

## 02441 Applied Statistics and Statistical Software

### Exercise 3A - Kali

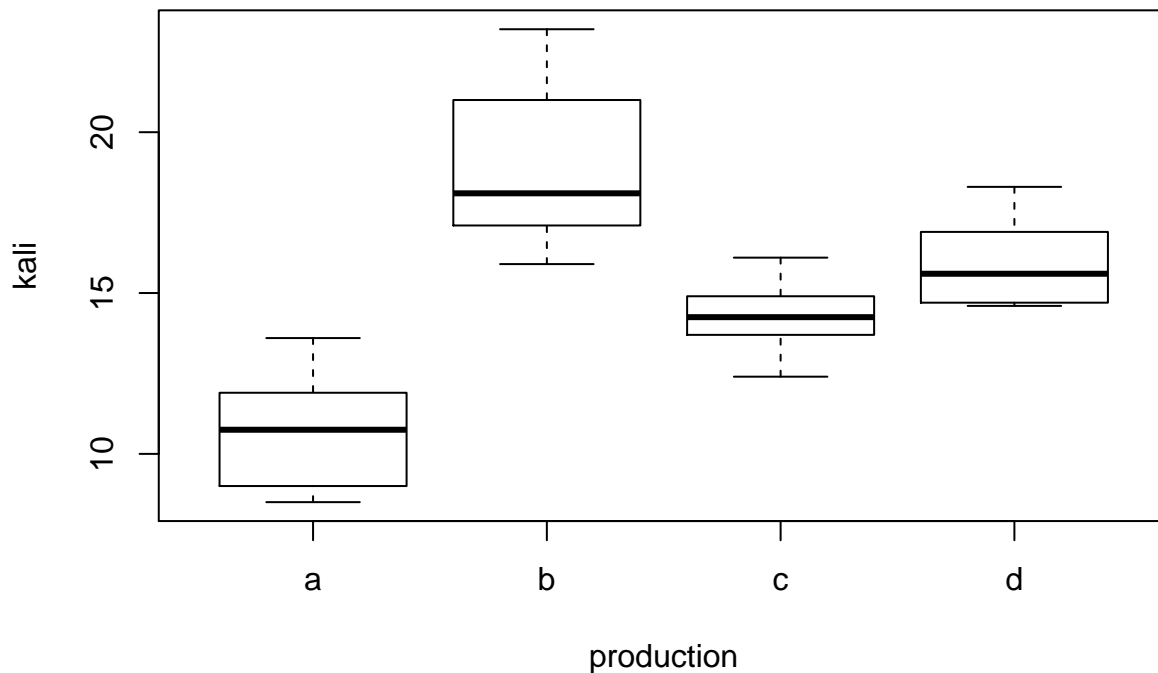
The dataset kali contains measurements of the content of kali (K<sub>2</sub>O) for four different productions

Variable name	Description
production	production unit
kali	content of kali

#### 1. Use a non-parametric test to examine if the content of kali depends on the different productions

Start by loading and plotting the data

```
ka <- read.table("kali.txt", header = TRUE)
plot(kali~production, data = ka)
```



Use Kruskal Wallis test

```
kruskal.test(ka$kali, ka$production)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  ka$kali and ka$production
## Kruskal-Wallis chi-squared = 17.853, df = 3, p-value = 0.0004717
```

2. Use a one-way ANOVA to examine if the content of kali depends on the different productions

```
lm1 <- lm(kali~production, data = ka)
anova(lm1)

## Analysis of Variance Table
##
## Response: kali
##           Df Sum Sq Mean Sq F value    Pr(>F)
## production  3 207.978   69.326   17.815 9.485e-06 ***
## Residuals  19  73.936    3.891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. If the content of kali depends on the different productions, which of the production(s) yield the highest content?

```
library(kableExtra)
kable(tapply(ka$kali, ka$production, mean), col.names = 'kali production mean')
```

	kali production mean
a	10.75000
b	18.90000
c	14.26667
d	16.02000