LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Statistische Software (R)

Paul Fink, M.Sc.

Institut für Statistik
Ludwig-Maximilians-Universität München

Verteilungen und Zufallszahlen

Übersicht Statistik-Funktionen

Funktion	Beschreibung
`mean()`	artihmetische Mittel
`median()`	Median
`exp(mean(log( )))`	Geometrisches Mittel
`quantile()`	empirische Quantile
`var()`	Stichproben-Varianz
`sd()`	Stichproben-Standardabweichung
`range()`	Minimum und Maximum
`diff(range())`	Spannweite
`cov()`	Stichproben-Kovarianz
`cor()`	Korrelation (Spearman, Bravais–Pearson)
`density()`	Kerndichteschätzer
`ecdf()`	Empirische Verteilungsfunktion

Nützliche Funktionen – Übersicht

Funktion	Beschreibung
`cumsum()`	Kumulierte Summe
`cumprod()`	Kumuliertes Produkt
`gamma()`	Gamma-Funktion: $\Gamma(n) = (n-1)!$
`beta()`	Betafunktion: $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
`sort()`	Sortieren eines Vektors
`rank()`	Bestimmung der Ränge
`factorial()`	Fakultät
`choose()`	Binomialkoeffizient $\binom{n}{k}$
`unique()`	Entfernung von Duplikaten
`duplicated()`	Indizierung von Duplikaten

Nützliche Funktionen – Diskretisieren

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] (8, Inf] (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
```

(Pseudo-) Zufallszahlen

- Statistik verwendet Zufallszahlen
- Generierbar nur durch deterministischen Algorithmus (Generator)
  $\Longrightarrow$ Pseudo-Zufallszahlen
- Problem: Vergleichbarkeit von zufälligen Ergebnissen
- Lösung: Zufallszahlengenerator initialisieren mit `set.seed()`

```
> set.seed(12345)
> rnorm(n = 2)
[1] 0.5855288 0.7094660
> rnorm(n = 2)
[1] -0.1093033 -0.4534972
> set.seed(12345)
> rnorm(n = 2)
[1] 0.5855288 0.7094660
```

Verteilungen und Zufallszahlen

Funktionen zur Berechnung von Dichten, Verteilungsfunktionen, Quantilen und Erzeugung von (Pseudo–) Zufallszahlen bekannter Verteilungen

Funktionsnamen-Schema

Funktionsanfang	Art der Funktion
`d`...	Dichte (**d**ensity)
`p`...	Verteilungsfunktion (**p**robability)
`q`...	Quantilsfunktion (**q**uantiles)
`r`...	Zufallszahl (**r**andom number)

Übersicht Modellverteilungen

Funktionsende	Verteilungsname
...`beta`	Beta-Verteilung
...`binom`	Binomial-Verteilung
...`cauchy`	Cauchy-Verteilung
...`exp`	Exponential-Verteilung
...`gamma`	Gamma-Verteilung
...`geom`	Geometrische-Verteilung
...`hyper`	Hypergeometrische-Verteilung
...`lnorm`	Log-Normal-Verteilung
...`norm`	Normal-Verteilung
...`pois`	Poisson-Verteilung
...`unif`	Gleich-/ Rechtecks-Verteilung
...`mvnorm`	Multivariate Normal-Verteilung (package `mvtnorm`)

Übersicht Prüfverteilungen

Funktionsende	Verteilungsname
...`chisq`	χ^2-Verteilung
...`f`	F-Verteilung
...`signrank`	Verteilung der Wilcoxon Vorzeichen-Rangsummen (1 Stichprobe)
...`t`	t-Verteilung
...`wilcox`	Verteilung der Wilcoxon Rangsummen (2 Stichproben)

Funktionsnamen-Schema – Beispiel Std.-NV

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

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

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

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

Funktionsnamen-Schema – Beispiel Std.-NV

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

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

- Stichprobe vom Umfang $n = 5$ aus $N(0,1)$–Verteilung

```
> X <- rnorm(n = 5)
> X
[1] -0.1093033 -0.4534972  0.6058875 -1.8179560  0.6300986
```

Die `sample()` Funktion

Ziehen einer Stichprobe

- mit festem Umfang (Argument `size`)
- aus endlich diskreten Mengen (Argument `x`)
- mit Zurücklegen (Argument `replace = TRUE`)
- oder ohne Zurücklegen (Argument `replace = FALSE`)
- und optional mit bestimmten Wahrscheinlichkeiten (Argument `prob`).

Argument `replace` ist auf `FALSE` voreingestellt.

Die `sample()` Funktion – Beispiele

- 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] 2 3 2 3 1 5 3 2 5 4
```

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

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

zmzv
  1   2   3   4   5
 98 112 383 306 101
```

Aufgaben

1. Erzeugen Sie Stichproben aus verschiedenen Verteilungen (Poisson, Binomial, χ^2, Exponential) mit verschiedenen Parametern und den Stichprobenumfängen $n = 20$, $n = 50$, $n = 100$ und $n = 1000$.
2. Berechnen Sie jeweils das arithmetische Mittel und vergleichen Sie es mit dem dazugehörigen theoretischen Erwartungswert.
3. Permutieren Sie den Vektor der Großbuchstaben `LETTERS`.