STAT 351 Homework 3

Charles Hwang

2/11/2020

Charles Hwang

Professor Matthews

STAT 351-001

21 February 2020

Problem 1

```r
rm(list=ls())
library(perm)
dat <- data.frame(kg=c(1.5,2.1,1.9,2.8,1.4,1.8,1.8,2.0,2.0,2.7,1.6,2.3,1.9,2.5,2.5,2.6,2.1,2.4),date=fa
dat
```

```
##     kg    date block
## 1  1.5  Sept 1     1
## 2  2.1  Sept 1     2
## 3  1.9  Sept 1     3
## 4  2.8  Sept 1     4
## 5  1.4  Sept 1     5
## 6  1.8  Sept 1     6
## 7  1.8 Sept 15     1
## 8  2.0 Sept 15     2
## 9  2.0 Sept 15     3
## 10 2.7 Sept 15     4
## 11 1.6 Sept 15     5
## 12 2.3 Sept 15     6
## 13 1.9 Sept 30     1
## 14 2.5 Sept 30     2
## 15 2.5 Sept 30     3
## 16 2.6 Sept 30     4
## 17 2.1 Sept 30     5
## 18 2.4 Sept 30     6
```

```r
permKS(dat$kg~dat$date,exact=TRUE,control=permControl(nmc=90000,p.conf.level=.95)) # Problem 1a
```

```
##
##  K-Sample Exact Permutation Test Estimated by Monte Carlo
##
## data:  dat$kg by dat$date
## p-value = 0.2222
##
## p-value estimated from 90000 Monte Carlo replications
## 95 percent confidence interval on p-value:
```

```
##  0.2194763 0.2249193
```

```r
friedman.test(dat$kg,dat$date,dat$block)                                     # Problem 1b
```

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

```r
anova(lm(kg~date,data=dat))                                                  # Problem 1c
```

```
## 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
```

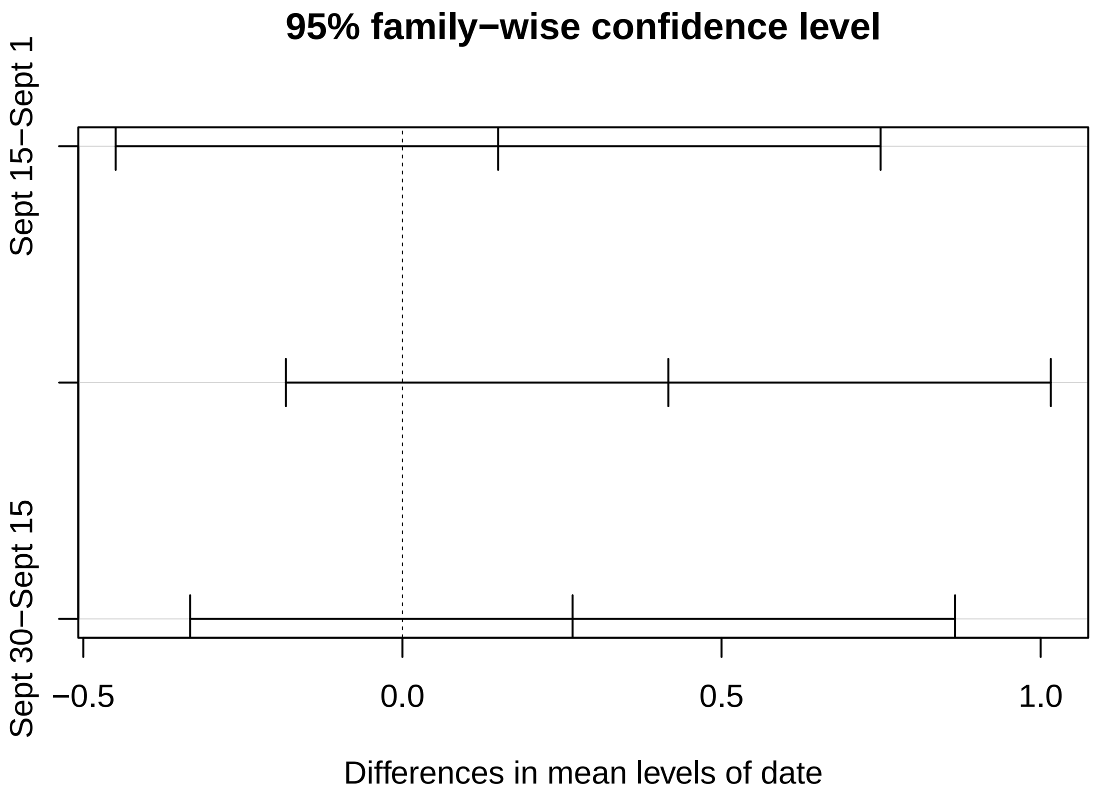
```r
cat("The p-value for ANOVA (p = ",anova(lm(kg~date,data=dat))["date","Pr(>F)"],") is higher than the p-
```

```
## The p-value for ANOVA (p = 0.2207506) is higher than the p-value for the Friedman's test (p = 0.11455
```

```r
# H0: There is no difference between the two means being compared             # Problem 1d
# HA: There is a difference between the two means being compared
a = .05
TukeyHSD(aov(lm(kg~date,data=dat)))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = lm(kg ~ date, data = dat))
##
## $date
##                      diff        lwr       upr     p adj
## Sept 15-Sept 1  0.1500000 -0.4492349 0.7492349 0.7950832
## Sept 30-Sept 1  0.4166667 -0.1825682 1.0159015 0.2012314
## Sept 30-Sept 15 0.2666667 -0.3325682 0.8659015 0.4962114
```

```r
plot(TukeyHSD(aov(lm(kg~date,data=dat))))
```

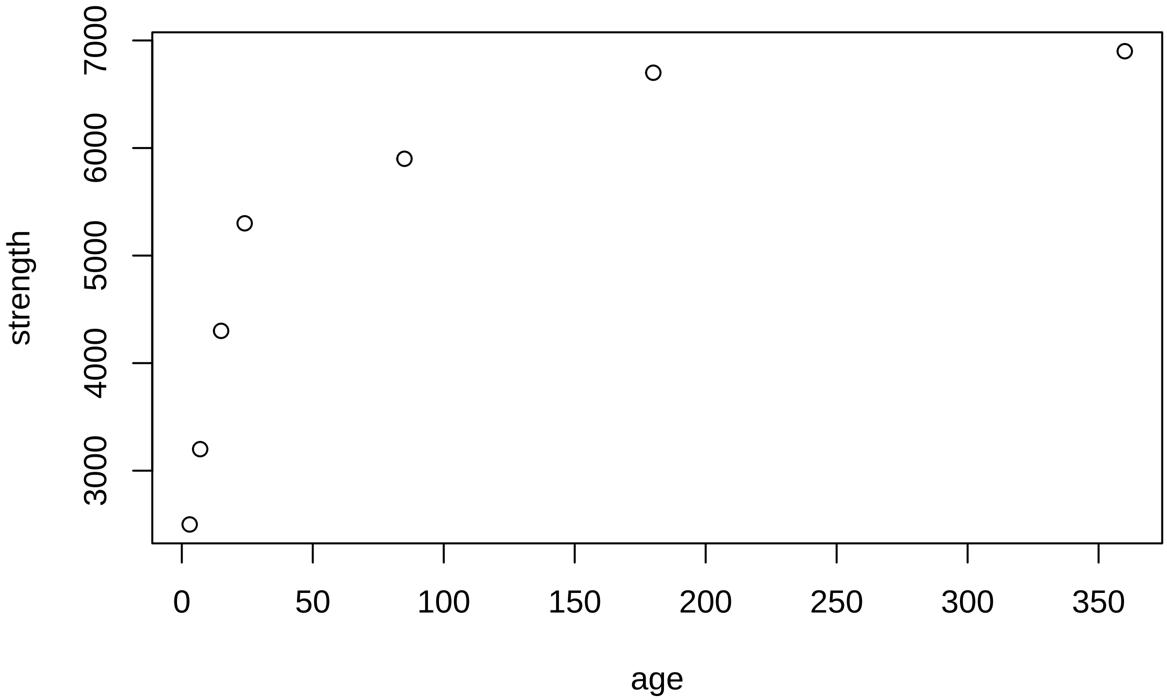



```
cat("The differences between Sept 1 and Sept 15 (p = ",TukeyHSD(aov(lm(kg~date,data=dat)))$date[1,"p ad
```

```
## The differences between Sept 1 and Sept 15 (p = 0.7950832), Sept 1 and Sept 30 (p = 0.2012314), and S
```

Problem 3

```
rm(list=ls())
age <- c(3,7,15,24,85,180,360)
strength <- c(2500,3200,4300,5300,5900,6700,6900)
plot(age,strength)                                                    # Problem 3a
```




```
cor(age,strength,method="pearson")
```

```
## [1] 0.7858418
```

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

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

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

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

```
# H0: There is no or an insignificant association between age and strength # Problem 3b
# HA: There is a significant association between age and strength
a = .05
cor.test(age,strength,method="spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  age and strength
## S = 1.2434e-14, p-value = 0.0003968
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
##   1
```

```
cat("We reject H0 at a = ",a,". There is sufficient evidence (p = ",cor.test(age,strength,method="spearm
```

```
## We reject H0 at a = 0.05. There is sufficient evidence (p = 0.0003968254) that there is a significan
```

Problem 4

```r
rm(list=ls())
set.seed(2102)
eosinophil <- c(55,140,91,122,111,185,203,101,76,145,95,101,196,45,299,226,65,70,196,72,121,171,151,113
n <- 10000 # Problem 4a
BSm <- rep(NA,n)
for (i in 1:n){
  BSm[i] <- mean(eosinophil[sample(1:length(eosinophil),length(eosinophil),replace=TRUE)])
}
mean((BSm-mean(eosinophil))^2) # Bootstrap estimate of MSE
```

```
## [1] 84.88472
```

```r
var(eosinophil)/length(eosinophil) # Compare bootstrap estimate to true value
```

```
## [1] 85.39678
```

```r
sd(BSm) # Bootstrap estimate of standard error
```

```
## [1] 9.213737
```

```r
BSs <- rep(NA,n) # Problem 4b
for (i in 1:n){
  BSs[i] <- sd(eosinophil[sample(1:length(eosinophil),length(eosinophil),replace=TRUE)])
}
mean((BSs-sd(eosinophil))^2) # Bootstrap estimate of MSE
```

```
## [1] 67.34701
```

```r
sd(eosinophil) # Compare bootstrap estimate to true value
```

```
## [1] 58.44545
```

```r
sd(BSs) # Bootstrap estimate of standard error
```

```
## [1] 8.10894
```

```r
BS95 <- rep(NA,n) # Problem 4c
for (i in 1:n){
  BS95[i] <- quantile(eosinophil[sample(1:length(eosinophil),length(eosinophil),replace=TRUE)],.95)
}
mean((BS95-quantile(eosinophil,.95))^2) # Bootstrap estimate of MSE
```

```
## [1] 1363.025
```

```r
quantile(eosinophil,.95) # Compare bootstrap estimate to true value
```

```
##   95%
## 228.5
```

```r
sd(BS95) # Bootstrap estimate of standard error
```

```
## [1] 36.66569
```

Problem 5

```r
rm(list=ls())
set.seed(2102)
n = 15 # Problem 5a
m = 5
```

```
v = 36
sample <- rnorm(n,m,v)
mean(sample) # Problem 5b
```

```
## [1] 11.41861
```

```
v/n # Problem 5c
```

```
## [1] 2.4
```

```
sim <- 10000 # Problem 5d
BSv <- rep(NA,sim)
for (i in 1:sim){
  BSv[i] <- var(sample[sample(1:n,n,replace=TRUE)])/n
}
mean((BSv-var(sample)/n)^2)
```

```
## [1] 112.7424
```