Assignment 1

Gijs Smeets, Daan Wijnhorst, Moos Middelkoop, group 35

February 2023

EXERCISE 4A

To randomize the distribution of soil additives, first we replace all values in the data frame for zeroes.

```
my.data = npk;
my.data$yield = NULL
my.vec = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
my.data$N = my.vec
my.data$P = my.vec
my.data$K = my.vec
```

Then, we randomly sample two integers per additive from a discrete uniform distribution, which correspond to given plots in the data frame where the additives will be inserted.

```
for (i in 1:6) {
  nitr = sample.int(4,2); nitr
  phoss = sample.int(4,2); phoss
  k = sample.int(4,2); k

  nitr = nitr + (4 * (i-1)); nitr
  phoss = phoss + (4 * (i-1)); phoss
  k = k + (4 * (i-1));k
  my.data$N[nitr] = 1; my.data$N[nitr]
  my.data$P[phoss] = 1;
  my.data$K[k] = 1;

}
# randomized soil additives per block
my.data
```

```
##    block N P K
## 1      1 1 0 1
## 2      1 0 1 1
## 3      1 1 0 0
```

```
## 4       1 0 1 0
## 5       2 0 1 0
## 6       2 1 1 0
## 7       2 0 0 1
## 8       2 1 0 1
## 9       3 1 0 0
## 10      3 0 1 0
## 11      3 1 0 1
## 12      3 0 1 1
## 13      4 0 1 1
## 14      4 1 0 0
## 15      4 1 1 1
## 16      4 0 0 0
## 17      5 1 0 1
## 18      5 0 0 0
## 19      5 1 1 0
## 20      5 0 1 1
## 21      6 1 1 0
## 22      6 0 1 1
## 23      6 0 0 1
## 24      6 1 0 0
```

This will result in 2 of each soil additive per block of 4 plots.

EXERCISE 4B

```
nitr = numeric(6)
nonitr = numeric(6)
for (i in 1:6){

  for (j in 0:1){
    cat("block", i, "N", j)
    mean = mean(npk[npk$block == i & npk$N == j,]$yield)
    print(mean)

    if (j == 0){
      nonitr[i] = mean
    }
    else{
      nitr[i] = mean
    }


  }
}
```

```
## block 1 N 0[1] 48.15
## block 1 N 1[1] 59.9
## block 2 N 0[1] 55.75
## block 2 N 1[1] 59.15
## block 3 N 0[1] 58.9
## block 3 N 1[1] 62.65
## block 4 N 0[1] 44.85
## block 4 N 1[1] 55.4
## block 5 N 0[1] 50.15
## block 5 N 1[1] 50.9
## block 6 N 0[1] 54.6
## block 6 N 1[1] 58.1
```

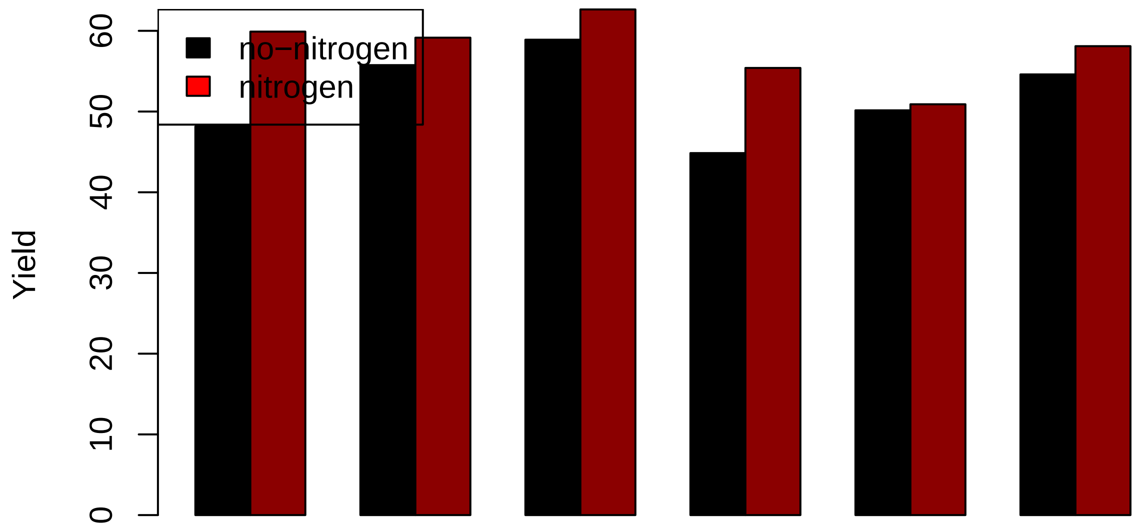
```r
nitr
```

```
## [1] 59.90 59.15 62.65 55.40 50.90 58.10
```

```r
nonitr
```

```
## [1] 48.15 55.75 58.90 44.85 50.15 54.60
```

```r
blocks = seq(1:6)
test = rbind(nonitr,nitr)
barplot(test,
        beside=T,
        xlab = "No-nitrogen vs Nitrogen",
        ylab = "Yield",
        col = c("black","darkred")
)
legend("topleft", c("no-nitrogen", "nitrogen"), fill = c("black","red"))
```



No−nitrogen vs Nitrogen

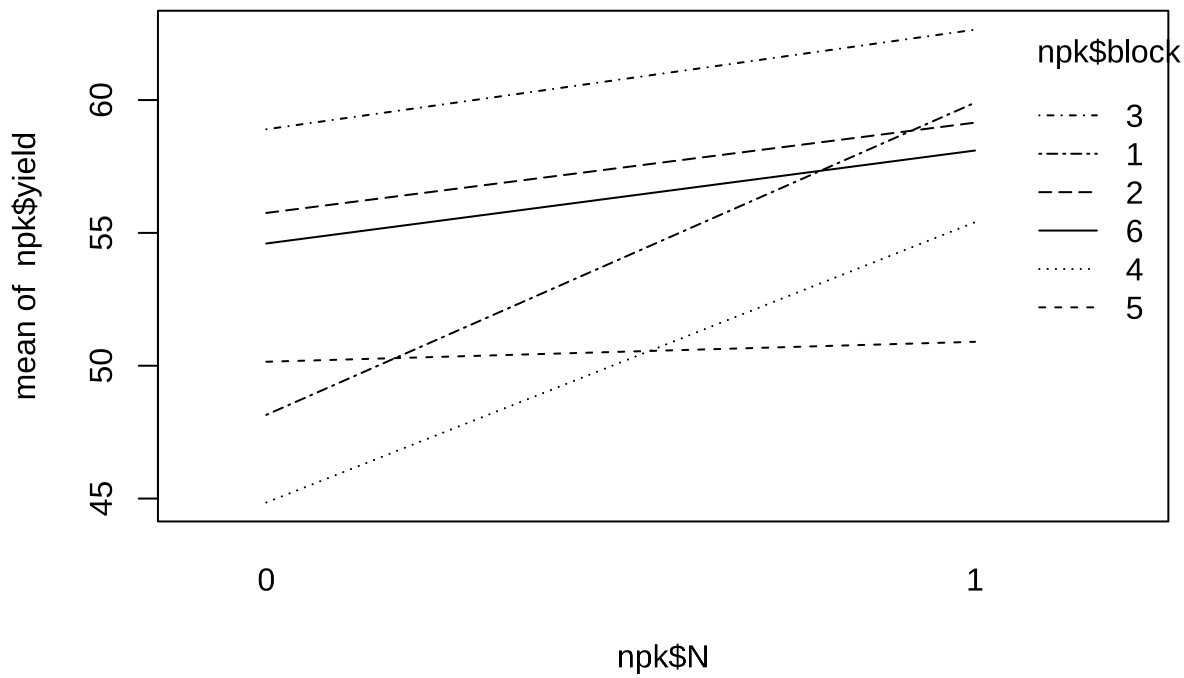
We want to understand the dependence of average yield on treatment factor ‘nitrogen’. We know the blocks to be balanced within each block and different in the sense that it is not necessarily the case that every plot receiving treatment is the same plot in each block. Taking the factor block into account can lead to removing variation and thus drawing more precise conclusions becomes possible. ## EXERCISE 4C We conduct the two-way ANOVA with two factors ‘block’ and ‘N’ and response variable ‘yield’ where we test the three hypotheses: 1. There is no main effect of factor ‘block’. 2. There is no main effect in factor ‘N’ and 3. There is no interaction between the two factors ‘N’ and ‘block’ as followed:

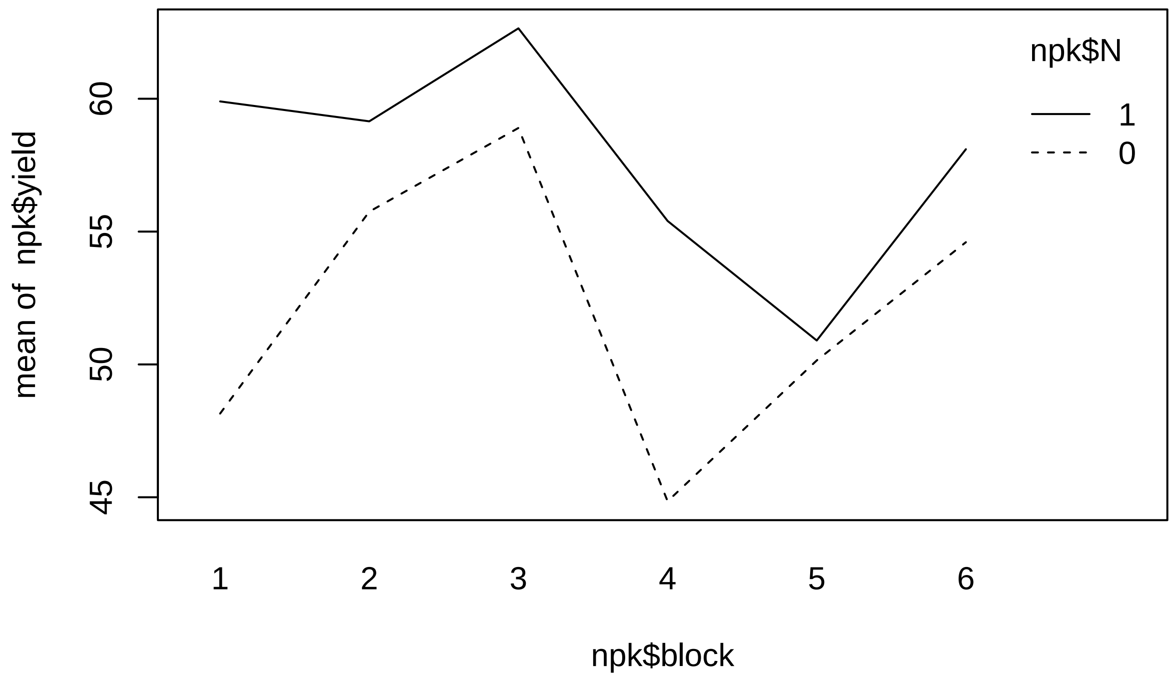
```
npk$block = as.factor(npk$block)
npk$N = as.factor(npk$N)
df = npk[c("yield","block","N")]
# two way anova
twoway = lm(yield~block*N,data=df)
anova(twoway)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value  Pr(>F)
## block      5 343.29  68.659  3.3592 0.03967 *
## N          1 189.28 189.282  9.2607 0.01021 *
## block:N    5  98.52  19.704  0.9640 0.47690
```

```
## Residuals 12 245.27  20.439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# not sign. interaction parameter, also interaction plots as evidence for gamma_ij = 0 (parall
interaction.plot(npk$N,npk$block,npk$yield); interaction.plot(npk$block,npk$N,npk$yield)
```





As seen in the results, we do not have enough evidence to reject hypothesis 3. as we did not obtain a P-value lower than the significance level. Therefore we do not have enough evidence to conclude there to be significant interaction between the two factors and we should use the additive model to see about the interaction effect.

```r
# -> compute anova on additive model
twoway2  = lm(yield~block+N,data=df)

# we obtain sign. values for block and N.
anova(twoway2)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## block      5 343.29  68.659  3.3951 0.026173 *
## N          1 189.28 189.282  9.3598 0.007095 **
## Residuals 17 343.79  20.223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

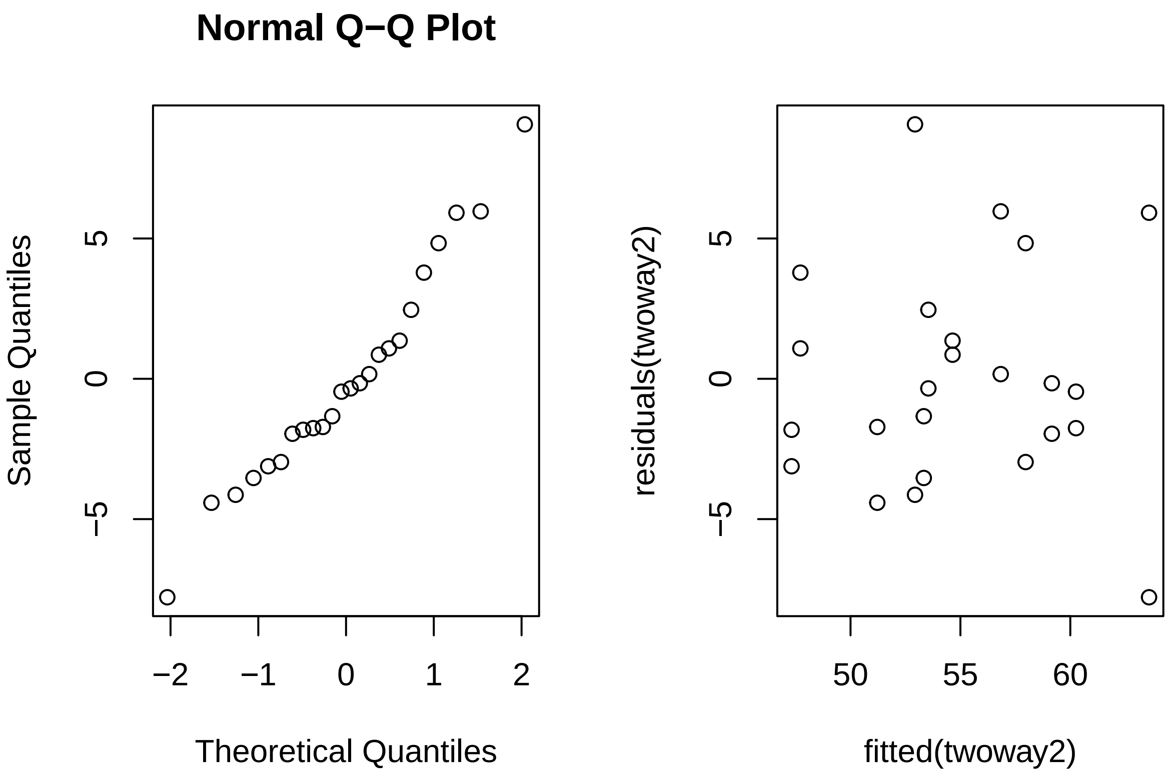
```r
summary(twoway2)
```

```
##
## Call:
## lm(formula = yield ~ block + N, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.783 -2.210 -0.400  1.633  9.067
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   51.217      2.429  21.089 1.26e-13 ***
## block2         3.425      3.180   1.077   0.2965
## block3         6.750      3.180   2.123   0.0488 *
## block4        -3.900      3.180  -1.226   0.2367
## block5        -3.500      3.180  -1.101   0.2864
## block6         2.325      3.180   0.731   0.4746
## N1             5.617      1.836   3.059   0.0071 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.497 on 17 degrees of freedom
## Multiple R-squared:  0.6077, Adjusted R-squared:  0.4693
## F-statistic: 4.389 on 6 and 17 DF,  p-value: 0.007404
```

```r
confint(twoway2)
```

```
##                   2.5 %    97.5 %
## (Intercept)  46.09266809 56.340665
## block2       -3.28388895 10.133889
## block3        0.04111105 13.458889
## block4      -10.60888895  2.808889
## block5      -10.20888895  3.208889
## block6       -4.38388895  9.033889
## N1            1.74328782  9.490046
```

```r
# seems normally distributed residuals
par(mfrow=c(1,2)); qqnorm(residuals(twoway2))
plot(fitted(twoway2),residuals(twoway2))
```



From the additive model we see a main-effect in factor factors considered, therefore we reject hypotheses 1. and 2. and we have enough evidence to conclude that there exists a main effect for both factors, but no interaction between the factors. It was sensible to include the factor 'block', as we could see that there exists a main effect in

was it sensible to also include 'block' as factor in this model? –> comment

yes, because significant (?)

or no, because it changes nothing (except for location?) -> as N(PK) is distr. differently across every block and

there is no way of describing the difference of how this distr. is affecting the yield

no friedman test! we need 24 blocks for this, to differentiate. see slide 37 of lec. 5 for explanation of this

#friedman.test(npk*yield*, *npk*N,npk$block)

```
## EXERCISE 4D
```

```
‘‘‘r
npk
```

```
##    block N P K yield
## 1      1 0 1 1  49.5
## 2      1 1 1 0  62.8
## 3      1 0 0 0  46.8
## 4      1 1 0 1  57.0
## 5      2 1 0 0  59.8
## 6      2 1 1 1  58.5
## 7      2 0 0 1  55.5
## 8      2 0 1 0  56.0
## 9      3 0 1 0  62.8
## 10     3 1 1 1  55.8
## 11     3 1 0 0  69.5
## 12     3 0 0 1  55.0
## 13     4 1 0 0  62.0
## 14     4 1 1 1  48.8
## 15     4 0 0 1  45.5
## 16     4 0 1 0  44.2
## 17     5 1 1 0  52.0
## 18     5 0 0 0  51.5
## 19     5 1 0 1  49.8
## 20     5 0 1 1  48.8
## 21     6 1 0 1  57.2
## 22     6 1 1 0  59.0
```

```
## 23     6 0 1 1  53.2
## 24     6 0 0 0  56.0
```

```
pairwise_N = lm(yield ~ N*block + P + K,data=npk)
pairwise_P = lm(yield ~ P*block + N + K,data=npk)
pairwise_K = lm(yield ~ K*block + P + N,data=npk)

anova(pairwise_N)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## N          1 189.28 189.282 13.3611 0.004423 **
## block      5 343.29  68.659  4.8465 0.016439 *
## P          1   8.40   8.402  0.5931 0.459045
## K          1  95.20  95.202  6.7201 0.026843 *
## N:block    5  98.52  19.704  1.3908 0.306583
## Residuals 10 141.67  14.167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(pairwise_P)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## P          1   8.40   8.402  0.4978 0.496588
## block      5 343.29  68.659  4.0678 0.028234 *
## N          1 189.28 189.282 11.2143 0.007381 **
## K          1  95.20  95.202  5.6404 0.038947 *
## P:block    5  71.40  14.280  0.8460 0.547341
## Residuals 10 168.79  16.879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(pairwise_K)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## K          1  95.20  95.202  5.6028 0.039477 *
## block      5 343.29  68.659  4.0407 0.028799 *
```

```
## P           1   8.40   8.402  0.4945 0.497989
## N           1 189.28 189.282 11.1397 0.007521 **
## K:block     5  70.27  14.054  0.8271 0.558263
## Residuals 10 169.92  16.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
additive = lm(yield ~ N+P+K+block, data=npk)
anova(additive)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value  Pr(>F)
## N          1 189.28 189.282 11.8210 0.00366 **
## P          1   8.40   8.402  0.5247 0.47999
## K          1  95.20  95.202  5.9455 0.02767 *
## block      5 343.29  68.659  4.2879 0.01272 *
## Residuals 15 240.19  16.012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

make 3 models -> two way anova

lm(yield ~ N*block + P + K), etc...

favorite model is without pairwise interaction term (so no block), just additive

EXERCISE 4E

To test if nitrogen has an effect on

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
mer1 = lmer(yield ~ N+P+K+(1|block),data=npk,REML=FALSE)
mer2 = lmer(yield ~ P+K+(1|block),data=npk,REML=FALSE)
anova(mer1,mer2)
```

```
## Data: npk
## Models:
## mer2: yield ~ P + K + (1 | block)
```

```
## mer1: yield ~ N + P + K + (1 | block)
##      npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## mer2    5 159.49 165.38 -74.745   149.49
## mer1    6 151.03 158.10 -69.514   139.03 10.46  1    0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

we see that anova of two models results in sign. p value. s.t. N has a significant impact on the yield.

compare to 4C!