Lab04 - multiple comparisons

Goals

The goal in this lab is to practice finding familywise conidence intervals for multiple comparisons.

Loading packages

Here are some packages with functionality you may need for this lab. Run this code chunk now.

```r
library(readr)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```r
library(gridExtra)
library(mosaic)
```

```
## Warning: package 'mosaic' was built under R version 3.5.2

## Loading required package: dplyr

## Warning: package 'dplyr' was built under R version 3.5.2

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##     combine

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggformula

## Warning: package 'ggformula' was built under R version 3.5.2

## Loading required package: ggstance

##
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh

##
## New to ggformula?  Try the tutorials:
##  learnr::run_tutorial("introduction", package = "ggformula")
##  learnr::run_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData
```

```
## Loading required package: Matrix

##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
##
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.

##
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':
##
##     mean

## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following object is masked from 'package:ggplot2':
##
##     stat

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median,
##     prop.test, quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```
library(dplyr)
library(gmodels)

options("pillar.sigfig" = 10) # print 10 significant digits in summarize output

longevity <- read_csv("http://www.evanlray.com/data/sleuth3/ex0501_longevity.csv") %>%
  mutate(
    Diet = factor(Diet, levels = c("NP", "N/N85", "N/R50", "R/R50", "lopro", "N/R40"))
  )
```

```
## Parsed with column specification:
## cols(
##   Lifetime = col_double(),
##   Diet = col_character()
## )
```

```
nrow(longevity)
```

```
## [1] 349
```

Find the multiplier that would be used for Bonferroni 95% familywise intervals for 5 comparisons based on this data set.

```
1 - 0.05/(2 * 5)
```

```
## [1] 0.995
```

```r
qt(0.995, df = 349 - 6)
```

```
## [1] 2.590239
```

Find the multiplier that would be used for Scheffe 95% familywise intervals for 5 comparisons based on this data set.

```r
sqrt((6-1) * qf(0.95, df1 = 6-1, df2 = 349 - 6))
```

```
## [1] 3.346868
```

Find familywise 95% Confidence Intervals

The code below finds individual 95% confidence intervals for the 5 differences in group means the researchers planned for this study. For each, add calculations of appropriately adjusted Bonferroni and Scheffe familywise intervals.

```r
anova_fit <- lm(Lifetime ~ Diet, data = longevity)
```

(a) $H_0 : \mu_2 = \mu_3$ **vs** $H_A : \mu_2 \neq \mu_3$**. Are the population mean lifetimes the same for the N/N85 and N/R50 groups?**

```r
fit.contrast(anova_fit, "Diet", c(0, 1, -1, 0, 0, 0), conf.int = 0.95)
```

```
##                            Estimate Std. Error   t value     Pr(>|t|)
## Diet c=( 0 1 -1 0 0 0 ) -9.605955   1.187682 -8.087982 1.057397e-14
##                            lower CI   upper CI
## Diet c=( 0 1 -1 0 0 0 ) -11.94201 -7.269897
## attr(,"class")
## [1] "fit_contrast"
```

```r
# Bonferroni intervals
-9.605955 - 2.59 * 1.187682
```

```
## [1] -12.68205
```

```r
-9.605955 + 2.59 * 1.187682
```

```
## [1] -6.529859
```

```r
# Scheffe intervals
-9.605955 - 3.35 * 1.187682
```

```
## [1] -13.58469
```

```r
-9.605955 + 3.35 * 1.187682
```

```
## [1] -5.62722
```

Here's an easier way to find the Bonferroni interval. To get a familywise 95% CI for five intervals, each confidence interval will have an individual confidence level of $(1 - 0.05/5) \times 100\%$, or 99%. Therefore, the following code to get an individual 99% CI will give us the first of 5 intervals with a familywise confidence level of 95%:

```r
fit.contrast(anova_fit, "Diet", c(0, 1, -1, 0, 0, 0), conf.int = 0.99)
```

```
##                            Estimate Std. Error   t value     Pr(>|t|)
## Diet c=( 0 1 -1 0 0 0 ) -9.605955   1.187682 -8.087982 1.057397e-14
##                            lower CI   upper CI
```

```
## Diet c=( 0 1 -1 0 0 0 ) -12.68234 -6.529574
## attr(,"class")
## [1] "fit_contrast"
```

Note that this result agrees with our manual calculation above, up to rounding errors.

(b) $H_0 : \mu_3 = \mu_4$ vs $H_A : \mu_3 \neq \mu_4$. Are the population mean lifetimes the same for the N/R50 and R/R50 groups?

```
fit.contrast(anova_fit, "Diet", c(0, 0, 1, -1, 0, 0), conf.int = 0.95)
```

```
##                             Estimate Std. Error   t value  Pr(>|t|) lower CI
## Diet c=( 0 0 1 -1 0 0 ) -0.5885312    1.19355 -0.493093 0.6222624 -2.93613
##                         upper CI
## Diet c=( 0 0 1 -1 0 0 ) 1.759068
## attr(,"class")
## [1] "fit_contrast"
```

```
# Bonferroni intervals
-0.5885312 - 2.59 * 1.19355
```

```
## [1] -3.679826
```

```
-0.5885312 + 2.59 * 1.19355
```

```
## [1] 2.502763
```

```
# Scheffe intervals
-0.5885312 - 3.35 * 1.19355
```

```
## [1] -4.586924
```

```
-0.5885312 + 3.35 * 1.19355
```

```
## [1] 3.409861
```

Or, the easier way for the Bonferroni interval:

```
fit.contrast(anova_fit, "Diet", c(0, 0, 1, -1, 0, 0), conf.int = 0.99)
```

```
##                             Estimate Std. Error   t value  Pr(>|t|)
## Diet c=( 0 0 1 -1 0 0 ) -0.5885312    1.19355 -0.493093 0.6222624
##                          lower CI upper CI
## Diet c=( 0 0 1 -1 0 0 ) -3.680111 2.503048
## attr(,"class")
## [1] "fit_contrast"
```

(c) $H_0 : \mu_3 = \mu_6$ vs $H_A : \mu_3 \neq \mu_6$. Are the population mean lifetimes the same for the N/R50 and N/R40 groups?

```
fit.contrast(anova_fit, "Diet", c(0, 0, 1, 0, 0, -1), conf.int = 0.95)
```

```
##                            Estimate Std. Error   t value   Pr(>|t|)
## Diet c=( 0 0 1 0 0 -1 ) -2.819484   1.171097 -2.407558 0.01658711
##                          lower CI   upper CI
## Diet c=( 0 0 1 0 0 -1 ) -5.122919 -0.5160481
## attr(,"class")
## [1] "fit_contrast"
```

```r
# Bonferroni intervals
-2.819484 - 2.59 * 1.171097
```

```
## [1] -5.852625
```

```r
-2.819484 + 2.59 * 1.171097
```

```
## [1] 0.2136572
```

```r
# Scheffe intervals
-2.819484 - 3.35 * 1.171097
```

```
## [1] -6.742659
```

```r
-2.819484 + 3.35 * 1.171097
```

```
## [1] 1.103691
```

Or, the easier way for the Bonferroni interval:

```r
fit.contrast(anova_fit, "Diet", c(0, 0, 1, 0, 0, -1), conf.int = 0.99)
```

```
##                            Estimate Std. Error   t value   Pr(>|t|)
## Diet c=( 0 0 1 0 0 -1 ) -2.819484   1.171097 -2.407558 0.01658711
##                            lower CI   upper CI
## Diet c=( 0 0 1 0 0 -1 ) -5.852904 0.2139368
## attr(,"class")
## [1] "fit_contrast"
```

(d) $H_0 : \mu_3 = \mu_5$ vs $H_A : \mu_3 \neq \mu_5$. Are the population mean lifetimes the same for the N/R50 and N/R50 lopro groups?

```r
fit.contrast(anova_fit, "Diet", c(0, 0, 1, 0, -1, 0), conf.int = 0.95)
```

```
##                           Estimate Std. Error  t value   Pr(>|t|)  lower CI
## Diet c=( 0 0 1 0 -1 0 ) 2.611469    1.19355 2.187984 0.02934503 0.2638701
##                           upper CI
## Diet c=( 0 0 1 0 -1 0 ) 4.959068
## attr(,"class")
## [1] "fit_contrast"
```

```r
# Bonferroni intervals
2.611469 - 2.59 * 1.19355
```

```
## [1] -0.4798255
```

```r
2.611469 + 2.59 * 1.19355
```

```
## [1] 5.702763
```

```r
# Scheffe intervals
2.611469 - 3.35 * 1.19355
```

```
## [1] -1.386924
```

```r
2.611469 + 3.35 * 1.19355
```

```
## [1] 6.609862
```

Or, the easier way for the Bonferroni interval:

```r
fit.contrast(anova_fit, "Diet", c(0, 0, 1, 0, -1, 0), conf.int = 0.99)
```

```
##                           Estimate Std. Error  t value   Pr(>|t|)   lower CI
## Diet c=( 0 0 1 0 -1 0 ) 2.611469    1.19355 2.187984 0.02934503 -0.4801107
##                           upper CI
## Diet c=( 0 0 1 0 -1 0 ) 5.703048
## attr(,"class")
## [1] "fit_contrast"
```

(e) $H_0 : \mu_2 = \mu_1$ **vs** $H_A : \mu_2 \neq \mu_1$. **Are the population mean lifetimes the same for the N/N85 and NP groups?**

```
fit.contrast(anova_fit, "Diet", c(-1, 1, 0, 0, 0, 0), conf.int = 0.95)
```

```
##                           Estimate Std. Error  t value     Pr(>|t|) lower CI
## Diet c=( -1 1 0 0 0 0 ) 5.289187   1.301006 4.065458 5.949477e-05 2.730232
##                           upper CI
## Diet c=( -1 1 0 0 0 0 ) 7.848142
## attr(,"class")
## [1] "fit_contrast"
```

```
# Bonferroni intervals
5.289187 - 2.59 * 1.301006
```

```
## [1] 1.919581
```

```
5.289187 + 2.59 * 1.301006
```

```
## [1] 8.658793
```

```
# Scheffe intervals
5.289187 - 3.35 * 1.301006
```

```
## [1] 0.9308169
```

```
5.289187 + 3.35 * 1.301006
```

```
## [1] 9.647557
```

Or, the easier way for the Bonferroni interval:

```
fit.contrast(anova_fit, "Diet", c(-1, 1, 0, 0, 0, 0), conf.int = 0.99)
```

```
##                           Estimate Std. Error  t value     Pr(>|t|) lower CI
## Diet c=( -1 1 0 0 0 0 ) 5.289187   1.301006 4.065458 5.949477e-05  1.91927
##                           upper CI
## Diet c=( -1 1 0 0 0 0 ) 8.659104
## attr(,"class")
## [1] "fit_contrast"
```

What is the interpretation of the individual 95% confidence interval obtained in part (a) from the fit.contrast function?

We are 95% confident that the difference in population mean lifetimes for the **N/N85** and **N/R50** groups is between -11.942 and -7.270. For 95% of samples, an interval calculated in this way will contain the difference in population mean lifetimes for the **N/N85** and **N/R50** groups.

What is the interpretation of the Bonferroni intervals?

We are 95% confident that the difference in population mean lifetimes for the **N/N85** and **N/R50** groups is between -12.68 and -6.53, for the **N/R50** and **R/R50** groups is between -3.68 and 2.50, for the **N/R50** and

N/R40 groups is between -5.85 and 0.21, for the **N/R50** and **N/R50** lopro groups is between -0.48 and 5.70, and for the **N/N85** and **NP** groups is between 1.92 and 8.66. For 95% of samples, all 5 of the intervals constructed in this way will simultaneously contain the difference in means they are estimating.