

# QSCI 482 Story 6

Simon Hans Edasi

2023-11-20

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\nTotal SS: {signif(tot_ss,2)}\nGroup SS: {signif(g_ss,2)}\nError SS: {signif(e_ss,2)}\nTotal DF: {signif(tot_dof,2)}\nGroup DF: {signif(g_dof,2)}\nError DF: {signif(e_dof,2)}\nGroup MS: {signif(g_MS,2)}\nError MS: {signif(e_MS,2)}\nF: {signif(f,2)}\nFcrit: {signif(f_crit,2)}\np-value: {signif(p,2)}')
```

```
## ANOVA
## Total SS: 110
## Group SS: 30
## Error SS: 80
## Total DF: 29
## Group DF: 2
## Error DF: 27
## Group MS: 15
## Error MS: 2.9
## F: 5
## Fcrit: 3.4
## p-value: 0.014
```

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

glue('Groups MS: 14.800\nError MS: 2.948\np-value: 0.014')

## Groups MS: 14.800
## Error MS: 2.948
## p-value: 0.014
```

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
```

```
# 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
se <- sqrt(e_MS / length(df[,1]))
gmdiff$se <- signif(se,2)

# 5. calculate q for each pairwise difference of means
q <- gmdiff$MeanDiff / gmdiff$se
gmdiff$q <- signif(q,2)

# 6. Find critical value
qcrit <- qtkey(1-alpha,k,e_dof,lower.tail = F)
gmdiff$qcrit <- signif(qcrit,2)

# 7. find p value
for (i in 1:3) {
  gmdiff[i,'p'] <- signif(ptukey(q[i],k,e_dof,lower.tail = F),2)
}
rownames(gmdiff) <- letters[1:3]
gmdiff

```

##	Treatment A	Treatment B	MeanDiff	se	q	qcrit	p
## a	Earphones	Music	2.2	0.54	4.10	0.43	0.020
## b	Earphones	Audiobook	2.0	0.54	3.70	0.43	0.037
## c	Audiobook	Music	0.2	0.54	0.37	0.43	0.960

It appears groups a (Earphones & Music) and b (Earphones & Audiobook) both stand out with having a significant difference in mean level of pain. We can group them separate from group c (Audiobook & Music).

a      b      c (1)

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                # Specified significance
k <- 4                       # Num groups
d <- 0.5                     # Detectable difference
e_MS <- 6.5                  # A priori Error MS
dp <- 0.9                    # Desired power
stop_difference <- 0.001     # Break loop if calc power is close to dp

n <- 1000                    # initial guess for each group n
for (i in 1:100){
  N <- n * k                  # find total N for Degrees of Freedom
                              # calculate phi for NCP
  phi <- sqrt( (n * d^2) / (2 * k * e_MS) )

  ncp <- k * phi^2            # calculate NCP for alt F dist

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

                              # check if we're close to desired power
  if (abs(power - dp) <= stop_difference){
```

```

    break
  }
  # If we're not close, adjust n for next guess
  else if (power < dp){
    n <- ceiling(n + (n/5))
  }
  else{
    n <- ceiling(n - (n/5))
  }
}
glue('ANOVA power: {signif(power,2)}\nRequired total N samples: {N}')

```

```

## ANOVA power: 0.9
## Required total 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
n_samp <- b / c

glue('Yes, we can sample {n_samp} fish for {b} and come in ${b-(N*c)} under budget')

```

```

## Yes, we can sample 4000 fish for 1e+05 and come in $26200 under budget

```