

Statistik in R

Dirk Seidensticker/Clemens Schmid

7. Juli 2017

Deskriptive Statistik

```
atlant <- read.csv(  
  "../data/AtlantData1.csv",  
  sep = "\\t",  
  header = TRUE)
```

```
head(atlant[,1:9])
```

##	site	X	Y	archont	feature	object	class	sherd	qty
## 1	A	472564	3939619	Atlas	surface		K	G	1
## 2	A	472564	3939619	Atlas	1	2	K	G	1
## 3	A	472564	3939619	Atlas	1	-1,2:2	K	R	1
## 4	A	472564	3939619	Atlas	1	3, -1-3-1:3	K	R	1
## 5	A	472564	3939619	Atlas	1	2, -1:2	K	G	1
## 6	A	472564	3939619	Atlas	1	-1:4, -2:4	K	G	1

```
min(atlant$muendungsD)
```

```
## [1] NA
```

Bei Variablen mit NaN muss dies R mitgeteilt werden!

```
min(atlant$muendungsD, na.rm = TRUE)
```

```
## [1] 5
```

```
max(atlant$muendungsD, na.rm = TRUE)
```

```
## [1] 35
```

```
mean(atlant$wt)
```

```
## [1] 395.3333
```

```
median(atlant$wt)
```

```
## [1] 153.5
```

Eine wichtige Eigenschaft des Medians ist Robustheit gegenüber Ausreißern!

```
summary(atlant$wt)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	6.0	44.0	153.5	395.3	501.5	4500.0

Spannweite

```
max(atlant$wt) - min(atlant$wt)
```

```
## [1] 4494
```

Varianz

```
var(atlant$wt)
```

```
## [1] 392716.4
```

Standardabweichung

```
sd(atlant$wt)
```

```
## [1] 626.6709
```

!!! R Funktionen beruhen auf der korrigierten Stichprobenvarianz !!!

```
var(atlant$wt)
```

```
## [1] 392716.4
```

empirische Varianz

$$var(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

```
(sum( (atlant$wt - mean(atlant$wt) ) ^2 ))/(length(atlant$wt) )
```

```
## [1] 389870.6
```

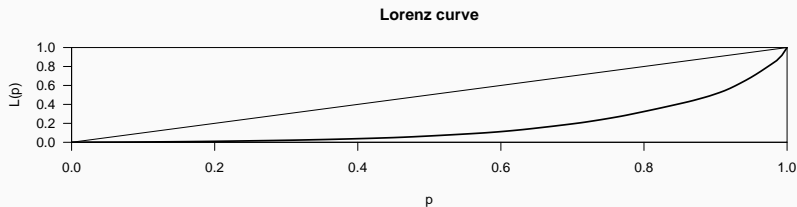
Gini-Koeffizient aus der Lorenz-Kurve (Paket: ineq)

```
ineq(atlant$wt,type="Gini")
```

```
## [1] 0.6580198
```

0 = gleichmäßigen Verteilung, 1 = maximaler Ungleichverteilung

```
plot(Lc(atlant$wt))
```



Schließende Statistik

```
t.test(atlant$wt)

##
##  One Sample t-test
##
## data:  atlant$wt
## t = 7.4108, df = 137, p-value = 1.171e-11
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  289.8458 500.8209
## sample estimates:
## mean of x
##  395.3333
```

Hypothesentests: *Chi-Quadrat-Unabhängigkeitstest*

prüft, ob zwei Merkmale stochastisch unabhängig sind.

a

```
##           sherd vessel
## excavation    77     40
## surface       7     14
```

```
chisq.test(a)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  a
## X-squared = 6.5805, df = 1, p-value = 0.01031
```

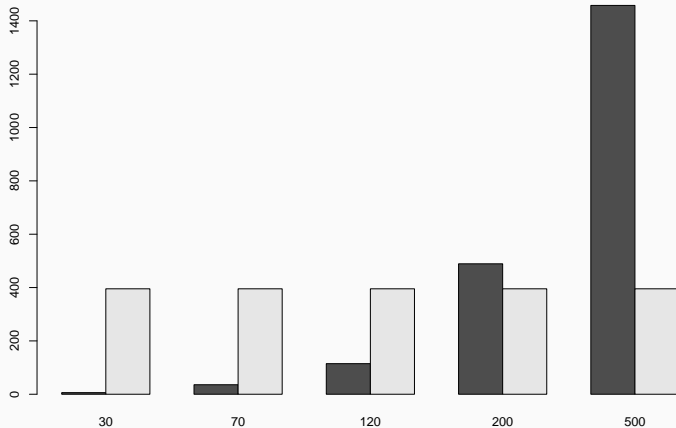
prüft, ob vorliegende Daten auf eine bestimmte Weise verteilt sind

##	size	wt.obs	wt.exp
## 1	30	6.00000	395.3333
## 2	70	35.61111	395.3333
## 3	120	114.64103	395.3333
## 4	200	489.00000	395.3333
## 5	500	1457.80952	395.3333

Mittleres Scherbengewicht

Hypothesentests: *Chi-Quadrat-Verteilungs-/Anpassungstest*

prüft, ob vorliegende Daten auf eine bestimmte Weise verteilt sind



prüft, ob vorliegende Daten auf eine bestimmte Weise verteilt sind

```
chisq.test(e$wt.obs,  
           p = e$wt.exp,  
           rescale.p=TRUE)
```

```
##  
## Chi-squared test for given probabilities  
##  
## data:  e$wt.obs  
## X-squared = 3552.4, df = 4, p-value < 2.2e-16
```