PROJECT 1 REPORT

Nonparametric Statistics

Data Description

The data includes the number of glucocorticoid receptor (GR) sites per leukocyte cell for five groups. The five groups are as follows: normal subjects, hairy-cell leukemia, chronic lymphatic leukemia, chronic myelocytic leukemia, and acute leukemia. In this case, the normal subjects will be the control group, and the patients with hairy-cell leukemia, chronic lymphatic leukemia, chronic myelocytic leukemia, or acute leukemia will be the four treatment groups. There are fourteen observations for normal subjects, five observations for hairy-cell leukemia, six observations for chronic lymphatic leukemia, four observations for chronic myelocytic leukemia, and eight observations for acute leukemia.

Table 6.4 Number of Glucocorticoid Receptor (GR) Sites per Leukocyte Cell

Normal subjects	Hairy-cell anemia	Chronic lymphatic leukemia	Chronic myelocytic leukemia	Acute leukemia
3,500	5,710	2,930	6,320	3,230
3,500	6,110	3,330	6,860	3,880
3,500	8,060	3,580	11,400	7,640
4,000	8,080	3,880	14,000	7,890
4,000	11,400	4,280		8,280
4,000		5,120		16,200
4,300				18,250
4,500				29,900
4,500				
4,900				
5,200				
6,000				
6,750				
8,000				

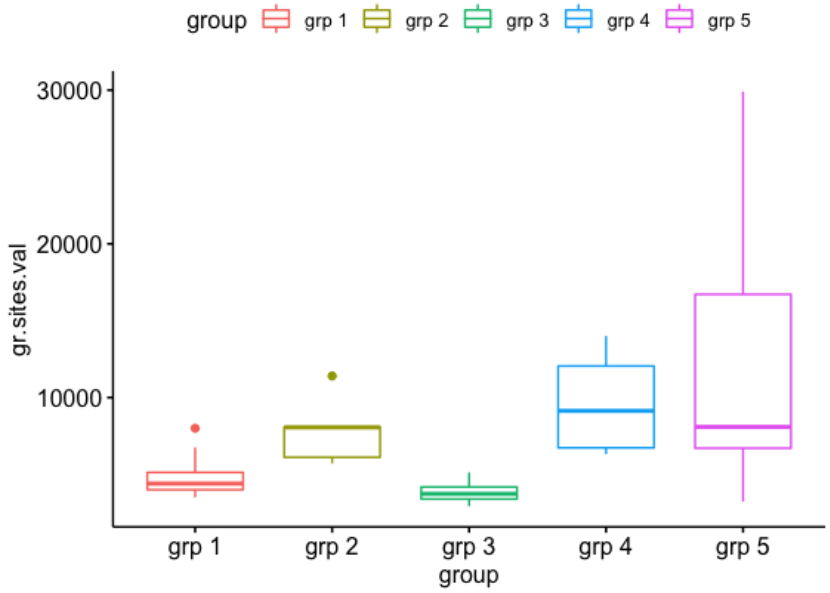
Source: K. Kontula, L. C. Andersson, T. Paavonen, G. Myllyla, L. Teerenhovi, and P. Vuopio (1980) and K. Kontula, T. Paavonen, R. Vuopio, and L. C. Andersson (1982).

For each group, the mean and variance are determined. These two descriptive statistics are shown in the chart below.

group	mean	variance
grp 1	4760.714	1773145.6
grp 2	7872.000	5076070.0
grp 3	3853.333	597666.7
grp 4	9645.000	13619300.0
grp 5	11908.750	81376183.9

grp 1: normal subjects
grp 2: hairy-cell leukemia patients
grp 3: chronic lymphatic leukemia patients
grp 4: chronic myelocytic leukemia patients
grp 5: acute leukemia patients

Additionally, box plots are used to visualize the data.



Research Questions

In regards to the data, the research questions I am investigating include:
Research Question 1: Are there any differences in the number of GR sites per leukocyte cell for the five subject populations?
Research Question 2: Do patients with leukemia have a larger number of GR sites per leukocyte cell than normal subjects?
Research Question 3: If there exists a significant difference in the number of GR sites per leukocyte cell for the five subject populations, can we draw conclusions about specific pairs of treatment effects?
Research Question 4: Do patients with hairy-cell leukemia, or patients with chronic lymphatic leukemia, or patients with chronic myelocytic leukemia, or patients with acute leukemia, have a larger number of GR sites per leukocyte cell than normal subjects?

Nonparametric Methods for Two or More Sample Location Problems

For the four research questions, the null hypothesis is $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ (no treatment effects). For research question 1, it involves determining if the treatment effects are not all equal, so I will be using the Kruskal-Wallis test. For research question 2, it involves treatment groups versus a control group, so I will be using the Fligner-Wolfe test. For research question 3, it involves pairwise multiple comparisons, so I will be using the Steel-Dwass-Critchlow-Fligner test. For research question 4, it involves treatment groups versus a control group with multiple comparisons, so I will be using the Nemenyi-Damico-Wolfe test. The four tests do not have a normality assumption, which makes them robust.

Performing the Analysis

For the analysis, I am using R to perform the following tests: Kruskal-Wallis test, Fligner-Wolfe test, Steel-Dwass-Critchlow-Fligner test, and Nemenyi-Damico-Wolfe test. For each test, I include the R code along with the output. The outputs will be used to draw conclusions about the research questions. Furthermore, the tests will be conducted at 5% level of significance.

R code:

```
gr1 <- c(3500, 3500, 3500, 4000, 4000, 4000, 4300, 4500, 4500, 4900, 5200, 6000, 6750, 8000)
gr2 <- c(5710, 6110, 8060, 8080, 11400)
gr3 <- c(2930, 3330, 3580, 3880, 4280, 5120)
gr4 <- c(6320, 6860, 11400, 14000)
gr5 <- c(3230, 3880, 7640, 7890, 8280, 16200, 18250, 29900)
n1 <- length(gr1)
n2 <- length(gr2)
n3 <- length(gr3)
n4 <- length(gr4)
n5 <- length(gr5)

gr.sites <- data.frame(gr.sites.val=c(gr1, gr2, gr3, gr4, gr5),
                       group=rep(c("grp 1", "grp 2", "grp 3", "grp 4", "grp 5"), c(n1, n2, n3, n4, n5)))
```

Kruskal-Wallis test:

$H_1: \tau_1, \dots, \tau_5 \; not \; all \; equal$

```
kruskal.test(gr.sites.val ~ as.factor(group), data = gr.sites)
```

```
        Kruskal-Wallis rank sum test

data:  gr.sites.val by as.factor(group)
Kruskal-Wallis chi-squared = 16.668, df = 4, p-value = 0.002242
```

Fligner-Wolfe test:

$H_2: \tau_2 \geq \tau_1, \tau_3 \geq \tau_1, \tau_4 \geq \tau_1, \tau_5 \geq \tau_1 \; with \; at \; least \; one \; strict \; inequality$

```
gr.sites$trt <- gr.sites$group != "grp 1"
library(coin)
wilcox_test(gr.sites.val~as.factor(trt), data=gr.sites, alternative="less")
```

```
        Asymptotic Wilcoxon-Mann-Whitney Test

data:  gr.sites.val by as.factor(trt) (FALSE, TRUE)
Z = -2.0056, p-value = 0.02245
alternative hypothesis: true mu is less than 0
```

Steel-Dwass-Critchlow-Fligner test:

H_3: *at least one of* $\tau_1 \neq \tau_2, \tau_1 \neq \tau_3, \tau_1 \neq \tau_4, \tau_1 \neq \tau_5, \tau_2 \neq \tau_3, \tau_2 \neq \tau_4, \tau_2 \neq \tau_5, \tau_3 \neq \tau_4, \tau_3 \neq \tau_5, \tau_4 \neq \tau_5$

```
library("PMCMRplus")
dscfAllPairsTest(gr.sites.val~as.factor(group), data=gr.sites)
```

```
        Pairwise comparisons using Dwass-Steele-Critchlow-Fligner all-pairs test

data: gr.sites.val by as.factor(group)

       grp 1  grp 2  grp 3  grp 4
grp 2  0.042  -      -      -
grp 3  0.462 0.049  -      -
grp 4  0.059 0.911 0.078   -
grp 5  0.274 0.992 0.206 0.997

P value adjustment method: single-step
```

Nemenyi-Damico-Wolfe test:

H_4: *at least one of* $\tau_2 > \tau_1, \tau_3 > \tau_1, \tau_4 > \tau_1, \tau_5 > \tau_1$

```
library("PMCMRplus")
kwManyOneNdwTest(gr.sites.val~as.factor(group), data=gr.sites, alternative="greater")
```

```
        Pairwise comparisons using Nemenyi-Damico-Wolfe many-to-one test

data: gr.sites.val by as.factor(group)

       grp 1
grp 2  0.054
grp 3  0.995
grp 4  0.038
grp 5  0.043

P value adjustment method: single-step
alternative hypothesis: greater
```

Conclusions

For the **Kruskal-Wallis test**, since p-value=0.002242 < α=0.05, reject H_0. There is enough evidence to conclude that there are some differences in the number of GR sites per leukocyte cell for the five subject populations at 0.05 significance level.
For the **Fligner-Wolfe test**, since p-value=0.02245 < α=0.05, reject H_0. There is enough evidence to conclude that patients with leukemia have a larger number of GR sites per leukocyte cell than normal subjects at 0.05 significance level.
For the **Steel-Dwass-Critchlow-Fligner test**, there were 10 pairwise comparisons:

1. p-value=0.042 < α=0.05; At an experiment-wise error rate of α=0.05, there is a significant difference between group 1 and group 2.
2. p-value=0.462 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 1 and group 3.
3. p-value=0.059 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 1 and group 4.
4. p-value=0.274 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 1 and group 5.
5. p-value=0.049 < α=0.05; At an experiment-wise error rate of α=0.05, there is a significant difference between group 2 and group 3.
6. p-value=0.911 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 2 and group 4.
7. p-value=0.992 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 2 and group 5.
8. p-value=0.078 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 3 and group 4.
9. p-value=0.206 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 3 and group 5.
10. p-value=0.997 > α=0.05; At an experiment-wise error rate of α=0.05, there is no significant difference between group 4 and group 5.

For the **Nemenyi-Damico-Wolfe test**, there were four comparisons of treatment groups versus a control group:

1. p-value=0.054 > α=0.05; At an experiment-wise error rate of α=0.05, there is not enough evidence to conclude that patients with hairy-cell leukemia have a larger number of GR sites per leukocyte cell than normal subjects.
2. p-value=0.995 > α=0.05; At an experiment-wise error rate of α=0.05, there is not enough evidence to conclude that patients with chronic lymphatic leukemia have a larger number of GR sites per leukocyte cell than normal subjects.
3. p-value=0.038 < α=0.05; At an experiment-wise error rate of α=0.05, there is enough evidence to conclude that patients with chronic myelocytic leukemia have a larger number of GR sites per leukocyte cell than normal subjects.
4. p-value=0.043 < α=0.05; At an experiment-wise error rate of α=0.05, there is enough evidence to conclude that patients with acute leukemia have a larger number of GR sites per leukocyte cell than normal subjects.