-ext 369.2500
## 8  1m16   obj-ext 231.6667
## 9  1m17   obj-ext 285.8571
## 10 1m18   obj-ext 642.8571
```

```
## 11  1m2   obj-ext  405.1429
## 12  1m3   obj-ext  456.6250
## 13  1m4   obj-ext  388.8571
## 14  1m5   obj-ext  410.2500
## 15  1m6   obj-ext  351.4286
## 16  1m7   obj-ext  300.0000
## 17  1m8   obj-ext  358.0000
## 18  1m9   obj-ext  509.2500
## 19  2m1   obj-ext  311.0000
## 20 2m10   obj-ext  600.0000
## 21 2m11   obj-ext  884.7500
## 22 2m12   obj-ext  260.6667
## 23 2m13   obj-ext  925.1250
## 24 2m14   obj-ext  277.8571
## 25 2m15   obj-ext  339.7500
## 26 2m16   obj-ext  381.5714
## 27 2m17   obj-ext  260.2500
## 28 2m18   obj-ext  435.7143
## 29 2m19   obj-ext  545.6250
## 30  2m2   obj-ext  565.1429
## 31 2m20   obj-ext  377.2857
## 32 2m21   obj-ext  706.5000
## 33 2m22   obj-ext  442.8571
## 34  2m3   obj-ext  732.7500
## 35  2m4   obj-ext  261.2857
## 36  2m5   obj-ext  383.3750
## 37  2m6   obj-ext  473.5714
## 38  2m7   obj-ext  253.5000
## 39  2m8   obj-ext  312.8571
## 40  2m9   obj-ext  469.0000
## 41  1m1  subj-ext  847.2857
## 42 1m10  subj-ext  233.7143
## 43 1m11  subj-ext  524.1429
## 44 1m12  subj-ext  472.1250
## 45 1m13  subj-ext  426.7143
## 46 1m14  subj-ext  347.7500
## 47 1m15  subj-ext  531.0000
## 48 1m16  subj-ext  395.8750
## 49 1m17  subj-ext 1886.2857
## 50 1m18  subj-ext 1611.2500
## 51  1m2  subj-ext  475.2500
## 52  1m3  subj-ext  698.0000
## 53  1m4  subj-ext  371.1250
## 54  1m5  subj-ext  726.4286
## 55  1m6  subj-ext  520.2500
## 56  1m7  subj-ext  336.0000
## 57  1m8  subj-ext  317.1250
## 58  1m9  subj-ext  566.0000
## 59  2m1  subj-ext  325.5714
## 60 2m10  subj-ext  433.1250
## 61 2m11  subj-ext  320.2857
## 62 2m12  subj-ext  609.2500
## 63 2m13  subj-ext  993.1429
## 64 2m14  subj-ext  272.2500
```

```
## 65 2m15  subj-ext  429.5714
## 66 2m16  subj-ext  363.7500
## 67 2m17  subj-ext  338.5714
## 68 2m18  subj-ext  424.5000
## 69 2m19  subj-ext 1034.1429
## 70  2m2  subj-ext  432.1250
## 71 2m20  subj-ext  518.6250
## 72 2m21  subj-ext  614.5714
## 73 2m22  subj-ext  621.2500
## 74  2m3  subj-ext  718.0000
## 75  2m4  subj-ext  262.6250
## 76  2m5  subj-ext  432.5714
## 77  2m6  subj-ext  790.7500
## 78  2m7  subj-ext  306.4286
## 79  2m8  subj-ext  390.0000
## 80  2m9  subj-ext  394.8571
```

```
# one-sample t-test
result<- t.test(rt~condition,paired=TRUE,bysubj)$statistic
print(result)
```

```
##         t
## -2.191819
```

```
new_ttest<-t.test(rt~condition,bysubj,mu=17, paired=TRUE)
new_ttest
```

```
##
##  Paired t-test
##
## data:  rt by condition
## t = -2.5144, df = 39, p-value = 0.01616
## alternative hypothesis: true difference in means is not equal to 17
## 95 percent confidence interval:
##  -222.092796   -8.912561
## sample estimates:
## mean of the differences
##               -115.5027
```

```
df<-39
abs_critical_t<-abs(qt(0.025,df))
abs_critical_t
```

```
## [1] 2.022691
```

```
SE<-90/sqrt(40)
SE
```

```
## [1] 14.23025
```

```
x_bar<-mean(df_gibsonwu2$rt)
x_bar
```

```
## [1] 501.0017
```

```
observed_t<-(x_bar-0)/SE
observed_t
```

```
## [1] 35.20681
```

in this case i suppose that we can reject the null hypothesis. Regarding the plausibility of the effect estimate from this study compared to the meta-analysis estimate. Several reasons could contribute to this difference: Methodological differences, Sample characteristics, Contextual factors.