Assignment 3 Question 4 Presentation

Fenglin Chen

15/11/2021

Table of Contents

Summary

4 steps to finding the significance level:

1. State the null hypothesis H0
2. Formulate a discrepancy measure $D(\mathcal{P}_1, \mathcal{P}_2)$
3. Calculate the observed discrepancy $d_{obs} = D(\mathcal{P}_1, \mathcal{P}_2)$
4. Apply D to randomly shuffled sub-populations

The proportion of random discrepancies greater than d_{obs} is the p-value.

Mathematical Formula

Observed discrepancy:

$$d_{obs} = D(\mathcal{P}_1, \mathcal{P}_2)$$

M randomly shuffled sub-populations:

$$(\mathcal{P}^{\star}_{1,1}, \mathcal{P}^{\star}_{2,1}), (\mathcal{P}^{\star}_{1,2}, \mathcal{P}^{\star}_{2,2}), \ldots, (\mathcal{P}^{\star}_{1,M}, \mathcal{P}^{\star}_{2,M})$$

p-value:

$$\text{p-value} = \frac{1}{M} \sum_{i=1}^{M} I\left(D(\mathcal{P}^{\star}_{1,i}, \mathcal{P}^{\star}_{2,i}) \geq d_{obs}\right)$$

General Code

```
# Generate randomly shuffled sub-populations
mixRandomly <- function(pop) {
  pop1 <- pop$pop1
  n_pop1 <- nrow(pop1)

  pop2 <- pop$pop2
  n_pop2 <- nrow(pop2)

  mix <- rbind(pop1,pop2)
  select4pop1 <- sample(1:(n_pop1 + n_pop2), n_pop1,
                        replace = FALSE)

  new_pop1 <- mix[select4pop1,]
  new_pop2 <- mix[-select4pop1,]
  list(pop1=new_pop1, pop2=new_pop2)
}
```

General Code

```
# Define the discrepancy measure
getDiscrepancy <- function(variate, attr) {
  function(pop) {attr(pop$pop1[, variate]) -
                 attr(pop$pop2[,variate])}
}
```

Code Example

Gen Ed class size vs. CTT class size

```
boroughs <- read.csv("Borough_Summary.csv", header=TRUE)
pop <- list(pop1 = subset(boroughs,
                          boroughs$PROGRAM.TYPE == "GEN ED"),
            pop2 = subset(boroughs,
                          boroughs$PROGRAM.TYPE == "CTT"))
set.seed(341)

discFn <- getDiscrepancy("AVERAGE.CLASS.SIZE", mean)
disc.random <- sapply(1:1000, FUN = function(...){
  discFn(mixRandomly(pop))
})

disc.prop <- mean( abs(disc.random) >= discFn(pop) )
disc.prop
```

```
## [1] 0.29
```

Code Example