

lab2: Linear Regression Analysis:

Use the following command lines to install and load package “gamair”. In order to activate the installing line, you need to remove symbol “#” appearing at the beginning of that line.

```
# install.packages("gamair")  
library(gamair)
```

---

Hubble Data:

Assigning different values to plotting character, pch, you are able to create plots with various symbols. Alternatively, you can use alphabetic characters:

```
data(hubble)  
names(hubble)  
hubble$x  
plot(hubble$x,hubble$y, pch = 12)  
  
plot(hubble$x,hubble$y, pch = 128)  
  
plot(hubble$x,hubble$y, pch = "A")
```

We now fit a linear model to predict velocity from distance. More precisely, we are aiming to set  $\beta_0 = 0$  and fit model  $y = \beta_1 * x + error$

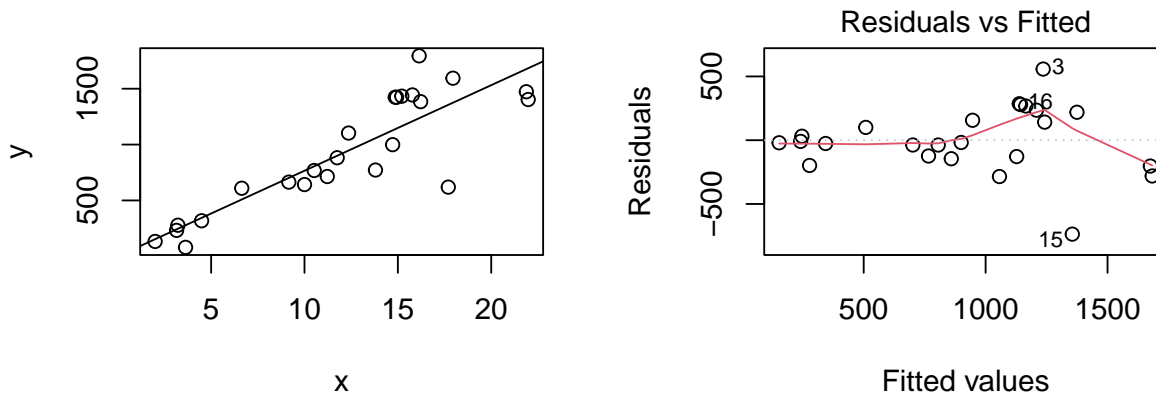
```
hmod <- lm(y ~ x - 1, data=hubble)  
#summary(hmod)  
coef(hmod)
```

```
##          x  
## 76.58117
```

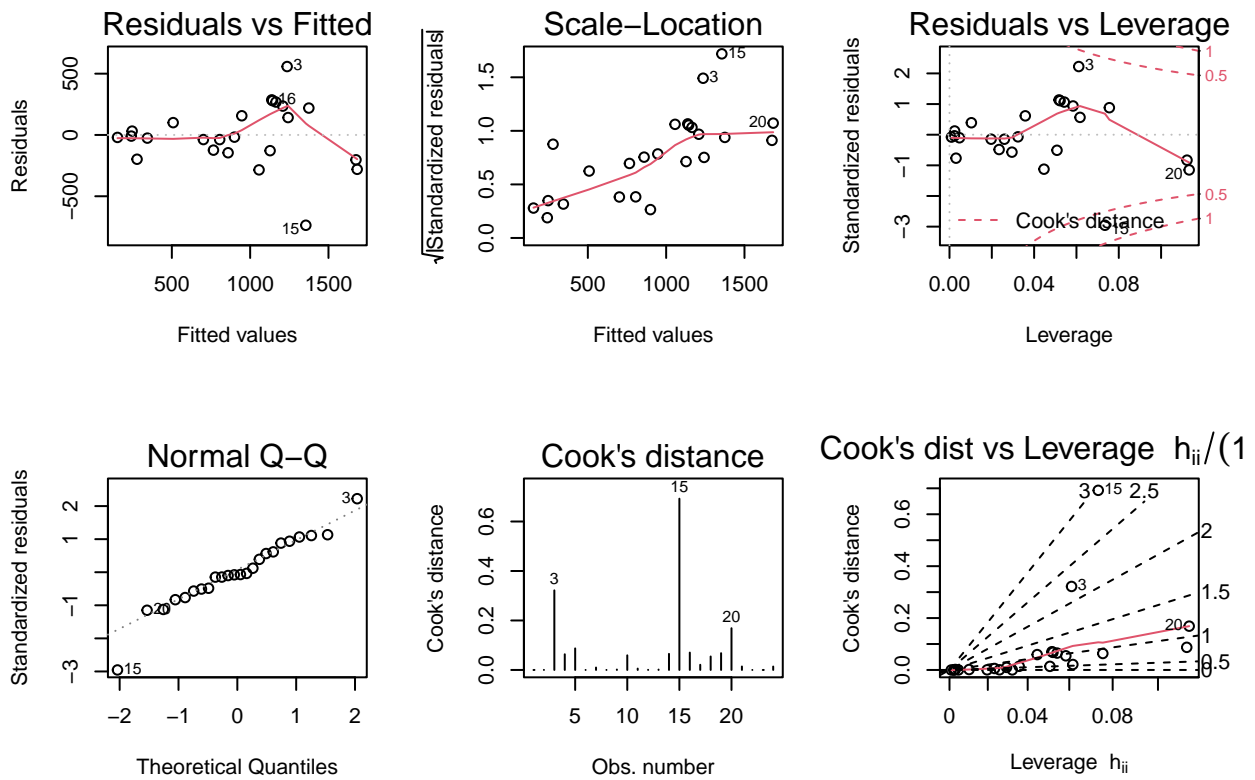
So, the output suggests:  $y = 76.58117 * x + error$ .

In the following, we use command “layout()” to create an array with arbitrary dimension and place the consecutive plots as the elements of that array. For better understanding, make comparison between the arguments of the layout command in the following two r chunks as well as corresponding outputs. Note that “ncol” stands for “number of columns”, similarly “nrow” stands for “number of rows”.

```
layout(matrix(1:2,ncol=2))  
plot(y ~ x, data = hubble)  
abline(hmod)  
plot(hmod, which = 1)
```



```
layout(matrix(1:6,ncol=3))
plot(hmod, which = c(1,2,3,4,5,6))
```



Plots above suggest the model assumptions hold approximately true: absolute value of standardized residuals do not exceed 3 (or 2) and similarly cook distances do not exceed 1. According to QQ-plot, there is no typical deviation from theoretical normal distribution. Although, it seems the assumption “Constant Error Variance” is mildly violated.

Cloud Seeding Data:

```
#install.packages("HSAUR3")
library(HSAUR3)
```

```
## Loading required package: tools
```

We are aiming to fit the following linear model :

$$\text{rainfall} = \beta_0 + \beta_1 * \text{seeding} + \beta_2 * \text{time} + \beta_3 * (1 - \text{seeding}) * \text{sne} + \beta_4 * \text{seeding} * \text{sne} + \beta_5 * (1 - \text{seeding}) * \text{cloudcover} + \beta_6 * \text{seeding} * \text{cloudcover} + \beta_7 * (1 - \text{seeding}) * \text{prewetness} + \beta_8 * \text{seeding} * \text{prewetness} + \beta_9 * (1 - \text{seeding}) * \text{echomotion} + \beta_{10} * \text{seeding} * \text{echomotion}$$

In other words, we wish to fit two different models corresponding to different levels of variable “seeding”. We call these levels “0” and “1” corresponding to “seedingo” and “seedinyes” respectively. Assign value “0” to variable seeding to obtain one of the models. Do the same for “seeding = 1” and obtain the other model.

```
clouds.formula <- rainfall ~ seeding + seeding:(sne + cloudcover + prewetness + echomotion) + time
clouds.lm <- lm(clouds.formula, data = clouds)
```

```
summary(clouds.lm)
```

```
##
## 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
## seedinyes     15.68293    4.44627    3.527  0.00372 **
## time          -0.04497    0.02505   -1.795  0.09590 .
## seedingo:sne   0.41981    0.84453    0.497  0.62742
## seedinyes:sne -2.77738    0.92837   -2.992  0.01040 *
## seedingo:cloudcover 0.38786    0.21786    1.780  0.09839 .
## seedinyes:cloudcover -0.09839    0.11029   -0.892  0.38854
## seedingo:prewetness 4.10834    3.60101    1.141  0.27450
## seedinyes:prewetness 1.55127    2.69287    0.576  0.57441
## seedingo:echomotionstationary 3.15281    1.93253    1.631  0.12677
## seedinyes: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
```

The above summary result suggests:

if seeding = 0, then:  $\text{rainfall} = -0.34624 - 0.04497 * \text{time} + 0.41981 * \text{sne} + 0.38786 * \text{cloudcover} + 4.10834 * \text{prewetness} + 3.15281 * \text{echomotion}$

if seeding = 1, then:  $\text{rainfall} = -0.34624 + 15.68293 - 0.04497 * \text{time} - 2.77738 * \text{sne} - 0.09839 * \text{cloudcover} + 1.55127 * \text{prewetness} + 2.59060 * \text{echomotion}$

If you fit a similar model with  $\beta_0 = 0$ , you will see intercepts  $-0.34624$  and  $-0.34624 + 15.68293$  directly in

the output table.

You can deduce a similar result by including variable “sne” (one of the interaction effect parents) in the model above.

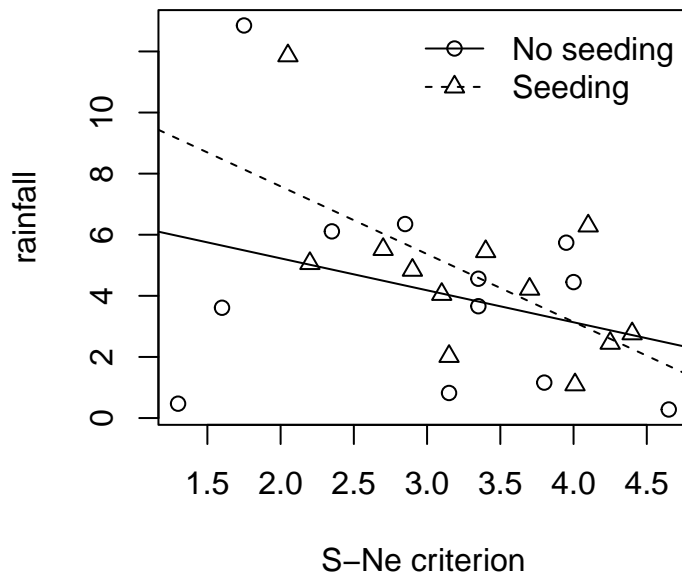
We also conclude that:  $R^2 = 0.7158$ ,  $R_{adj}^2 = 0.4972$ ,

p-value for rejecting “ $H_0 : \beta_0 = \beta_1 = \dots \beta_{10} = 0$ ” againsts “ $H_1 : \text{there is at least one non - zero parameter}$ ” equals 0.02431. So, we reject  $H_0$  at the level 0.05.

---

We can also look at the regression relationship between S-Ne and rainfall with and without seeding:

```
layout(1)
psymb <- as.numeric(clouds$seeding)
plot(rainfall ~ sne, data = clouds, pch = psymb, xlab = "S-Ne criterion")
abline(lm(rainfall ~ sne, data = clouds, subset = seeding == "no"))
abline(lm(rainfall ~ sne, data = clouds, subset = seeding == "yes"), lty = 2)
legend("topright", legend = c("No seeding", "Seeding"), pch = 1:2, lty = 1:2, bty = "n")
```



interpretation of this plot:

Similar to the previous part, we have two different regression lines corresponding to “seeding = 0” and “seeding = 1”, see solid line and dashed line.

Is there an interaction between seeding and S-Ne? The lines are not parallel, that means the coefficient of variable “sne” in one level is significantly different from the coefficient of variable “sne” in the other model. Consequently there is interaction between variables seeding and "sne".