Testen

Aufgabe 1

Einlesen der Daten

```
atem <- read.table("http://www.stat.uni-muenchen.de/service/datenarchiv/atem/atemwege.asc"
                   ,header=T,na.strings="-1")
```

a) **Einstichprobenfall: Test**

```
summary(atem)

##        nr              zone          aller            kehle
##  Min.   :   1.0   Min.   :1.00   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 413.0   1st Qu.:2.00   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 821.0   Median :2.00   Median :0.0000   Median :0.0000
##  Mean   : 824.8   Mean   :2.36   Mean   :0.1207   Mean   :0.1078
##  3rd Qu.:1239.0   3rd Qu.:3.00   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1652.0   Max.   :3.00   Max.   :1.0000   Max.   :1.0000
##
##      raumu             rauva            sozio           schnu
##  Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :2.000   Median :0.0000
##  Mean   :0.2021   Mean   :0.2718   Mean   :1.943   Mean   :0.1821
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :3.000   Max.   :1.0000
##
##      huste             gebmo            gebtg           gebja
##  Min.   :0.0000   Min.   : 1.000   Min.   : 1.00   Min.   :73.0
##  1st Qu.:0.0000   1st Qu.: 3.000   1st Qu.: 8.00   1st Qu.:76.0
##  Median :0.0000   Median : 6.000   Median :15.00   Median :79.0
##  Mean   :0.1039   Mean   : 6.394   Mean   :15.64   Mean   :78.4
##  3rd Qu.:0.0000   3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:80.0
##  Max.   :1.0000   Max.   :12.000   Max.   :31.00   Max.   :82.0
##
##      gross            sex            untmo           unttg           untja
##  Min.   :105.0   Min.   :1.000   Min.   :1.000   Min.   : 1.00   Min.   :1989
##  1st Qu.:130.0   1st Qu.:1.000   1st Qu.:1.000   1st Qu.: 9.00   1st Qu.:1989
##  Median :140.0   Median :1.000   Median :2.000   Median :20.00   Median :1989
##  Mean   :141.3   Mean   :1.481   Mean   :1.977   Mean   :18.43   Mean   :1989
##  3rd Qu.:152.0   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:27.00   3rd Qu.:1989
##  Max.   :188.0   Max.   :2.000   Max.   :4.000   Max.   :31.00   Max.   :1989
##
##       gewi            fvc             pef              fef50
##  Min.   :16.00   Min.   :1.010   Min.   : 1.760   Min.   :1.110
##  1st Qu.:27.00   1st Qu.:1.910   1st Qu.: 4.050   1st Qu.:2.390
##  Median :33.00   Median :2.330   Median : 4.930   Median :2.930
##  Mean   :36.04   Mean   :2.511   Mean   : 5.161   Mean   :3.100
##  3rd Qu.:43.00   3rd Qu.:3.000   3rd Qu.: 5.990   3rd Qu.:3.643
##  Max.   :95.00   Max.   :6.050   Max.   :12.840   Max.   :8.760
```

```
##  NA's   :10      NA's   :221     NA's   :221      NA's   :221
##      fef75           lubro
##  Min.   :0.440   Min.   :0.0000
##  1st Qu.:1.110   1st Qu.:0.0000
##  Median :1.400   Median :0.0000
##  Mean   :1.498   Mean   :0.3932
##  3rd Qu.:1.760   3rd Qu.:1.0000
##  Max.   :5.210   Max.   :1.0000
##  NA's   :221
```

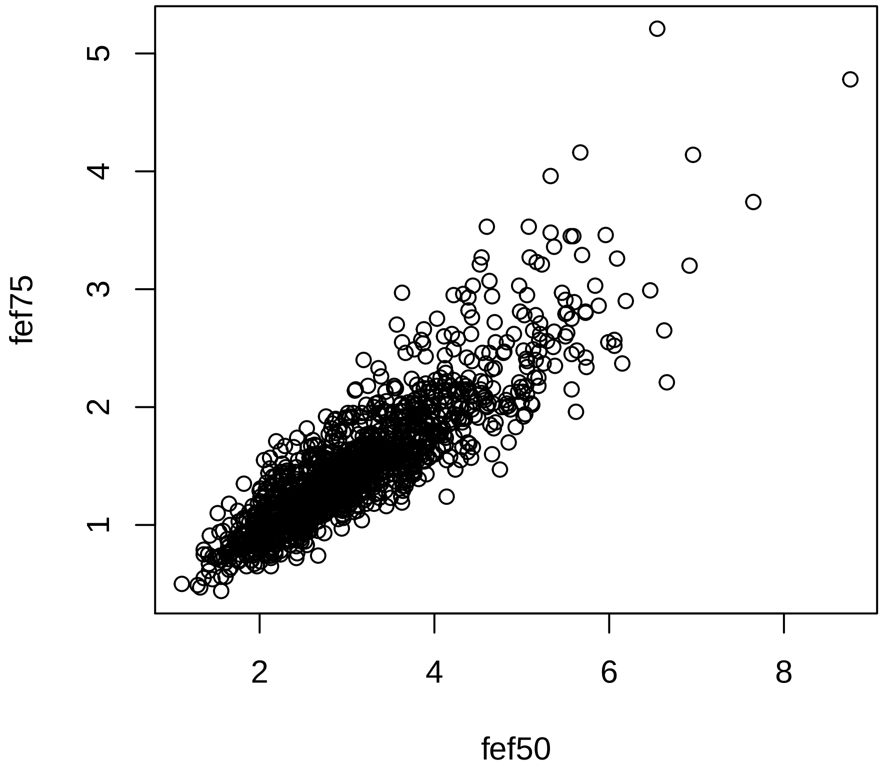
```
atem.a <- atem[complete.cases(atem$pef,atem$fef50,atem$fef75,atem$sex),]
summary(atem.a)
```

```
##        nr              zone          aller            kehle
##  Min.   :   1.0   Min.   :1.00   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 381.8   1st Qu.:2.00   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 748.5   Median :2.00   Median :0.0000   Median :0.0000
##  Mean   : 782.4   Mean   :2.35   Mean   :0.1227   Mean   :0.1054
##  3rd Qu.:1177.2   3rd Qu.:3.00   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1651.0   Max.   :3.00   Max.   :1.0000   Max.   :1.0000
##
##      raumu            rauva            sozio           schnu
##  Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :2.000   Median :0.0000
##  Mean   :0.2108   Mean   :0.2718   Mean   :1.932   Mean   :0.1657
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :3.000   Max.   :1.0000
##
##      huste            gebmo            gebtg           gebja
##  Min.   :0.0000   Min.   : 1.000   Min.   : 1.00   Min.   :73.00
##  1st Qu.:0.0000   1st Qu.: 3.000   1st Qu.: 8.00   1st Qu.:76.00
##  Median :0.0000   Median : 6.000   Median :16.00   Median :79.00
##  Mean   :0.1002   Mean   : 6.422   Mean   :15.74   Mean   :78.26
##  3rd Qu.:0.0000   3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:80.00
##  Max.   :1.0000   Max.   :12.000   Max.   :31.00   Max.   :82.00
##
##      gross            sex            untmo           unttg           untja
##  Min.   :105.0   Min.   :1.000   Min.   :1.000   Min.   : 1.00   Min.   :1989
##  1st Qu.:130.0   1st Qu.:1.000   1st Qu.:1.000   1st Qu.: 9.00   1st Qu.:1989
##  Median :141.0   Median :1.000   Median :2.000   Median :20.00   Median :1989
##  Mean   :142.1   Mean   :1.466   Mean   :1.835   Mean   :18.21   Mean   :1989
##  3rd Qu.:153.0   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:25.00   3rd Qu.:1989
##  Max.   :188.0   Max.   :2.000   Max.   :4.000   Max.   :31.00   Max.   :1989
##
##       gewi            fvc             pef              fef50
##  Min.   :16.00   Min.   :1.010   Min.   : 1.760   Min.   :1.110
##  1st Qu.:27.00   1st Qu.:1.910   1st Qu.: 4.050   1st Qu.:2.390
##  Median :34.00   Median :2.330   Median : 4.930   Median :2.930
##  Mean   :36.64   Mean   :2.511   Mean   : 5.161   Mean   :3.100
##  3rd Qu.:44.00   3rd Qu.:3.000   3rd Qu.: 5.990   3rd Qu.:3.643
##  Max.   :95.00   Max.   :6.050   Max.   :12.840   Max.   :8.760
##  NA's   :6
##      fef75           lubro
##  Min.   :0.440   Min.   :0.0000
##  1st Qu.:1.110   1st Qu.:0.0000
##  Median :1.400   Median :0.0000
##  Mean   :1.498   Mean   :0.3833
##  3rd Qu.:1.760   3rd Qu.:1.0000
##  Max.   :5.210   Max.   :1.0000
##
```

```
attach(atem.a)
```

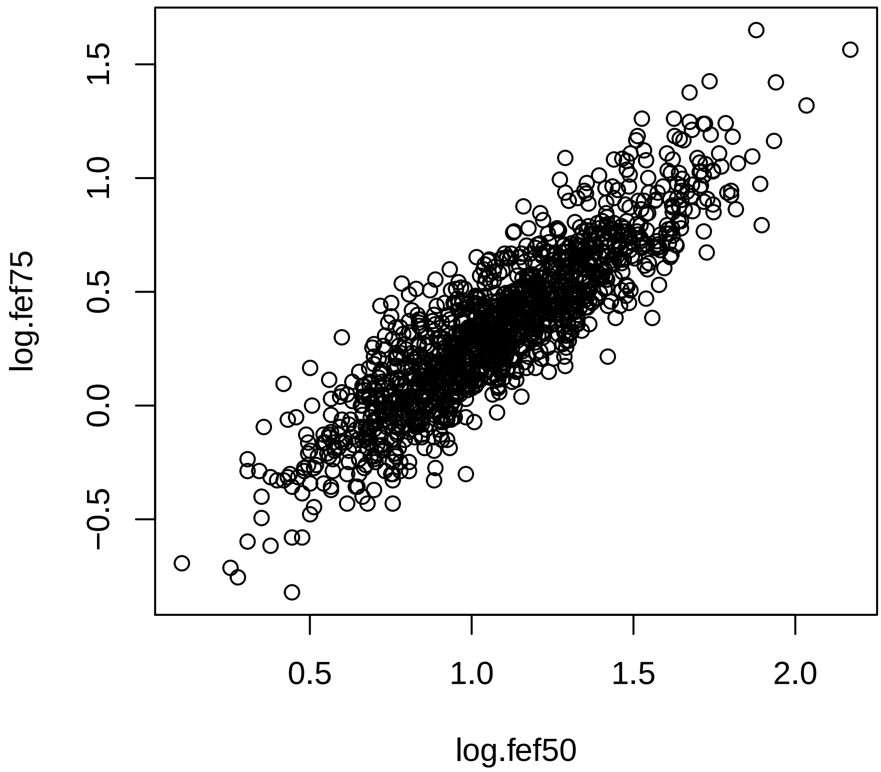
Originalvariablen fef50 und fef75 genügen nicht der Normalverteilung. Die marginalen Verteilungen erscheinen eher linkssteil:

```
plot(fef50,fef75)
```



Logarithmierte Variablen:

```
log.fef50<-log(fef50)
log.fef75<-log(fef75)
log.fef<-cbind(log.fef50,log.fef75)
plot(log.fef50,log.fef75)
```



```r
summary(log.fef)
```

```
##    log.fef50        log.fef75
##  Min.   :0.1044   Min.   :-0.8210
##  1st Qu.:0.8713   1st Qu.: 0.1044
##  Median :1.0750   Median : 0.3365
##  Mean   :1.0843   Mean   : 0.3424
##  3rd Qu.:1.2927   3rd Qu.: 0.5653
##  Max.   :2.1702   Max.   : 1.6506
```

```r
cor(log.fef)
```

```
##           log.fef50 log.fef75
## log.fef50 1.0000000 0.8846577
## log.fef75 0.8846577 1.0000000
```

Beträchtliche Korrelation zwischen den beiden Variablen!

i) **Kovarianz unbekannt**
Univariate t-Tests:

```r
t.test(log.fef50,mu=log(3),conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  log.fef50
## t = -1.7144, df = 1327, p-value = 0.08669
## alternative hypothesis: true mean is not equal to 1.098612
## 99 percent confidence interval:
##  1.062643 1.105859
## sample estimates:
## mean of x
##  1.084251
```

```r
t.test(log.fef75,mu=log(1.4),conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  log.fef75
## t = 0.623, df = 1327, p-value = 0.5334
## alternative hypothesis: true mean is not equal to 0.3364722
## 99 percent confidence interval:
##  0.3177716 0.3670822
## sample estimates:
## mean of x
## 0.3424269
```

Multivariater Test für Erwartungswert im Ein-Stichproben-Fall (Kovarianz unbekannt):

```r
n <- dim(atem.a)[1]
p <- 2
x.quer <- c(mean(log.fef50),mean(log.fef75)) # Erwartungswertvektor
mu0 <- c(log(3),log(1.4))
S <- var(log.fef)                            # geschätzte Kovarianzmatrix
```

Teststatistik:

```
T.s <- n*(n-p)/(p*(n-1))*(x.quer-mu0)%*%solve(S) %*%(x.quer-mu0)
print(T.s)

##          [,1]
## [1,] 11.99079

print(qf(0.99,p,n-p))

## [1] 4.621201

# p-Wert:
print(pf(T.s,p,n-p,lower.tail=F))  # lower.tail=F bedetet P[X > x]

##              [,1]
## [1,] 6.902319e-06
```

⇒ Mittelwertsvektor signifikant verschieden von mu0 Man erkennt: deutlich geringerer p-Wert des multivariaten Tests als bei univariaten Tests (hier wird Information über starke Korrelation ausgenutzt)

ii) **Kovarianz bekannt**
Univariate t-Tests:

```
T1 <- (mean(log.fef50,na.rm=T)-log(3))*sqrt(n)/sqrt(0.09)
print(abs(T1))      # Betrag der Teststatistik

## [1] 1.744487

print(qt(0.99,n-1)) # Quantil der t-Verteilung

## [1] 2.329161

print(pt(T1,n-1,lower.tail=F))  # p-Wert

## [1] 0.9593471
```

```
T2 <- (mean(log.fef75,na.rm=T)-log(1.4))*sqrt(n)/sqrt(0.12)
print(abs(T2))      # Betrag der Teststatistik

## [1] 0.6264178

print(qt(0.99,n-1)) # Quantil der t-Verteilung

## [1] 2.329161

print(pt(T2,n-1,lower.tail=F))  # p-Wert

## [1] 0.2655743
```

Multivariater Test für Erwartungswert im Ein-Stichproben-Fall (Kovarianz bekannt): Komponenten n,p,x.quer,mu0 wie bei i)

```
Sigma <- matrix(c(0.09,0.07,0.07,0.12),2,2)
Z.s<-n*(x.quer-mu0)%*%solve(Sigma) %*%(x.quer-mu0)
print(Z.s)

##          [,1]
## [1,] 8.983716

print(qchisq(0.99,p))

## [1] 9.21034

print(pchisq(Z.s,p,lower.tail=F))
```

```
##             [,1]
## [1,] 0.01119981
```

⇒ Mittelwertsvektor nicht verschieden von mu0

b) **Einstichprobenfall: Konfidenzbereich**

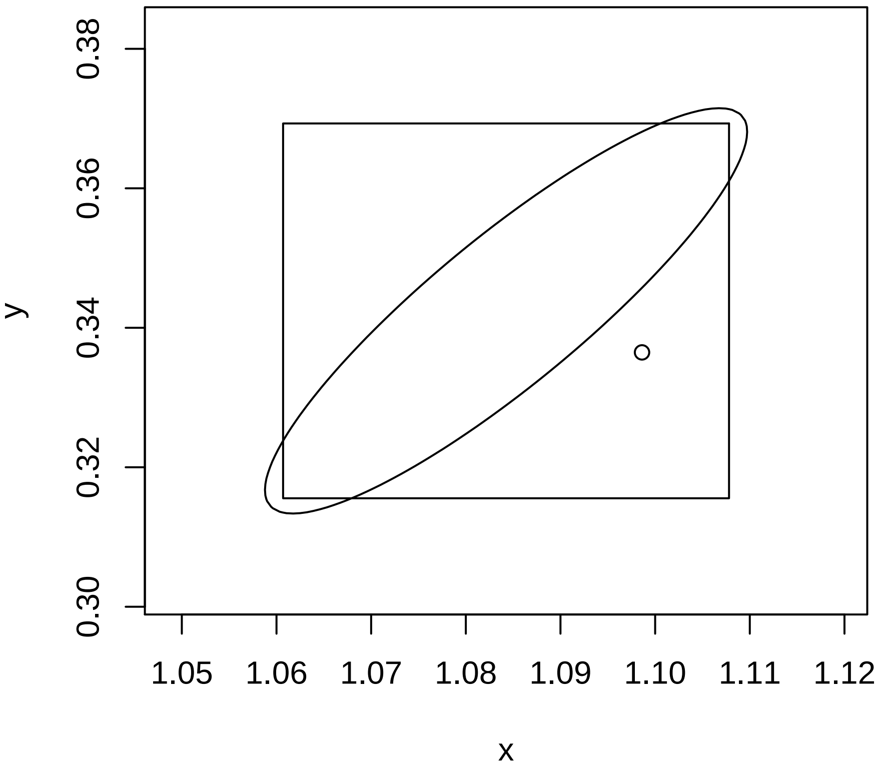
Funktion zur Berechnung der Mahalanobis-Distanz D^2 (Ellipse):

```
d2.function<-function(mu1,mu2,S,xquer){

  Sinv <- solve(S)
  res <-  Sinv[1,1]*(xquer[1]-mu1)^2+
          (Sinv[1,2]+ Sinv[2,1])*(xquer[2]-mu2)*(xquer[1]-mu1)+
          Sinv[2,2]*(xquer[2]-mu2)^2
  return(res)
}
```

Funktion zum Plotten der Konfidenzbereiche:

```
draw.confi<-function(S=diag(rep(1,2)),n,alpha,xquer=rep(0,2),bonf=TRUE,...){

  p <- 2
  if (bonf){
  # Bonferroni: xi.quer+-t(n,1-alpha_i/2)*sqrt(Sii/n)
  x <- qt(1-alpha/(p*2),n-1)*sqrt(S[1,1]/n)
  y <- qt(1-alpha/(p*2),n-1)*sqrt(S[2,2]/n)
  }
  else
  {
  x <- sqrt(p*(n-1)*qf(1-alpha/(p*2),p,n-p)/(n-p))*sqrt(S[1,1]^2/n)
  y <- sqrt(p*(n-1)*qf(1-alpha/(p*2),p,n-p)/(n-p))*sqrt(S[2,2]^2/n)
  }
  # Bonferroni-Rechteck:
  plot(x=c(xquer[1]+x,xquer[1]+x,xquer[1]-x,xquer[1]-x,xquer[1]+x),
          y=c(xquer[2]+y,xquer[2]-y,xquer[2]-y,xquer[2]+y,xquer[2]+y),
          type="l",xlab="x",ylab="y",xlim=c(xquer[1]-1.5*x,xquer[1]+1.5*x),
          ylim=c(xquer[2]-1.5*y,xquer[2]+1.5*y),...)

  #Erstellen eines Grids für die Konfidenzellipsoide
  x.seq<-seq((xquer[1]-1.5*x),xquer[1]+1.5*x,length=100)
  y.seq<-seq((xquer[2]-1.5*y),(xquer[2]+1.5*y),length=100)

  #Berechnung des äußeren Produkts von x.seq und y.seq, Berechnung der
  #Mahalanobis-Distanz mit diesen Werten
  z<-outer(x.seq,y.seq,d2.function,S=S,xquer=xquer)

  #Einzeichnen der Ellipse
  ellipse<-contourLines(x.seq,y.seq,z,levels=(n-1)*p/((n-p)*n)*qf(1-alpha,p,n-p))
  lines(ellipse[[1]]$x,ellipse[[1]]$y)

}
```

```
n<-length(log.fef[complete.cases(log.fef),1])
p<-2
alpha<-0.01

draw.confi(S=S,n=n,alpha=alpha,xquer=x.quer)
points(mu0[1],mu0[2])
```



c) **Einstichprobenfall: Symmetrie Test**

```
x3<-log(pef)
x2<-log(fef50)
x1<-log(fef75)

y<-cbind(x1-x3,x2-x3)
yquer<- c(mean(y[,1],na.rm=T),mean(y[,2],na.rm=T))
yquer

## [1] -1.2580644 -0.5162401

S <- var(y,use="complete.obs")
S

##            [,1]       [,2]
## [1,] 0.06934678 0.03775723
## [2,] 0.03775723 0.03254871

T.s<-(n-p)*n/(p*(n-1))*(yquer)%*%solve(S) %*%(yquer)
print(T.s)                                # Betrag der Teststatistik

##          [,1]
## [1,] 16718.75

print(qf(0.99,p-1,n-p+1))                 # Quantil der F-Verteilung

## [1] 6.654023

print(pf(T.s,p-1,n-p-1,lower.tail=F)) # p-Wert

##      [,1]
## [1,]    0
```

$\Rightarrow H_0$ wird abgelehnt

d) **Zweistichprobenfall: Test**

```
atemvar<-cbind(fvc,pef,fef50,fef75)
summary(atemvar)
```

```
##       fvc             pef             fef50           fef75
##  Min.   :1.010   Min.   : 1.760   Min.   :1.110   Min.   :0.440
##  1st Qu.:1.910   1st Qu.: 4.050   1st Qu.:2.390   1st Qu.:1.110
##  Median :2.330   Median : 4.930   Median :2.930   Median :1.400
##  Mean   :2.511   Mean   : 5.161   Mean   :3.100   Mean   :1.498
##  3rd Qu.:3.000   3rd Qu.: 5.990   3rd Qu.:3.643   3rd Qu.:1.760
##  Max.   :6.050   Max.   :12.840   Max.   :8.760   Max.   :5.210
```

Funktion zur Berechnung eines multivariaten Tests im Zwei-Stichproben-Fall für unabhängige Stichproben.

```
T2test <- function(varmat,gruppe,gruppenwerte=c(1,0)){

  #Argumente:
  #varmat = Matrix der Variablen
  #gruppe = Variable, nach der gruppiert werden soll
  #gruppenwerte = Ausprägung der Gruppenwerte

  #Reduktion auf vollständige Beobachtungen
  varmatgruppe<-cbind(varmat,gruppe)
  varmat<-varmat[complete.cases(varmatgruppe),]
  gruppe<-gruppe[complete.cases(varmatgruppe)]
  #Anzahl der Komponenten der Erwartungswertvektoren \mu_1 und \mu_2
  p<-dim(varmat)[2]

  #Berechnung der Mittelwertvektoren in den Gruppen
  x1.quer<-apply(varmat[gruppe==gruppenwerte[1],],MARGIN=2,FUN=mean)
  x2.quer<-apply(varmat[gruppe==gruppenwerte[2],],MARGIN=2,FUN=mean)

  #Berechnung Stichprobenumfänge in den Gruppen
  n1<-sum(gruppe==gruppenwerte[1])
  n2<-sum(gruppe==gruppenwerte[2])

  #Berechnung der Kovarianzmatrizen in den Gruppen
  S1<-var(varmat[gruppe==gruppenwerte[1],])
  S2<-var(varmat[gruppe==gruppenwerte[2],])

  #Berechnung der gepoolten Kovarianzmatrix
  S<-1/(n1+n2-2)*((n1-1)*S1+(n2-1)*S2)

  #Teststatistik T^2 (vgl. Aufgabe 1)
  T2<-n1*n2/(n1+n2)*(x1.quer-x2.quer)%*%solve(S)%*%(x1.quer-x2.quer)
  # p-Wert
  pwert<-pf((n1+n2-p-1)/((n1+n2-2)*p)*T2,p,n1+n2-p-1,lower.tail=F)
  erg<-list("T2"=T2,"pwert"=pwert)
  return(erg)
}
```

Test für Gruppenunterschiede bzgl. Geschlecht (1: männlich, 2: weiblich)

```
summary.factor(sex)
```

```
##   1   2
## 709 619
```

Univariate Tests

```
t.test(fvc[sex==1],fvc[sex==2],var.equal=T)

##
##  Two Sample t-test
##
## data:  fvc[sex == 1] and fvc[sex == 2]
## t = 8.1272, df = 1326, p-value = 9.978e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2740413 0.4484332
## sample estimates:
## mean of x mean of y
##  2.679379  2.318142

t.test(pef[sex==1],pef[sex==2],var.equal=T)

##
##  Two Sample t-test
##
## data:  pef[sex == 1] and pef[sex == 2]
## t = 5.8953, df = 1326, p-value = 4.74e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3204372 0.6400602
## sample estimates:
## mean of x mean of y
##  5.384401  4.904152

t.test(fef50[sex==1],fef50[sex==2],var.equal=T)

##
##  Two Sample t-test
##
## data:  fef50[sex == 1] and fef50[sex == 2]
## t = 2.4227, df = 1326, p-value = 0.01554
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.02493319 0.23717016
## sample estimates:
## mean of x mean of y
##  3.161100  3.030048

t.test(fef75[sex==1],fef75[sex==2],var.equal=T)

##
##  Two Sample t-test
##
## data:  fef75[sex == 1] and fef75[sex == 2]
## t = 0.7599, df = 1326, p-value = 0.4475
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03677108  0.08326634
## sample estimates:
## mean of x mean of y
##  1.508886  1.485638
```

$\Rightarrow$ bis auf fef75 signifikante Unterschiede zum 0.05-Niveau!

Multivariate Tests

```
T2test(atemvar[,3:4],gruppe=sex,gruppenwerte=c(1,2))
```

```
## $T2
##          [,1]
## [1,] 13.93985
##
## $pwert
##              [,1]
## [1,] 0.0009796528
```

⇒ Unterschiede 'hochsignifikant'!

```
detach(atem.a)
```