Home Work 3

Rick Glabo, slowly getting rid of loops

February 25, 2016

Problem 1

a)

Using block permutations, test to see if the null hypothesis Ho: Mu0 = … = Mu5, that all the means of the groups are the same with the alternative hypothesis being that one or more are different. The blocks are used here instead of the dates because this could eliminate some noise caused from taking the data from different timeframes, or alot of balanced variance within a treatment. If all the blocks have the same mean, then the dates will as well.

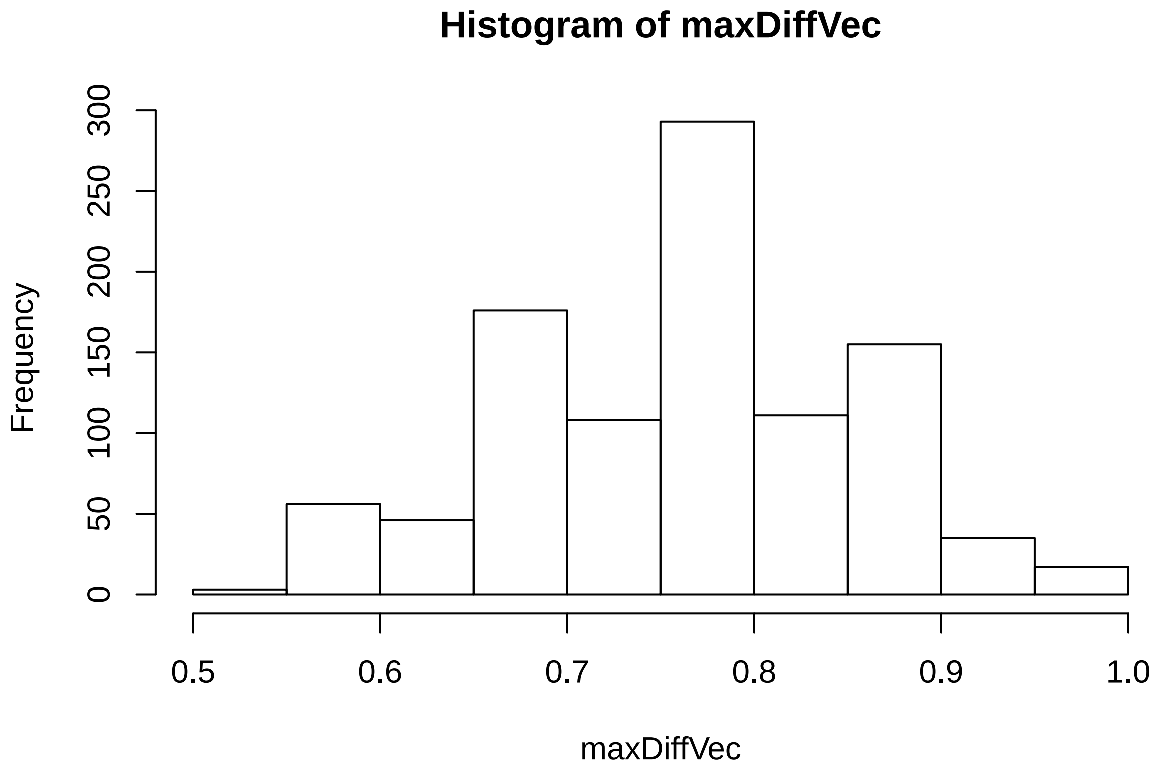
```r
#a)
#randomized block design with permutation
diffs = tapply(dat$kg,dat$date,mean)
tst_stat = abs(max(diffs)-min(diffs))

nsims = 1000
maxDiffVec = rep(NA,nsims)

for (i in 1:nsims){
  datPermute<-dat
  #Shuffle WITHIN blocks
  datPermute$kg = unlist(tapply(datPermute$kg,datPermute$block,
                                function(x){sample(x,length(x),replace=FALSE)}))

  #fit the linear model
  diffs = tapply(datPermute$kg,datPermute$block, mean)

  #Compute the max absolute difference
  maxDiffVec[i] = abs(max(diffs)-min(diffs))
}
hist(maxDiffVec)
abline(v=tst_stat,col="red",lwd=5)
```



```
p_val = sum(maxDiffVec <= tst_stat)/nsims
```

b)

Preform a Friedman test

```
#b)
friedman.test(dat$kg,dat$date,dat$block)
```

```
##
##  Friedman rank sum test
##
## data:  dat$kg, dat$date and dat$block
## Friedman chi-squared = 4.3333, df = 2, p-value = 0.1146
```

```
#not significant, fail to reject
```

This test doesn't have the significance to reject the null hypotheis.

c)

```
#c)
#randomized block design with anova

f = c("Sept 1","Sept 15","Sept30")
k = 3
n = 6

kg = dat$kg

#matching treatment
tm = gl(k, 1, n*k, factor(f))

#blocking factor
blk = gl(n, k, k*n)

#creat anova model using vectors
av = aov(kg ~ tm + blk)
summary(av)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## tm           2 0.2311  0.1156   0.575  0.580
## blk          5 0.6894  0.1379   0.686  0.645
## Residuals   10 2.0089  0.2009
```

```
#since the p-vales are all insignificant we can reject the null hypothesis that the means are the same
```

d)

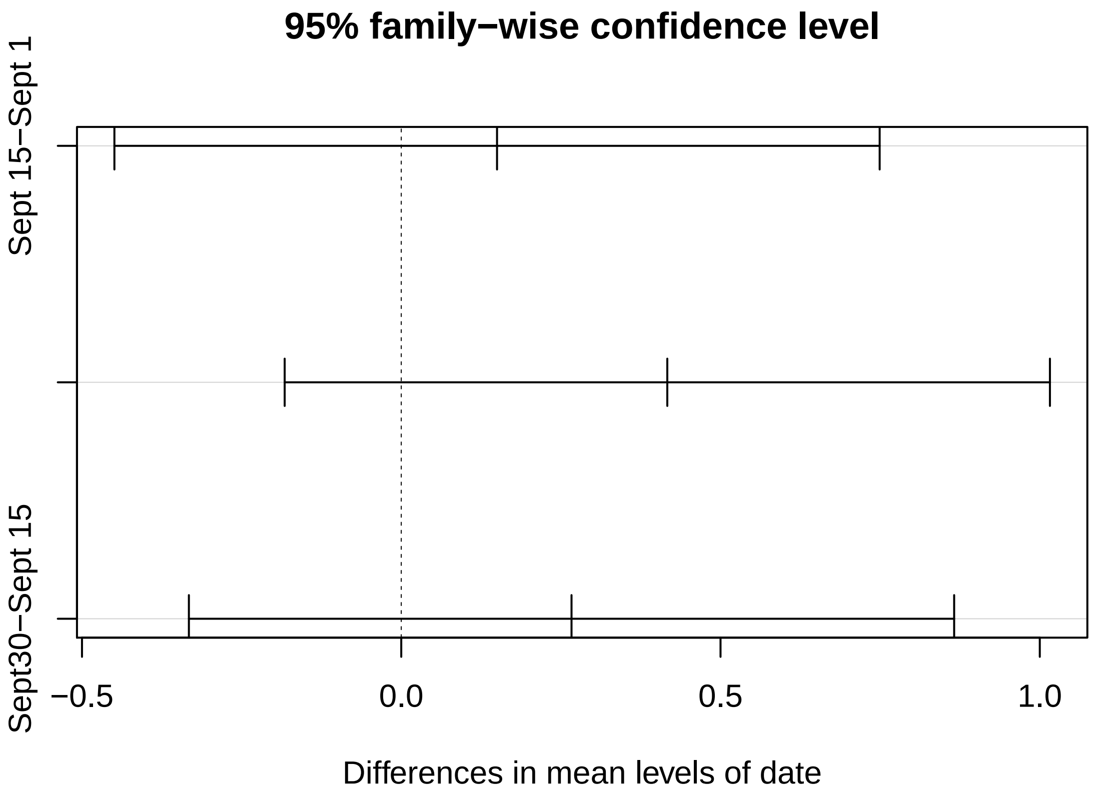
```
#d)
l1 = lm(kg ~ date, data=dat)
anova(l1)
```

```
## Analysis of Variance Table
##
## Response: kg
##           Df  Sum Sq Mean Sq F value Pr(>F)
## date       2 0.53444 0.26722  1.6736 0.2208
## Residuals 15 2.39500 0.15967
```

```
#calculate the means
tapply(dat$kg, dat$date, mean)
```

```
##   Sept 1  Sept 15   Sept30
## 1.916667 2.066667 2.333333
```

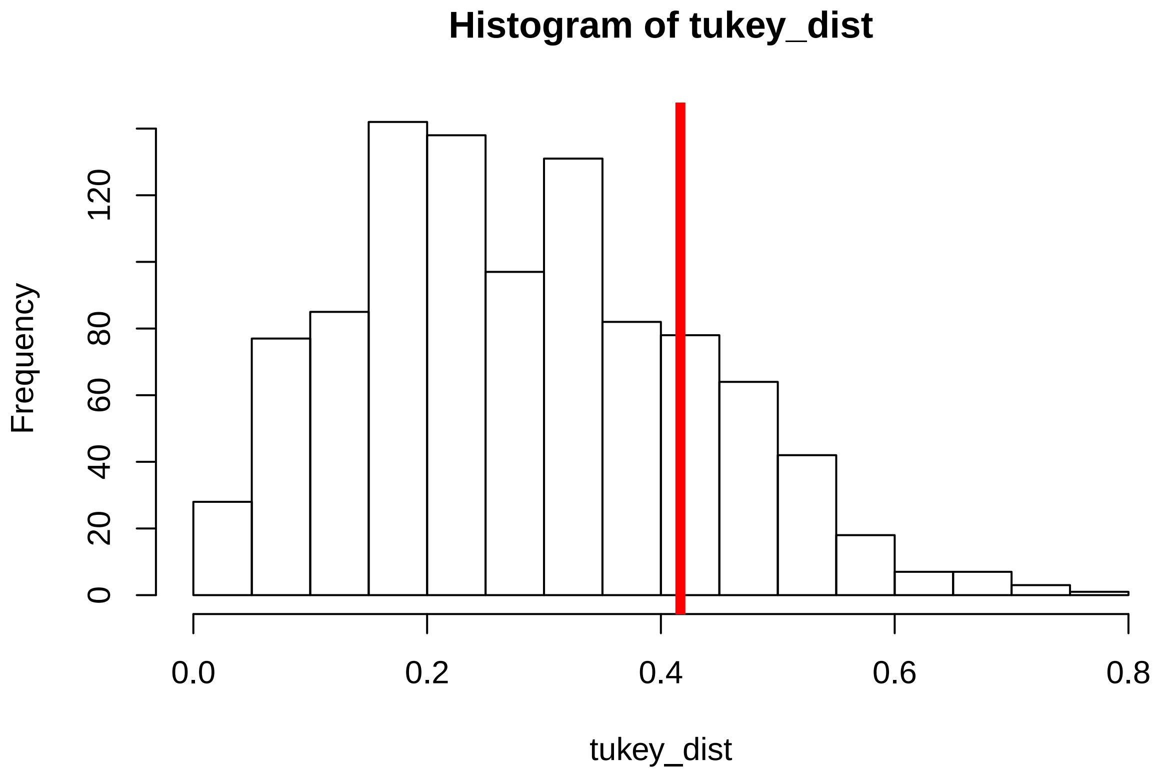
```
#use the aov function as input for tukey
a1 = aov(kg ~ date, data = dat)
#calculate diffs using tukey
posthoc = TukeyHSD(x = a1, which = 'date', conf.level=0.95)
#plot the tukey results
plot(posthoc)
```



```
#get test stat for permutation test
max_diff = max(posthoc$date[1:3])

#tukey permutation test with for loop
n = length(dat$kg)
nsim = 1000
tukey_dist = rep(NA,nsim)
for (i in 1:nsim){
  dat_cp = dat
  perm = sample(dat$kg, n, replace = FALSE)
  dat_cp$kg = perm
  tukey_dist[i] = max(abs(TukeyHSD(aov(kg ~ date, data = dat_cp), which = 'date', conf.level=0.95)$date
}

#plot the distribution of max pairwise mean differences
hist(tukey_dist)
abline(v=max_diff,col="red",lwd=5)
```



```
#calculate the pvalue for tukey
p_val = sum(tukey_dist>=max_diff)/nsim
#can see that this test is not significant



#tukey_test function
tukeyPerm = function(var, group, nsims = 10000){

  #create function for generating test statistics
  max_difference = function(var, group){
    dist_i = max(abs(apply(combn(tapply(var, group, mean),2), 2, diff)))
    return(dist_i)
  }

  #test statistic
  tst_stat = max_difference(var,group)

  #create a function for permuting the mean
  permute_means = function(var,group){
    new = sample(var,length(var),replace = FALSE)
    return(max_difference(new, group))
  }

  #create the distribution
  tukey_distribution = replicate(nsim,permute_means(var,group))


  p_val = sum(tukey_distribution > tst_stat)/nsims

  return(p_val)
}

tukeyPerm(dat$kg,dat$date,1000)
```

```
## [1] 0.197
```

Problem 2

```
#2

ozone = read.csv("C:\\Users\\rgalbo\\Documents\\NonParametric\\HW3\\ozone.csv")
ozone
```

```
##   Subject After.1ppm After.6ppm After1.0ppm
## 1       1       0.08       0.01        0.06
## 2       2       0.21       0.17        0.19
## 3       3       0.50       0.11        0.34
## 4       4       0.14       0.07        0.14
```

Problem 3

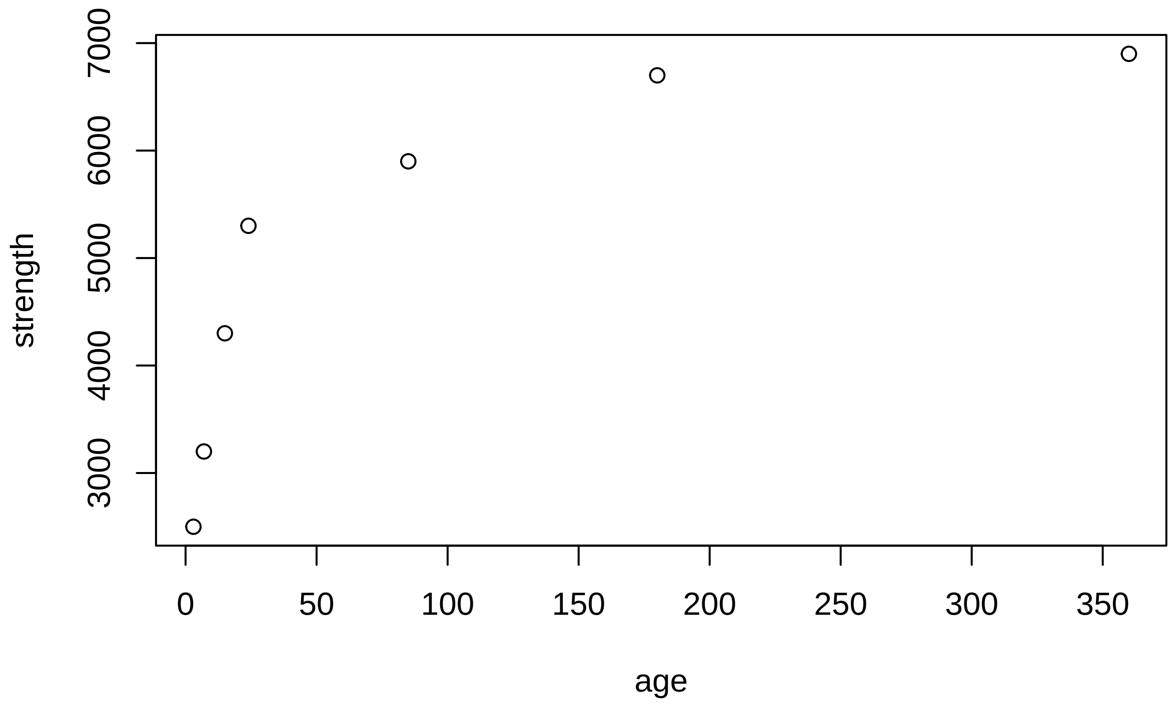
Problem 6

Using the cor function in R

a)

```
#6
#load data for correlation
age = c(3,7,15,24,85,180,360)
strength = c(2500,3200,4300,5300,5900,6700,6900)

#a)
#plot the non-linear data
plot(age,strength)
```



```
#vector of correlation methods
corTypes = c("pearson", "kendall", "spearman")

#vector for storing correlation
corr_coefs = list()

for (method in corTypes){
  corr_coefs[method]=cor(age, strength, method = method)
}
```

b)

```
#conduct all correlation tests
cor.test(age,strength, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  age and strength
## t = 2.8414, df = 5, p-value = 0.03619
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.08030986 0.96677652
## sample estimates:
##       cor
## 0.7858418
```

```
cor.test(age,strength, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  age and strength
## S = 0, p-value = 0.0003968
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
##   1
```

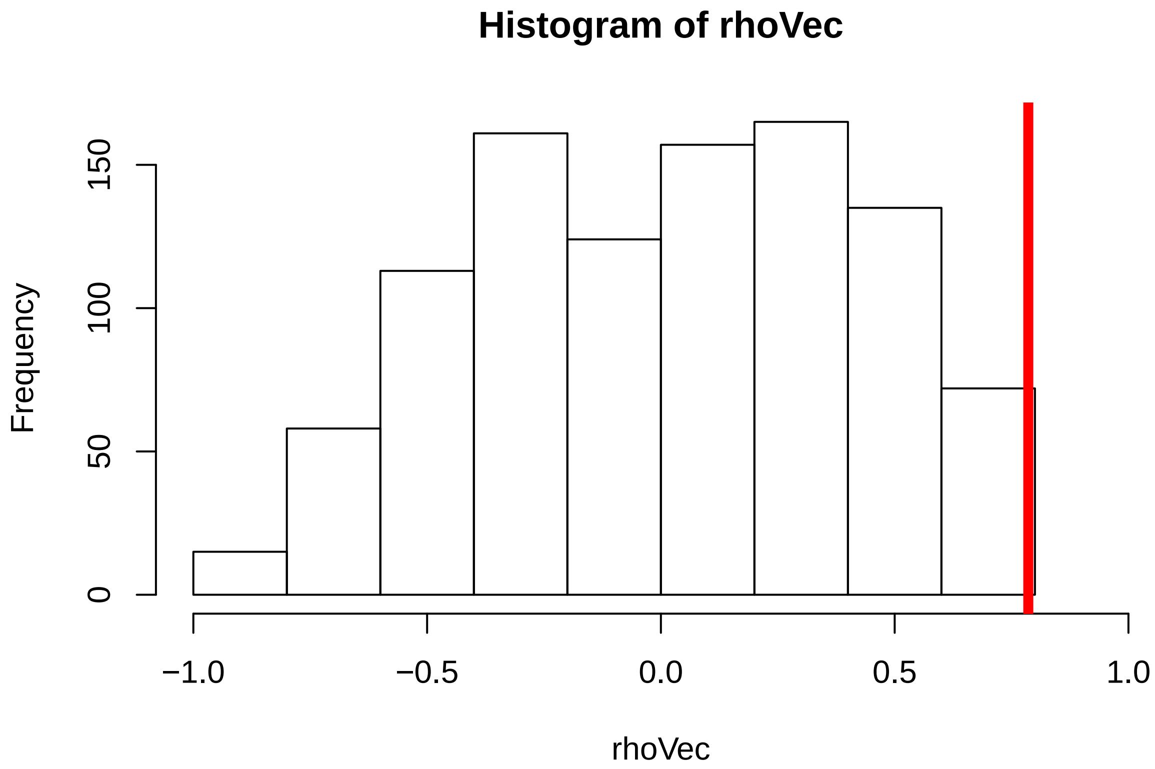
```
cor.test(age,strength, method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  age and strength
## T = 21, p-value = 0.0003968
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
##   1
```

```
#permutation test for pearson
rho = cor(age,strength)

#compute probability using permutation test
nsims = 1000
rhoVec<-rep(NA,nsim)
  for (i in 1:nsim){
    strengthPerm<-sample(strength,length(strength),replace=FALSE)
    rhoVec[i]<-cor(age,strengthPerm)
  }

#plot histogram and test statistic
hist(rhoVec,xlim=c(-1,1))
abline(v=rho,col="red",lwd=5)
```



```
#calc p-val
sum(rhoVec>=abs(rho))/nsim+sum(rhoVec<=-abs(rho))/nsim
```

```
## [1] 0.02
```

```
#significant at alpha =0.05, reject null hypothesis
```