

## Statistische Software (R)

Paul Fink, M.Sc.

Institut für Statistik  
Ludwig-Maximilians-Universität München

*Pseudo Zufallszahlen, Dichten, Verteilungsfunktionen, etc.*



## Nützliche Funktionen

Weitere nützliche Funktionen im Zusammenhang der deskriptiven Beschreibung von Daten:

- Sortieren eines Vektors:

```
> x <- c(1, 3, 2, 5)
> sort(x)
[1] 1 2 3 5
> sort(x, decreasing = TRUE)
[1] 5 3 2 1
```

- Bestimmung der Ränge:

```
> x <- c(1, 3, 2, 5, 2)
> rank(x)
[1] 1.0 4.0 2.5 5.0 2.5
```

## Übersicht deskriptiver Maße & anderer nützlicher Funktionen

Funktion	Maß
<code>mean()</code>	arithmetische Mittel
<code>median()</code>	Median
<code>exp(mean(log( )))</code>	Geometrisches Mittel
<code>quantile()</code>	Quantile
<code>var()</code>	Varianz (Version mit $1/(n-1)$ )
<code>sd()</code>	Standardabweichung (Version mit $1/(n-1)$ )
<code>range()</code>	Minimum und Maximum
<code>diff(range())</code>	Spannweite
<code>cov()</code>	Kovarianz (Version mit $1/(n-1)$ )
<code>cor()</code>	Korrelation (Spearman, Bravais-Pearson)
<code>density()</code>	Kerndichteschätzer
<code>ecdf()</code>	Empirische Verteilungsfunktion

Fink: Statistische Software (R) SoSe 2014

1

## Nützliche Funktionen

- Indizierung mehrfach vorkommender Werte in einem Vektor:

```
> x <- c(1, 3, 2, 5, 2)
> duplicated(x)
[1] FALSE FALSE FALSE FALSE TRUE
```

- Entfernung von Duplikaten (z.B. Bestimmung aller vorkommenden Merkmalsausprägungen in einer Stichprobe):

```
> x <- c(1, 3, 2, 5, 2)
> unique(x)
[1] 1 3 2 5
```

## Nützliche Funktionen

- Diskretisierung einer (quasi-)stetigen Variable:

```
> x <- c(1.3, 1.5, 2.5, 3.8, 4.1, 5.9, 7.1, 8.4, 9.0)
> xdiscrete <- cut(x, breaks=c(-Inf, 2, 5, 8, Inf) )
> is.factor(xdiscrete)
[1] TRUE
> xdiscrete
[1] (-Inf,2] (-Inf,2] (2,5] (2,5] (2,5] (5,8] (5,8] (8, Inf]
[9] (8, Inf]
Levels: (-Inf,2] (2,5] (5,8] (8, Inf]
> table(xdiscrete)
xdiscrete
(-Inf,2] (2,5] (5,8] (8, Inf]
      2      3      2      2
```

## Nützliche Funktionen

- Gammafunktion: Für natürliche Zahlen  $n$  gilt:  $\Gamma(n) = (n-1)!$

```
> c(gamma(5), factorial(4))
[1] 24 24
> c(gamma(0.5), sqrt(pi))
[1] 1.772454 1.772454
```

- Betafunktion:  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ :

```
> c(gamma(5) * gamma(3) / gamma(5 + 3), beta(5, 3))
[1] 0.00952381 0.00952381
```

## Nützliche Funktionen

- Kumulierte Summe und Produkt:

```
> x <- c(1, 3, 2, 5)
> cumsum(x)      # 1, 1+3, 1+3+2, 1+3+2+5
[1] 1 4 6 11
> cumprod(x)     # 1, 1*3, 1*3*2, 1*3*2*5
[1] 1 3 6 30
```

- Fakultät:

```
> factorial(5)
[1] 120
```

- Binomialkoeffizient  $\binom{n}{k}$ :

```
> choose(4,2)
[1] 6
```

## Verteilungen und Zufallszahlen

R bietet Funktionen zur Berechnung von Dichten, Verteilungsfunktionen und Quantilen, sowie zur Erzeugung von (Pseudo-) Zufallszahlen gebräuchlicher Verteilungen.

Der Funktionsname folgt dabei folgendem Schema:

Anfangsbuchstabe	Art der Funktion
d	Dichte (density)
p	Verteilungsfunktion (probability)
q	Quantilsfunktion (quantiles)
r	Zufallszahl (random number)

## Verteilungen und Zufallszahlen

Beispiele:

- Die Dichte der  $N(0, 1)$ -Verteilung an der Stelle  $x = 0$  ( $1/\sqrt{2\pi}$ ):

```
> c(dnorm(x = 0), 1/sqrt(2*pi))  
[1] 0.3989423 0.3989423
```

- Die Verteilungsfunktion der  $N(0, 1)$ -Verteilung an der Stelle  $q$ , also  $\Phi(q) = P(X \leq q)$ :

```
> pnorm(q = 0)  
[1] 0.5  
> pnorm(q = 1.96)  
[1] 0.9750021
```

- Das  $p$ -Quantil der  $N(0, 1)$ -Verteilung, also  $z_p$ :

```
> qnorm(p=0.95)  
[1] 1.644854
```

## Verteilungen und Zufallszahlen

Diese Funktionen stehen u.a. für folgende Verteilungen zur Verfügung:

- Modellverteilungen**

Funktion	Verteilung
beta	Beta
binom	Binomial
cauchy	Cauchy
exp	Exponential
gamma	Gamma
geom	Geometrische
hyper	Hypergeometrische
lnorm	Log-Normal
norm	Normal
pois	Poisson
unif	Gleichverteilung, Rechtecksverteilung
mvnorm	Multivariate Normalverteilung, package <code>mvtnorm</code>

## Verteilungen und Zufallszahlen

- Eine Stichprobe vom Umfang  $n = 5$  einer  $N(0, 1)$ -Verteilung, sogenannte Pseudo-Zufallszahlen.

```
> X <- rnorm(n = 5)  
> X  
[1] 1.05192864 0.03978332 -0.45625959 -0.34689411 -1.13714623
```

## Verteilungen und Zufallszahlen

- Prüfverteilungen**

Funktion	Verteilung
chisq	$\chi^2$
f	$F$
signrank	Wilcoxon Vorzeichen-Rangsummen (1 Stichprobe)
t	$t$
wilcox	Wilcoxon Rangsummen (2 Stichproben)

## Die sample Funktion

Mit der Funktion `sample()` kann eine Stichprobe aus einer endlichen diskreten Mengen gezogen werden.

Sie erlaubt das Ziehen einer Stichprobe aus einem Vektor vom Umfang `size` mit oder ohne Zurücklegen (`replace=TRUE` oder `replace=FALSE`) und optional mit bestimmten Wahrscheinlichkeiten (Argument `prob`).

Das `replace` – Argument ist auf `FALSE` voreingestellt.

Beispiel:

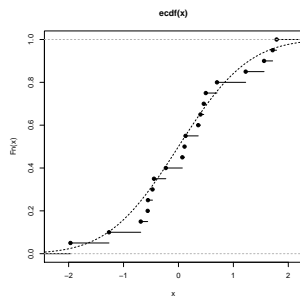
- Ziehen mit Zurücklegen aus einer Gleichverteilung über  $\{1, 2, 3, 4, 5\}$

```
> sample(x = c(1, 2, 3, 4, 5), size = 10, replace = TRUE)
[1] 1 1 5 2 4 2 3 1 4 2
```

## Empirische Verteilungsfunktion

Empirische und theoretische Verteilungsfunktion der Standardnormalverteilung:

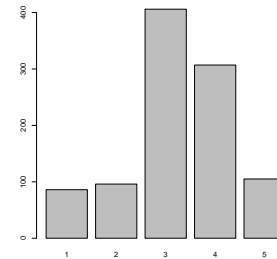
```
> set.seed(123)
> x <- rnorm(20)
> plot(ecdf(x))
> lines(seq(-4, 4, by = .1), pnorm(seq(-4, 4, by = .1)), lty = "dashed")
```



## Die sample Funktion

- Ziehen mit Zurücklegen aus einer vorgegebenen Verteilung (`prob` gesetzt):

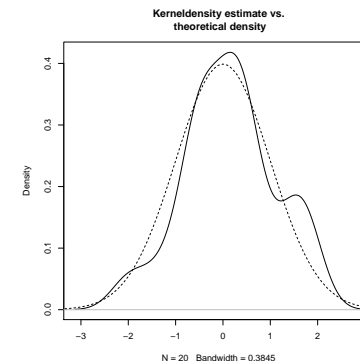
```
> y <- sample(x = c(1, 2, 3, 4, 5),
+           size = 1000,
+           replace = TRUE,
+           prob = c(0.1, 0.1, 0.4, 0.3, 0.1))
> barplot(table(y))
```



## Kerndichteschätzer

Kerndichtschätzer und theoretische Verteilungsfunktion der Standardnormalverteilung:

```
> fn <- density(x)
> plot(fn, main = "Kerndensity estimate vs.\n theoretical density")
> lines(seq(-4, 4, by = .1), dnorm(seq(-4, 4, by = .1)), lty = "dashed")
```



# Aufgaben

---

1. Erzeugen Sie Stichproben aus verschiedenen Verteilungen (Poisson, Binomial,  $\chi^2$ , Exponential) mit verschiedenen Parametern und Stichprobenumfängen. Visualisieren Sie die Approximation der standardisierten Summen durch die Normalverteilung (zentraler Grenzwertsatz).
2. Zeigen Sie, dass das Vorgehen wie in 1. für die Cauchy-Verteilung nicht klappt.