HW2.R

sanghyun

2025-04-12

```r
set.seed(12345)
y <- matrix(rbinom(10000 * 1000, 1, 0.1), 10000, 1000)
x <- matrix(rnorm(10000 * 1000), 10000, 1000)
x[y==1] <- rnorm(sum(y==1), mean=2)


# 1 ------------------------------------------------------------------------

alpha = 0.05

pvalue_t_20 = 2 * (1 - pt(abs(x), df = 20))
pvalue_t_50 = 2 * (1 - pt(abs(x), df = 50))
pvalue_z = 2 * (1 - pnorm(abs(x)))

#apply(pvalue_t_20 < alpha, 2, sum)
#apply(pvalue_t_50 < alpha, 2, sum)
#apply(pvalue_z   < alpha, 2, sum)

prop_t_20 = apply(pvalue_t_20 < alpha, 2, mean)
prop_t_50 = apply(pvalue_t_50 < alpha, 2, mean)
prop_z = apply(pvalue_z   < alpha, 2, mean)

mean_t_20 = mean(prop_t_20)
mean_t_50 = mean(prop_t_50)
mean_z = mean(prop_z)

ans_1 = data.frame(
  t_20 = mean_t_20,
  t_50 = mean_t_50,
  z = mean_z
)
rownames(ans_1) = "prp_mean"

ans_1
```

```
##               t_20    t_50         z
## prp_mean 0.0799926 0.08995 0.0967012
```

```r
# 2 ------------------------------------------------------------------------

alphas = c(0.01, 0.05, 0.1)
```

```r
p_v = list(T20 = pvalue_t_20, T50 = pvalue_t_50, Z = pvalue_z)

ans_2 = matrix(NA, nrow = length(p_v), ncol = length(alphas))
rownames(ans_2) = c("t_20", "t_50", "z")
colnames(ans_2) = paste("alpha", alphas, sep = "_")


for (i in 1:length(alphas)) {
  alpha = alphas[i]
  for (j in 1:length(p_v)) {
    pv = p_v[[j]]
    tmp = NULL
    for(k in 1:10000) {
      tmp_y = y[k, ]
      tmp[k] = sum(tmp_y[pv[k,] < alpha] == 0) / sum(tmp_y == 0)
    }
    ans_2[j, i] = mean(tmp)
  }
}

ans_2
```

```
##       alpha_0.01 alpha_0.05  alpha_0.1
## t_20 0.004425059 0.03703651 0.08462776
## t_50 0.007396965 0.04466808 0.09379949
## z    0.009996210 0.05003354 0.10006209
```

```r
# 3 ------------------------------------------------------------------------

ans_3 = matrix(NA, nrow = length(p_v), ncol = length(alphas))
rownames(ans_3) = c("t_20", "t_50", "z")
colnames(ans_3) = paste("alpha", alphas, sep = "_")

for (i in 1:length(alphas)) {
  alpha = alphas[i]
  for (j in 1:length(p_v)) {
    pv = p_v[[j]]
    tmp = NULL
    for (k in 1:10000) {
      tmp_y = y[k, ]
      tmp[k] = sum(tmp_y[pv[k,] < alpha] == 1) / sum(tmp_y == 1)
    }
    ans_3[j, i] = mean(tmp)
  }
}

ans_3
```

```
##      alpha_0.01 alpha_0.05 alpha_0.1
## t_20  0.1993276  0.4658230 0.6082822
## t_50  0.2496102  0.4966718 0.6268997
## z     0.2828594  0.5159006 0.6387333
```

```r
# 4 ------------------------------------------------------------------------

alphas = c(0.01, 0.05, 0.1)
p_v = list(T20 = pvalue_t_20, T50 = pvalue_t_50, Z = pvalue_z)

ans_4 = matrix(NA, nrow = length(p_v), ncol = length(alphas))
rownames(ans_4) = c("t_20", "t_50", "z")
colnames(ans_4) = paste("FWER", alphas, sep = "_")

for (i in 1:length(alphas)) {
  alpha = alphas[i]

  for (j in 1:length(p_v)) {
    pv = p_v[[j]]
    fp = NULL

    for (k in 1:1000) {
      multi_pvals = pv[, k]
      adj = p.adjust(multi_pvals, method = "bonferroni")
      signifi = which(adj < alpha)
      fp[k] = ifelse((length(signifi) > 0) & any(y[signifi, k] == 0), 1, 0)

    }
    ans_4[j, i] = mean(fp)
  }
}
ans_4
```

```
##      FWER_0.01 FWER_0.05 FWER_0.1
## t_20     0.000     0.000    0.000
## t_50     0.000     0.004    0.008
## z        0.009     0.042    0.084
```

```r
# 5 ------------------------------------------------------------------------

fdr_levels = c(0.01, 0.05, 0.1, 0.2, 0.3)
p_v = list(T20 = pvalue_t_20, T50 = pvalue_t_50, Z = pvalue_z)

ans_5 = matrix(NA, nrow = length(p_v), ncol = length(fdr_levels))
rownames(ans_5) = c("t_20", "t_50", "z")
colnames(ans_5) = paste("FDR", fdr_levels, sep = "_")

for (j in 1:length(p_v)) {
  pv = p_v[[j]]

  for (i in 1:length(fdr_levels)) {
    fdr_level = fdr_levels[i]
    fdr = NULL

    for (k in 1:1000) {
      multi_pvals = pv[, k]
      adj = p.adjust(multi_pvals, method = "BH")
      signifi = which(adj < fdr_level)
```

```
      fdr[k] = ifelse(length(signifi) == 0, NA, sum(y[signifi, k] == 0) / length(signifi))
    }
    ans_5[j, i] = mean(fdr, na.rm = TRUE)
  }
}

ans_5
```

```
##        FDR_0.01    FDR_0.05    FDR_0.1     FDR_0.2    FDR_0.3
## t_20 0.00000000 0.000000000 0.00000000 0.008248924 0.05540648
## t_50 0.00000000 0.008877082 0.03430960 0.107847040 0.19615274
## z    0.01057778 0.047749745 0.09048853 0.179639239 0.26967016
```