

Week 06 Good statistical practice II

Open and reproducible science: dependable computations and statistics

Homework - Solution

We will continue to work on the data collected for the article “Avoiding overhead aversion in charity”
<https://www.science.org/doi/10.1126/science.1253932>

The complete data that were produced are available here: <https://dataverse.harvard.edu/dataverse/AvoidOHAversion>

```
data_lab_experiment <-  
  read.table("https://dataverse.harvard.edu/api/access/datafile/:persistentId?persistentId=doi:10.7910/  
            header = TRUE, fill = TRUE)
```

First we complete all necessary calculations for the replication of the results of the lab experiment in the paper.

```
proportions <- data_lab_experiment %>%  
  mutate(overhead_level = case_when(noover == 1 ~ 0,  
                                     (high == 1) | (highcover == 1) ~ 50,  
                                     (low == 1) | (lowcover == 1) ~ 5),  
         cover = case_when(noover == 1 ~ "control",  
                           (highcover == 1) | (lowcover == 1) ~ "covered",  
                           TRUE ~ "not covered")) %>%  
  group_by(overhead_level, cover) %>%  
  count(allocation) %>%  
  group_by(overhead_level, cover) %>%  
  mutate(sum = sum(n),  
         freq = n/sum(n)) %>%  
  ungroup() %>%  
  filter(allocation == 1) %>%  
  group_by(overhead_level, cover) %>%  
  mutate(CI_lower = prop.test(n, sum)$conf.int[1],  
         CI_upper = prop.test(n, sum)$conf.int[2])  
  
proportions <- proportions[, -3]  
proportions$Treatment <- c(1,4,2,5,3)  
  
test1 <- prop.test(x = as.numeric(c(proportions[proportions$Treatment == 1,3],  
                                     proportions[proportions$Treatment == 2,3])),  
                  n = as.numeric(c(proportions[proportions$Treatment == 1,4],  
                                     proportions[proportions$Treatment == 2,4])),  
                  correct = FALSE)  
  
test2 <- prop.test(x = as.numeric(c(proportions[proportions$Treatment == 1,3],
```

```

                                proportions[proportions$Treatment == 3,3])),
n = as.numeric(c(proportions[proportions$Treatment == 1,4],
                    proportions[proportions$Treatment == 3,4])),
correct = FALSE)

test3 <- prop.test(x = as.numeric(c(proportions[proportions$Treatment == 2,3],
                                    proportions[proportions$Treatment == 3,3])),
n = as.numeric(c(proportions[proportions$Treatment == 2,4],
                    proportions[proportions$Treatment == 3,4])),
correct = FALSE)

test4 <- prop.test(x = as.numeric(c(proportions[proportions$Treatment == 1,3],
                                    proportions[proportions$Treatment == 5,3])),
n = as.numeric(c(proportions[proportions$Treatment == 1,4],
                    proportions[proportions$Treatment == 5,4])),
correct = FALSE)

test5 <- prop.test(x = as.numeric(c(proportions[proportions$Treatment == 3,3],
                                    proportions[proportions$Treatment == 5,3])),
n = as.numeric(c(proportions[proportions$Treatment == 3,4],
                    proportions[proportions$Treatment == 5,4])),
correct = FALSE)

mytests <- c("1 vs 2", "1 vs 3", "2 vs 3", "1 vs 5", "3 vs 5")
mypvals <- c(test1$p.value, test2$p.value, test3$p.value, test4$p.value, test5$p.value)
myzvals <- c(sqrt(test1$statistic), sqrt(test2$statistic), sqrt(test3$statistic),
              sqrt(test4$statistic), sqrt(test5$statistic))

data.frame(Combination = mytests,
            "p-value" = format.pval(mypvals, 1, .001),
            "z-value" = round(myzvals, 2)) %>%
  kableExtra::kable() %>%
  kable_classic(full_width = F) %>%
  kable_styling(latex_options = "HOLD_position")

```

Combination	p.value	z.value
1 vs 2	0.329	0.98
1 vs 3	0.001	3.27
2 vs 3	0.020	2.32
1 vs 5	0.774	0.29
3 vs 5	0.003	3.00

Task 1 Multiple comparisons

1. Which two research questions are answered using the lab experiment?

The paper states the following two research questions: “First, we sought empirical support for our assertion above that an increase in overhead costs associated with a donation decreases giving. Existing evidence is limited, and it is important to document it in a scientific investigation. Second, we wanted to gain insight into what drives overhead aversion.”

2. Which tests are performed for which research question?

For the first research question the tests of treatment 2 vs 1, 3 vs 1 and 3 vs 2 were performed.

For the second research question the tests of 5 vs 1 and 5 vs 3 were performed.

3. Would you have suggested to compare more or other groups?

Treatment 4 was not tested at all, a comparison version treatment 1 and versus treatment 5 may have been interesting for the second research question.

4. Would you suggest to adjust for multiple comparisons?

Yes, within each research question.

5. Adjust the p-values for each of the two research questions with the Bonferroni-Holm procedure using the function `p.adjust` and the method `holm`. Additionally, adjust all p-values of the paper together.

```
format.pval(p.adjust(mypvals[1:3],method = "holm"),3)
```

```
## [1] "0.32911" "0.00324" "0.04015"
```

```
format.pval(p.adjust(mypvals[4:5],method = "holm"),3)
```

```
## [1] "0.77446" "0.00533"
```

```
format.pval(p.adjust(mypvals,method = "holm"),3)
```

```
## [1] "0.65823" "0.00539" "0.06023" "0.77446" "0.01065"
```

6. Would the conclusions qualitatively change after adjustment?

Only when adjusting all p-values in the paper a qualitative change occurs in the comparison of 3 vs 4 which becomes nonsignificant.

Task 2 Repeated inspection

1. For the tests of treatment 1 vs. 2 and treatment 1 vs 3 repeatedly inspect significance in the sample by considering that pairs of participants have been assessed at the same time and in the order in which they are given in the data set. Start the inspection after 20 pairs have been assessed and inspect after each pair until the length of the shorter sample is reached. Count how often a significant result would have been observed at each inspection step and plot the curve of p-values versus the observation pairs.

```
samplestreatment <- data_lab_experiment %>%
  mutate(overhead_level = case_when(noover == 1 ~ 0,
                                     (high == 1) | (highcover == 1) ~ 50,
                                     (low == 1) | (lowcover == 1) ~ 5),
         cover = case_when(noover == 1 ~ "control",
                           (highcover == 1) | (lowcover == 1) ~ "covered",
```

```

TRUE ~ "not covered"))

sampletreatment1 <- samplestreatment$allocation[samplestreatment$cover == "control"]
sampletreatment2 <- samplestreatment$allocation[samplestreatment$cover == "not covered" &
  samplestreatment$overhead_level == 5]
sampletreatment3 <- samplestreatment$allocation[samplestreatment$cover == "not covered" &
  samplestreatment$overhead_level == 50]

pvals1 <- NULL
for (i in 20:90){
pvals1 <- c(pvals1, prop.test(x = c(sum(sampletreatment1[1:i]), sum(sampletreatment2[1:i])),
  n = c(length(sampletreatment1[1:i]), length(sampletreatment2[1:i])),
  correct = FALSE)$p.value))
sum(pvals1 < 0.05)

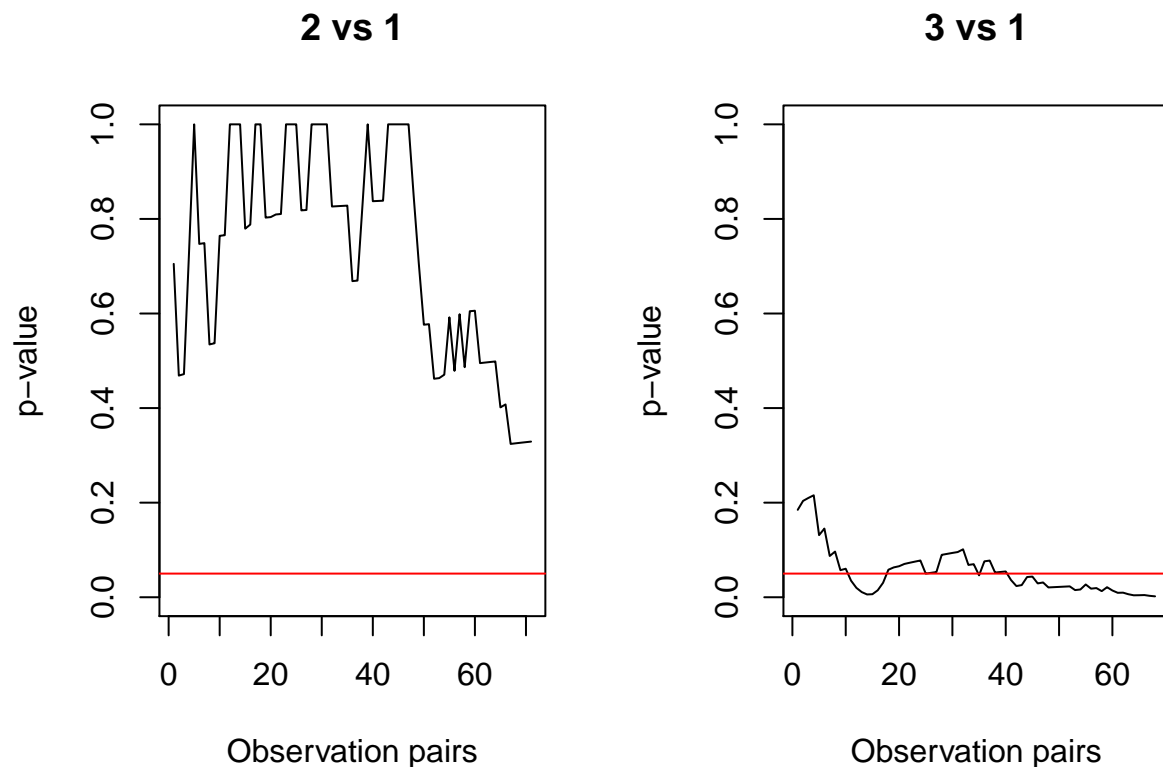
## [1] 0

pvals2 <- NULL
for (i in 20:87){
pvals2 <- c(pvals2, prop.test(x = c(sum(sampletreatment1[1:i]), sum(sampletreatment3[1:i])),
  n = c(length(sampletreatment1[1:i]), length(sampletreatment3[1:i])),
  correct = FALSE)$p.value))
sum(pvals2 < 0.05)

## [1] 37

par(mfrow=c(1,2))
plot(pvals1, type = 'l', xlab = "Observation pairs", ylab = "p-value", ylim = c(0,1))
title("2 vs 1")
abline(h=0.05, col='red')
plot(pvals2, type = 'l', xlab = "Observation pairs", ylab = "p-value", ylim = c(0,1))
title("3 vs 1")
abline(h=0.05, col='red')

```



- Using the function `rbinom` simulate 10 samples of length 90 in both situations (73.3% vs 66.7% and 73.3% vs 49.4%), repeatedly inspect significance as above, count and plot for each of the samples. Discuss your observations within the samples of each situation and between the two situations.

```
set.seed(2022)

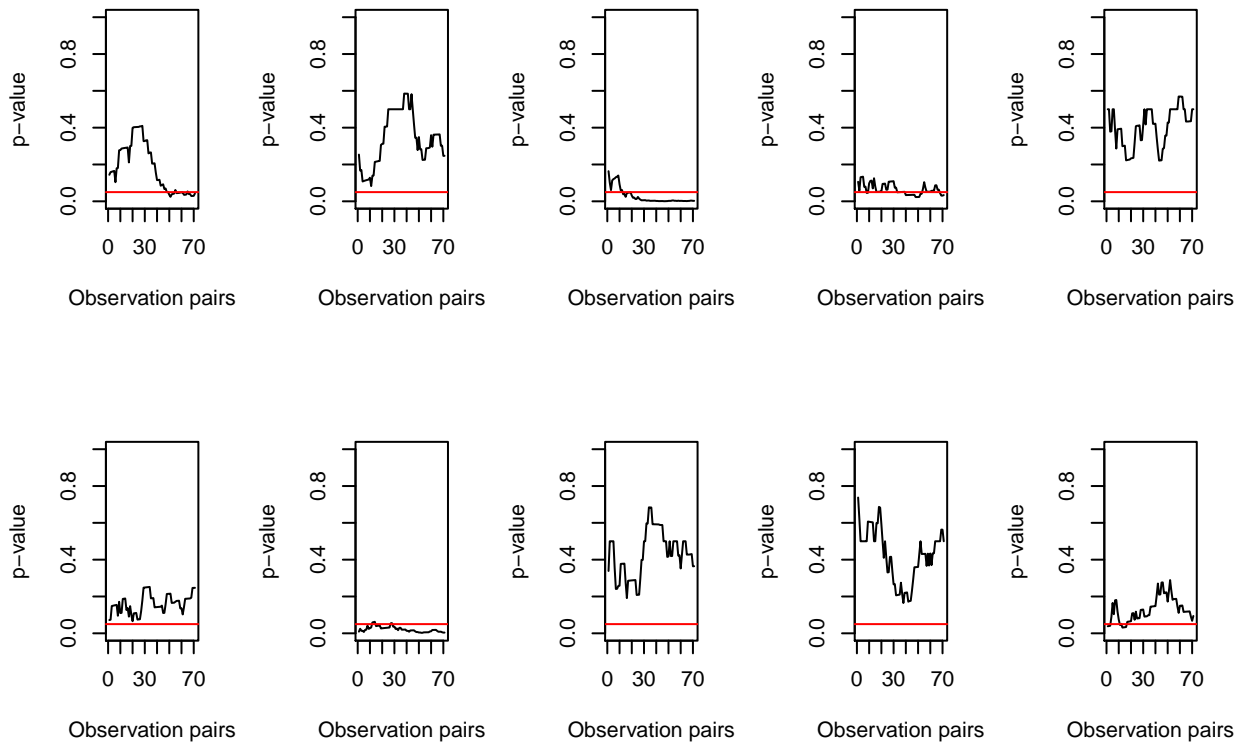
p1 <- .733
p2 <- .667
samplesize <- 90
nsamp <- 10

pvals12 <- matrix(NA, ncol = samplesize-19, nrow = nsamp)
suppressWarnings(for (k in 1:nsamp) {
  s1 <- rbinom(samplesize,1,p1)
  s2 <- rbinom(samplesize,1,p2)
  for (i in 20:(samplesize)) {
    pvals12[k, i-19] <-
      prop.test(
        x = c(sum(s1[1:i] == 1), sum(s2[1:i] == 1)),
        n = c(length(s1[1:i]), length(s2[1:i])),
        correct = FALSE,
        alternative = "greater"
      )$p.value
  }
})
```

```

par(mfrow=c(2,5))
for (i in 1: 10){
  plot(pvals12[i,], type = 'l', xlab = "Observation pairs", ylab = "p-value", ylim = c(0,1))
  abline(h=0.05, col='red')
}

```



```

nosign12 <- apply(pvals12,1, function(x) sum(x < 0.05))
nosign12

```

```
## [1] 21 0 59 23 0 0 66 0 0 8
```

```

p1 <- .733
p3 <- .494
samplesize <- 90
nsamp <- 10

pvals13 <- matrix(NA, ncol = samplesize-19, nrow = nsamp)
suppressWarnings(for (k in 1:nsamp) {
  s1 <- rbinom(samplesize,1,p1)
  s3 <- rbinom(samplesize,1,p3)
  for (i in 20:(samplesize)) {
    pvals13[k, i-19] <-
      prop.test(
        x = c(sum(s1[1:i] == 1), sum(s3[1:i] == 1)),

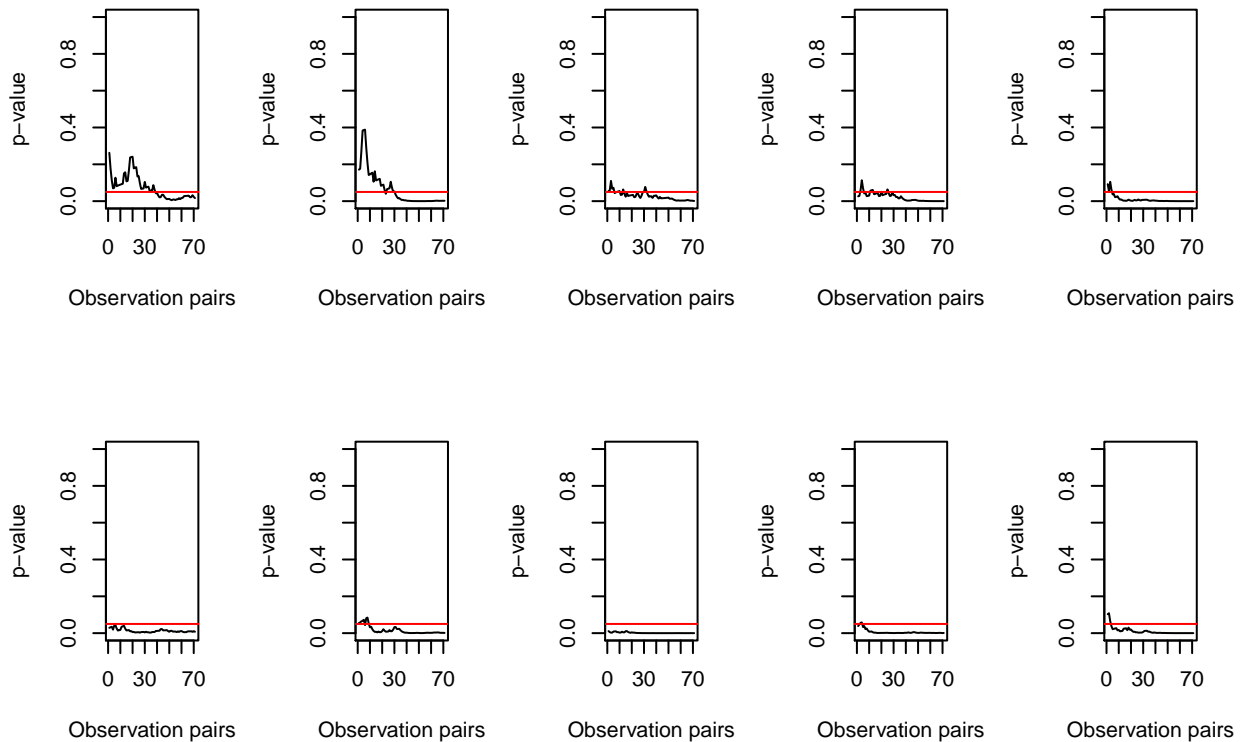
```

```

      n = c(length(s1[1:i]), length(s3[1:i])),
      correct = FALSE,
      alternative = "greater"
    )$p.value
  }
})

par(mfrow=c(2,5))
for (i in 1: 10){
  plot(pvals13[i,], type = 'l', xlab = "Observation pairs", ylab = "p-value", ylim = c(0,1))
  abline(h=0.05, col='red')
}

```



```

nosign13 <- apply(pvals13,1, function(x) sum(x < 0.05))
nosign13

```

```
## [1] 33 42 59 63 67 71 63 71 69 68
```

For the first situation, corresponding to treatment 2 vs 1, there are some simulated cases that are similar to the observed one, but there are also cases that resemble more the 3 vs 1 comparison. For the simulated samples mimicking treatment 3 vs 1 one can observe that all graphs are during long stretches below the threshold.