

```

---
title: "Kausalanalyse Resistenz"
author: ''
date: "15.03.2022"
output:
  pdf_document:
    keep_tex: yes
  word_document: default
geometry: margin=0.5cm
---

# Bibliotheken laden, Hilfsfunktion

```{r, warning=F, error=F, message=F}
library(stringr) # String-verarbeitung
library(ggplot2) # moderne plots
library(gridExtra)

debug <- T # debug printout
debug <- F # kein debug printout
Log <- function(string) {
 if(debug){print(string)}
}
```

# Resistenzen.Rmd erzeugte Resistenzen.csv, dieses einlesen
Und evtl. ansehen

```{r}
Resistenzen <- read.csv("Resistenzen.csv")

csv schreiben fügt vorne Index-Spalte an; diese entfernen :
Resistenzen[,1] <- NULL

View(Resistenzen)
```

## Verteilungen

```{r}
ResistenzenWM1 <- Resistenzen[Resistenzen["WM.group"] == "1",] # waste
milk Group
ResistenzenWM2 <- Resistenzen[Resistenzen["WM.group"] == "2",] # no waste
milk Group
#View(ResistenzenWM2)

ResistenzenOLS0 <- Resistenzen[Resistenzen["OLS.group"] == "0",] # other
livestock Group
ResistenzenOLS1 <- Resistenzen[Resistenzen["OLS.group"] == "1",] # no other
livestock Group
#View(ResistenzenOLS0);View(ResistenzenOLS1)

ResistenzenIAC0 <- Resistenzen[Resistenzen["IAC.group"] == "0",] # ill
animals in calving box Group

```

```

ResistenzenIAC1 <- Resistenzen[Resistenzen["IAC.group"] == "1",] # no ill
animals in calving box Group
#View(ResistenzenIAC0);View(ResistenzenIAC1)

ResistenzenHSC0 <- Resistenzen[Resistenzen["HSC.group"] == "0",] # stable
w\o outlet
ResistenzenHSC1 <- Resistenzen[Resistenzen["HSC.group"] == "1",] # stable
with outlet
ResistenzenHSC2 <- Resistenzen[Resistenzen["HSC.group"] == "2",] # outdoors
ResistenzenHSC3 <- Resistenzen[Resistenzen["HSC.group"] == "3",] # 0+1
ResistenzenHSC4 <- Resistenzen[Resistenzen["HSC.group"] == "4",] # 1+2
ResistenzenHSC5 <- Resistenzen[Resistenzen["HSC.group"] == "5",] # 0+2
#View(ResistenzenHSC0);View(ResistenzenHSC1);View(ResistenzenHSC2);View(ResistenzenHSC3);
```

```

Graphiken und Deskriptive Analyse: Für diesen Fall analysieren wir die (meist links und/oder rechts abgeschnittenen) Verteilungen

```

```{r, warning=F, error=F, message=F}
graphisch <- function(indep, antib, anfang,ende, schrittBin,schrittLab) {

 if (ende < 0) { # kleiner Trick um zusätzliches
Funktionsargument zu vermeiden
 Ende=F
 ende = -ende
 } else{
 Ende=T
 }
 Log(paste("Ende, ende =",Ende,ende))

 if(indep == "WM.group"){
 listdfs <- list(Resistenzen , ResistenzenWM1 , ResistenzenWM2) #
Vektor klappt hier nicht!
 Titel <- c("WM oder nicht", "WM ", "keine WM ")
 }
 if(indep == "OLS.group"){
 listdfs <- list(Resistenzen , ResistenzenOLS1 , ResistenzenOLS0)
 Titel <- c("OLS oder nicht", "OLS ", "kein OLS ")
 }
 if(indep == "IAC.group"){
 listdfs <- list(Resistenzen , ResistenzenIAC1 , ResistenzenIAC0)
 Titel <- c("IAC oder nicht", "IAC ", "kein IAC ")
 }
 if(indep == "HSC.group"){
 listdfs <- list(Resistenzen , ResistenzenHSC0, ResistenzenHSC1,
ResistenzenHSC2, ResistenzenHSC3, ResistenzenHSC4,
ResistenzenHSC5)
 Titel <- c("HSC beliebig ",
"0: stable w\o outlet", "1: stable with outlet", "2:
outdoors ", "0+1 ", "1+2 ", "0+2
")
 }

 DF <- Resistenzen # listdfs[[i]]

```

```

numstrings <- str_replace(DF[[antib]], paste0("<=", anfang),
as.character(anfang))
z.B. "1" als numerischer Platzhalter für "<=1"

numstrings <- str_replace(numstrings, paste0(">", ende),
as.character(ende)) #+1))
z.B. "33" als numerischer Platzhalter für ">32"

numbers <- as.numeric(numstrings) # jetzt alles als Zahlen
DF$numbers <- numbers
Log("i, numbers="); Log(i); Log(numbers)

https://stackoverflow.com/questions/23944355/r-hist-right-left-clump-
binning war hier eine Inspiration,
fieseln für die tick labels weil R sonst aus Platzproblemen das wichtige
letzte tick label unterdrückt:
if(Ende) {
 seqAt <- seq(schrittBin, ende+schrittBin, by=schrittLab) #
kleineres ende+... klappt nicht
 seqAt[length(seqAt)] <- seqAt[length(seqAt)]-schrittBin # also
zurückkorrigieren
 #seqAt <- seq(schrittBin, ende+0.001, by=schrittLab)
} else {
 seqAt <- seq(schrittBin, ende, by=schrittLab) # 1 Bin weniger
}
Log("seqAt:"); Log(seqAt)
seqLab <- seq(schrittBin+schrittLab, ende, by=schrittLab)
seqLab <- seq(schrittBin, ende+schrittBin, by=schrittLab)

Log("seqLab:"); Log(seqLab)

seqLab_cutR <- seqLab[1:length(seqLab)-1]
Log("seqLab_cutR:"); Log(seqLab_cutR)

seqLab_cutLR <- seqLab_cutR[2:length(seqLab_cutR)] # ... [2:...] klappt
nicht
Log("seqLab_cutLR:"); Log(seqLab_cutLR)

Labels <- c(paste0("<=", anfang), seqLab_cutLR)
Log("Labels, 1:"); Log(Labels)

if(Ende) { Labels <- c(Labels, paste0(">", ende)) }
Log("Labels, 2:"); Log(Labels)

#Ylab <- paste("Frequenz", antib, Titel[i])
Ylab <- "Frequency"
Xlab <- "Resistency"

Log("breaks="); Log(seqAt)
Log("labels ="); Log(Labels)

plot <- ggplot(DF, aes(x=numbers)) +
 scale_x_continuous(trans = "log10", breaks=seqAt, labels=Labels,
limits=c(anfang*.8, ende*1.2)) +
 # limits=anfang...ende aber bisschen mehr um dort zu plotten

```

```

-0.01 o. *0.9; +1 o. *1.1 genügt nicht
geom_histogram() +
theme(axis.text.x = element_text(size=12)) +
theme(axis.text.y = element_text(size=12)) +
xlab(Xlab) + ylab(Ylab) +
facet_grid(IAC.group ~ .) + ggtitle(antib)
print(plot)

numbers <- na.omit(numbers) # für deskriptive Beschreibung(?)

Median könnte im "<=" Bereich liegen oder im ">=", entsprechend
reagieren:
median <- median(numbers, n.rm=T) # (n.rm=T fürs Lesen aus file, vorher
war das "NA")
rel <- "=" # Relations-Symbol
Log(paste("median, anfang =",median, anfang))
if(median == anfang){
 rel <- "<="
}
if(Ende && median == ende){
 rel <- ">"
}

for (i in 2:length(Titel)){ # nicht 1. plot "XY oder nicht"
deskriptive Statistik - geht sicher o. eigene DFs!!?
 DF <- listdfs[[i]] # kürzer: nur hier explizit?

 print(paste(antib,"- Resistenz,", Titel[i], ":"))
 print(paste(" Median ", rel, median))

 if(Ende && (max(numbers) > ende)){ # gibt overflow bin, ist nicht leer:
Verteilung nicht nach oben beschränkt

 # kleinste Werte kleinstmöglich und grösste Werte kleinstmöglich
ergibt Mindestwert des Mittelwertes
 mean <- mean(replace(numbers, numbers==anfang, 0))

 print(paste(" Mittelwert >= ", mean))
 print("")
 } else { # Verteilung nach oben beschränkt
 if (anfang %in% numbers) { # Verteilung nach oben beschränkt, nicht
nach unten
 # (underflow bin gibt's FAST immer)
 mean1 <- mean(numbers) # kleinste Werte grösstmöglich gibt
Höchstwert des Mittelwertes

 numbers0 <- replace(numbers, numbers==anfang, 0) # kleinste Werte
kleinstmöglich
 mean0 <- mean(numbers0) # ergibt
Mindestwert des Mittelwertes
 print(paste(" Mittelwert in ", sprintf("%.3f",mean0),
"...", sprintf("%.3f",mean1)))
 print("")
 } else { # Verteilung nach oben und unten
beschränkt : einfachster Fall

```



```

```{r, warning=F, error=F, message=F}
  graphisch("IAC.group", "TET" , 2,32, 2,8 )
```
```{r, warning=F, error=F, message=F}
  graphisch("IAC.group", "TMP" , 0.25 , 16 , 0.25,8 )
```
```{r, warning=F, error=F, message=F}
  graphisch("IAC.group", "SMX" , 8 , 512 , 8,256 )

```

```

  #stop the script - by error
```

```

Die Mittelwerte der Resistenz sind für 5 Antibiotika vergleichbar (AMP, MERO, TGC, TAZ, CHL), für GEN tendenziell grösser im Fall \*Ill Animals in Calving box\*, für 3 Antibiotika tendenziell kleiner in diesem Fall (ZIP, AZI, NAL), für TET definitiv grösser in diesem Fall und für 3 Antibiotika definitiv kleiner in diesem Fall (FOT, TMP, SMX). Diese Relationen sind im wesentlichen gleich gerichtet wie in WM - keine WM.

Der Vergleich des Medians der 2 Gruppen zeigt Unterschiede nur für TET und SMX, in der gleichen Richtung wie der Mittelwert. Deshalb diskutiere ich den Median nicht weiter.

## ## Other Live Stock - Gruppen

Mit "OLS" abgekürzt.

```

```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "AMP", 1,32, 1,8)
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "MERO", 0.03 , -0.06, 0.015, 0.015 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "CIP" , 0.015, 8 , 0.015, 4 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "AZI" , 2 , 64 , 1 , 8 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "GEN" , 0.5 , 16 , 0.5 , 4 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "TGC" , 0.25 , -0.5 , 0.25 , 0.25 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "TAZ" , 0.25,-1 , 0.25,0.25 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "FOT" , 0.25 , 4 , 0.25 , 1 )
```
```{r, warning=F, error=F, message=F}
  graphisch("OLS.group", "CHL" , 8 , 64 , 8,16 )
```
```{r, warning=F, error=F, message=F}

```

```

graphisch("OLS.group", "NAL" , 4      , 64      , 4,16      )
```
graphisch("OLS.group", "TET" , 2 , 32 , 2,8)
```
graphisch("OLS.group", "TMP" , 0.25   , 16      , 0.25,8    )
```
graphisch("OLS.group", "SMX" , 8 , 512 , 8,256)

#stop the script

```

Die Mittelwerte der Resistenz sind für MERO, GEN und TAZ vergleichbar, für 5 Antibiotika tendenziell grösser im Fall \*Other Livestock\* (CIP, FOT, CHL, NAL, SMX), für TGC tendenziell kleiner in diesem Fall und für 4 Antibiotika definitiv kleiner in diesem Fall (AMP, AZI, TET, TMP). Diese Relationen sind im wesentlichen entgegengesetzt zu WM - keine WM!

# Waste Milk - Gruppen

```

```{r, warning=F, error=F, message=F}
graphisch("WM.group", "AMP", 1,32, 1,8)
```

Der Mittelwert ist höher ohne WM.
```{r, warning=F, error=F, message=F}
graphisch("WM.group", "MERO", .03,-0.06, .015,.015)
```

```

Der Mittelwert ist vergleichbar ohne WM (tatsächlich tendenziell minimal höher - das ist leicht zu kontrollieren: MERO ist immer  $\leq 3$  - ausser einmal 0.06 für Betrieb 4 und der ist WM group 2).

```

```{r, warning=F, error=F, message=F}
graphisch("WM.group", "CIP", 0.015,8, .015,4)
```

```

Der Mittelwert ist tendenziell höher ohne WM.

```

```{r, warning=F, error=F, message=F}
graphisch("WM.group", "AZI", 2,64, 1,16)
```

```

Der Mittelwert ist höher ohne WM.

```

```{r, warning=F, error=F, message=F}
graphisch("WM.group", "GEN", 0.5,16, 0.5,4)
```

```

Der Mittelwert ist vergleichbar ohne WM.

```

```{r, warning=F, error=F, message=F}
graphisch("WM.group", "TGC", 0.25,-0.5, 0.25,0.25)
```

```

Der Mittelwert ist vergleichbar ohne WM.

```

```{r, warning=F, error=F, message=F}
graphisch("WM.group", "TAZ", 0.25, -1, .25,.25)
```

```

Der Mittelwert ist vergleichbar ohne WM. Genauer: tendenziell höher - das kann man auch noch per Hand kontrollieren: TAZ ist immer  $\leq 0.25$  ausser für:

- Waste Milk: 0.5 für Betriebe 11 und 15
- Keine Waste Milk: 0.5 für Betriebe 12, 59 und 3\*1 für Betrieb 52

(Betrieb 30 wurde ganz am Anfang schon gelöscht)

Die Werte 0.5 balancieren sich also aus für Waste Milk oder nicht, und der Unterschied kommt von den 3 Werten 1: Ohne WM ist resistenter.

```
```{r, warning=F, error=F, message=F}
  graphisch("WM.group", "FOT", 0.25, 4, .25, 1)
```
```

Der Mittelwert ist tendenziell höher ohne WM.

```
```{r, warning=F, error=F, message=F}
  graphisch("WM.group", "CHL", 8,64, 8,16)
```
```

Der Mittelwert ist tendenziell höher ohne WM.

```
```{r, warning=F, error=F, message=F}
  graphisch("WM.group", "NAL", 4,64, 4,16)
```
```

Der Mittelwert ist tendenziell höher ohne WM.

```
```{r, warning=F, error=F, message=F}
  graphisch("WM.group", "TET", 2,32, 2,8)
```
```

Der Mittelwert ist tendenziell höher ohne WM.

```
```{r, warning=F, error=F, message=F}
  graphisch("WM.group", "TMP", 0.25,16, .25,8)
```
```

Der Mittelwert ist höher ohne WM.

```
```{r, warning=F, error=F, message=F}
  graphisch("WM.group", "SMX", 8,512, 8,256)
```
```

Der Mittelwert ist vergleichbar ohne WM.

Die Mittelwerte der Resistenz sind für 5 Antibiotika vergleichbar (MERO, GEN, TGC, TAZ, SMX), für 3 Antibiotika tendenziell grösser im Fall \*WM\* (CIP, FOT, NAL) und für 5 Antibiotika definitiv grösser in diesem Fall (AMP, AZI, HCL, TET, TMP).

# Husbandry System Calves - Gruppen

Mit "HSC" abgekürzt.

```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "AMP", 1,32, 1,8)
```
```

```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "MERO", 0.03 , -0.06, 0.015,0.015)
```
```

```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "CIP" , 0.015, 8 , 0.015, 4 )
```
```

```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "AZI" , 2 , 64 , 1 , 8 )
```
```

```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "GEN" , 0.5 , 16 , 0.5 , 4 )
```
```



```

```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "TGC" , 0.25 , -0.5 , 0.25 , 0.25 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "TAZ" , 0.25 , -1 , 0.25 , 0.25 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "FOT" , 0.25 , 4 , 0.25 , 1 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "CHL" , 8 , 64 , 8,16 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "NAL" , 4 , 64 , 4,16 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "TET" , 2 , 32 , 2,8 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "TMP" , 0.25 , 16 , 0.25,8 )
```
```{r, warning=F, error=F, message=F}
  graphisch("HSC.group", "SMX" , 8 , 512 , 8,256 )
```

```

Es ist kein sehr ausgeprägtes Muster für grösste/kleinste Resistenzen zu erkennen. Tendenziell ergeben 1 und 1+2 die grössten Resistenzen, 2 und vor allem 0+1 die kleinsten.

# Vollständigkeit

Jetzt sind alle Verteilungen geplotted und deskriptiv analysiert, ausser:

- AMI: alle Proben sensitiv  $\leq 4$
- COL: alle Proben sensitiv  $\leq 1$

# Weitere Schritte

## Technischer Natur

\* je 2 plots kombinieren übereinander in einem Quadrat

## Fundamentalener Natur

\* Kausalitäten studieren mittels Regressionen