

R Exercise #2

Greg Ceccarelli

June 23, 2015

Question 1

Write a function in R that does the following: * Takes sample data (e.g., as a vector), mean of the null hypothesis, and population standard deviation, and a Boolean variable indicating whether we want a one-tailed or two-tailed test. Returns p-values for the required test (Null is “population mean is the one that the function is fed”) , as well as a Boolean value showing whether the test passes a conventional 5% level or not.

- Choose a sample size and then use `rnorm(n, mean, sd)` to generate a random sample and test your function.

```
##import plyr to use rdply
library(plyr)

testfun <- function(data=c(), mu, pop.sd, two.tailed=TRUE, significance=.05) {
  ##computed sample mean
  xbar <- mean(data)
  ##computed sample size
  n <- length(data)
  ##compute z score
  z <- (xbar - mu) / (pop.sd /sqrt(n))

  #modify and compute pvalue depending on type of test required
  if (two.tailed==TRUE) {
    p <- 2 * pnorm(z, lower.tail = FALSE)
  } else if (two.tailed==FALSE & xbar > mu) {
    p <- pnorm((z), lower.tail = FALSE)
  } else if (two.tailed==FALSE & xbar < mu) {
    p <- pnorm((z))
  }

  #create boolean value depending on
  if (p>significance) {
    pass <- FALSE
  } else {
    pass <- TRUE
  }

  ##return a dataframe to make storing replicated results easy
  df <- data.frame(p, pass)
  colnames(df) <- c("pvalue","pass")
  return(df)
}

##test function with rnorm
testfun(rnorm(200,197.3,25.6),191,25.6,FALSE,.05)
```

```
##      pvalue pass
## 1 0.00417458 TRUE
```

- Use replicate or sapply to generate a sample and do the test multiple, say 1000, times. Plot the histogram of p-values that you are getting when the Null is true. Is your function calculating Type-I errors correctly?

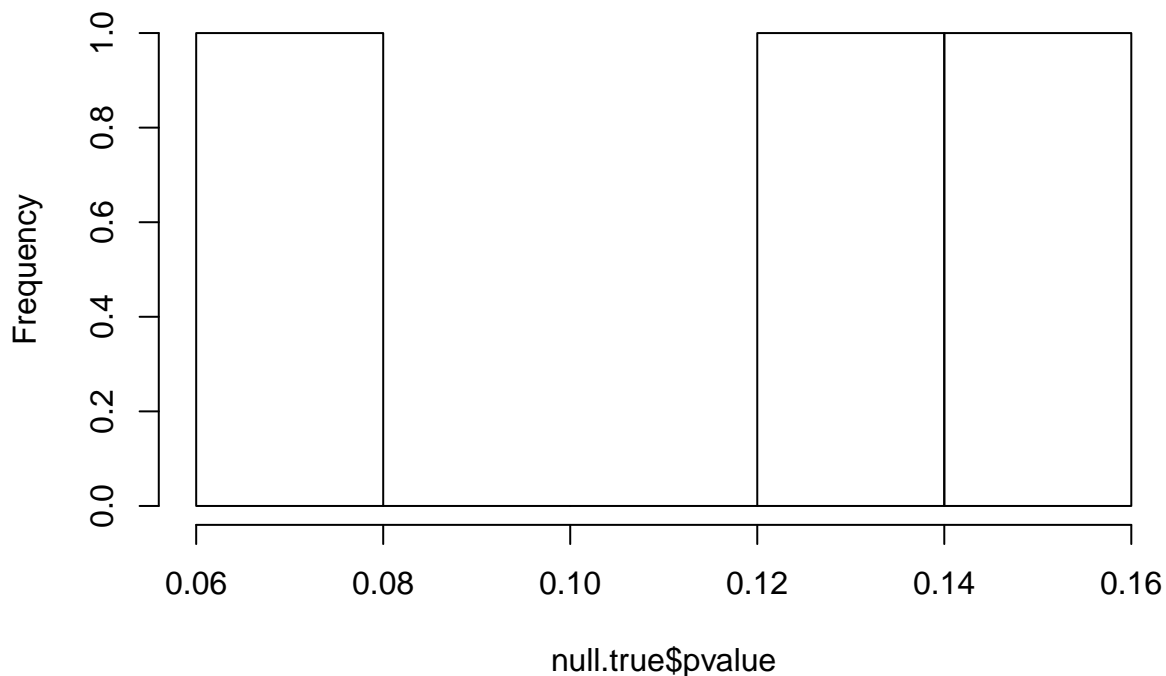
```
##import plyr to use rdply
library(plyr)

##repeat previous example with test function with rnorm using rdply to replicate 100 times
result <- rdply(100, testfun(rnorm(200,197.3,25.6),191,25.6,FALSE,.05))

##save pvalues only when NULL is true (ie, pass==FALSE)
null.true <- result[result$pass==FALSE,]

##plot p-values to understand type 1 error
hist(null.true$pvalue)
```

Histogram of null.true\$pvalue



- Now assume your Null is false. Note: For type-II error calculation, you need a specific assumption about the mean of the population from which the sample is taken. Assume that this mean is one standard deviation above the Null mean.
- Calculate type-II errors both theoretical and by simulation as in step 3. Then increase the sample size and redo this. What happens to statistical power when you increase sample size?

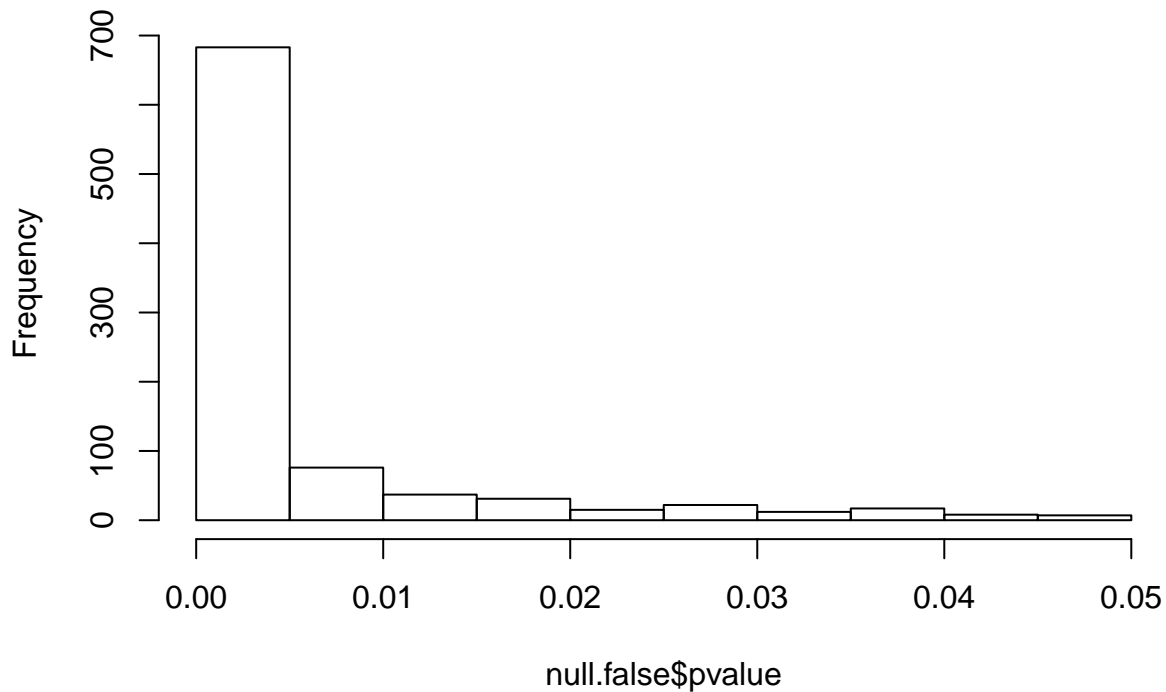
```
##repeat previous example with test function with rnorm using rdply to replicate 1000 times
result <- rdply(1000, testfun(rnorm(70,2,.25),2.09,.25,FALSE,.05))
```

```

result$power <- 1 - (result$pvalue)
##save pvalues only when NULL is FALSE (ie, pass==TRUE)
null.false <- result[result$pass==TRUE,]
##plot p-values to understand type 2 error
hist(null.false$pvalue)

```

Histogram of null.false\$pvalue



```
summary(result$pass)
```

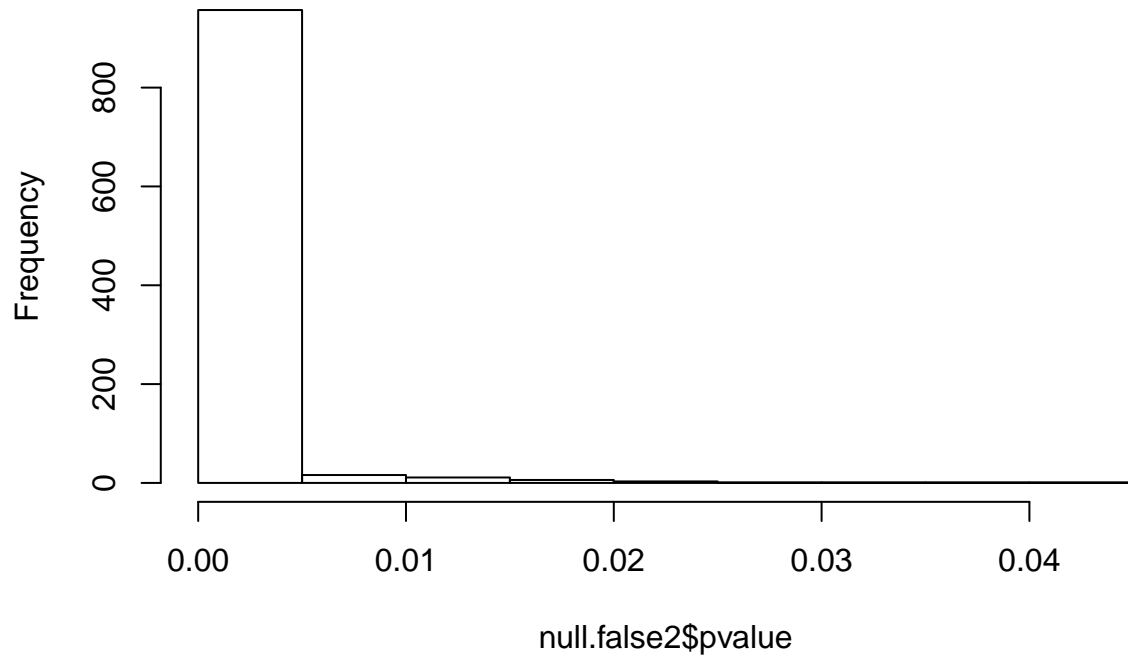
```
##      Mode  FALSE   TRUE  NA's
## logical    92   908    0
```

```

## doubling sample sizes
result <- rdply(1000, testfun(rnorm(140,2,.25),2.09,.25,FALSE,.05))
result$power <- 1 - (result$pvalue)
##save pvalues only when NULL is FALSE (ie, pass==TRUE)
null.false2 <- result[result$pass==TRUE,]
##plot p-values to understand type 2 error
hist(null.false2$pvalue)

```

Histogram of null.false2\$pvalue



```
summary(result$pass)
```

##	Mode	FALSE	TRUE	NA's
## logical		3	997	0