CUE\_04

Klemens Bayr

# Ausganssituation

In einem Rechenzentrum stehen für einen größeren Serverraum mehrere Messdaten (Stichprobe) für einen bestimmten Zeitraum zur Verfügung. Jeder Eintrag steht für einen Server zu einem bestimmten Zeitpunkt und beinhaltet: • Servertemperatur (Grad Celsius) • Prozessorauslastung (in %) • Plattenauslastung (in %) • Lüfterstatus (“original”, “getauscht”) Analysieren Sie zunächst die Servertemperaturen und schätzen Sie die Situation in der Grundgesamtheit ab. Gibt es einen Unterschied zwischen den Servern mit alten und jenen mit getauschten Lüftern? Gibt es Zusammenhänge mit den anderen Messwerten (Prozessor- und Plattenauslastung)?

# Datenmanagement

daten = "/Users/klemensbayr/Dropbox/FH Technikum/Datenanalyse/cue4/server.csv"  
  
tabl = read.table(daten, header = TRUE, sep=" ", dec = "." , na.strings = "XXX", fileEncoding = "utf-8")  
head(tabl)

## temperatur prozessor platten Lüfter  
## 1 34.5 50.1 46.8 getauscht  
## 2 20.0 27.7 68.6 original  
## 3 63.2 58.2 44.6 getauscht  
## 4 NA 17.4 41.6 getauscht  
## 5 63.4 53.3 36.5 getauscht  
## 6 44.7 33.1 60.2 getauscht

summary(tabl)

## temperatur prozessor platten Lüfter   
## Min. : 20.00 Min. : 5.90 Min. :23.20 getauscht:50   
## 1st Qu.: 37.10 1st Qu.: 39.42 1st Qu.:41.17 original :50   
## Median : 59.20 Median : 52.35 Median :49.85   
## Mean : 57.62 Mean : 51.98 Mean :49.41   
## 3rd Qu.: 73.20 3rd Qu.: 62.62 3rd Qu.:57.85   
## Max. :126.40 Max. :100.00 Max. :76.40   
## NA's :1

Die Anzahl der Beobachtungen

beo = dim(tabl)  
beo

## [1] 100 4

nabeo = sum(is.na(tabl))  
nabeo

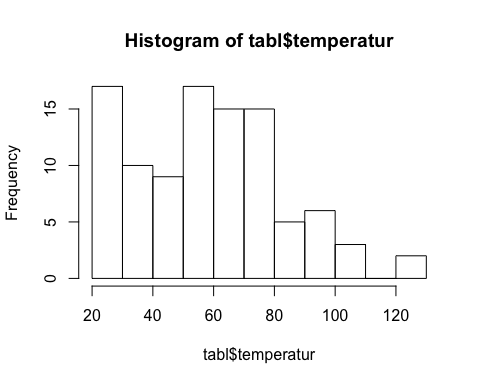
## [1] 1

In der Stichprobe finden sich insgesamt 100 Werte, die Aussagen über Servertemperatur (Grad Celsius), Prozessorauslastung (in %), Plattenauslastung (in %) und Lüfterstatus treffen. Also zählen wir gesamt 4 metrische Variablen.

In der Gesamt befindet sich ein fehlender Wert.

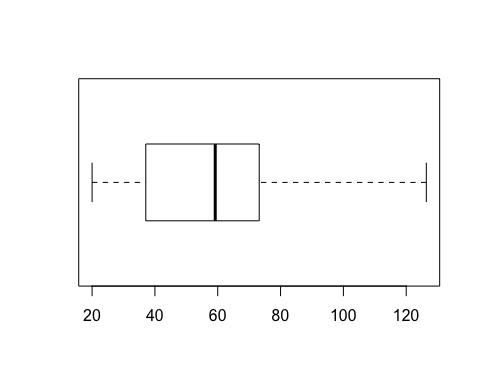
# Deskriptive Beurteilung

hist.default(tabl$temperatur)



Die Verteilung ist eher rechtsschief zu beurteilen.

boxplot(tabl$temperatur, horizontal = TRUE)



Beim Boxplot wird klar ersichtlich, dass die meisten in der Stichprobe enthaltenen Server Temperaturen unter 80°C aufweisen und hohe Temperaturen einen eher geringen Anteil ausmachen.

summary(tabl$temperatur)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 20.00 37.10 59.20 57.62 73.20 126.40 1

sd(tabl$temperatur)

## [1] NA

mad(tabl$temperatur)

## [1] NA

Die Temperaturwerte der Server liegen zwischen 20°C und 126.40°C, das arithmeitsche Mittel ist bei 57.62°C. Der Median liegt hingegen bei 59.2°C. Mittelwert und Standardabweichung sind durch die schiefe Verteilung verzerrt.

# Rückschluss auf die Grundgesamtheit

Ein 95%-Konfidenzintervall für die Temperaturen kann mittels wilcox.test() ermittelt werden:

wilcox.test(tabl$temperatur, conf.int = TRUE)$conf.int

## [1] 51.75003 62.50000  
## attr(,"conf.level")  
## [1] 0.95

Das Konfidenzintervall für die Temperaturen beträgt laut wilcox.test [51,8; 62,5]. Anhand dieser Ergebnisse kann man behaupten, dass die Temperaturen der Server etwa im Bereich 51,8°C und 62,5 °C liegen.

# Zusammenhänge

Alle Analysen der Zusammenhänge werden mittels Streudiagramm dargestellt.

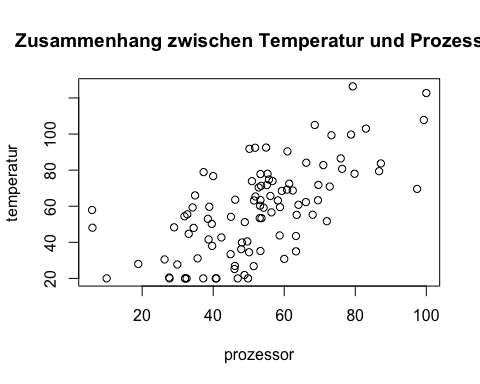
## Zusammenhang zwischen Temperatur und Status des Lüfters

Bevor die Daten analyisert werden können, müssen erst die Daten des Lüfters in seperiert werden.

alt = subset(tabl, Lüfter == "original")  
neu = subset(tabl, Lüfter == "getauscht")

## Zusammenhang zwischen Temperatur und Prozessorauslastung

plot(temperatur~prozessor, data = tabl, main = "Zusammenhang zwischen Temperatur und Prozessor")



Die Struktur dieses Streudiagrammes lässt darauf schließen, dass eine relativ gute Korrelation vermutet wird.

Korrelation nach Pearson:

round(cor(tabl$temperatur,tabl$prozessor, use = "pairwise.complete.obs"),2)

## [1] 0.65

Korrelation nach Spearman

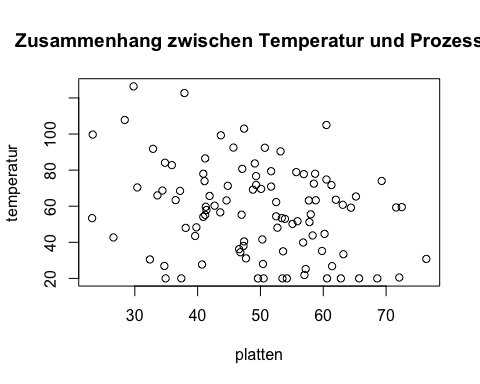
round(cor(tabl$temperatur,tabl$prozessor, use = "pairwise.complete.obs", method="spearman"),2)

## [1] 0.67

Es gibt eine eindeutige Korrealtion zwischen der Temperatur und der Prozessorauslastung.

## Zusammenhang zwischen Temperatur und Plattenauslastung

plot(temperatur~platten, data = tabl, main = "Zusammenhang zwischen Temperatur und Prozessor")



Korrelation nach Pearson:

round(cor(tabl$temperatur,tabl$platten, use = "pairwise.complete.obs"),2)

## [1] -0.29

Korrelation nach Spearman

round(cor(tabl$temperatur,tabl$platten, use = "pairwise.complete.obs", method="spearman"),2)

## [1] -0.24

Es lässt sich keine eindeutige Korrealtion zwischen der Temperatur und der Plattenauslastung der Server feststellen.

# Fazit

Generell weisen die Daten auf Zusammenhänge zwischen der Servertemperatur und der Prozessorauslastung hin.

Deutlich zu erkennen ist, das zwischen Temperatur und Plattenauslastung kein Zusammenhang besteht.

Die Temperatur liegt weiters im Intervall 51,8°C und 62,5°C