A2- Exercise1 and 2

geoffrey

February 29, 2020

Exercise 1

a. Randomized design

A randomized design with two categorical factors, with

1. the first factor having three categorical levels and
2. the second factor having two levels and
3. having three samples for each unique categorie

can be produced with the following R code:

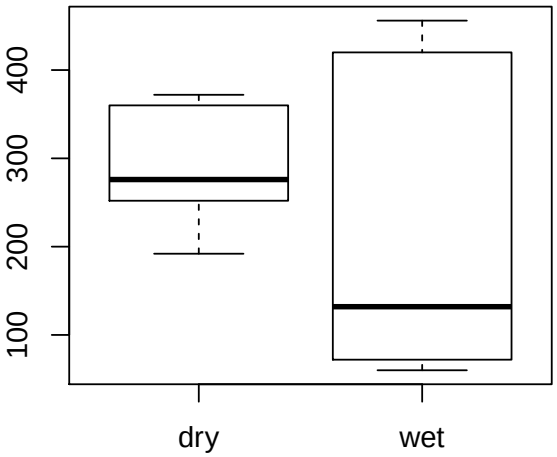
```
I=3; J=2; N=3
rbind(rep(1:I,each=N*J),rep(1:J,N*I),sample(1:(N*I*J)))
```

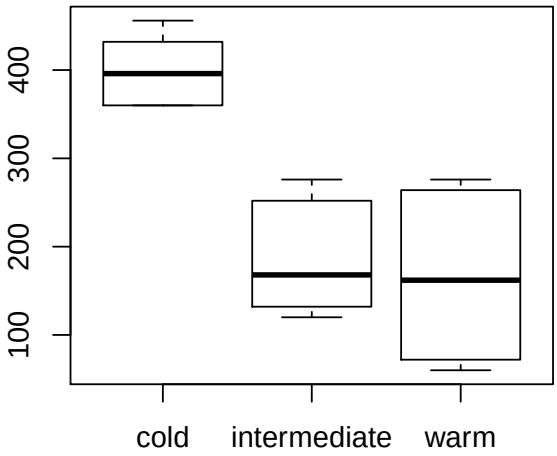
b. Plotting

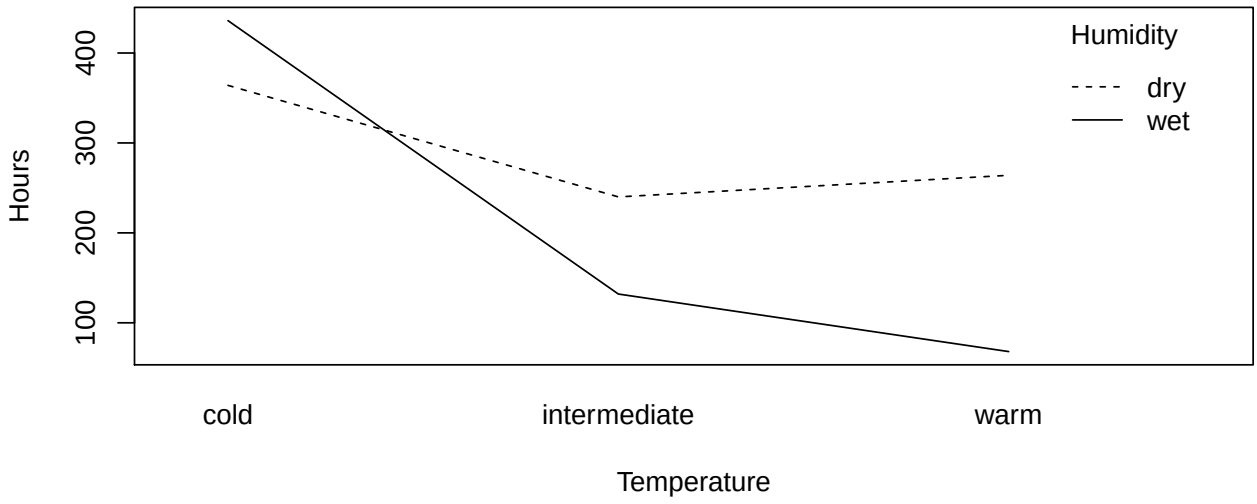
The boxplot and interaction plot below confirms our intuition:

1. A cold environment causes a much slower decay
2. Wet bread has a much wider distribution (variance)
3. On average dry bread decays slower than wet bread
4. However, wet and cold (frozen) bread has the slowest decay

From the non-parallel lines in the interaction plot and the wide distribution of the wet sample we conclude that the (wet) humidity amplifies the effect of the temperature and thus it can be explained by the strong interaction between the two factors (opposed to the errors in the measurement).







c. Two way ANOVA

From the two-way anova resulst below, we can conclude that both factors have a main effect on the decay time of bread. It can also be concluded that the factors have an interaction effect.

```
## Analysis of Variance Table
##
## Response: hours
##                       Df Sum Sq Mean Sq F value    Pr(>F)
## humidity               1  26912   26912  62.296 4.316e-06 ***
## environment            2 201904  100952 233.685 2.461e-10 ***
## humidity:environment   2  55984   27992  64.796 3.705e-07 ***
## Residuals             12   5184     432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

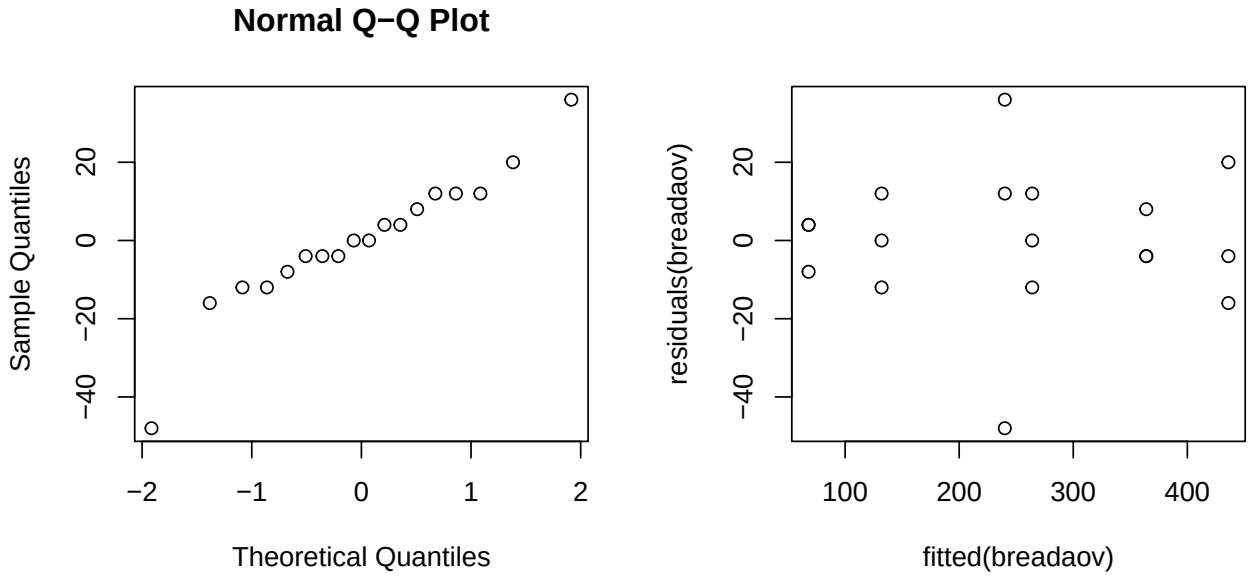
d. Coefficients

TODO

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

e. Diagnostics

The first requirements is that for each unique categorie, there should be atleast 2 samples, which is the case. Then the most important requirement is that the data among the factors should approximiatly have equal variances. This has been tested in b. and the conclusions was that they approximitally were the same. A different test we can do after the ANOVA test, is check whether the error is normally distributed, which is to be expected of a random variable. In the following QQplot it can be seen that the residuals are approximiatly normally distributed. And in the fitted residuals plot it can be seen that the spread is approxomitelly the same among the fitted values, however there are 2 outliers in the middle.



Exercise 2

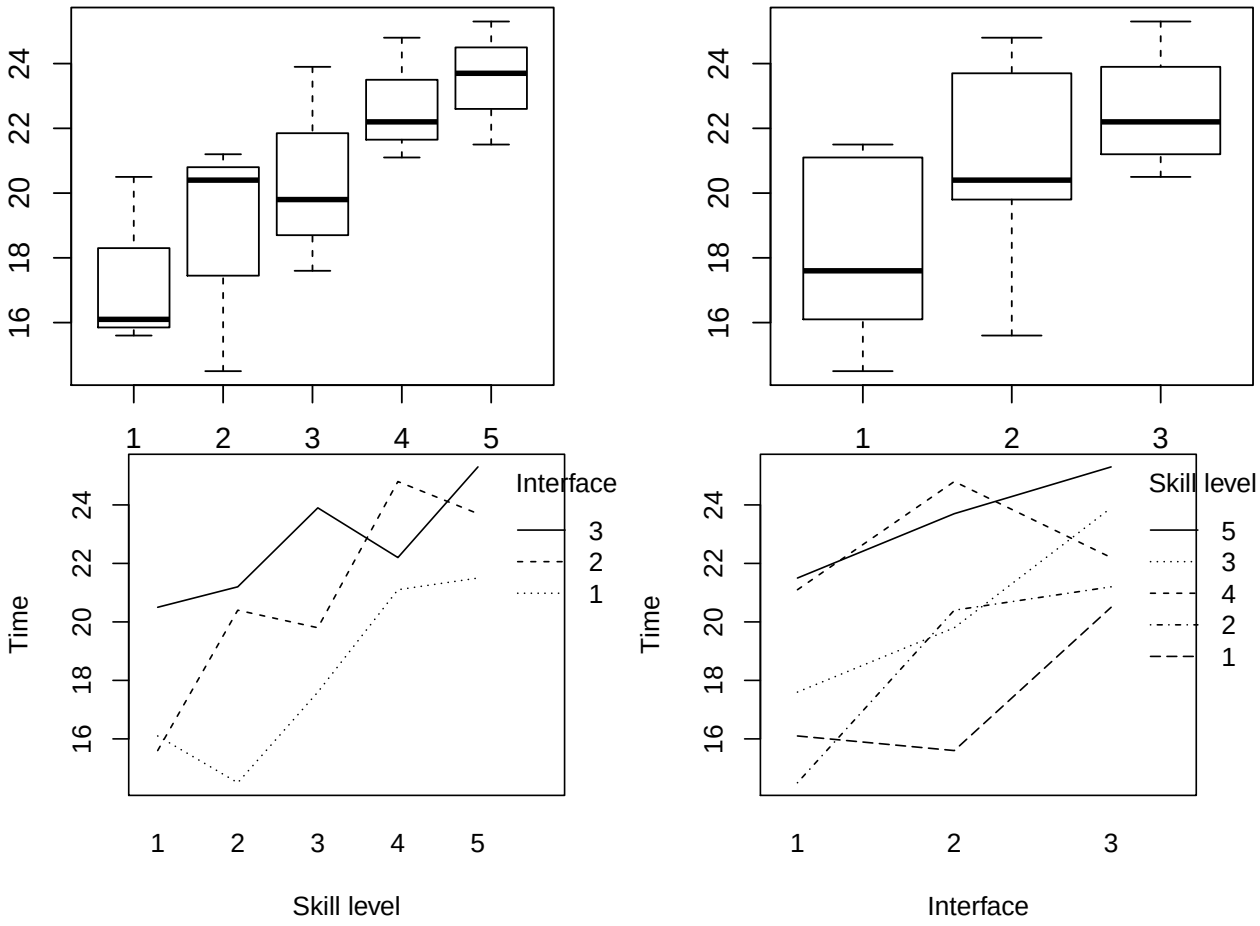
a. Randomized block design

The following code generates a random block design with five blocks, a factor with three levels, and one sample per unique categorie.

```
B=5;
if1 = sample(1:5)
if2 = sample(6:10)
if3 = sample(11:15)
for (i in 1:B) print(c(if1[i], if2[i], if3[i]))
```

b. Graphical pre-analysis

The boxplots below suggest that indeed the skil level and the interfaces matter for the search time. Where skill level 1 is indeed the fastest, and interface 1 is the fastest from the three interfaces. And from the interaction plots below it can be observed that overall the factors have the same pattern; all lines start in the lower left corner and end towwards the upper right corner. However, they are not perfectly parallel, this can be explained by the small sample sizes which causes local irregularities. Thus we conclude that there is no interaction between the two factors.



c. Anova with 1 block factor

From the ANOVA results below it can be concluded that the search time is not the same for all interfaces. Futhermore, we can estimate the time it takes for a user with skill level 3 to find a product using interface 2 by looking at the summary table and adding the coeffcients of these two categories to the incercept. Thus that would be 15.015+3.033+2.7=20.748.

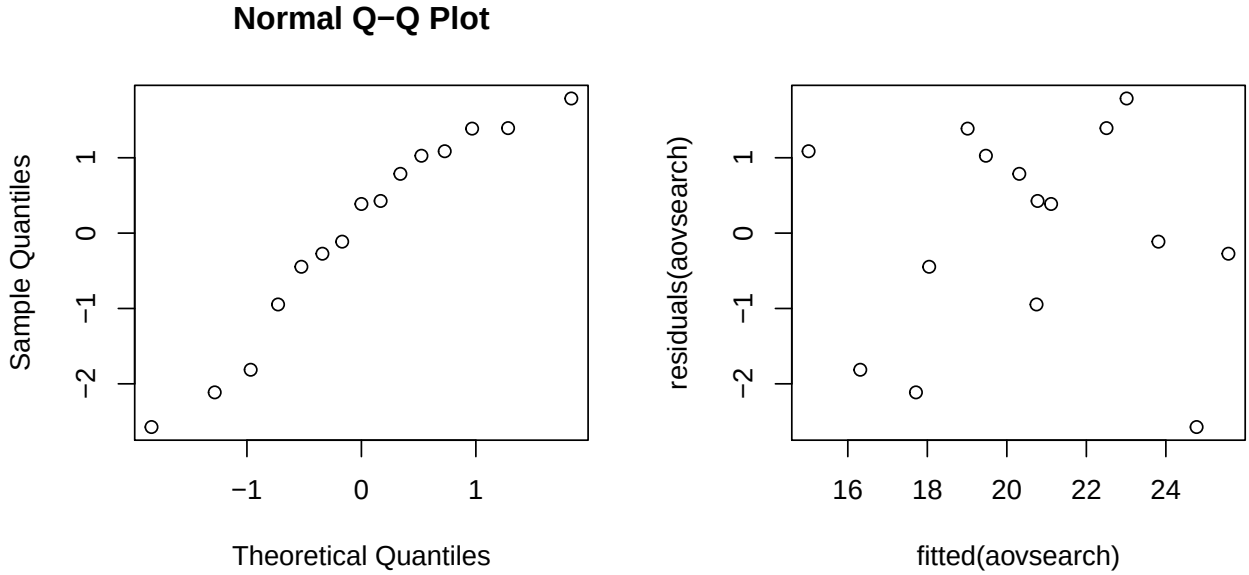
```
## Analysis of Variance Table
##
## Response: time
##           Df Sum Sq Mean Sq F value  Pr(>F)
## interface  2 50.465 25.2327  7.8237 0.01310 *
## skill      4 80.051 20.0127  6.2052 0.01421 *
## Residuals  8 25.801  3.2252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = time ~ interface + skill, data = search)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5733 -0.6967  0.3867  1.0567  1.7867
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   15.013      1.227  12.238 1.85e-06 ***
## interface2     2.700      1.136   2.377  0.04474 *
## interface3     4.460      1.136   3.927  0.00438 **
## skill2         1.300      1.466   0.887  0.40118
## skill3         3.033      1.466   2.069  0.07238 .
## skill4         5.300      1.466   3.614  0.00684 **
## skill5         6.100      1.466   4.160  0.00316 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.796 on 8 degrees of freedom
## Multiple R-squared:  0.8349, Adjusted R-squared:  0.7111
## F-statistic: 6.745 on 6 and 8 DF,  p-value: 0.008395
```

d. Diagnostics

The QQ-plot of the residuals below looks normally distributed, which is good. The fitted residuals do not depict any outliers.



e. Friedmann test

The result of the Friedman test is the same as the ANOVA: we reject the H0, thus there is a difference in sesarch times.

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

f. One-way ANOVA

The one-way ANOVA returns no significant difference in the search time between the interfaces. This result is not very usefull, because 1) we removed a lot of information from the model and 2) the model now assumes that the block is a random selection of all available blocks, which is not the case because the blocks were fixed/predetermined.

```
## Analysis of Variance Table
##
## Response: time
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## interface  2  50.465  25.233  2.8605 0.09642 .
## Residuals 12 105.852   8.821
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```