

CS130 - Assignment 1: R Competency

Minerva University

CS130: Statistical Modeling: Prediction and Causal Inference

Prof. A. Diamond

January 20, 2023

Data Analysis of Scurvy Dataset

1. Data summary (data was analyzed using R, code can be found in Appendix A):

- Sample size: 12 observations (rows):
- 8 variables (columns): study_id, treatment, dosing_regimen_for_scurvy, gum_rot_d6, skin_sores_d6, weakness_of_the_knees_d6, lassitude_d6, fit_for_duty_d6
- 6 types of treatment: cider, purgative_mixture, sea_water, vinegar, citrus, dilute_sulfuric_acid. Each of the treatments was tested on 2 observations.
- 11 seamen do not fit for duty, 1 seaman fits for duty
- 10 seamen were weak in the knees for each treatment which is 83% of the sample

2. Symptoms analysis

The main symptoms the seamen experienced included: gum rot, skin sores, weakness of knees, and lassitude.

A mild level of gum rot was experienced by 2 patients #3 and #9. Another 3 patients #1, #2, and #4 experienced a moderate level of this symptom. The severe level had 6 other patients. Only patient #10 had no symptoms of gum rot.

A total of 9 people, namely patients #2 #3 #4 #5 #6 #7 #8 #11 #12, experienced severe levels of lassitude. There was only one patient without this symptom, it was the same one who did not have symptoms of gum rot, patient #10. Patient #9 had a mild level of symptoms.

Regarding skin sores, patient #10 was the only one without the symptoms. Patients #2 and #9

had mild levels of skin sores.

The majority of the patients experienced a severe level of knee weakness. Only 2 patients, #9 and #10, did not experience this symptom while everyone else suffered from either severe or moderate levels of this symptom.

3. Conclusion:

Based on the data analysis we can see that patient #10 experienced no symptoms. Also when I checked who was the only person fit for duty it was patient #10. When I checked which treatment that patient had it was citrus. Since I know that each treatment was received by 2 people I checked the number of the second patient that was cured by the same treatment and checked their symptoms. Based on the previous analysis we know that patient #9 had mild skin sores and weakness of knee levels. I checked the remaining 2 symptoms: lassitude and gum rot and their levels were mild as well. There were no other people that would show a similar pattern. Therefore, the treatment that included citrus is likely to be the most effective.

Word Count: 521 words

Appendix A: code for the data analysis. The file with the code is attached to this assignment

```
#obligatory part

install.packages("medicaldata")
library("medicaldata")
#load data
scurvy <- medicaldata::scurvy
ls()

#get information about the shape and column names
head(scurvy)
dim(scurvy)

#create a table for treatment and duty fitness
treatment <- scurvy$treatment
table(treatment)

duty <- scurvy$fit_for_duty_d6
table(duty)

#filter vector based on the knee weakness and count result
weak <- scurvy$weakness_of_the_knees_d6
result <- length(which(weak != "0_none"))
result

#filter data based on the gum_rot and lassitude
which(scurvy$gum_rot_d6=="1_mild")
which(scurvy$lassitude_d6=="3_severe")
which(scurvy$gum_rot_d6=="1_mild" & scurvy$lassitude_d6=="3_severe")

#create an empty vector to store modified variables
weak_knees <- c()

for (i in weak) {
  #select the first character and convert it to an integer
  k<- strtoi(substr(i,1,1))
  weak_knees <- append(weak_knees, k)
}

#Find the quartiles (10th, 50th, and 90th percentiles)
quantile(weak_knees, probs = c(.1, .5, .9))

#added by student
which(scurvy$gum_rot_d6=="2_moderate")
which(scurvy$gum_rot_d6=="2_moderate")
which(scurvy$gum_rot_d6=="3_severe")
length(which(scurvy$gum_rot_d6=="3_severe"))
```

```
which(scurvy$gum_rot_d6=="0_none")
length(which(scurvy$lassitude_d6=="3_severe"))
length(which(scurvy$lassitude_d6=="0_none"))
which(scurvy$lassitude_d6=="3_severe")
which(scurvy$lassitude_d6=="0_none")
which(weak == "0_none")
which(scurvy$skin_sores_d6 == "0_none")
which(scurvy$skin_sores_d6 == "1_mild")
which(scurvy$fit_for_duty == "1_yes")
treatment[10]
which(scurvy$treatment == "citrus")
scurvy$lassitude_d6[9]
which(scurvy$gum_rot_d6=="1_mild")
which(weak == "1_mild")
which(scurvy$gum_rot_d6=="1_mild")
which(scurvy$lassitude_d6=="1_mild")
```

References

GeeksforGeeks. (2020, May 10). Adding elements in a vector in R programming – append() method.

<https://www.geeksforgeeks.org/adding-elements-in-a-vector-in-r-programming-append-method/>

R For Loop. (n.d.). https://www.w3schools.com/r/r_for_loop.asp

Cn, P. (2022, August 3). *The which() function in R programming*. DigitalOcean Community.

<https://www.digitalocean.com/community/tutorials/which-function-in-r>

R length function -- EndMemo. (n.d.). <http://www.endmemo.com/r/length.php>

Z., & posts by Zach, V. A. (2022, July 13). *How to Filter a Vector in R (4 Examples)* - Statology.

Statology. Retrieved January 20, 2023, from <https://www.statology.org/filter-vector-in-r/>