

```

#####
#
# Reinhard Hoepfner, Statistik mit Rechneruebungen, SoSe19
#
# SANDSTEINPOROSITAET      (57 Daten)
#
# Datensatz stammt aus
#
# Marsal, Statistische Methoden fuer Erdwissenschaftler,
# Schweizerbart-Verlag 1979, Beispiel 2.5, S.7
#
# und der Kontext stammt aus
#
# Engelhard, Der Porenraum der Sedimente, Springer 1960
# Abbildung 11, S. 18
#
# Porositaet von Sandstein aus einem Bohrkern (Emsland)
# Teilbereich entspricht ca. 1100 - 1105 m Bohrtiefe
#
# Porositaet ist definiert als der prozentuale Anteil
# des Porenraums eines Gesteins am Gesamtvolumen
#
# 25.04.19
#
#####

# datensatz
porositaet <- c(
  22.1, 23.5, 25.3, 26.6, 23.9, 26.0, 22.8, 22.3, 23.1, 23.0, 21.0, 21.8,
  22.0, 22.2, 22.3, 22.4, 22.4, 22.4, 22.3, 21.6, 22.1, 22.6, 22.1, 21.9,
  22.3, 23.9, 23.2, 22.5, 23.7, 23.3, 24.4, 22.6, 23.9, 24.2, 27.6, 27.9,
  25.2, 21.7, 20.0, 19.8, 21.5, 25.6, 25.3, 24.1, 28.6, 23.7, 24.0, 21.8,
  24.9, 24.2, 25.0, 23.7, 27.3, 23.0, 23.8, 21.2, 21.1
)
dat <- porositaet ;

#####

dat ;
# elemente aus der liste oder
# teilbereiche des datensatzes herausgreifen
dat[13] ;
dat[24] ;
dat[13:24] ;

sort(dat) ;
# sortiert die daten aufsteigend
rev( sort(dat) ) ;
# sortiert die daten absteigend

summary(dat) ;
# gibt median, mittelwert und quartile
# help(quantile)

c( quantile(dat,0.1) , quantile(dat,0.9) ) ;
# obere und untere 10% quantile

#####

# boxplots zur darstellung des datensatzes
boxplot( dat , col=5 ) ;
title( main = "datensatz sandsteinporositaet" ) ;

# was zeigt der plot :
abline( h = min(dat), lty=2, col=4 ) ;

```

```

abline( h = max(dat), lty=2, col=4 ) ;

abline( h = median(dat), lty=2, col=4 ) ;
abline( h = quantile(dat,0.75), lty=2, col=4 ) ;
abline( h = quantile(dat,0.25), lty=2, col=4 ) ;

# 'range' skaliert einen bereich,
# jenseits dessen 'extreme' beobachtungen einzeln angezeigt werden
# definiert in abhaengigkeit von median und quartilsabstand
# (mathematisch wenig befriedigend)
boxplot( dat, range=0.7, col=4 ) ;
boxplot( dat, range=0.5, col=4 ) ;
boxplot( dat, range=0.3, col=4 ) ;
title( main = "datensatz sandsteinporositaet" ) ;

# verglichen mit normalverteilten daten
# mit vergleichbarem mittelwert und varianz
# ist der datensatz 'sandsteinporositaet' einseitig schief
datneu <- rnorm( length(dat), mean(dat), sqrt(var(dat)) ) ;
# R verlangt eingabe des skalenfaktors, also der standardabweichung
# help( rnorm ) ;

boxplot( datneu, range=0.3, col=4 ) ;
title( main = "normalverteilter vergleichsdatensatz" ) ;

# im paarweisen vergleich koennte der datensatz 'sandsteinporositaet'
# wohl als normalverteilter datensatz durchgehen
# mehrmals wiederholen
par( mfrow=c(1,2) ) ;
#
boxplot( dat, range=0.3, col=4 ) ;
title( main = "datensatz sandsteinporositaet" ) ;
#
datneu <- rnorm( length(dat), mean(dat), sqrt(var(dat)) ) ;
boxplot( datneu, range=0.3, col=4 ) ;
title( main = "normalverteilter vergleichsdatensatz" ) ;
#
par( mfrow=c(1,1) ) ;
# mehrmals wiederholen

##### ende 25.04.19 #####

```