Legal Data Science: Der moderne Weg zur Wahrheit

Diagramme

Seán Fobbe

2023-05-08

Inhaltsverzeichnis

1 Vorbereitung

1.1 Bericht erstellen

Diesen Befehl ausführen um den gesamten Bericht zu erstellen.

```
rmarkdown::render("Fobbe_2023-05-03_Diagramme.R")
```

1.2 Packages laden

```
library(datasauRus)
library(ggplot2)
library(viridis)
library(data.table)
library(knitr)
```

1.3 Palette einlesen

Color Blind Friendly Palette with Black from http://jfly.iam.u-tokyo.ac.jp/color/ as presented at http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/

```
col <- c("#000000", # 1: Black
         "#E69F00", # 2: Orange
         "#56B4E9", # 3: Sky Blue
         "#009E73", # 4: Bluish Green
         "#F0E442", # 5: Yellow
         "#0072B2", # 6: Blue
         "#D55E00", # 7: Vermillion,
         "#CC79A7") # 8: Reddish Purple
```

2 Normalverteilung

2.1 Berechnungen

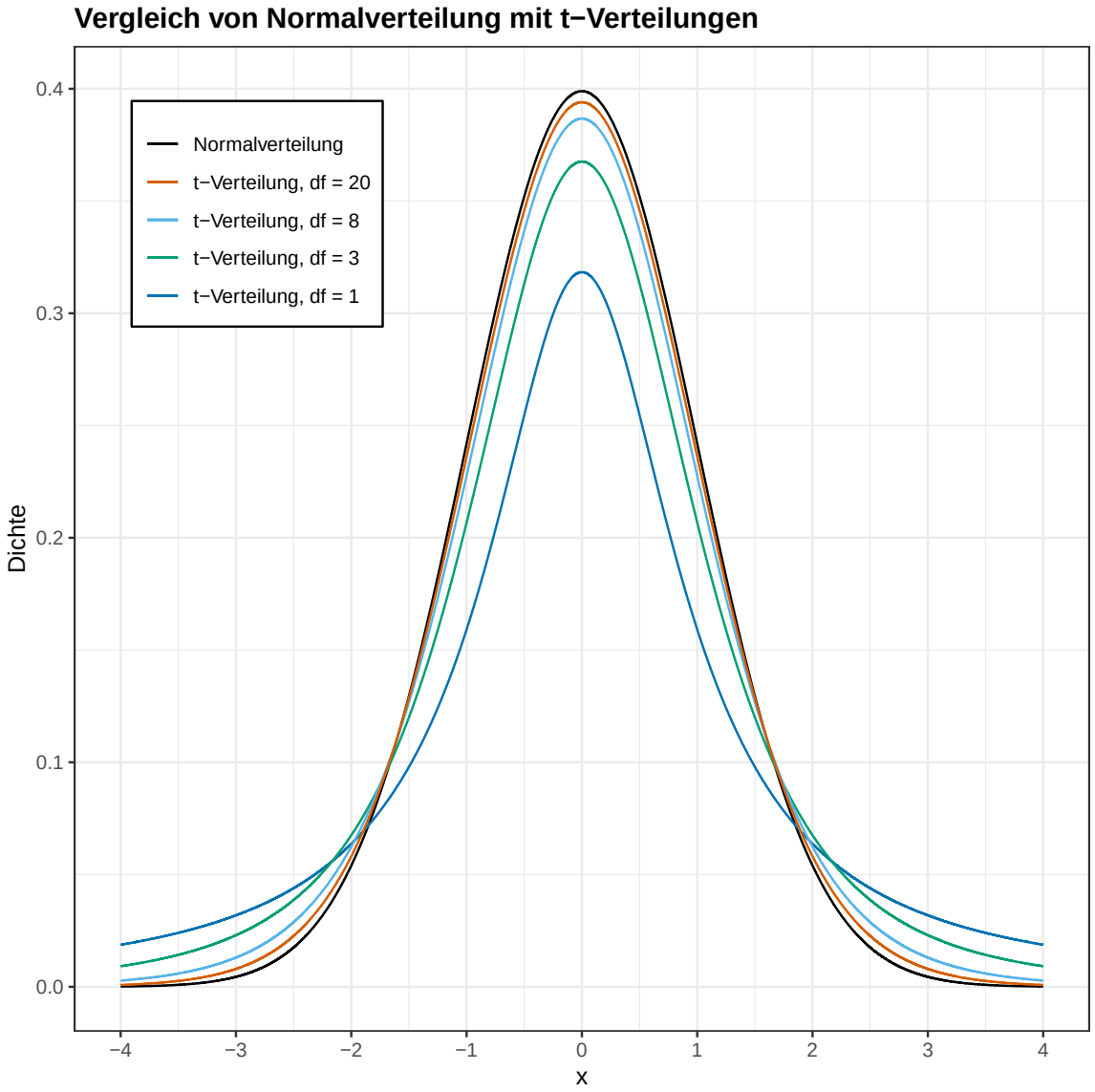
```
x <- seq(-4, 4, length = 50000)

dnorm <- dnorm(x)
tdeg1 <- dt(x,1)
tdeg3 <- dt(x,3)
tdeg8 <- dt(x,8)
tdeg20 <- dt(x,20)


dt <- data.table(x,
                 dnorm,
                 tdeg1,
                 tdeg3,
                 tdeg8,
                 tdeg20)
```

2.2 Visualisierung

```
ggplot(data = dt, aes(x = x))+
    geom_line(aes(y = dnorm, color = col[1]))+
    geom_line(aes(y = tdeg1, color = col[6]))+
    geom_line(aes(y = tdeg3, color = col[4]))+
    geom_line(aes(y = tdeg8, color = col[3]))+
    geom_line(aes(y = tdeg20, color = col[7]))+
    scale_x_continuous(breaks = seq(-4,4,1))+
    scale_y_continuous(breaks = seq(0,0.5,0.1))+
    theme_bw()+
    theme(plot.title = element_text(lineheight=.8, face="bold"),
          legend.position=c(.18, .83),
          legend.direction = "vertical",
          legend.background = element_rect(size=0.5, linetype="solid", colour ="
    black"),
          legend.title=element_blank())+
    labs(x = "x",
         y = "Dichte",
         title = "Vergleich von Normalverteilung mit t-Verteilungen")+
    scale_color_identity(breaks = c(col[1], col[7], col[3], col[4], col[6]),
                         labels = c("Normalverteilung",
                                    "t-Verteilung, df = 20",
                                    "t-Verteilung, df = 8",
                                    "t-Verteilung, df = 3",
                                    "t-Verteilung, df = 1"),
                         guide = "legend")
```

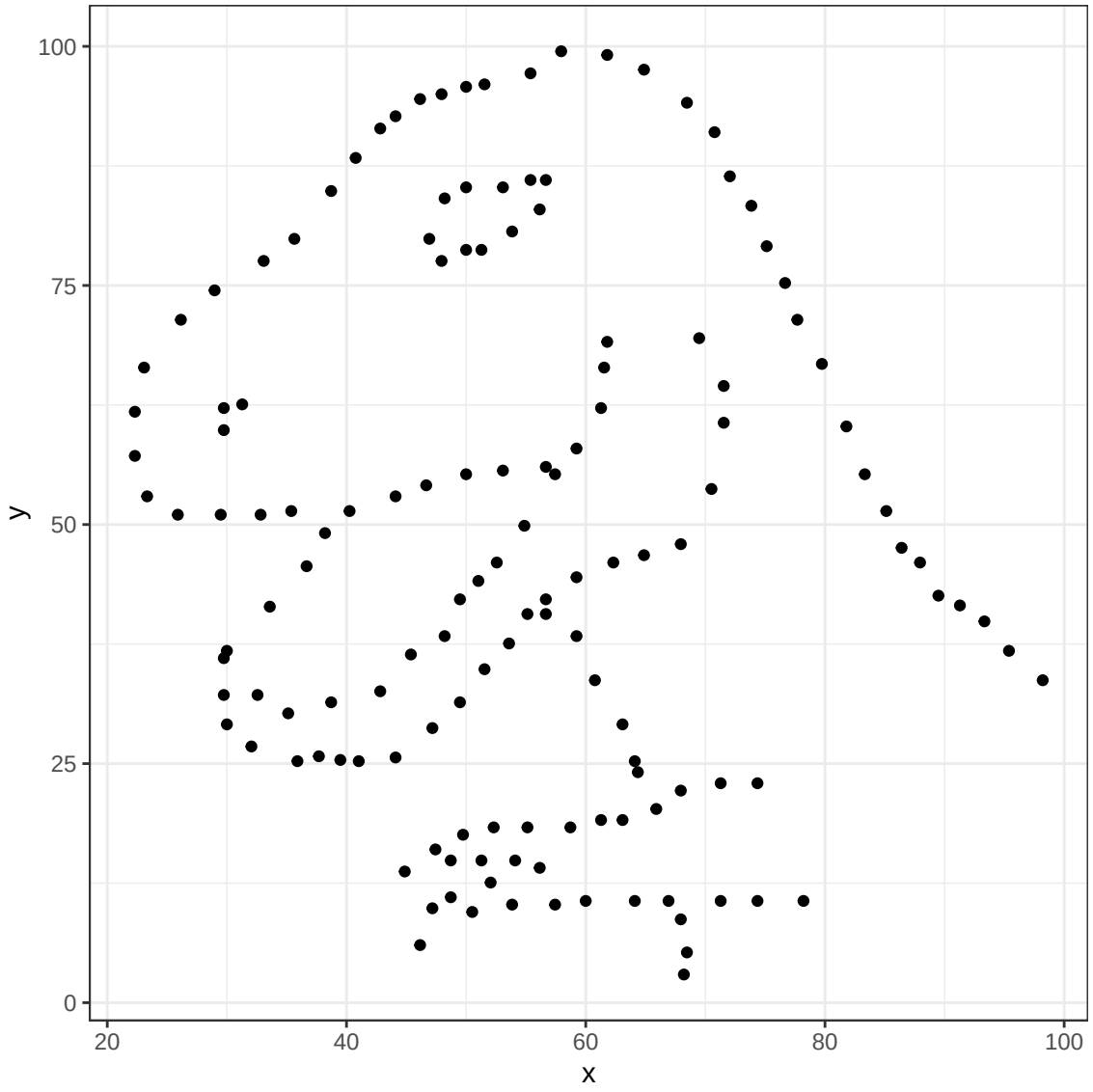

Vergleich von Normalverteilung mit t−Verteilungen
Normalverteilung
t−Verteilung, df = 20
t−Verteilung, df = 8
t−Verteilung, df = 3
t−Verteilung, df = 1
Dichte
0.0
0.1
0.2
0.3
0.4
−4
−3
−2
−1
0
1
2
3
4
x

3 Datasaurus

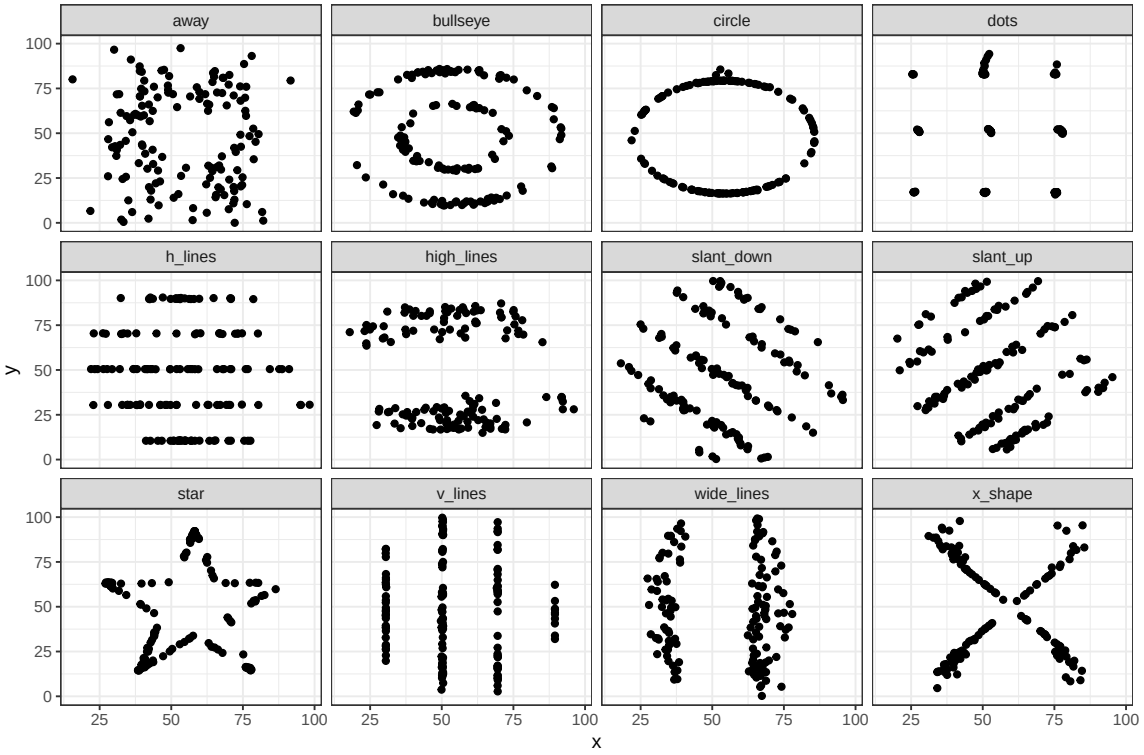
```
dt <- datasaurus_dozen
setDT(dt)
```

3.1 Visualisierung

```
ggplot(dt[dataset == "dino"],
       aes(x = x,
           y = y)) +
    geom_point() +
    theme_bw() +
    theme(legend.position = "none")
```



```
ggplot(dt[dataset != "dino"],
       aes(x = x,
           y = y)) +
    geom_point() +
    theme_bw() +
    theme(legend.position = "none") +
    facet_wrap(~dataset, ncol = 4)
```



3.2 Beschreibende Statistik

3.2.1 Zusammenfassungen berechnen

```
summaries <- dt[,{

    x_mean  <- mean(x)
    x_sd <- sd(x)
    y_mean <- mean(y)
    y_sd <- sd(y)
    corr <- cor(x, y)
    list(x_mean, x_sd, y_mean, y_sd, corr)

}, by = "dataset"]
```

3.2.2 Zusammenfassungen anzeigen

```
kable(summaries)
```

dataset	x_mean	x_sd	y_mean	y_sd	corr
dino	54.26327	16.76514	47.83225	26.93540	-0.0644719
away	54.26610	16.76983	47.83472	26.93974	-0.0641284
h_lines	54.26144	16.76590	47.83025	26.93988	-0.0617148
v_lines	54.26993	16.76996	47.83699	26.93768	-0.0694456
x_shape	54.26015	16.76996	47.83972	26.93000	-0.0655833
star	54.26734	16.76896	47.83955	26.93027	-0.0629611
high_lines	54.26881	16.76670	47.83545	26.94000	-0.0685042
dots	54.26030	16.76774	47.83983	26.93019	-0.0603414
circle	54.26732	16.76001	47.83772	26.93004	-0.0683434
bullseye	54.26873	16.76924	47.83082	26.93573	-0.0685864
slant_up	54.26588	16.76885	47.83150	26.93861	-0.0686092
slant_down	54.26785	16.76676	47.83590	26.93610	-0.0689797
wide_lines	54.26692	16.77000	47.83160	26.93790	-0.0665752

3.2.3 Zusammenfassungen als TEX schreiben

```
writeLines(kable(summaries,
      format = "latex",
      align = 'r',
      booktabs = TRUE,
      longtable = TRUE),

      "datasaurus.tex")
```