R Code: Multiple Comparisons (Sleuth3 Sections 6.3 and 6.4)

2019-09-25

R code, Handicaps Study

```
handicaps <- read_csv("http://www.evanlray.com/data/sleuth3/ex0601_handicaps.csv") %>%
  mutate(
    Handicap = factor(Handicap, levels = c("None", "Amputee", "Crutches", "Hearing", "Wheelchair"))
  )

nrow(handicaps)
```

```
## [1] 70
```

```
head(handicaps)
```

```
## # A tibble: 6 x 2
##   Score Handicap
##   <dbl> <fct>
## 1   1.9 None
## 2   2.5 None
## 3   3   None
## 4   3.6 None
## 5   4.1 None
## 6   4.2 None
```

```
handicaps %>%
  count(Handicap)
```

```
## # A tibble: 5 x 2
##   Handicap       n
##   <fct>      <int>
## 1 None          14
## 2 Amputee       14
## 3 Crutches      14
## 4 Hearing       14
## 5 Wheelchair    14
```

To keep things shorter, suppose we just want to find 4 confidence intervals:

- $\mu_1 - \mu_2$
- $\mu_1 - \mu_3$
- $\mu_1 - \mu_4$
- $\mu_1 - \mu_5$

Bonferroni adjustment

Step 1: Figure out the individual confidence levels that get you the desired familywise confidence level

- Can always work through the table
- General formula is $1 - \alpha/k$ where:
  - k is number of intervals we're making (4 in our example)
  - If we want a 95% familywise confidence level, $\alpha = 0.05$

```
1 - 0.05/4
```

```
## [1] 0.9875
```

Step 2: Make the confidence intervals

```
anova_fit <- lm(Score ~ Handicap, data = handicaps)
fit.contrast(anova_fit, "Handicap", c(1, -1, 0, 0, 0), conf.int = 0.9875)
```

```
##                             Estimate Std. Error   t value  Pr(>|t|)
## Handicap c=( 1 -1 0 0 0 ) 0.4714286  0.6171922 0.7638278 0.4477337
##                             lower CI upper CI
## Handicap c=( 1 -1 0 0 0 ) -1.114207 2.057064
## attr(,"class")
## [1] "fit_contrast"
```

```
fit.contrast(anova_fit, "Handicap", c(1, 0, -1, 0, 0), conf.int = 0.9875)
```

```
##                             Estimate Std. Error  t value  Pr(>|t|)
## Handicap c=( 1 0 -1 0 0 ) -1.021429  0.6171922 -1.65496 0.1027537
##                             lower CI  upper CI
## Handicap c=( 1 0 -1 0 0 ) -2.607064 0.5642072
## attr(,"class")
## [1] "fit_contrast"
```

```
fit.contrast(anova_fit, "Handicap", c(1, 0, 0, -1, 0), conf.int = 0.9875)
```

```
##                            Estimate Std. Error  t value  Pr(>|t|)
## Handicap c=( 1 0 0 -1 0 )     0.85  0.6171922 1.377205 0.1731733
##                              lower CI upper CI
## Handicap c=( 1 0 0 -1 0 ) -0.7356358 2.435636
## attr(,"class")
## [1] "fit_contrast"
```

```
fit.contrast(anova_fit, "Handicap", c(1, 0, 0, 0, -1), conf.int = 0.9875)
```

```
##                              Estimate Std. Error    t value  Pr(>|t|)
## Handicap c=( 1 0 0 0 -1 ) -0.4428571  0.6171922 -0.7175352 0.4756148
##                             lower CI upper CI
## Handicap c=( 1 0 0 0 -1 ) -2.028493 1.142779
## attr(,"class")
## [1] "fit_contrast"
```

We could also calculate by hand based on the estimates and standard errors from R:

- The multiplier is the quantile of the t distribution at the point $1 - \alpha/2k$

```
bonferroni_multiplier <- qt(1 - 0.05/(2*4), df = 70 - 5)
bonferroni_multiplier
```

```
## [1] 2.569112
```

```
c(-0.471 - bonferroni_multiplier * 0.617, -0.471 + bonferroni_multiplier * 0.617)
```

```
## [1] -2.056142  1.114142
```

```
c(1.021 - bonferroni_multiplier * 0.617, 1.021 + bonferroni_multiplier * 0.617)
```

```
## [1] -0.564142  2.606142
```

```
c(0.85 - bonferroni_multiplier * 0.617, 0.85 + bonferroni_multiplier * 0.617)
```

```
## [1] -0.735142  2.435142
```

```
c(-0.443 - bonferroni_multiplier * 0.617, -0.443 + bonferroni_multiplier * 0.617)
```

```
## [1] -2.028142  1.142142
```

Scheffe 95% familywise CIs

- Use Multiplier = $\sqrt{(I-1)F_{(I-1),(n-1)}(1-\alpha)}$

First, find the multiplier

```
scheffe_multiplier <- sqrt((5 - 1) * qf(0.95, df1 = 5 - 1, df2 = 70 - 5))
scheffe_multiplier
```

```
## [1] 3.170514
```

```
c(-0.471 - scheffe_multiplier * 0.617, -0.471 + scheffe_multiplier * 0.617)
```

```
## [1] -2.427207  1.485207
```

```
c(1.021 - scheffe_multiplier * 0.617, 1.021 + scheffe_multiplier * 0.617)
```

```
## [1] -0.9352073  2.9772073
```

```
c(0.85 - scheffe_multiplier * 0.617, 0.85 + scheffe_multiplier * 0.617)
```

```
## [1] -1.106207  2.806207
```

```
c(-0.443 - scheffe_multiplier * 0.617, -0.443 + scheffe_multiplier * 0.617)
```

```
## [1] -2.399207  1.513207
```

What is the interpretation of the Scheffe intervals above in context? As part of your answer, explain what it means that they have a 95% familywise confidence level.