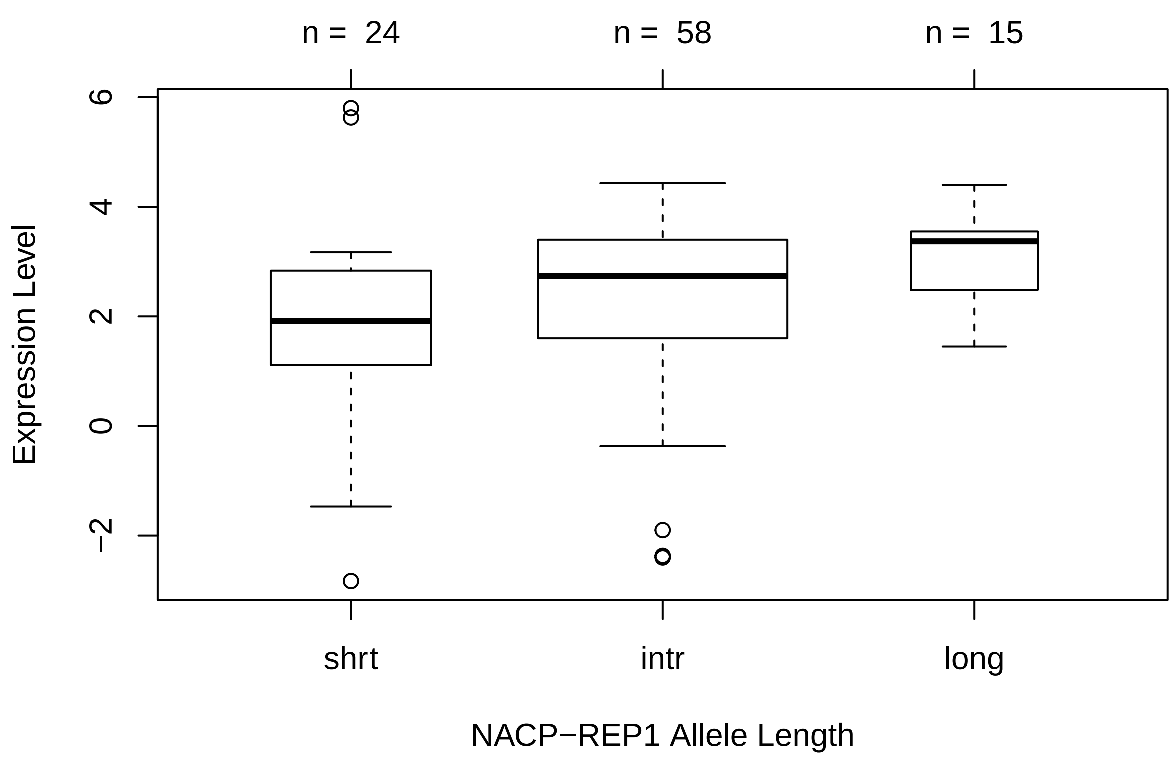
Rainfall with/without Cloud Seeding

Thomas Pattara

6th November 2018

```
library(HSAUR3)
library(dplyr)
library(tidyr)
library(ggplot2)
library(gee)
library(Matrix)
library(lme4)
library(multcomp)
library(Hmisc)
library(vcd)
library(coin)
```

1. Consider alpha dataset from the coin package. Compare the results when using glht and TukeyHSD (Refer to Chap 5 review for TukeyHSD).




```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = elevel ~ alength, data = alpha)
```

```
##
## Linear Hypotheses:
##                   Estimate Std. Error t value Pr(>|t|)
## intr - shrt == 0   0.4342     0.3836   1.132   0.4924
## long - shrt == 0   1.1888     0.5203   2.285   0.0614 .
## long - intr == 0   0.7546     0.4579   1.648   0.2271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = elevel ~ alength, data = alpha)
##
## $alength
##                diff         lwr      upr     p adj
## intr-shrt 0.4341523 -0.47943766 1.347742 0.4970962
## long-shrt 1.1887500 -0.05017513 2.427675 0.0628589
## long-intr 0.7545977 -0.33575201 1.844947 0.2307995
```

Answer:

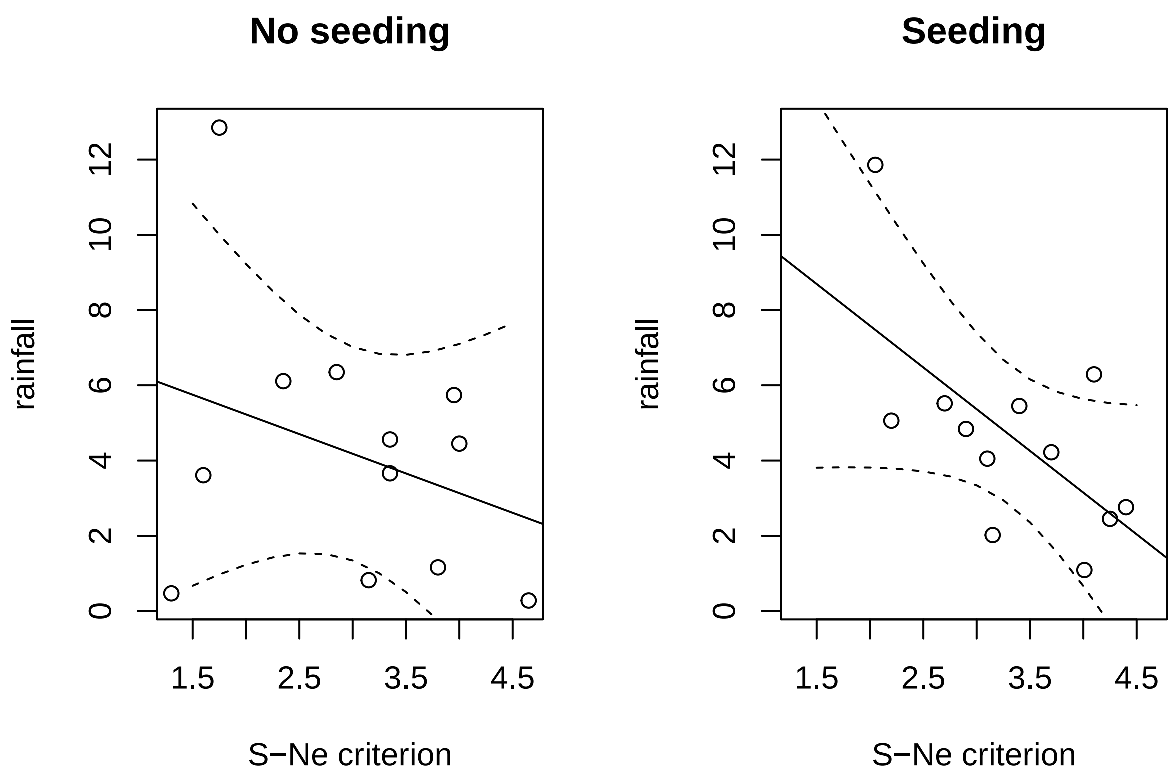
We can see the results of the alpha data above with the General Linear Hypotheses function and the TukeyHSD function. We get the adjusted p values, but none of the elements are significant at the 0.05 significance level.

2. Consider clouds data from HSAUR3 package

a. Read and write a report (no longer than one page) on the clouds data given in Chapter 15 section 15.3.3 from Handbook Ed 3.

We studied the dependency of rainfall on S-Ne values by means of linear models in Chapter 6. Since the number of observations is small, an additional assessment of the variability of the fitted regression lines is necessary. This can be achieved with the help of confidence band around some estimated regression line.

We will begin with a model for simple linear regression as : rainfalli = β 0 + β 1 snei + ϵ i. Where we will calculate the confidence band for the predicted rainfall. We can formulate the problem as a linear combination of the regression coefficients by multiplying a matrix K to a grid of S-Ne values (ranging from 1.5 to 4.5).



Confidence bands for the estimated effects help to identify days where the uncertainty about rainfall is largest.The function confband basically fits a linear model using lm to a subset of the data, sets up the matrix K and nicely plots both the regression line and the confidence band. For the days without seeding, there is more uncertainty about the true regression line compared to the days with cloud seeding. Clearly, this is caused by the larger variability of the observations in the left part of the figure.

b. Consider the linear model fitted to the clouds data as summarised in Chapter 6, Figure 6.5. Set up a matrix K corresponding to the global null hypothesis that all interaction terms present in the model are zero. Test both the global hypothesis and all hypotheses corresponding to each of the interaction terms.

```
##
## Call:
## lm(formula = clouds_formula, data = clouds)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5259 -1.1486 -0.2704  1.0401  4.3913
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -0.34624    2.78773  -0.124  0.90306
## seedingyes                      15.68293    4.44627   3.527  0.00372 **
## time                            -0.04497    0.02505  -1.795  0.09590 .
## seedingno:sne                    0.41981    0.84453   0.497  0.62742
## seedingyes:sne                  -2.77738    0.92837  -2.992  0.01040 *
## seedingno:cloudcover             0.38786    0.21786   1.780  0.09839 .
## seedingyes:cloudcover           -0.09839    0.11029  -0.892  0.38854
## seedingno:prewetness             4.10834    3.60101   1.141  0.27450
```

```
## seedingyes:prewetness           1.55127    2.69287   0.576  0.57441
## seedingno:echomotionstationary   3.15281    1.93253   1.631  0.12677
## seedingyes:echomotionstationary  2.59060    1.81726   1.426  0.17757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.205 on 13 degrees of freedom
## Multiple R-squared:  0.7158, Adjusted R-squared:  0.4972
## F-statistic: 3.274 on 10 and 13 DF,  p-value: 0.02431
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = rainfall ~ seeding + seeding:(sne + cloudcover +
##     prewetness + echomotion) + time, data = clouds)
##
## Linear Hypotheses:
##               Estimate Std. Error t value Pr(>|t|)
## yes - no == 0   15.683      4.446   3.527  0.00372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: aov(formula = rainfall ~ seeding + seeding:(sne + cloudcover +
##     prewetness + echomotion) + time, data = clouds)
##
## Linear Hypotheses:
##        Estimate Std. Error t value Pr(>|t|)
## 1 == 0  0.41981    0.84453   0.497   0.9984
## 2 == 0 -2.77738    0.92837  -2.992   0.0668 .
## 3 == 0  0.38786    0.21786   1.780   0.4681
## 4 == 0 -0.09839    0.11029  -0.892   0.9520
## 5 == 0  4.10834    3.60101   1.141   0.8573
## 6 == 0  1.55127    2.69287   0.576   0.9959
## 7 == 0  3.15281    1.93253   1.631   0.5603
## 8 == 0  2.59060    1.81726   1.426   0.6928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
## [1] "seedingno:sne"                    "seedingyes:sne"
## [3] "seedingno:cloudcover"             "seedingyes:cloudcover"
## [5] "seedingno:prewetness"             "seedingyes:prewetness"
## [7] "seedingno:echomotionstationary"  "seedingyes:echomotionstationary"
```

c. How does adjustment for multiple testing change which interactions are significant?**

The adjusted p value remain the same when we take the General Linear Hypotheses on the model, however, when we apply the K matrix in the glht function, the p values does not remain the same for each interations in multiple testing.

3. For the logistic regression model presented in Chapter 7 in Figure 7.7, perform a multiplicity adjusted test on all regression coefficients (except for the intercept) being zero. Do the conclusions drawn in Chapter 7 remain valid?

```
##
## Call:
## glm(formula = fm2, family = binomial(), data = womensrole)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.39097  -0.88062   0.01532   0.72783   2.45262
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)             2.09820    0.23550   8.910  < 2e-16 ***
## genderFemale            0.90474    0.36007   2.513  0.01198 *
## education              -0.23403    0.02019 -11.592  < 2e-16 ***
## genderFemale:education -0.08138    0.03109  -2.617  0.00886 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 451.722  on 40  degrees of freedom
## Residual deviance:  57.103  on 37  degrees of freedom
## AIC: 203.16
##
## Number of Fisher Scoring iterations: 4
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: glm(formula = fm2, family = binomial(), data = womensrole)
##
## Linear Hypotheses:
##        Estimate Std. Error z value Pr(>|z|)
## 1 == 0 -0.23403    0.02019 -11.592   <0.001 ***
## 2 == 0  2.09820    0.23550   8.910   <0.001 ***
## 3 == 0 -0.08138    0.03109  -2.617   0.0185 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
## [1] "(Intercept)"            "genderFemale"
## [3] "education"              "genderFemale:education"
```

Answer

Yes the conclusion in chapter 7 still remain valid, as all the adjusted p values are still significant, but have closer value to the significant level compared to the results obtained in chapter 7.