CAS Datenanalyse HS16 - DeskStat

Wahrscheinichkeitsverteilungen

Zufallsvariablen

Definition

Eine Variable X ist eine Zufallsvariable, wenn der Wert, den X annimmt, von dem Ausgang eines Zufallsexperiments abhängt. Eine Zufallsvariable ordnet jedem Ergebniss eines Zufallsexperiments einen numerischen Wert zu.

Zufallsvariablen werden meist mit Großbuchstaben geschrieben.

Zufallsvariablen

Bemerkung: Zufallsvariablen sind daher Funktionen, die jedem Ergebnis eine (reelle) Zahl zuordnen. Sie haben also nicht direkt etwas mit Zufall zu tun. Da nun Ergebnisse durch Zahlen repräsentiert werden, kann mit ihnen gerechnet werden.

Wahrscheinlichkeitsverteilungen

Definition

Eine Wahrscheinlichkeitsverteilung beschreibt, wie sich die Werte einer Zufallsvariablen verteilen.

Binomialverteilung

Definition

Die Binomialverteilung beschreibt die Anzahl der Erfolge in einer Serie von gleichartigen und unabhängigen Versuchen, die jeweils genau zwei mögliche Ergebnisse haben („Erfolg“ oder „Misserfolg“). Solche Versuchsserien werden auch Bernoulli-Prozesse genannt.

Bezeichnet p die Wahrscheinlichkeit eines erfolgreichen Versuchs, so bestimmt sich die Wahrscheinlichkeit für x erfolgreiche Ergebnisse in n unabhängigen Versuchen folgendermassen:

$$B(x|p,n) = \binom{n}{x} p^x (1-p)^{n-x} \text{ für } x \in \mathbb{N}$$

Binomialverteilung

Problem: Eine Multiple-Choise-Prüfungen besteht aus 12 Fragen. Jede Fragen gibt 5 verschiedenen Antworten, von denen aber nur jeweils eine Antwort richtig ist. Ein Student löst die Aufgaben nach dem Zufallsprinzip. Bestimmen Sie die Wahrscheinlichkeit dafür, dass der Student maximal vier korrekte Antworten gibt.

Binomialverteilung

Antwort: Für eine korrekten Antwort gilt $p = 0.2$. Die Wahrscheinlichkeit für genau 4 richtige Antworten finden wir mit:

```
dbinom(4, size=12, prob=0.2)
```

```
## [1] 0.1328756
```

Die Wahrscheinlichkeit für maximal 4 korrekte Antworten ist somit:

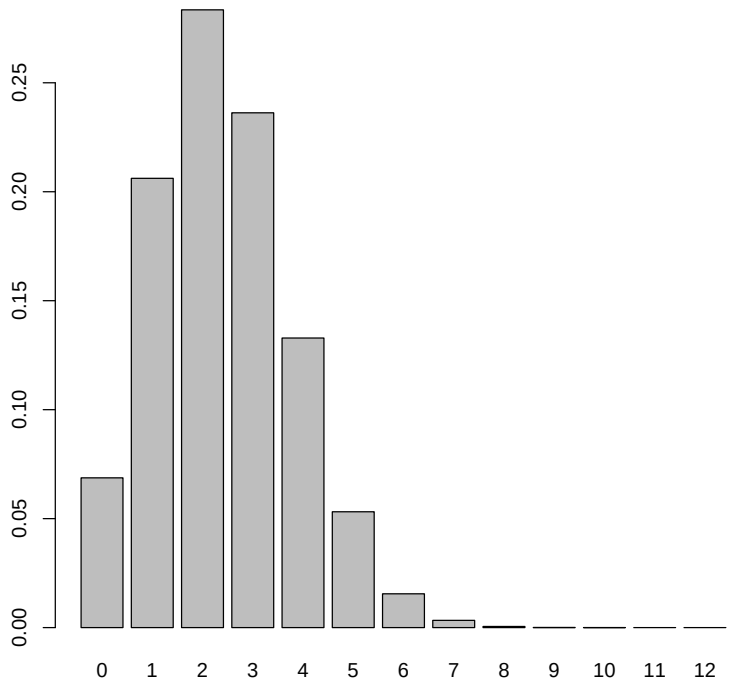
```
dbinom(4, size=12, prob=0.2) +
+ dbinom(3, size=12, prob=0.2) +
+ dbinom(2, size=12, prob=0.2) +
+ dbinom(1, size=12, prob=0.2) +
+ dbinom(0, size=12, prob=0.2)
```

```
## [1] 0.9274445
```

Binomialverteilung

Erweiterte Antwort:

```
yprob <- dbinom(0:12,size=length(0:12)-1,prob = 1/5)
names(yprob) <- 0:12
barplot(yprob)
```



Binomialverteilung

Erweiterte Antwort: Alternativ können wir die kummulierte Wahrscheinlichkeit direkt berechnen mit:

```
pbinom(4, size=12, prob=0.2)
```

```
## [1] 0.9274445
```

Die Wahrscheinlichkeit für vier oder weniger korrekte Antworten beträgt damit 92.7%.

Hypergeometrische Verteilung

Definition

Die Hypergeometrische Verteilung beschreibt eine Stichprobe, die ohne Zurücklegen gezogen wird. Die einzelnen Versuche sind dann nicht unabhängig.

Sei N die Anzahl der Elemente in der Grundgesamtheit; M die Anzahl der Elemente, die für uns günstig sind; n sei die Grösse der Stichprobe; k die Anzahl der Elemente aus M, die in n enthalten sind; $\binom{n}{k}$ ist der Binomialkoeffizient.

$$Hyper(k|M,N,n) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

Hypergeometrische Verteilung

Problem: Beim Schweizer Zahlenlotto sind 6 Zahlen aus 42 zu ziehen.

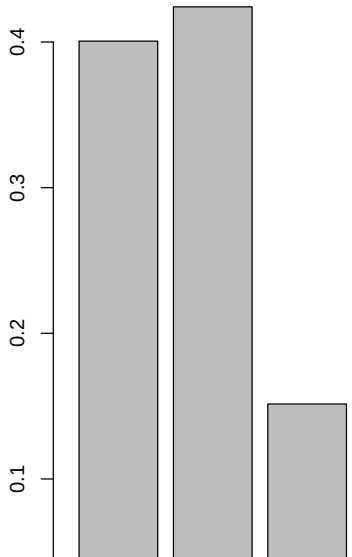
Wir bezeichnen mit x die Anzahl der richtig angekreutzten Zahlen.

Bestimmen Sie die Wahrscheinlichkeitsverteilung und stellen Sie diese grafisch dar.

Hypergeometrische Verteilung

Antwort:

```
ylotto <- dhyper(0:6,m=6,n=39,k=6)

names(ylotto) <- 0:6

barplot(ylotto)
```



Poissonverteilung

Definition

Die Poissonverteilung ist eine diskrete Verteilung, mit der man die Anzahl von Ereignissen in einem gegebenen Zeitintervall modelliert.

Ihr einziger Parameter λ bezeichnet die durchschnittlich zu erwartende Anzahl an Ereignissen.

$$Pois(x|\lambda) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \text{ mit } x \in \mathbb{N}$$

Poissonverteilung

Beispiel:

- Die Anzahl der Tore, die eine Fussballmannschaft während eines Spiels erzielt.

Poissonverteilung

Beispiel:

- Die Anzahl der Tore, die eine Fussballmannschaft während eines Spiels erzielt.
- Die Anzahl der Kunden, die während eines Tages am Postschalter auftauchen.

Poissonverteilung

Beispiel:

- Die Anzahl der Tore, die eine Fussballmannschaft während eines Spiels erzielt.
- Die Anzahl der Kunden, die während eines Tages am Postschalter auftauchen.
- Die Anzahl der SMS, die Handynutzer während eines Tages verschicken.

Poissonverteilung

Beispiel:

- Die Anzahl der Tore, die eine Fussballmannschaft während eines Spiels erzielt.
- Die Anzahl der Kunden, die während eines Tages am Postschalter auftauchen.
- Die Anzahl der SMS, die Handynutzer während eines Tages verschicken.
- Die Anzahl der Gäste, die ein Restaurant zwischen 20 Uhr und 22 Uhr besuchen.

Poissonverteilung

Problem: Eine Brücke wird durchschnittlich von 12 Autos pro Minute passiert. Wie gross ist die Wahrscheinlichkeit, dass sich in einer Minute mehr als 17 Autos auf der Brücke befinden?

Poissonverteilung

Antwort: Die Wahrscheinlichkeit für weniger als 16 Autos auf der Brücke finden wir mit der Funktion `ppoiss`.

```
ppois(16, lambda=12) # lower tail
```

```
## [1] 0.898709
```

Die Wahrscheinlichkeit für 17 und mehr Autos ist somit:

```
1-ppois(16, lambda=12) # oder
```

```
## [1] 0.101291
```

```
ppois(16, lambda=12, lower=FALSE)
```

```
## [1] 0.101291
```

Stetige Gleichverteilung

Definition

Die stetige Gleichverteilung ist eine Verallgemeinerung der diskreten Gleichverteilung. Während bei der diskreten Gleichverteilung jede ganze Zahl zwischen a und b möglich ist (beim Würfelwurf ist z.B. $a = 1$ und $b = 6$), so ist bei der stetigen Gleichverteilung nun jede reelle Zahl im Intervall von a bis b ein mögliches Ergebnis. Ihre Dichtefunktion lautet:

$$Uni(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{für } a \leqslant x \leqslant b \\ 0 & \text{für } x < a \text{ oder } x > b \end{cases}$$

Stetige Gleichverteilung

Beispiel:

- Zufallszahlen.

Stetige Gleichverteilung

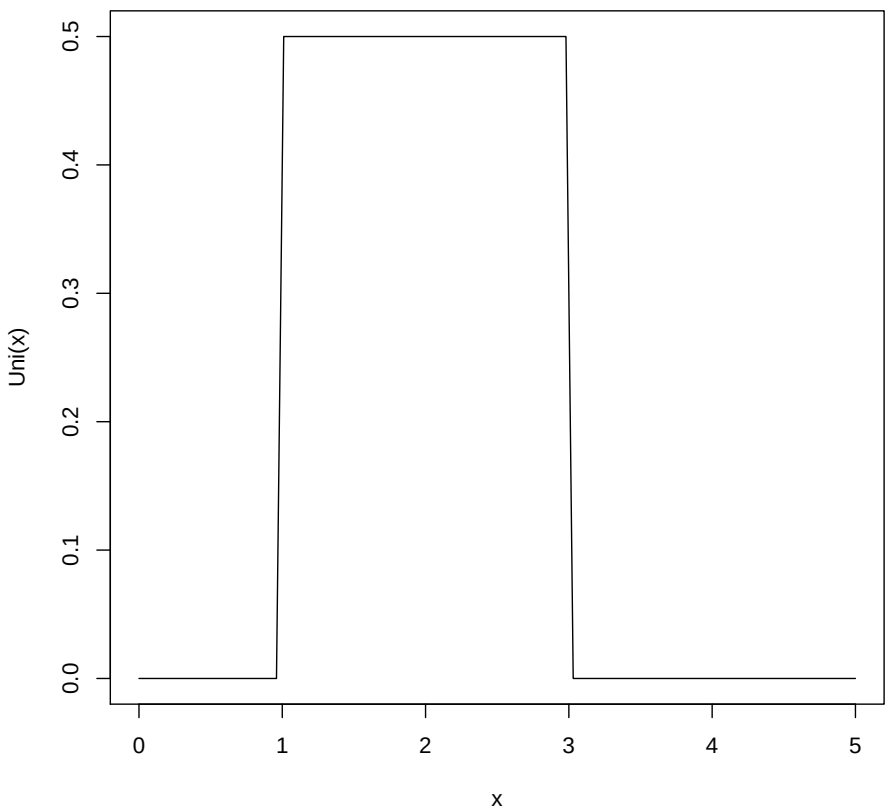
Beispiel:

- Zufallszahlen.
- Wartezeiten auf den Bus.

Stetige Gleichverteilung

Beispiel:

```
xv <- seq(0,5,length=100)
plot(xv, dunif(xv,1,3), type = "l", ylab = "Uni(x)", xlab = "x")
```



Stetige Gleichverteilung

Problem: Bestimmen Sie 10 Zufallszahlen zwischen 1 und 3.

Stetige Gleichverteilung

Antwort: Wir verwenden die Zufallszahlfunktion `runif` der stetigen Gleichverteilung.

```
runif(10, min=1, max=3)
```

```
##  [1] 2.115256 1.553116 2.338847 1.638142 2.121753 2.945975
##  [7] 2.722053 1.858388 2.954451 1.650964
```

Exponentialverteilung

Definition

Die Exponentialverteilung beschreibt die Dauer zwischen zufällig auftretenden Ereignissen. Der einzige Parameter λ steht für die Zahl der erwarteten Ereignisse pro Einheitsintervall. Ihre Dichtefunktion lautet:

$$Exp(x|\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{für } x \geqslant 0 \\ 0 & \text{für } x < 0 \end{cases}$$

Exponentialverteilung

Beispiel:

- Zeit zwischen zwei Anrufen.

Exponentialverteilung

Beispiel:

- Zeit zwischen zwei Anrufen.
- Lebensdauer von Atomen beim radioaktiven Zerfall.

Exponentialverteilung

Beispiel:

- Zeit zwischen zwei Anrufen.
- Lebensdauer von Atomen beim radioaktiven Zerfall.
- Lebensdauer von Bauteilen, Maschinen und Geräten, wenn Alterungserscheinungen nicht betrachtet werden müssen.

Exponentialverteilung

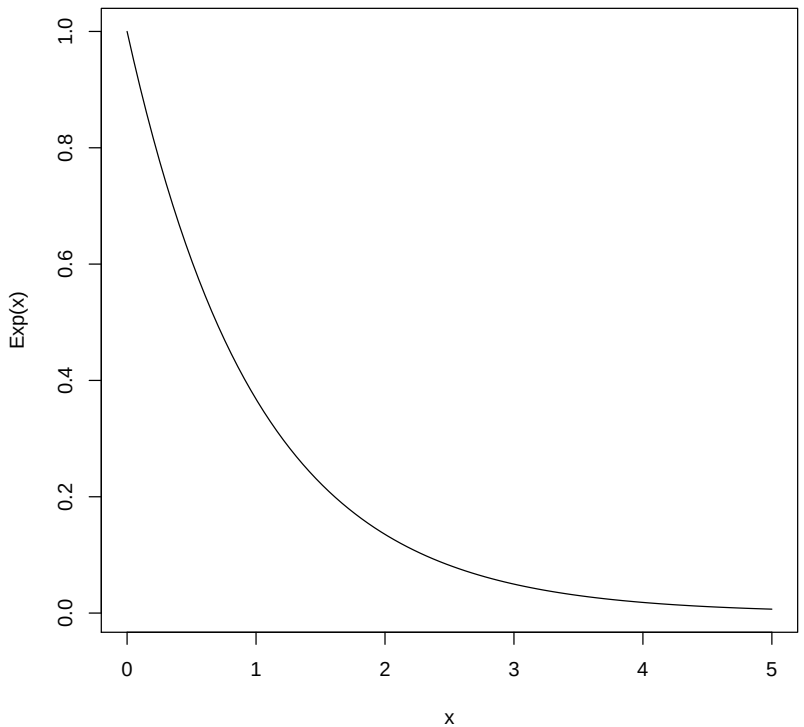
Beispiel:

- Zeit zwischen zwei Anrufen.
- Lebensdauer von Atomen beim radioaktiven Zerfall.
- Lebensdauer von Bauteilen, Maschinen und Geräten, wenn Alterungserscheinungen nicht betrachtet werden müssen.
- als grobes Modell für kleine und mittlere Schäden in Hausrat, Kraftfahrzeug-Haftpflicht, Kasko in der Versicherungsmathematik.

Exponentialverteilung

Beispiel:

```
xv <- seq(0,5,length=100)
plot(xv, dexp(xv, rate=1), type = "l", ylab = "Exp(x)",
xlab = "x")
```



Exponentialverteilung

Problem: Die durchschnittliche Abfertigungszeit an der Kasse eines Supermarktes betrage 3 Minuten. Mit welcher Wahrscheinlichkeit wird ein Kunde in weniger als 2 Minuten bedient?

Exponentialverteilung

Antwort: Die durchschnittliche Anzahl Kunden, die pro Minute bedient werden, beträgt $\lambda = \frac{1}{3}$.

```
pexp(2, rate=1/3)
```

```
## [1] 0.4865829
```

Der Kunde wird mit einer Wahrscheinlichkeit von 48.7% innerhalb von 2 Minuten bedient.

Normalverteilung

Definition

Die Normalverteilung ist wohl die wichtigste Verteilung in der Statistik.

Sie besitzt zwei Parameter, den Mittelwert μ und die

Standardabweichung σ. Ihre Dichtefunktion lautet:

$$N(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

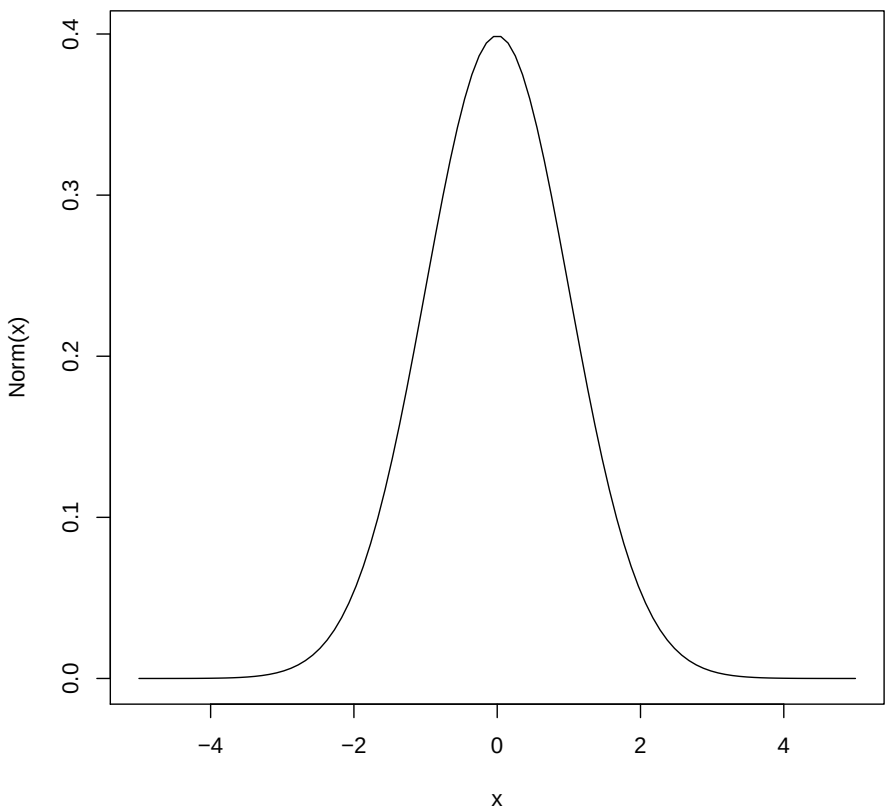
Für die Standardnormalverteilung gilt $\mu = 0$ und $\sigma = 1$,

d.h. $Z \sim N(0,1)$.

Normalverteilung

Beispiel:

```
xv <- seq(-5,5,length=100)

plot(xv, dnorm(xv), type = "l", ylab = "Norm(x)", xlab = "x")
```



Normalverteilung

Problem: Die Ergebnisse eines Abschlusstestes folgen einer Normalverteilung mit $\mu = 72$ und $\sigma = 15.2$. Welcher Anteil der Studierenden erreicht mindestens 84 Punkte?

Normalverteilung

Antwort:

```
pnorm(84, mean=72, sd=15.2, lower.tail=FALSE)
```

```
## [1] 0.2149176
```

Der Anteil der Studierenden, die mindestens 84 Punkte erzielen, beträgt 21.5%.

Chi-Quadrat-Verteilung

Definition

Die Chi-Quadrat-Verteilung wird in Zusammenhang mit Hypothesentest zu Kontingenztabellen und Verteilungsformen verwendet. Sie ist eine stetige Wahrscheinlichkeitsverteilung über der Menge der nicht-negativen reellen Zahlen. Der einzige Parameter ist die Anzahl der Freiheitsgrade df. Ist eine Zufallsvariable X chi-quadrat-verteilt, so gilt:

$$X \sim \chi^2(df)$$

Chi-Quadrat-Verteilung

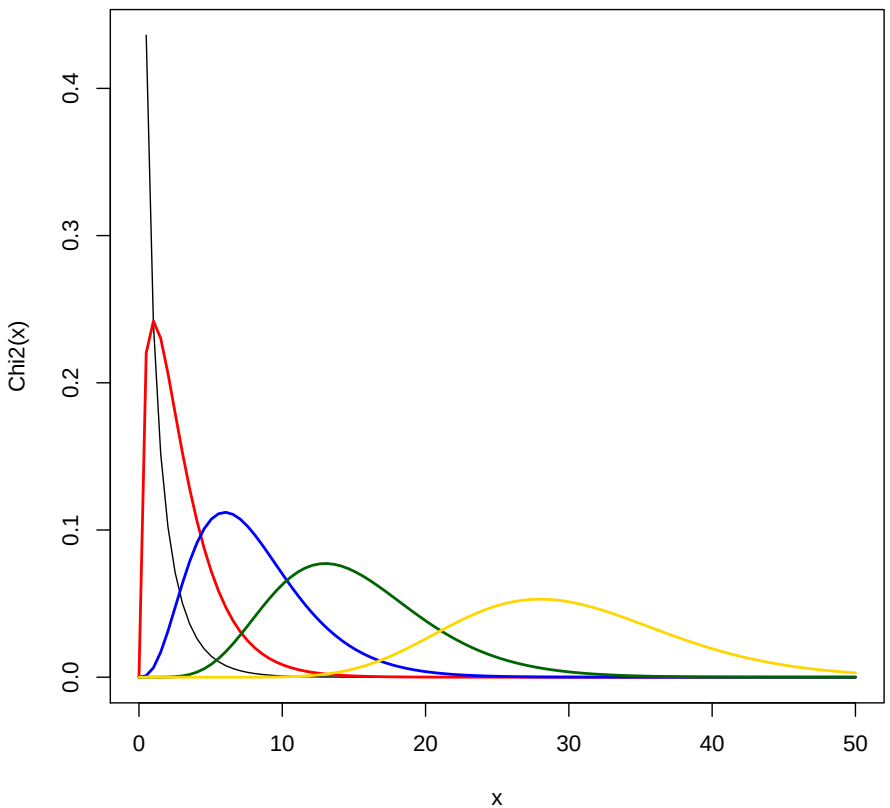
Beispiel:

```
xv <- seq(0,50,length=100)

degf <- c(3,8,15,30)

colors <- c("red", "blue", "darkgreen", "gold")
```

Chi-Quadrat-Verteilung

Beispiel:

```
plot(xv, dchisq(xv, df=1), type = "l", ylab = "Chi2(x)", xlab = "x")
for (i in 1:4){lines(xv, dchisq(xv,degf[i]), lwd=2, col=colors[i])}
```



Chi-Quadrat-Verteilung

Problem: Bestimmen Sie das 95%-Perzentil der χ^2-Verteilung mit Freiheitsgrad 7.

Chi-Quadrat-Verteilung

Antwort:

```
qchisq(.95, df=7)
```

```
## [1] 14.06714
```

Das 95%-Perzentil der χ^2-Verteilung mit $df = 7$ ist 14.067.

Studentsche t-Verteilung

Motivation

Wenn die Standardabweichung σ der Grundgesamtheit unbekannt ist, benutzt man die t-Verteilung (anstatt der Normalverteilung), vorausgesetzt die nötigen Bedingungen sind erfüllt. Die Variable X ist dann t-verteilt mit dem Freiheitsgrad $n - 1$.

$$X \sim t(df)$$

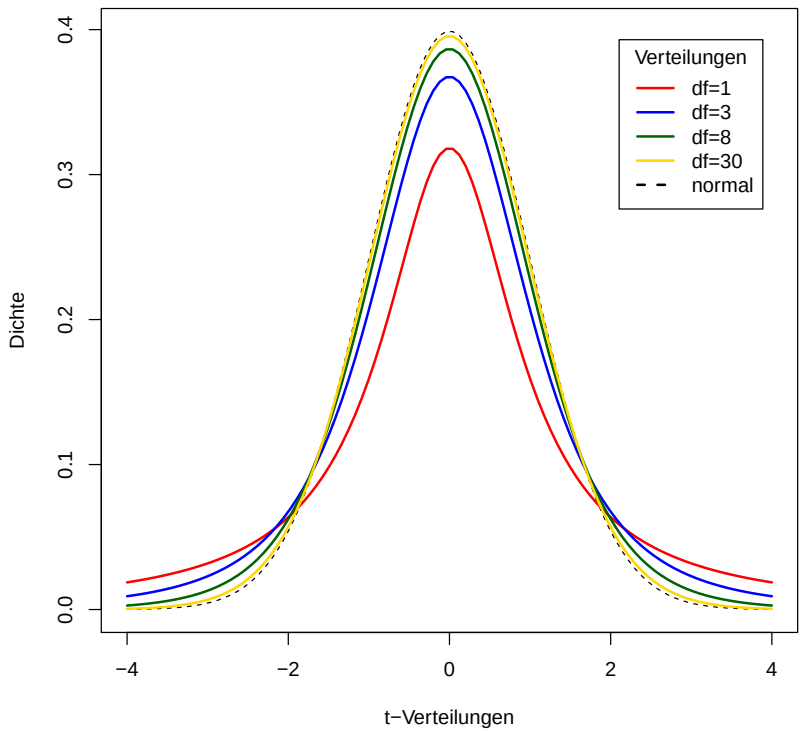
Studentsche t-Verteilung

Beispiel:

```
x <- seq(-4, 4, length=100)
hx <- dnorm(x)
degf <- c(1, 3, 8, 30)
colors <- c("red", "blue", "darkgreen", "gold", "black")
labels <- c("df=1", "df=3", "df=8", "df=30", "normal")
```

Studentsche t-Verteilung

```
plot(x, hx, type="l", lty=2, xlab="t-Verteilungen", ylab="Dichte")
for (i in 1:4){lines(x, dt(x,degf[i]), lwd=2, col=colors[i])}
legend("topright", inset=.05, title="Verteilungen",
labels, lwd=2, lty=c(1, 1, 1, 1, 2), col=colors)
```



Studentsche t-Verteilung

Problem: Bestimmen Sie das 2.5%- und das 97.5%-Perzentil der Studentschen t-Verteilung mit Freiheitsgrad 5.