assignment1

nihat uzunalioglu - 2660298, emiel kempen - 2640580, saurabh jain - 2666959

2/26/2020

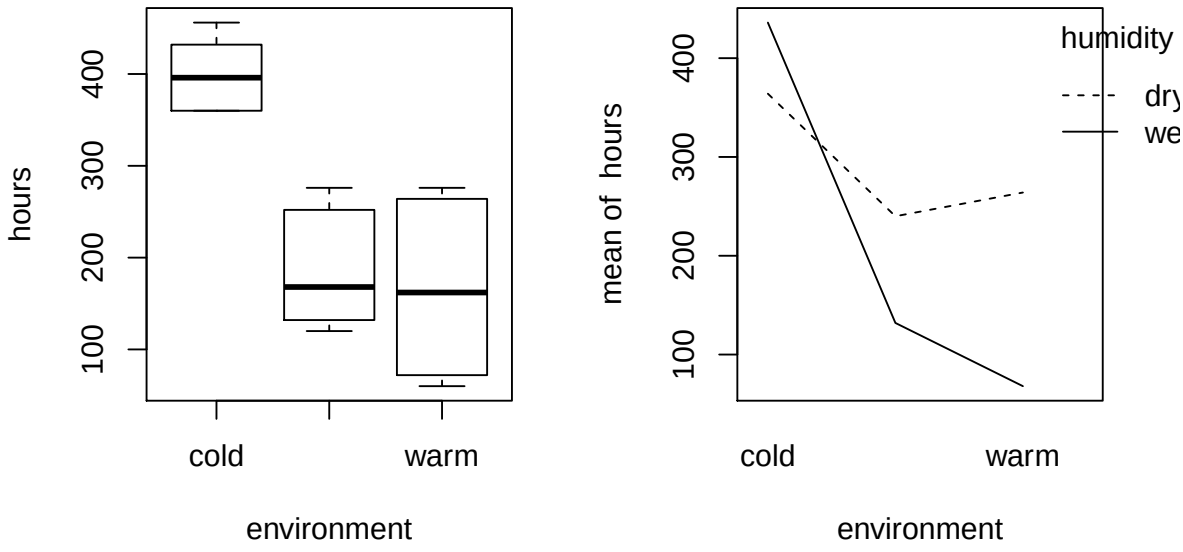
Exercise 1

a) The 18 slices came from a single loaf, but were randomized to the 6 combinations of conditions. Present an R-code for this randomization process.

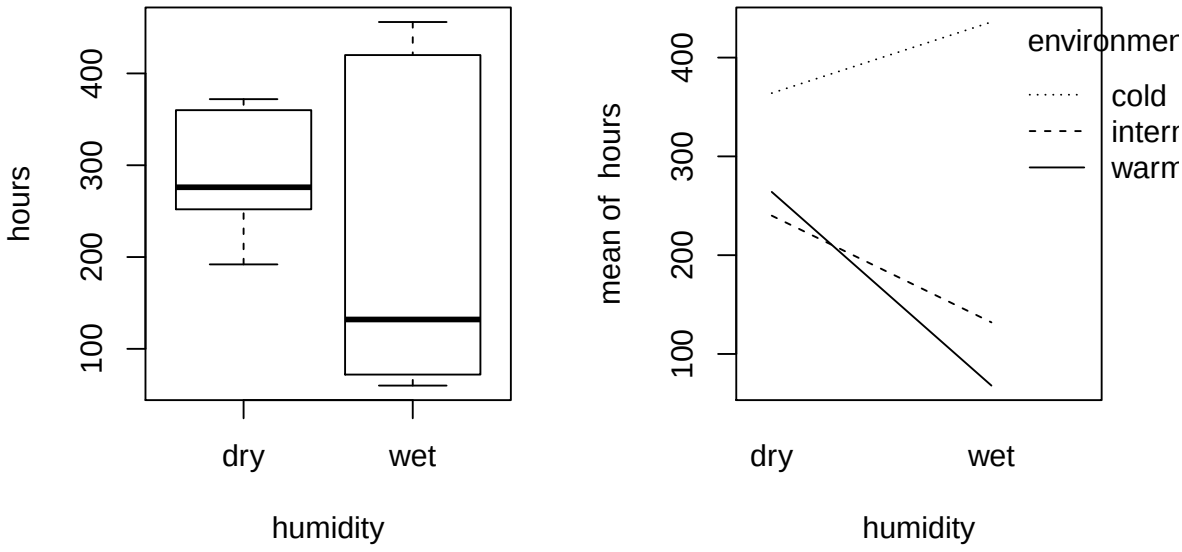
```
# The randomization process for 18 slices
# Take hours column from the data
hrs = as.vector(as.matrix(bread$hours))
# Create environment column
env = rep(c('cold', 'intermediate', 'warm'), each = 6)
# Create humidity column
humi = rep(c('dry', 'wet'), each = 3)
# Converting to data frame
head(data.frame(cbind(hrs, env, humi)))
```

```
##   hrs  env humi
## 1 360 cold  dry
## 2 360 cold  dry
## 3 372 cold  dry
## 4 420 cold  wet
## 5 456 cold  wet
## 6 432 cold  wet
```

b) Make two boxplots of hours versus the two factors and two interaction plots (keeping the two factors



fixed in turn).



c) Perform an analysis of variance to test for effect of the factors temperature, humidity, and their interaction. Describe the interaction effect in words.

```
# Creating linear model and ANOVA test
breadaov = lm(hours~environment*humidity, data = bread); anova(breadaov)
```

```
## Analysis of Variance Table
```

```
##
## Response: hours
##                       Df Sum Sq Mean Sq F value    Pr(>F)
## environment            2 201904  100952 233.685 2.461e-10
## humidity               1  26912   26912  62.296 4.316e-06
## environment:humidity   2  55984   27992  64.796 3.705e-07
## Residuals             12   5184     432
```

```
p_interaction = anova(breadaov)$Pr[3]
```

- The p-value for testing for H_0:$\gamma_{i,j} = 0$ for all i, j is 3.7054783×10^{-7}. Therefore, we reject the null hypothesis H_0 which means the interaction between environment and humidity is significant for this dataset.

d) Which of the two factors has the greatest (numerical) influence on the decay? Is this a good question?

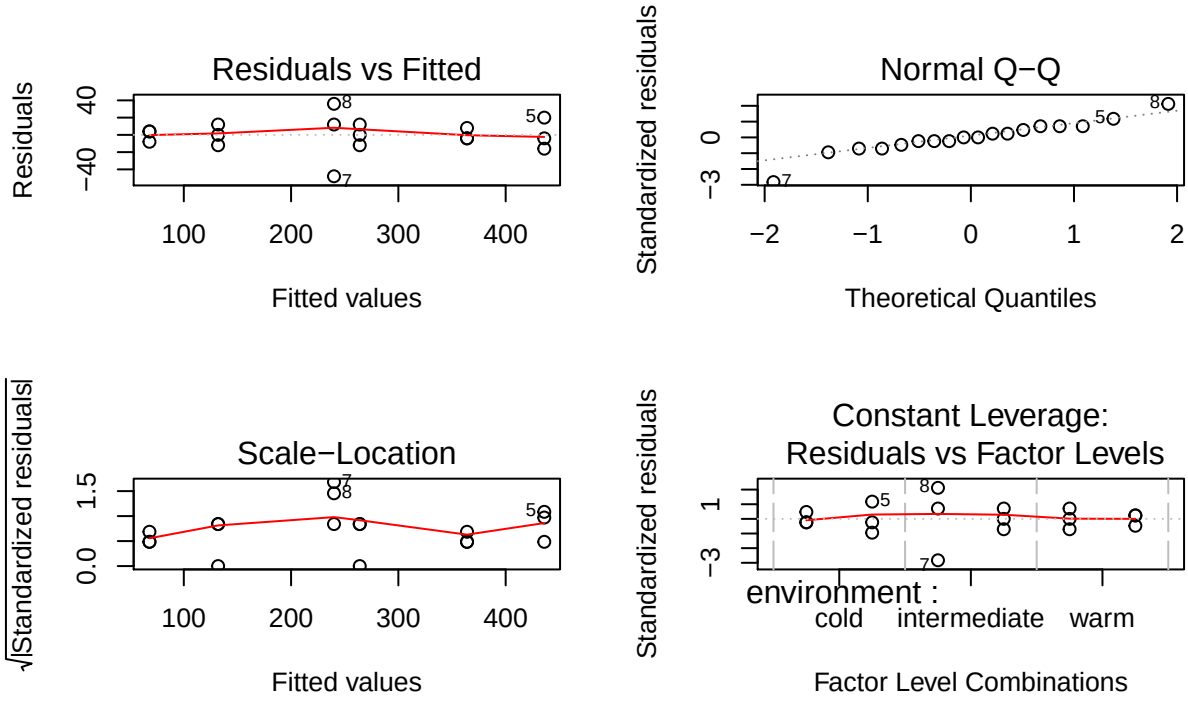
```
summary(breadaov)[4]
```

```
## $coefficients
##                                      Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)                               364   12.00000  30.333333 1.032769e-12
## environmentintermediate                  -124   16.97056  -7.306770 9.389760e-06
## environmentwarm                          -100   16.97056  -5.892557 7.336887e-05
## humiditywet                                72   16.97056   4.242641 1.142103e-03
## environmentintermediate:humiditywet      -180   24.00000  -7.500000 7.233671e-06
## environmentwarm:humiditywet              -268   24.00000 -11.166667 1.073751e-07
```

- When we look up to the variance analysis results, 192, 2, 1 which corresponds to an environment with intermediate and dry has the most decaying effect in the dataset.

e) Check the model assumptions by using relevant diagnostic tools. Are there any outliers?

```
par(mfrow=c(2, 2))
# Plot the linear fitted model graphs
plot(breadaov)
```



- According to the tables we can say that 192, 2, 1 and 276, 2, 1 are the two that can be considered as outliers.

Exercise 2

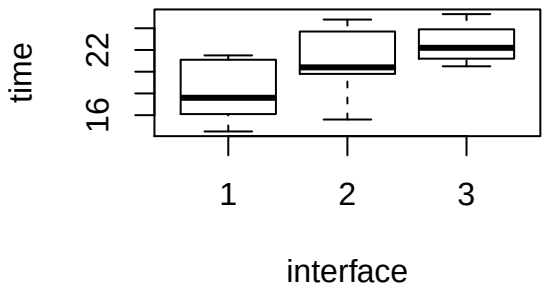
a) Number the selected students 1 to 15 and show how (by using R) the students could be randomized to the interfaces in a randomized block design.

```
# N <- as.vector(as.matrix(search_data$time))
# I <- as.vector(as.matrix(search_data$skill))
# B <- as.vector(as.matrix(search_data$interface))
# for (i in 1:B) print(sample(1:(N*I)))
xtabs(time~interface+skill,data=search_data)
```

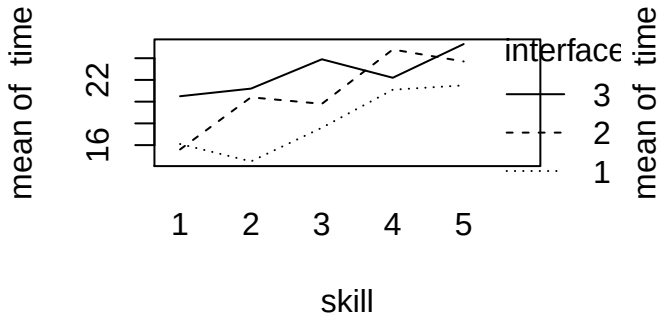
```
##           skill
## interface    1    2    3    4    5
##         1 16.1 14.5 17.6 21.1 21.5
##         2 15.6 20.4 19.8 24.8 23.7
##         3 20.5 21.2 23.9 22.2 25.3
```

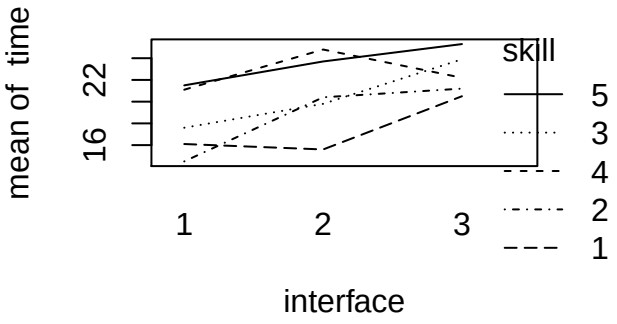
b) Make some graphical summaries of the data. Are any interactions between interface and skill apparent?

```
# interaction.plot(search_data$interface, search_data$skill, search_data$time)
# boxplot(search_data$interface, search_data$skill, search_data$time)
attach(search_data)
par(mfrow=c(1,2))
boxplot(time~interface)
boxplot(time~skill)
```




```
par(mfrow=c(1,2))
interaction.plot(skill,interface,time)
interaction.plot(interface,skill,time)
```





c) Test the null hypothesis that the search time is the same for all interfaces. Estimate the time it takes a typical user of skill level 3 to find the product on the website if the website uses interface 2.

d) Check the model assumptions by using relevant diagnostic tools.

```
aovsearch=lm(time~interface+skill)
anova(aovsearch)
```

```
## Analysis of Variance Table
##
## Response: time
##           Df Sum Sq Mean Sq F value    Pr(>F)
## interface  1 49.729  49.729  21.422 0.0005817
## skill      1 78.732  78.732  33.916 8.165e-05
## Residuals 12 27.856   2.321
```

```
summary(aovsearch)
```

```
##
## Call:
## lm(formula = time ~ interface + skill)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.19667 -0.73167 -0.05667  1.07333  2.63333
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.2267     1.3341   8.415 2.23e-06
## interface     2.2300     0.4818   4.628 0.000582
## skill         1.6200     0.2782   5.824 8.16e-05
##
## Residual standard error: 1.524 on 12 degrees of freedom
## Multiple R-squared:  0.8218, Adjusted R-squared:  0.7921
## F-statistic: 27.67 on 2 and 12 DF,  p-value: 3.203e-05
```

e) Perform the non-parametric Friedman test to test whether there is an effect of interface.

```
friedman.test(time,interface,skill)
```

```
##
##  Friedman rank sum test
##
## data:  time, interface and skill
## Friedman chi-squared = 6.4, df = 2, p-value = 0.04076
```

f) Test the null hypothesis that the search time is the same for all interfaces by a one-wayANOVA test, ignoring the variable skill. Is it right/wrong or useful/not useful to perform this test on this dataset? What assumption on the way the data were obtained is necessary for this test to be valid, and was this assumption met?

Exercise 3

a) Test whether the type of feedingstuffs innfluences milk production using an ordinary "fixed effects" model, fitted with `lm`. Estimate the difference in milk production.

b) Repeat a) and b) by performing a mixed effects analysis, modelling the cow effect as a random effect (use the function lmer). Compare your results to the results found by using a fixed effects model. (You will need to install the R-package lme4, which is not included in the standard distribution of R.)

c) Study the commands:

```
\> attach(cow)
\> t.test(milk[treatment=="A"],milk[treatment=="B"],paired=TRUE)
```

Does this produce a valid test for a difference in milk production? Is its conclusion compatible with the one obtained in a)? Why?