

MT.Zunqiu.Wang

Zunqiu Wang

10/15/2021

```
library(reshape2)
```

Q1

```
prime <- function(num) {  
  lst <- c(2) # initialize a vector containing 2  
  sq <- seq(2, num)  
  for (i in sq) {  
    div <- seq(2, sq[i-1]) # since starting 2 so the biggest divider less than  
    # i is i-1 position of sq  
    rem <- sq[i] %% div # test reminder to loop every sq elements  
    if ((0 %in% rem) == FALSE) { # check reminder if 0  
      lst[i] <- sq[i] # added to lst if it has reminder after division  
    } else {  
      next  
    }  
    lst <- lst[!is.na(lst)] # get rid of NAs where reminder is 0  
  }  
  return(lst)  
}  
  
prime(100)
```

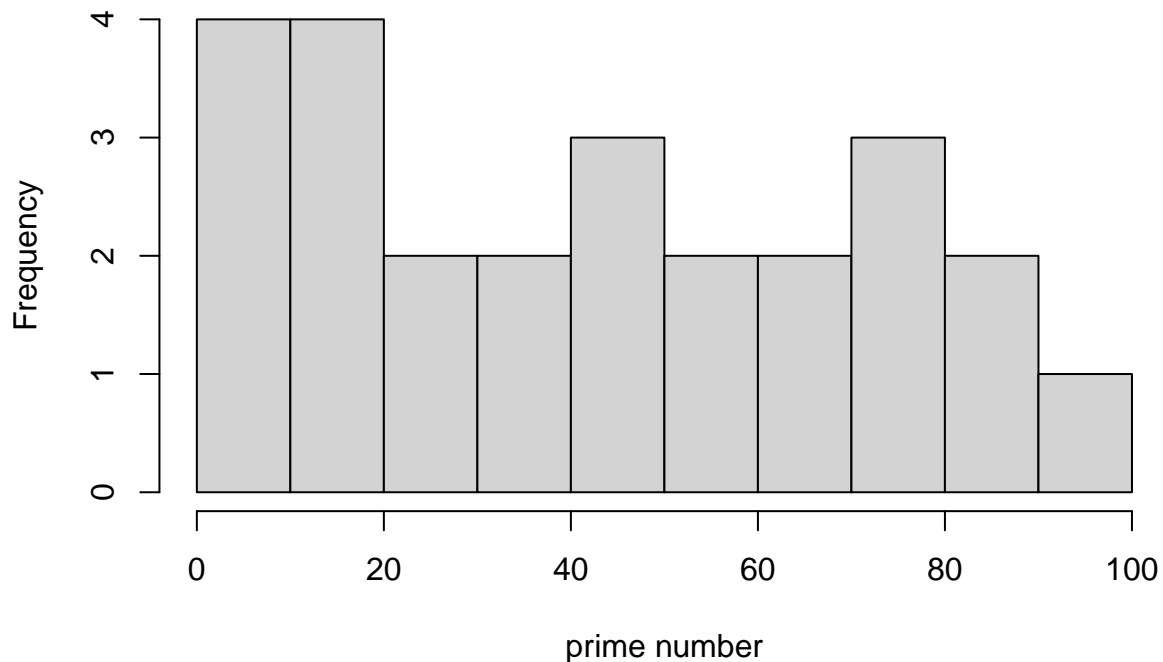
```
## [1] 2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73 79 83 89 97
```

```
prime.list <- prime(100)
```

Q2

```
# distribution of first 100 prime numbers  
hist(prime.list, breaks = 10, xlab = "prime number", main = "histogram of first 100 prime numbers")
```

histogram of first 100 prime numbers



Q3 a

sample space of all possible combinations (HHHHH), (HHHHT), (HHHHT), (HHHTT), (HHHTH), (HHTHT), (HHTTH), (HHTTT), (HTHHH), (HTHHT), (HTHTH), (HTHTT), (HTTHH), (HTTHT), (HTTTH), (HTTTT), (THHHH), (THHHT), (THHTH), (THHTT), (THTHH), (THTHT), (THTTH), (THTTT), (TTHHH), (TTHHT), (TTHTH), (TTHTT), (TTTHH), (TTTHT), (TTTTH), (TTTTT)
 total 32 possible combinations among which at least 3 or more heads in a row are: (HHHTT), (HHHHT), (HHHHT), (HTHHH), (HHHHH), (THHHT), (TTHHH), (THHHH)
 so Prob(more than 3 heads in a row)=8/32=1/4

b

among which given first is head at least 3 or more heads in a row are: (HHHTT), (HHHHT), (HHHHT), (HHHHH), (HTHHH)
 so prob(more than 3 heads in a row given first is head)=(5/32)/(1/2)=5/16

Q4

$$P(A|+) = \frac{P(+|A)P(A)}{P(+|A)P(A) + P(+|A')P(A')}$$

```
# A = hit
sensitivity <- 0.99
fp <- 0.01
pA <- 1/100000

prob <- function(sensitivity, fp, pA) {
  prb <- (sensitivity * pA) / (sensitivity * pA + fp * (1 - pA))
  return(prb)
}
prob(sensitivity, fp, pA)
```

```
## [1] 0.0009890307
```

Q5

```
# for 5 or more we want upper tail  
ppois(5,1,F)
```

```
## [1] 0.0005941848
```

Q6

H_0 =the average college student get 7 hours of sleep a night H_a =the average college student does not get 7 hours of sleep a night

```
# create a function to calculate all statistics  
sample <- c(7,6,5,8,6,6,4,5,8,7)  
x.bar <- mean(sample)  
mu <- 7  
sd <- sd(sample)  
CI <- 0.95  
n <- 10  
desc.val.table <- function(x.bar, mu, sd, n, CI) {  
  se <- sd/sqrt(n)  
  t.stat <- (x.bar-mu)/se # t stat  
  t.crit.perc.low <- (1-CI)/2 # % for lower bound 2 tail  
  t.crit.perc.high <- (1-CI)/2 + CI # % for higher bound 2 tail  
  t.crit.2.tail.low <- qt(t.crit.perc.low , n-1) # t critical for lower bound 2 tail  
  t.crit.2.tail.high <- qt(t.crit.perc.high, n-1) # t critical for higher bound 2 tail  
  CI.low.2.tail <- x.bar - t.crit.2.tail.high*se # CI lower bound for 2 tail  
  CI.high.2.tail <- x.bar + t.crit.2.tail.high*se # CI higher bound for 2 tail  
  p.val.2.tail <- 2*(1- pt(-1 * t.stat, n-1)) # negative t stat has to be  
  # converted to positive for two tail  
  df <- data.frame(t.stat=t.stat,  
                   t.crit.2.tail.low=t.crit.2.tail.low,  
                   t.crit.2.tail.high=t.crit.2.tail.high, CI.low.2.tail=CI.low.2.tail,  
                   CI.high.2.tail=CI.high.2.tail,  
                   p.val.2.tail=p.val.2.tail)  
  df <- melt(df)  
  return(df)  
}  
desc.val.table(x.bar, mu, sd, n, CI)
```

```
## No id variables; using all as measure variables
```

```
##           variable      value  
## 1           t.stat -1.92153785  
## 2 t.crit.2.tail.low -2.26215716  
## 3 t.crit.2.tail.high  2.26215716  
## 4      CI.low.2.tail  5.25818887  
## 5      CI.high.2.tail  7.14181113  
## 6      p.val.2.tail  0.08684229
```

at 95% CI, we fail to reject null hypothesis since $p\text{ val}(0.08) > p\text{ val threshold}(0.05)$

Q7

$$t_{stat} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} < -2.26215716 \quad -2.26215716 = \frac{6.2 - 7}{\frac{1.316561}{\sqrt{n}}} \quad n = 13.8595$$

so $14 - 10 = 4$, we need 4 more people

Q8

```
# paired t test
control <- c(7,6,5,8,6,6,4,5,8,7)
exp <- c(5,4,5,7,5,4,5,4,6,5)
n <- 10
mean.diff <- mean(control- exp)
sd.diff <- sd(control- exp)
se.diff <- sd.diff/sqrt(n)
t.stat <- mean.diff/se.diff
t.stat

## [1] 3.674235

t.crit <- qt(c(0.025,0.975), n-1)
t.crit

## [1] -2.262157  2.262157

p.val <- 2*pt(t.stat, n-1, lower.tail = FALSE)
p.val

## [1] 0.005121073

t.test(control, exp, paired = TRUE, alternative = "two.sided")

##
## Paired t-test
##
## data: control and exp
## t = 3.6742, df = 9, p-value = 0.005121
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4611826 1.9388174
## sample estimates:
## mean of the differences
##                1.2
```

t stat is greater than t crit, we reject null hypothesis

Q9

```
# chi sq calculations
df9 <- data.frame(condition=c("treatment", "control"), live=c(4,8), die=c(11,7))
sum.live <- colSums(df9[,2:3])[1]
sum.die <- colSums(df9[,2:3])[2]
sum.trt <- rowSums(df9[,2:3])[1]
sum.ctr <- rowSums(df9[,2:3])[2]
total <- sum.trt + sum.ctr
e.liv.trt <- sum.live * sum.trt/total
e.liv.ctr <- sum.live * sum.ctr/total
e.die.trt <- sum.die * sum.trt/total
e.die.ctr <- sum.die * sum.ctr/total
chisq.test(df9[,2:3])
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: df9[, 2:3]
## X-squared = 1.25, df = 1, p-value = 0.2636
```

```
qchisq(0.95,1)
```

```
## [1] 3.841459
```

```
1-pchisq(1.25,1)
```

```
## [1] 0.2635525
```

Based on 95% CI, $\chi^2 \text{ stat}(1.25) < \chi \text{ crit}(3.84)$ so we conclude that

plants live or die is not dependent on what they are watered with.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2 = \frac{(4-6)^2}{6} + \frac{(8-6)^2}{6} + \frac{(11-9)^2}{9} + \frac{(7-9)^2}{9} = 1.25$$

$$df = (r-1)(c-1) = 1 * 1 = 1$$

Q10 $H_0 = \mu_w = \mu_v = \mu_c$

H_a = at least one is different

F-stat = $\frac{\text{average variance between groups}}{\text{average variance within groups}}$

N = total number of observations = 40

G = Groups = 3

$$\text{Between variance} = \frac{n_1(\bar{y}_1 - \bar{y})^2 + \dots + n_G(\bar{y}_G - \bar{y})^2}{G-1}$$

$$\text{Within variance} = \frac{(n_1-1)s_1^2 + \dots + (n_G-1)s_G^2}{N-G}$$

$$df_1 = G - 1 \quad df_2 = N - G$$

```

#create a function to calculate F stat, p val based on formula
y <- c(50, 45, 55)
s <- c(10, 7, 4)
n <- c(20, 10, 10)
mu <- 50
anov <- function(y, mu, s, n) {
  bvec <- numeric()
  wvec <- numeric()
  for (i in 1:length(n)) {
    bvec[i] <- n[i] * (y[i]-mu)^2
    wvec[i] <- (n[i]-1) * s[i]^2
  }
  df1 <- length(n) - 1
  df2 <- sum(n) - length(n)
  BV <- sum(bvec)/df1
  WV <- sum(wvec)/df2
  fstat <- BV/WV
  fcrit <- qf(0.95, df1, df2)
  pval <- 1 - pf(fstat, df1, df2)
  df <- data.frame(param = c("f_stat", "f_crit", "p_val"), values = c(fstat, fcrit, pval))
  return(df)
}
anov(y, mu, s, n)

```

```

##      param      values
## 1 f_stat 3.72233400
## 2 f_crit 3.25192385
## 3 p_val 0.03365643

```

at 95% CI, $f \text{ stat}(3.7223) > f \text{ crit}(2.352)$, so we reject null hypothesis

and accept that there is significant difference among 3 groups