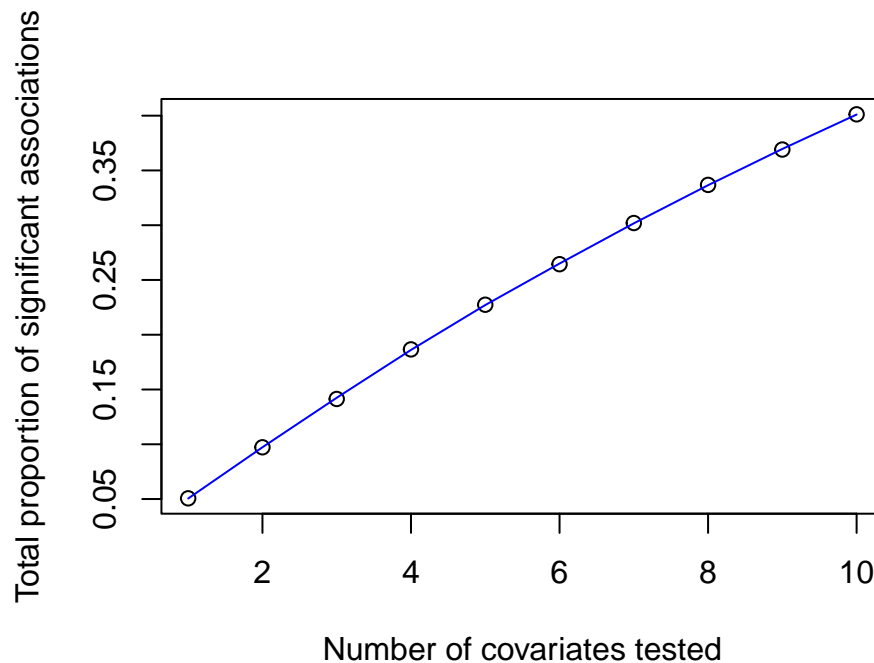


BIOST 505 Homework 2

Katie McFarlane

1/28/22

Question 1



We note a positive trend, but see that it is increasing slightly less each time: based on pure chance of finding a statistically significant association at $\alpha=0.05$, we expect a proportion of roughly

$\sum_{k=1}^{10} (0.95)^{k-1} * 0.05$ for each k covariate, as each proportion includes those that came before it.

Code Appendix

```
###packages
library(knitr)
library(tidyverse)
knitr::opts_chunk$set(echo = FALSE)
### -----
```

```

### Q1 code

#qk counter for number of significant associations for each covariate
qk <- rep(0, 10)

for(i in 1:50000){ #50000 datasets
  #pval holds ten p-values per each dataset
  #is rewritten each time
  pval <- rep(NA, 10)
  #piece holds all covariate and outcome values
  #is rewritten each time
  piece <- matrix(nrow=200, ncol=11, dimnames=list(c(1:200), c(LETTERS[1:11])))
  for(i in 1:200){ #200 patients per dataset
    piece[i, 11] <- rnorm(1) # creates outcome from N dist
    for(j in 1:10){ #creates 10 bernoulli variables
      piece[i, j] <- as.integer(rbernoulli(1, p=0.5))
    }
  }
  #t-test comparing the means of our outcome (piece[,11]) given our ith covariate column
  #finds and stores 10 p-values
  for(i in 1:10){
    k <- t.test(piece[,11]~piece[, i])
    pval[i] <- k$p.value
  }
  #sk initializes our smallest values as the first p-value (to test against)
  sk <- rep(pval[1], 10)
  j=0 #resets a counter j each loop; j goes from 1 to 10 inside for loop
  for(i in 1:10){ #stores smallest value among first j p-values
    j=j+1
    for(k in 1:j){
      if(pval[k]<sk[j]){
        sk[j] <- pval[k] #finds and stores smallest p-value
      }
    }
  }
  for(m in 1:10){ #looks at our smallest p-values and adds to counter qk if p-value<0.05
    if(sk[m]<0.05){
      qk[m] = qk[m]+1
    }
  }
}

qk <- qk/50000 #gives us proportion rather than total count
z<- c(1:10) #creates x-axis labeled 1 through 10

q <- loess(qk~z) #fits a loess curve through data
r <- predict(q)
### -----
### Q1 plot
plot(z,qk, xlab="Number of covariates tested",
      ylab="Total proportion of significant associations")
lines(r, lwd=1, col="blue")

```