problemset1_3.R

dehshini

2024-01-30

```
#problem set 1 term 3
#environment
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#read in dataset nepal621
setwd("/Users/dehshini/code/R/problem_sets")
nepal621  <-  read.csv("./datasets/nepal621.csv", header=TRUE)
str(nepal621)
```

```
## 'data.frame':    27121 obs. of  4 variables:
##  $ sex   : chr  "Male" "Male" "Male" "Male" ...
##  $ age   : chr  "3-4" "1-2" "3-4" "3-4" ...
##  $ trt   : chr  "Placebo" "Vit A" "Placebo" "Vit A" ...
##  $ status: chr  "Alive" "Alive" "Alive" "Alive" ...
```

```
head(nepal621)
```

```
##    sex age     trt status
## 1 Male 3-4 Placebo  Alive
## 2 Male 1-2   Vit A  Alive
## 3 Male 3-4 Placebo  Alive
## 4 Male 3-4   Vit A  Alive
## 5 Male  <1   Vit A  Alive
## 6 Male 3-4 Placebo  Alive
```

```
tail(nepal621)
```

```
##          sex age     trt status
```

```
## 27116 Female 1-2 Placebo  Alive
## 27117 Female 1-2 Placebo  Alive
## 27118 Female 1-2   Vit A  Alive
## 27119 Female 3-4 Placebo  Alive
## 27120 Female 1-2   Vit A  Alive
## 27121 Female 3-4 Placebo  Alive
```

```
#Suppose you are interested in choosing an appropriate sample size
# for estimating the 16- month mortality rate for children younger
# than 3 years of age in a developing country in
# which vitamin A supplementation is not currently available.

#subset children under 3 years
nepal_under3  <-  nepal621 %>%
    filter(age == "<1" | age == "1-2")

str(nepal_under3)
```

```
## 'data.frame':    16543 obs. of  4 variables:
##  $ sex   : chr  "Male" "Male" "Male" "Male" ...
##  $ age   : chr  "1-2" "<1" "1-2" "1-2" ...
##  $ trt   : chr  "Vit A" "Vit A" "Vit A" "Placebo" ...
##  $ status: chr  "Alive" "Alive" "Alive" "Alive" ...
```

```
#crosstabulate
ct <- table(nepal_under3$trt, nepal_under3$status)
addmargins(ct)
```

```
##
##           Alive  Died   Sum
##   Placebo  7880   239  8119
##   Vit A    8218   206  8424
##   Sum     16098   445 16543
```

```
# proportions of status for each treatment
ct_prop <- prop.table(ct, margin = 1)
addmargins(ct_prop)
```

```
##
##                Alive       Died        Sum
##   Placebo 0.97056288 0.02943712 1.00000000
##   Vit A   0.97554606 0.02445394 1.00000000
##   Sum     1.94610894 0.05389106 2.00000000
```

```
#the proprtion of children who died in the placebo group
# is 0.0294 while the proportion of children who died in
# the vitamin A group is 0.0245

#note:
#for the ct_prop table,
#row 1 is placebo, row 2 is vitamin A, col 2 is risk of death
#p1 is the risk of death in the placebo group
```

```r
#p2 is the risk of death in the vitamin A group

#calculate sample size using the nepal proportions
#power 80%, alpha = 0.05
power.prop.test(
    n = NULL,
    p1 = round(ct_prop[[1,2]], 4),
    p2 = round(ct_prop[[2,2]], 4),
    sig.level = 0.05,
    power = 0.8,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 17143.9
##              p1 = 0.0294
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
#with proportion in placebo 0.5% lower
power.prop.test(
    p1 = round(ct_prop[[1, 2]], 4) - 0.005,
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.8,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 37442605
##              p1 = 0.0244
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
# with proportion in placebo 0.5% higher
power.prop.test(
    p1 = round(ct_prop[[1, 2]], 4) + 0.005,
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.8,
```

```
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 4576.765
##              p1 = 0.0344
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# with risk in placebo at 1.2 times risk in vitamin A
power.prop.test(
    p1 = round(ct_prop[[2, 2]] * 1.2, 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.8,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 17833.49
##              p1 = 0.0293
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# with risk in placebo at 1.5 times risk in vitamin A
power.prop.test(
    p1 = round(ct_prop[[2, 2]] * 1.5, 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.8,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 3127.366
##              p1 = 0.0367
##              p2 = 0.0245
##       sig.level = 0.05
```

```
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# with risk in placebo at 1.75 times risk in vitamin A
power.prop.test(
    p1 = round(ct_prop[[2, 2]] * 1.75, 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.8,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 1523.066
##              p1 = 0.0428
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
#########################
# power 90%, alpha = 0.05
#########################

# calculate sample size using the nepal proportions
power.prop.test(
    p1 = round(ct_prop[[1, 2]], 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.9,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 22950.32
##              p1 = 0.0294
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
#with proportion in placebo 0.5% lower
power.prop.test(
    p1 = round(ct_prop[[1, 2]], 4) - 0.005,
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.9,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 50125024
##              p1 = 0.0244
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
# with proportion in placebo 0.5% higher
power.prop.test(
    p1 = round(ct_prop[[1, 2]], 4) + 0.005,
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.9,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 6126.491
##              p1 = 0.0344
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
# with risk in placebo at 1.2 times risk in vitamin A
power.prop.test(
    p1 = round(ct_prop[[2, 2]] * 1.2, 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.9,
    alternative = "two.sided"
)
```

```
##
```

```
##      Two-sample comparison of proportions power calculation
##
##               n = 23873.48
##              p1 = 0.0293
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# with risk in placebo at 1.5 times risk in vitamin A
power.prop.test(
    p1 = round(ct_prop[[2, 2]] * 1.5, 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.9,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 4186.157
##              p1 = 0.0367
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
# with risk in placebo at 1.75 times risk in vitamin A
power.prop.test(
    p1 = round(ct_prop[[2, 2]] * 1.75, 4),
    p2 = round(ct_prop[[2, 2]], 4),
    sig.level = 0.05,
    power = 0.9,
    alternative = "two.sided"
)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 2038.455
##              p1 = 0.0428
##              p2 = 0.0245
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```