

QSCI 482 Story 6

Simon Hans Edasi

2023-11-16

1a. The resulting data from 30 children is contained in the file “6 ReducingPain.csv”. Convert the data into long format data any way you wish, and upload the resulting long format data file “6 ReducingPainLong.csv” to the assignment hand in page.

```
df <- read.csv('6ReducingPain.csv')
Xlong <- gather(df, key = 'Treatment', value = 'Pain')
write.csv(df, '6ReducingPainLong.csv')
```

1b. Using either the original data file or the long format data file, read the data into R and conduct a one-way ANOVA, the long way around, to determine if there are significant differences among the means of the three groups of children. Note the null and alternate hypotheses. Calculate and report the Groups SS, Total SS, and Error SS, and their corresponding degrees of freedom and mean squares; the test statistic; and the resulting p-value.

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

H_A : At least one mean is different

```
alpha <- 0.05
K <- colnames(df)
k <- length(K)
N <- nrow(Xlong)
X <- mean(Xlong$Pain)

tot_dof <- N - 1
tot_ss <- sum( (Xlong$Pain - X)^2 )

g_dof <- k - 1
g_ss = 0 # group sum of squares
for (i in seq_along(K)){
  gm <- mean(df[,i])
  g_ss <- g_ss + ( length(df[,i]) * (gm - X)^2 )
}

e_dof <- tot_dof - g_dof
e_ss <- tot_ss - g_ss

g_MS <- g_ss / g_dof
e_MS <- e_ss / e_dof

f <- g_MS / e_MS

f_crit <- qf(alpha, g_dof, e_dof, lower.tail = F)
```

```
p <- pf(f, g_dof, e_dof, lower.tail = F)
glue('ANOVA\nGroup SS: {g_ss}\nTotal SS: {tot_ss}\nError SS: {e_ss}\nGroup DF: {g_dof}\nTotal DF: {tot_dof}\nError DF: {e_dof}\nF: {f}\nFcrit: {f_crit}\np-value: {p}')
```

```
## ANOVA
## Group SS: 29.6
## Total SS: 109.2
## Error SS: 79.6
## Group DF: 2
## Total DF: 29
## Error DF: 27
## F: 5.02010050251256
## Fcrit: 3.3541308285292
## p-value: 0.0140062152533203
```

1c. Run the ANOVA using the `aov()` function, and report the groups MS, error MS, and p-value.

```
aov.res <- aov(Pain~Treatment,Xlong)
summary(aov.res)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    2   29.6   14.800    5.02  0.014 *
## Residuals   27   79.6    2.948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1d. What do you conclude from running this test?

At least one treatment stands out as having a different mean level of pain.

1e. Run a Tukey multiple comparison test the long way around to determine which groups are significantly different. Report the difference in means, error MS, error DF, SE, q values, and p values. Use underlining and letters (a,b,c) to clearly outline which groups can be combined or separated.

```
glue('# of combos = {(k*(k-1)) / 2}')
```

```
## # of combos = 3
```

```
alpha <- 0.05
```

```
# 1: calculate means of each group in ascending order
```

```
gm <- data.frame()
```

```
for (i in seq_along(colnames(df))){
```

```
  gm[i,1] <- colnames(df[i])
```

```
  gm[i,2] <- mean(df[,i])
```

```
}
```

```
colnames(gm)[1] <- 'Treatment'
```

```
colnames(gm)[2] <- 'Mean'
```

```
gm <- gm[order(gm$Mean),]
```

```
# 2. Calculate pairwise differences for all combos starting w/ largest diff
```

```
gmdiff <- data.frame()
```

```

combs <- list()
n <- 1
for (i in rev(gm$Mean) ){
  for (j in gm$Mean){

    if (i == j){
      next
    }
    if (any(sapply(combs, function(x) identical(x, c(i, j))))) {
      next
    }
    gmdiff[n,1] <- as.character(gm$Treatment[which(gm$Mean == i)])
    gmdiff[n,2] <- as.character(gm$Treatment[which(gm$Mean == j)])
    gmdiff[n,3] <- i - j

    combs[[length(combs) + 1]] <- c(j,i)
    n <- n + 1
  }
}

colnames(gmdiff)[1] <- 'Treatment A'
colnames(gmdiff)[2] <- 'Treatment B'
colnames(gmdiff)[3] <- 'MeanDiff'

# 3. Calculate error MS -- already have it from ANOVA

# 4. calculate SE
gmdiff$se <- sqrt(e_MS / length(df[,1]))

# 5. calculate q for each pairwise difference of means
gmdiff$q <- gmdiff$MeanDiff / gmdiff$se

# 6. Find critical value
gmdiff$qcrit <- qtkey(1-alpha,k,e_dof,lower.tail = F)

# 7. find p value
for (i in 1:3){
  gmdiff[i,'p'] <- ptukey(gmdiff[i,'q'],k,e_dof,lower.tail = F)
}
gmdiff

##   Treatment A Treatment B MeanDiff      se      q      qcrit      p
## 1   Earphones      Music      2.2 0.5429685 4.0518003 0.4318156 0.02107552
## 2   Earphones  Audiobook      2.0 0.5429685 3.6834548 0.4318156 0.03805353
## 3   Audiobook      Music      0.2 0.5429685 0.3683455 0.4318156 0.96334443

```

1f. Run a Tukey multiple comparison test the short way around.

```

tukey.res <- TukeyHSD(aov.res,ordered = T,conf.level = 0.95)
print(tukey.res)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered

```

```
##
## Fit: aov(formula = Pain ~ Treatment, data = Xlong)
##
## $Treatment
##              diff          lwr          upr          p adj
## Audiobook-Music    0.2 -1.70387901  2.103879  0.9633444
## Earphones-Music    2.2  0.29612099  4.103879  0.0210755
## Earphones-Audiobook 2.0  0.09612099  3.903879  0.0380535
```

1g. What would you advise for hospitals taking care of children after surgery?

Don't give them earphones as the mean pain level is statistically significantly higher than audiobooks or music.

Question 2

2a. What is the total number of salmon that you need to tag with acoustic tags? Explain your workings, and do this the long way around.

```
alpha <- 0.05
k <- 4
d <- 0.5
e_MS <- 6.5
dp <- 0.9
stop_difference <- 0.001

n <- 1000
for (i in 1:100){
  N <- n * k
  phi <- sqrt( (n * d^2) / (2 * k * e_MS) )
  ncp <- k * phi^2

  f.crit <- qf(alpha, k-1, N-k, lower.tail = F)
  power <- pf(f.crit, k-1, N-k, ncp = ncp, lower.tail = F)

  if (abs(power - dp) <= stop_difference){
    break
  }
  else if (power < dp){
    n <- ceiling(n + (n/2))
  }
  else{
    n <- ceiling(n - (n/2))
  }
}
glue('ANOVA power: {power}\nRequired N samples: {N}')
```

```
## ANOVA power: 0.900044961063545
```

```
## Required N samples: 2952
```

2b. Your total budget is \$100,000 for acoustic tags, and each tag costs \$25. Will you be able to complete the experiment with the desired power?

```
b <- 100000
c <- 25
```

```
t_n <- b / c
glue('Yes, we can sample {t_n} fish with a budget of ${b} at an individual cost ${c},
     more than enough with our required {N} samples')
```

```
## Yes, we can sample 4000 fish with a budget of $1e+05 at an individual cost $25,
## more than enough with our required 2952 samples
```