Multiple Comparison, FDR, and Simultaneous Inference

Mohamed Ahmed

Collaboration:

Amin baabol and I collaborated in producing this report. We worked together and verified each other's work on every part of this assignment.

Exercises

1. (Question 15.1 on pg. 295 in HSAUR, modified for clarity) Consider **alpha** dataset from the **coin** package. Compare the results when using **glht** and TukeyHSD (Refer to Chapter 5 for TukeyHSD).

Discussion:

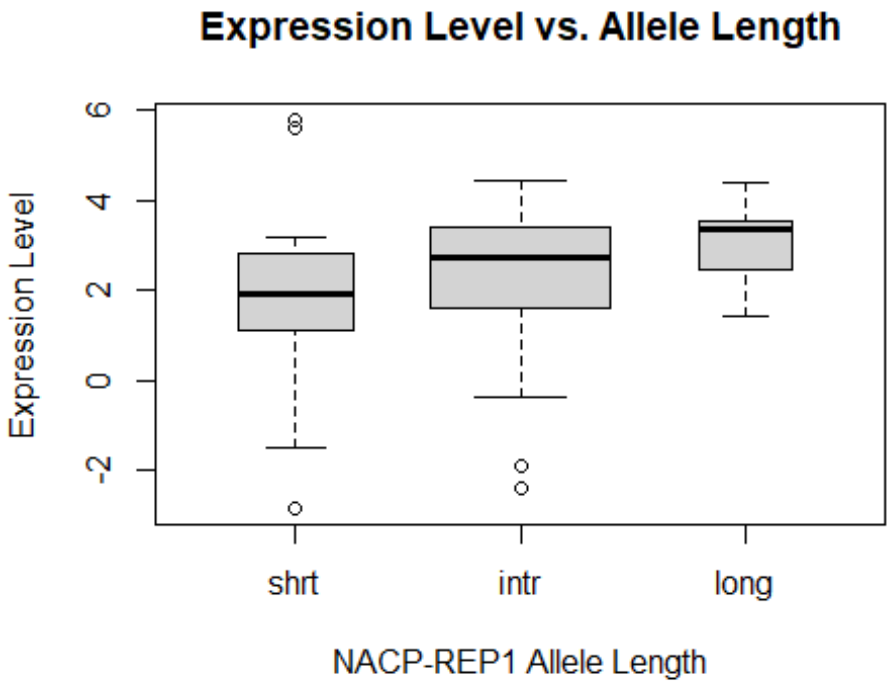
We started our analysis by manipulating the data to be into a model and for various levels of the variable alength to be compared. First, an **F-test** was conducted and we found that there is not difference between the expression levels of the variable alength. furthermore, a general linear hypothesis test(Tukey's) using the ordinary covariance matrix was conducted to verify if there actually difference between different levels. The test indicated there is no significant difference between the expression levels among allele lengths. it was important to inspect the variance homogeneity. Therefore, we plotted the mean differences of the expressions levels of the allele lengths and we found that variance homogeneity is violated. We decided to implement a **Sandwich estimator** of K matrix.

Figure 1b and figure 1c are two versions of the glht model, the former uses the "ordinary" tukey's,while the later and the more accurate one uses the sandwich estimator.According to the glht with the sandwich estimator, there is a significant difference between the mean expression levels of the long and short allele lengths of at least 1.1888.The p-value of this partial hypothesis is 0.0226 as indicated by figure 1c.

Further more, we ran a tukey's honest significant difference test on our anova model. According to the summary in figure 1d and its corresponding plot there is no significant difference between the mean expression levels of the various allele lengths. The lowest adjusted p-value for the mean expression levels of the allele lengths is 0.0628589 where as the glht model with the sandwich estimator has a p-value of 0.0226. Therefore, the "sandwich" glht() method is more reliable and accurate than the tukeyHSD() method in the presence of unbalanced data and when the variance homogeneity rule is violated.

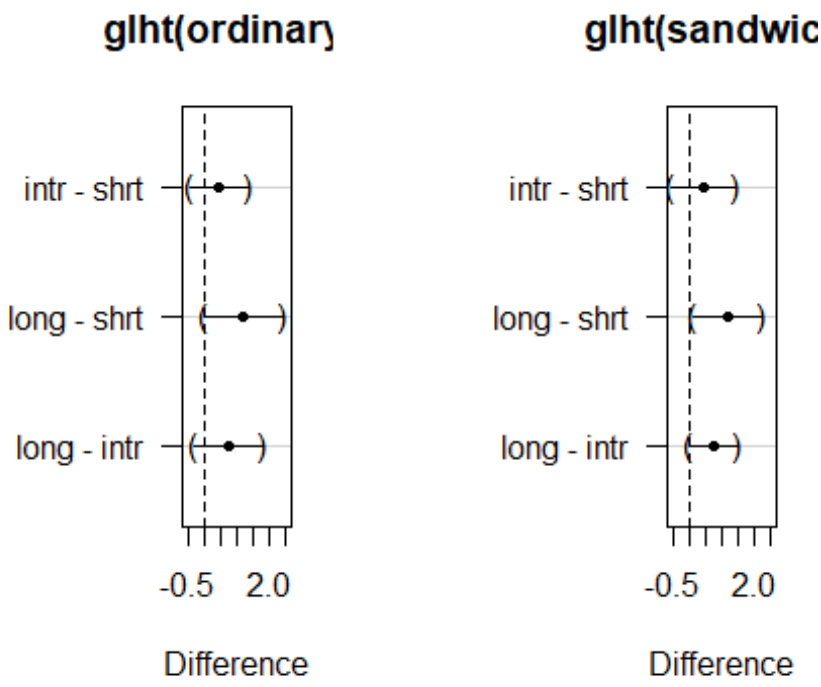
```
##        alength elevel
## 1        short   1.43
## 2 intermediate  -1.90
## 3 intermediate   1.55
## 4 intermediate   3.27
```

```
## 5 intermediate   0.30
## 6 intermediate   1.90

## [1] "F-test"

##             Df Sum Sq Mean Sq F value Pr(>F)
## alength      2  13.06   6.528   2.613 0.0786 .
## Residuals   94 234.85   2.498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

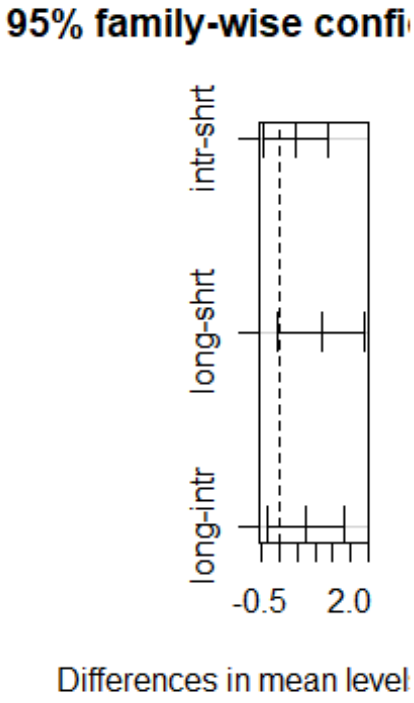



```
## [1] "glht summary"

##
##   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   0.2270
## ---
```

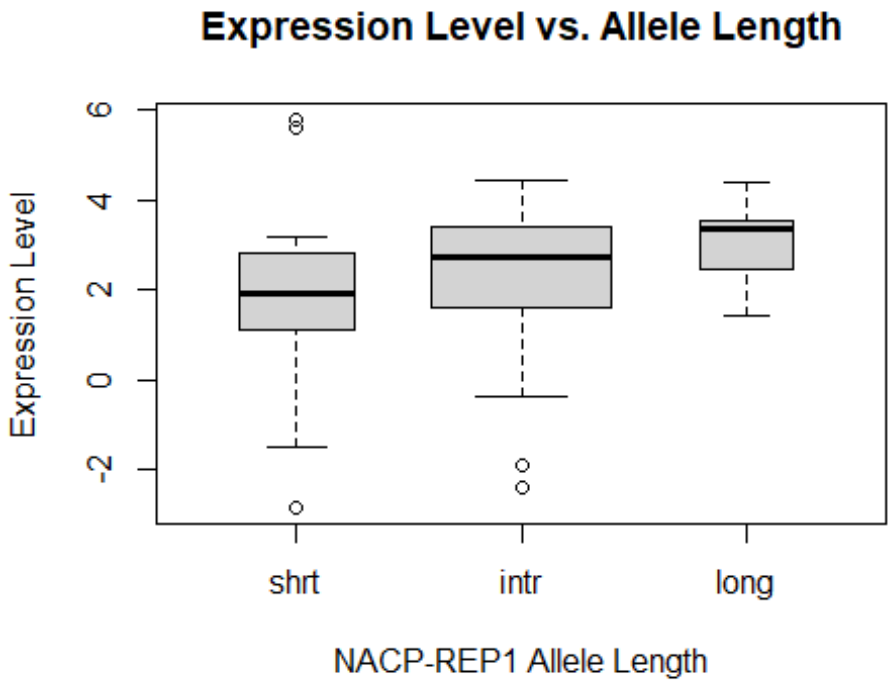
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

## [1] "glht(ordinary) coefficients"

## intr - shrt long - shrt long - intr
##   0.4341523   1.1887500   0.7545977

## [1] "glht(ordinary) covariances matrix"

##             intr - shrt long - shrt long - intr
## intr - shrt  0.14717604   0.1041001 -0.04307591
## long - shrt  0.10410012   0.2706603  0.16656020
## long - intr -0.04307591   0.1665602  0.20963611

## [1] "Figure 1c:glht(sandwich) summary"

##
##   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     0.4239   1.024   0.5594
## long - shrt == 0   1.1888     0.4432   2.682   0.0227 *
## long - intr == 0   0.7546     0.3184   2.370   0.0503 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

## [1] "glht(sandwich) coefficients"

## intr - shrt long - shrt long - intr
##   0.4341523   1.1887500   0.7545977

## [1] "glht(sandwich) covariances matrix"

##             intr - shrt long - shrt long - intr
## intr - shrt  0.17971983  0.13737638 -0.04234345
## long - shrt  0.13737638  0.19638853  0.05901215
## long - intr -0.04234345  0.05901215  0.10135559
```




```
## [1] "Figure 1d:TukeyHSD summary"

##   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
```

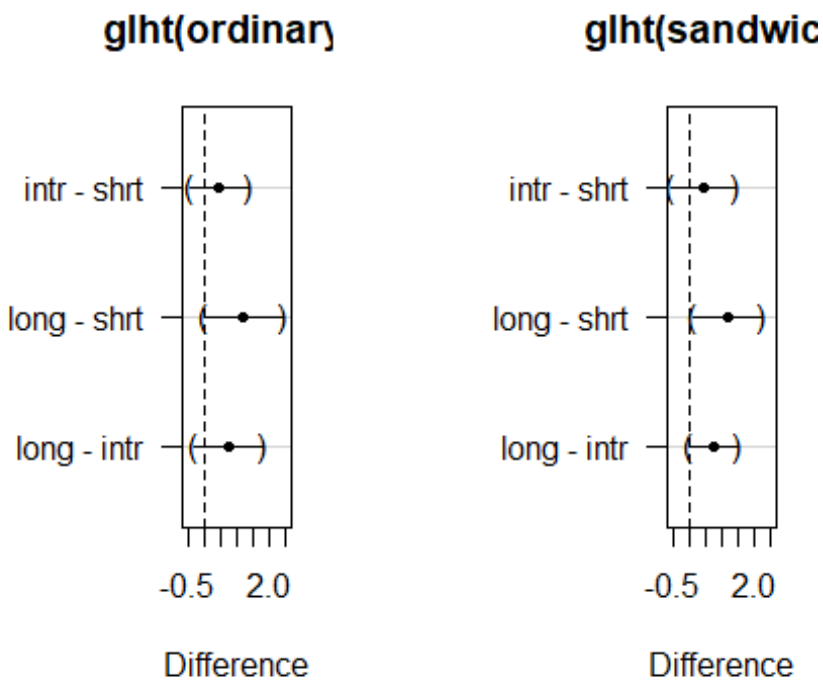



```
## [1] "Figure 1a: F-test"

##             Df Sum Sq Mean Sq F value Pr(>F)
## alength      2  13.06   6.528   2.613 0.0786 .
## Residuals   94 234.85   2.498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

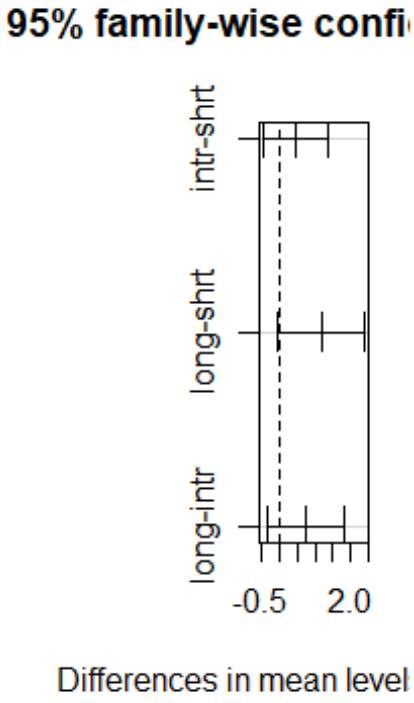



```
## [1] "Figure 1b:glht(ordinary) summary"

##
##   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   0.2270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)

## [1] "Figure 1c:glht(sandwich) summary"

##
##   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     0.4239   1.024   0.5594
## long - shrt == 0   1.1888     0.4432   2.682   0.0227 *
## long - intr == 0   0.7546     0.3184   2.370   0.0502 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```




```
## [1] "Figure 1d:TukeyHSD summary"

##   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
```



2. (Question 15.2 on pg. 296 in HSAUR, modified for clarity) 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 HSAUR Ed 3.

Data

Cloud seeding is a practice done to influence the weather,typically to increase precipitation or reduce hail, fog, or ice in high traffic areas such as airports and busy inter-states.The **Clouds** dataset was collected in an experiment that was conducted in Florida in 1975 to probe the use of large-scale silver iodide in individual cloud seeding to increase rainfall. The data contains details about cloud treatments with organic and inorganic materials, days the treatment was applied, suitability criterion(SNE), and other data that was collected in the experiment. Since the data is small, we can use it to evaluate the variability of the estimated regression line.

Method

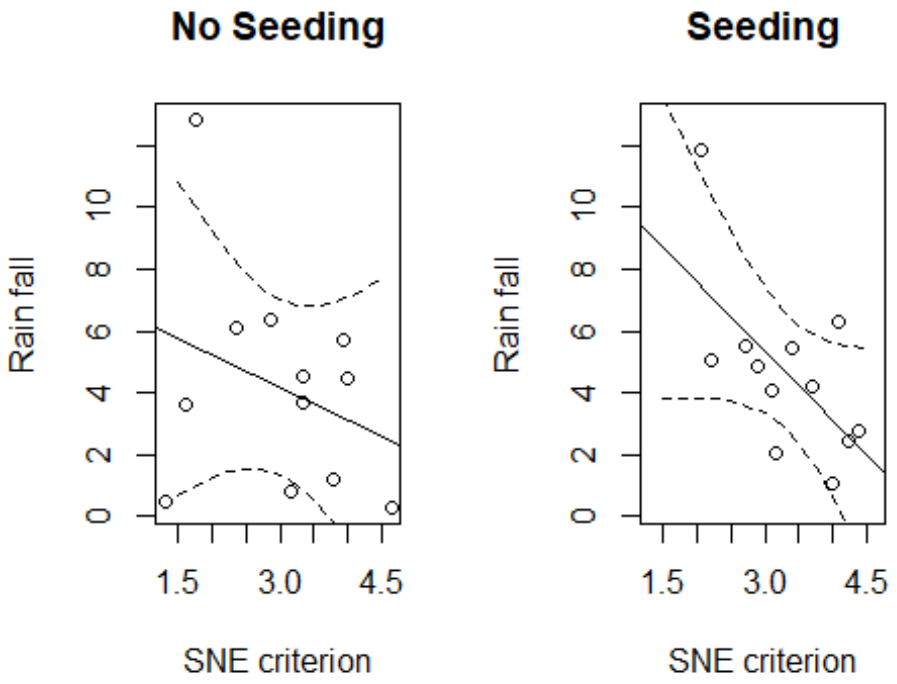
We first fitted a linear model to see if there is a linear relationship between the independent variable suitability criterion(SNE) and the response variable rainfall.We want to calculate the confidence interval region that has a probability $ => 1 – α$,for the estimated regression line. To do that, we used the **Pointwise** method, but we want to control TYPE I error for all predictors simultaneously. Therefore,the alternative method would be to restructure the linear model to include linear combination of the regression coefficients. Matrix K will be multiplied by θ which represents coefficients $\beta 0, \beta 1$ of interest which will make up the confidence interval for the fitted regression line. Finally,

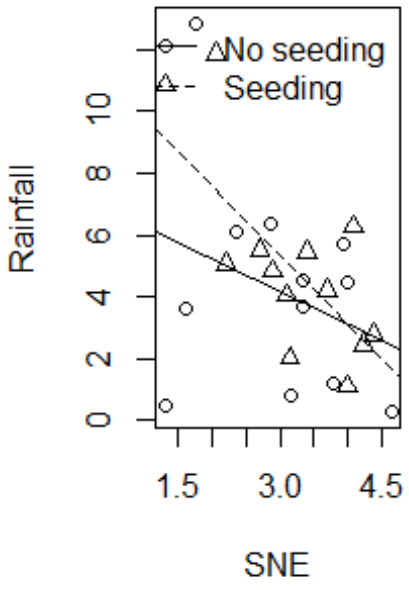
we plot two different graphs to observe the confidence bands for the rainfall when seeding is implemented and when seeding is not implemented.

Results

The linear regression plot shows that smaller S-Ne values along with seeding causes more rainfall than when seeding is not implemented. However, when we have high S-Ne values, the plot shows that we get less rainfall. The point where both lines intersect when SNE of 4 which indicates that rainfall can be maximized by applying seeding with SNE values lower than 4.Furthermore, the plots shows that we have more uncertainty in the regression line when seeding is not implemented than when seeding is implemented.This uncertainty is due to the substantial variability in the observations when seeding is not implemented. Since we have more certainty in seeding's true regression lines, we can conclude that seeding is a better fit than when seeding is not used up until SNE value of 4 at which point no seeding will yield more rainfall.

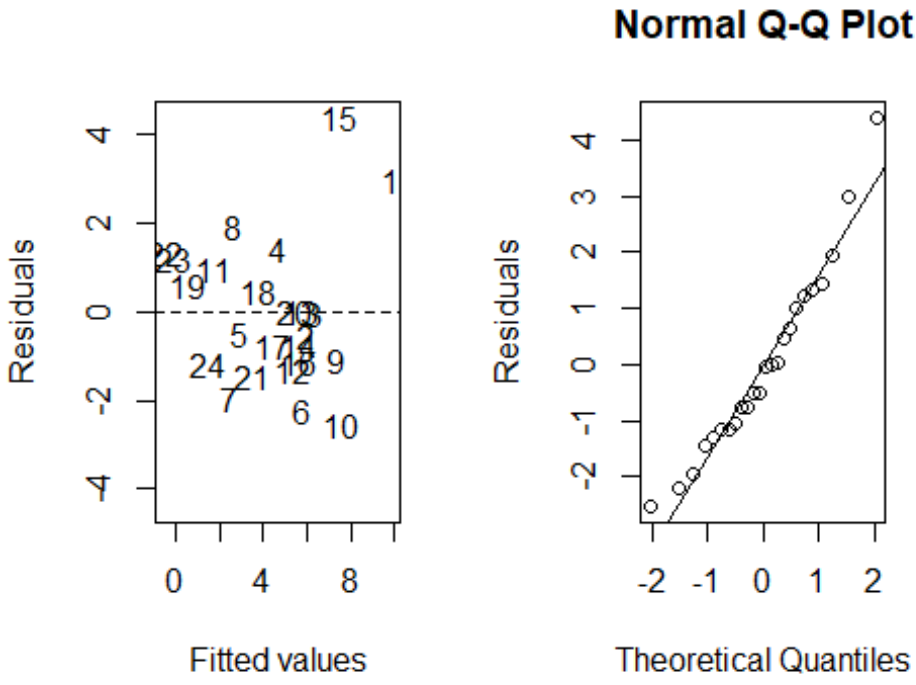
```
## [1] "Figure 2a"
```





b. Consider the linear model fitted to the clouds data as summarized 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.

```
## [1] "Figure 2b"
```



Discussion

We fit the linear regression model from chapter 6 and created matrix K with all of the interaction terms equal to zero. Then, we tested the global null hypotheses and the null partial hypotheses corresponding to each of the interaction terms using the **glht** function. The global hypotheses test is significant with p-value **0.02430934 < 0.05**. All hypotheses corresponding to each of the interaction terms were tested while set to zero. All the interaction term were insignificant at level of 0.05 except for for **seedingyes** with p-value **0.0293 < 0.05**.

```
## [1] "Simple multi-linear regression"

##
## 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

## [1] "General linear hypothesis test"

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = clouds_formula, data = clouds)
##
## Linear Hypotheses:
##                                      Estimate Std. Error t value Pr(>|t|)
## seedingyes == 0                      15.68293    4.44627   3.527   0.0294
*
## time == 0                            -0.04497    0.02505  -1.795   0.5006
## seedingno:sne == 0                    0.41981    0.84453   0.497   0.9992
## seedingyes:sne == 0                  -2.77738    0.92837  -2.992   0.0769
.
## seedingno:cloudcover == 0             0.38786    0.21786   1.780   0.5096
## seedingyes:cloudcover == 0           -0.09839    0.11029  -0.892   0.9657
## seedingno:prewetness == 0             4.10834    3.60101   1.141   0.8856
## seedingyes:prewetness == 0            1.55127    2.69287   0.576   0.9978
## seedingno:echomotionstationary == 0   3.15281    1.93253   1.631   0.6031
## seedingyes:echomotionstationary == 0  2.59060    1.81726   1.426   0.7331
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

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

Discussion

According to the multi-linear regression summary the interaction terms seedingyes and seedingyes:sne are significant at the 0.05 confidence interval level.seedingyes is the critically significant parameter in explaining rainfall followed seedingyes:sne. On the other hand, the multiple comparison testing raised the p-values of the interaction terms, thus changing their significance levels. According to the glht model, only seedingyes with an

adjusted p-value of 0.0296 is significant at the 0.05 alpha level which makes sense because glht is comparing partial hypothesis between pairs.

Given the small sample size of the original clouds dataset the multi-linear r egression model is susceptible to outlier influences.To check the validity of the simple linear model we check the fitted values against the residual and also checked the normality assumption through normal qq-plot. According to figure 2b,there are a couple of outliers that might be influencing the model performance.Practically speaking, if we were to only use the multi-linear regression model it is best to re-run the model with those out-liers removed to check if the model results improve. However, due to the rebustness against outliers of the multiple comparison testing method we can safely deduce that that seedingyes is the most important parameter in explaining the response variable rainfall.

3. (Question 15.3 on pg. 296 in HSAUR, modified for clarity) or 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?

Interpretations:

First,the Glm model shows that the interaction terms and the intercept are highly significant at level of **0.05**. The next step was to compare different variables that have different levels. We defined the contrast of interest and a ran generalized linear hypothesis test using the **glht** model object. We want to test the difference between the interaction terms. After running the model, it looks like that all the coefficients are significant with **p-value < 0.05**. Therefore, we conclude that chapter 7 conclusion remains valid at significance level of 0.05

```
##
## 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|)
## genderFemale == 0             0.90474    0.36007   2.513   0.0244 *
## education == 0               -0.23403    0.02019 -11.592   <0.001 ***
## genderFemale:education == 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)
```

#Citations

Michael, Semhar, and Christopher P. Saunders. “Simultaneous Inference and Multiple Comparisons” Chapter 15.8 Nov. 2020, South Dakota State University, South Dakota State University.

Michael, Semhar, and Christopher P. Saunders. “Analysis of Variance: Chapter 5 Review” Chapter 5.8 Nov. 2020, South Dakota State University, South Dakota State University.

Everitt, Brian, and Torsten Hothorn. A Handbook of Statistical Analyses Using n SECOND EDITION. Taylor and Francis Group LLC, 2010.

Hothorn, T., & Everitt, B. S. (n.d.). A Handbook of Statistical Analyses Using R — 3rd Edition. Retrieved 2020, from http://cran.uni-muenster.de/web/packages/HSAUR3/vignettes/Ch_simultaneous_inference.pdf