

Question 4

Jin Barai

Introduction

Suppose we have the population $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2\}$. We can now compare some attributes across the sub-populations of the above population. We need to understand if the difference between attributes of two sub-populations - $a(\mathcal{P}_1)$ and $a(\mathcal{P}_2)$ is unusual or expected.

We use a test of significance to understand this, relative to *randomly mixed* sub-populations.

Test of Significance Algorithm

1. Stating the **null hypothesis** denoted by H_0 : \mathcal{P}_1 and \mathcal{P}_2 are drawn from the same population
2. Construct a discrepancy measure $D = D(\mathcal{P}_1, \mathcal{P}_2)$ where the large values indicate evidence **against** H_0 . The discrepancy measure is also referred to as **test statistic**.
3. Calculate the **observed discrepancy** $d_{obs} = D(\mathcal{P}_1, \mathcal{P}_2)$
4. Shuffle the sub-populations M times and calculate the **observed p-value** given by:

$$p - value = Pr(D \geq d_{obs} \mid H_0 \text{ is true}) \approx \frac{1}{M} \sum_{i=1}^M I(D(\mathcal{P}_{1,i}^*, \mathcal{P}_{2,i}^*) \geq d_{obs})$$

which basically means finding the probability that a randomly shuffled sub-population has a discrepancy measure (D) **at least** as large as what we observed (d_{obs}), given that H_0 is true. The smaller the p -value, the greater the evidence against the null hypothesis.

Note: In order to calculate the p -value *exactly*, one must consider all $\binom{N_1+N_2}{N_1} = \binom{N_1+N_2}{N_2}$ possible permutations, but since it is too many permutations to consider we just use M of them.

Example

Let's use the TitanicSurvival data to find an estimate of the significance levels using test of significance algorithm. The two sub-populations would be the \mathcal{P}_{male} = survived male passengers and \mathcal{P}_{female} = survived female passengers. The attribute would be the average age.

H_0 : The two populations ($\mathcal{P}_{male}, \mathcal{P}_{female}$) are randomly drawn.

```
Titanic = Titanic[Titanic$survived == "yes",]
pop <- list(pop1 = subset(Titanic, Titanic$sex == "male"), pop2 = subset(Titanic, Titanic$sex=="female"))

# The difference in average age of the subpopulations
diffAveAge <- getAveDiffsFn("age")
# The ratio of the standard deviations between the subpopulations
ratioSDAge <- getSDRatioFn("age")
round(c(diffAveAge(pop), ratioSDAge(pop)),3)

## [1] -2.838 1.053
```

Randomly mixing the two sub-populations and calculating the above summaries. We use a sample of $M=5000$ shuffled pairs in this case.

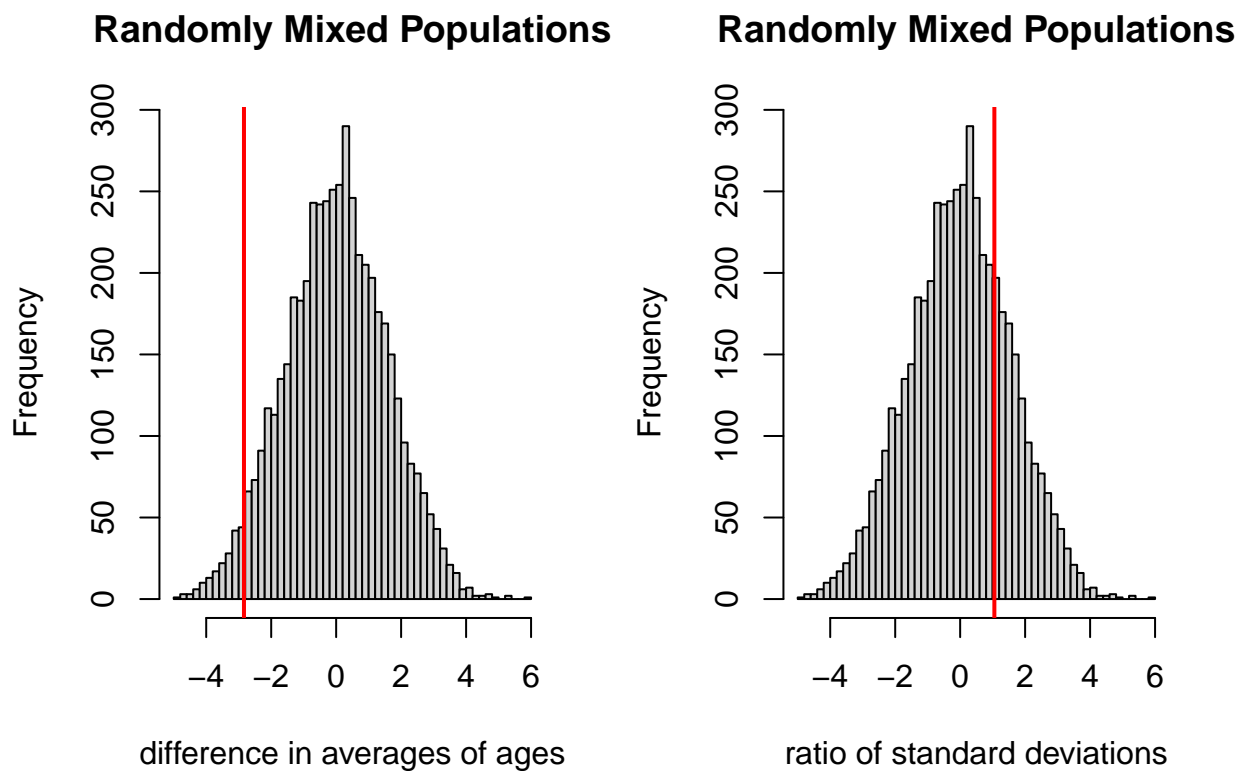
```
set.seed(341)
mixedPop <- mixRandomly(pop)
diffAge <- sapply(1:5000, FUN = function(...) {
  diffAveAge(mixRandomly(pop))
})
round(c(diffAveAge(mixedPop), ratioSDAge(mixedPop)), 3)
```

```
## [1] 0.333 1.061
```

Finding the p -value

```
sum(abs(diffAge) >= abs(diffAveAge(pop))) / length(diffAge)
```

```
## [1] 0.0704
```



Conclusions:

- p -value ≈ 0.0704 for the null hypothesis H_0
- From the p -value we can conclude that there is **weak evidence** against H_0
- Hence, we have very weak evidence against the null hypothesis that the pair $(\mathcal{P}_{male}, \mathcal{P}_{female})$ was randomly drawn

Remarks

- The observed p -value doesn't measure evidence **in favour** of H_0 .
- The observed p -value provides a common (probabilistic) scale on which to measure the **evidence against null hypothesis**
- A test of significance **neither accepts nor rejects a null hypothesis**, it provides a measure of evidence against it.