Honours Lecture Practical

Honours Lecture Practical

Objectives: Rough idea of

- What R is and how it works
- What simulation studies are and how they work

Small Print / Assumptions Two sample t-test

1. Normality: Errors are normally distributed (robust unless n very small)
2. Homogeneity of Variances: Variances of errors are equal across groups $\sigma_1^2 = \sigma_2^2$. (robust unless group sizes unequal $n_{\text{max}}\ n_{\text{min}} > 1.5$)
3. Idependence: Errors are indepdent from each other (not robust)

We are skeptical!

Research questions:

1. t-test valid when the assumptions are met?
2. What happens when the assumptions are violated?

Plan

Create artifical data and check what happens

Recap defintion 95% confidence interval estimator:

- Math: $\mathbb{P}(\theta^* \in \delta(D)) = 95\%$
- English: The true parameter is within 95% of the generated intervals

–> Generate many intervals (data sets) and check

Detailed Plan / Pseduocode

- Repeat often (Repeat)

1. Generate data (Data)
2. Calculate confidence interval (CI)
3. Check: True value in confidence interval? (Check)

- Count how often true value is in confidence interval (Count)

Data: Loading and Looking at Data

```
data(sleep)
head(sleep)
```

```
##   extra group ID
## 1   0.7     1  1
## 2  -1.6     1  2
## 3  -0.2     1  3
## 4  -1.2     1  4
## 5  -0.1     1  5
## 6   3.4     1  6
```

Data: Loading and Looking at Data

```
sleep[c(1:3,18:20),]
```

```
##    extra group ID
## 1    0.7     1  1
## 2   -1.6     1  2
## 3   -0.2     1  3
## 18   1.6     2  8
## 19   4.6     2  9
## 20   3.4     2 10
```

CI: Performing the *t*-test

```
tResult <- t.test(extra ~ group,data=sleep,var.equal=TRUE)
print(tResult)
```

```
##
##  Two Sample t-test
##
## data:  extra by group
## t = -1.8608, df = 18, p-value = 0.07919
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -3.363874  0.203874
## sample estimates:
## mean in group 1 mean in group 2
##            0.75            2.33
```

CI: Extracting the Confidence Interval

```
confInter <- tResult$conf.int
print(confInter)
```

```
## [1] -3.363874  0.203874
## attr(,"conf.level")
## [1] 0.95
```

Check: Check a value in CI

```
toCheck <- 0
if (toCheck > confInter[1] && toCheck < confInter[2]){
  isIn <- TRUE
}else{
  isIn <- FALSE
}
print(isIn)
```

```
## [1] TRUE
```

Data: Generating Artifical Data

5 Coin tosses from a fair coin

```
coinTosses <- rbinom(5,1,0.5)
print(coinTosses)
```

```
## [1] 1 1 0 1 1
```

Data: Generating Normal Data

```
set.seed(1903)
group1 <- rnorm(n=10,mean=20,sd=1)
group2 <- rnorm(n=10,mean=3,sd=1)
print(round(group1,2))
```

```
##  [1] 20.44 19.48 19.20 19.30 20.16 20.77 20.59 21.18 19.
```

```
print(round(group2,2))
```

```
##  [1] 2.31 2.01 3.61 1.87 4.02 4.29 3.48 3.03 2.92 2.97
```

Repeat: Repeating a Part

```
for (i in 1:3){
  variable <- rnorm(n=1,mean=20,sd=1)
  print(sprintf('Iteration: %d, Variable %.2f',i,variable))
}
```

```
## [1] "Iteration: 1, Variable 19.65"
## [1] "Iteration: 2, Variable 20.91"
## [1] "Iteration: 3, Variable 19.86"
```

Count: Check + Repeat

```
tails <- 0
for (i in 1:5){
  coinToss <- rbinom(1,1,0.5)
  if (coinToss==1){
    tails <- tails + 1
  }
}
```

Pair Exercise

Repeat often (for i in 1:100)

1. Generate data according to the assumptions (rnorm)
2. Calculate confidence interval (t.test)
3. Check and Count: True value in confidence interval? (if . . . ; count <- count+1)