Multivariate Statistics EX1

203832050 and 312584204

April 15, 2019

1

a. given: $n = 70, \alpha = 0.05, p = 10$

$$\begin{cases} H_0 : \mu = (-0.9, 0, -0.9, 1, -1, 1, -2, 0, 0, 0)^T \\ H_1 : \text{otherwise} \end{cases}$$

The statistic test for checking hypothesis on the expectation when the Covariance Matrix is unknown is:

$$T^2_{Hotteling} = n * (\bar{x} - \mu_0)^T S^{-1} (\bar{x} - \mu_0)$$

In addition we saw that $c * T^2_{Hotteling} \sim F_{1-\alpha,p,n-p}$

where $c = \frac{n-p}{p*(n-1}$

After sampling 70 observations we got:

$$F_{st} = c * T^2_{Hotteling} = 1.148508 < 1.99259 = F_{0.95,10,60} = F_{cr}$$

Therefore we won't reject H_0 and we say that: $\mu = (-0.9, 0, -0.9, 1, -1, 1, -2, 0, 0, 0)^T$

The $p - value = 0.3430566$

```
param<-read.csv(choose.files(),header=F)
mu<-param[,1]
sig<-matrix(nrow = 10,ncol=10)
for (i in 2:11) {
  sig[,i-1]<-param[,i]
}
set.seed(123)
p<-10
n<-70
data1<-rmvnorm(n,mu,sig)
mu0<-c(-0.9,0,-0.9,1,-1,1,-2,0,0,0)
xavg<-colMeans(data1)
s<-cov(data1)
Thotteling<-n*t((xavg-mu0))%*%inv(s)%*%(xavg-mu0)
c<-(n-p)/(p*(n-1))
```

```
(statisti<-c*Thotteling)
Fcritical<-qf(0.95,p,n-p)
statisti>Fcritical
(pval<-1-pf(statisti,p,n-p))
```

b. given: n=200

$H_1 : \mu = (1,1,1,1,1,1,1,1,1,1)^T$

We will estimate the power by the MLE (maximum likelihood estimator) which is the ratio between the number of rejection to number of trials:

$$\hat{\pi} = \frac{rejections}{trials}$$

now, $F_{cr} = F_{0.95,10,190} = 1.8808$

We can see that we rejected H_0 29 times, then:

$$\hat{\pi} = \frac{29}{100} = 0.29$$

```
set.seed(523511)
power<-0
n<-200
mu<-rep(1,10)
Fcritical<-qf(0.95,p,n-p)
c<-(n-p)/(p*(n-1))
for (i in 1:100) {
  data<-mvrnorm(n,mu,sig)
  xavg<-colMeans(data)
  s<-cov(data)
  Thotteling<-n*t((xavg-mu0))%*%solve(s)%*%(xavg-mu0)
  (statisti<-c*Thotteling)
  if(statisti>Fcritical){
    power<-power+1
  }
}
(power<-power/100)
```

c. given: same samples from task a

$$\begin{cases} H_0 : \mu_1 = \mu_2, \mu_4 = \mu_5 \\ H_1 : \text{otherwise} \end{cases}$$

We define a matrix B:

$$B = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Now, the hypothesis is equivalent to the following hypothesis:

$$\begin{cases} H_0 : B * \mu = 0 \\ H_1 : \text{otherwise} \end{cases}$$

As we saw in the class we need to do a linear transform to the sample: $Y_i = B * X_i$ and $Y \sim N_2(B\mu, B\Sigma B^T)$.

$$F_{cr} = F_{0.95,2,68} = 3.1316$$

$$F_{st} = c * T^2_{Hotteling} = \frac{70-2}{2*(70-1)} * T^2_{Hotteling} = 0.3617974$$

Therefore we won't reject H_0.

The $p - value = 0.6977559$.

```
n<-70
set.seed(123)
data1<-mvrnorm(n,mu,sig)
B<-matrix(NA,2,4)
B[1,]<-c(1,-1,0,0)
B[2,]<-c(0,0,1,-1)
mu0<-rep(0,2)
xavg<-colMeans(data1)[c(1,2,4,5)]
Sx<-cov(data1)[c(1,2,4,5),c(1,2,4,5)]
Sy<-B%*%Sx%*%t(B)
Thotteling<-n*t(B%*%xavg)%*%solve(Sy)%*%B%*%xavg
c<-(n-2)/(2*(n-1))
(statisti<-c*Thotteling)
Fcritical<-qf(0.95,2,n-2)
(statisti>Fcritical)
(pval<-1-pf(statisti,2,n-2))
```

2

a. given: n=100, p=3.

$$\begin{cases} H_0 : \Sigma \quad is \quad diagonal \\ H_1 : \text{otherwise} \end{cases}$$

As we saw in class, according to Willks theorem:

$$-n * log(det(R)) \sim \chi^2_{\frac{p(p-1)}{2}}$$

where R is the correlation matrix.

After we estimate the correlation matrix according to the samples we got:

$$\chi^2_{st} = -n * log(det(R)) = -100 * log(-0.981308) = 1.886895$$
$$\chi^2_{cr} = \chi^2_{0.95,3} = 7.814$$

$\chi^2_{st} < \chi^2_{cr}$ therefore we do not reject H_0 and we say that Σ is diagonal

The $p - value = 0.5962105$.

```
n<-100
p<-3
set.seed(124)
mu<-c(0,1,0)
sig<-matrix(c(16,1,1,1,16,2,1,2,16), nrow = 3, ncol = 3)
data<-mvrnorm(n,mu,sig)
s<-cov(data)
s0<-matrix(c(s[1,1],0,0,0,s[2,2],0,0,0,s[3,3]),3,3)
R<-cor(data)
(statisti<--n*log(det(R)))
chiCritical<-qchisq(0.95,p*(p-1)/2)
(statisti>chiCritical)
(pval<-1-pchisq(statisti,3))
```

b. We will estimate the power in the following point:

$$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

The process is similar to question 1 task b: MLE of the power. Now the comparison is between:

$$\chi^2_{st} = -n * log(det(R)) \quad to \quad \chi^2_{0.95,3}$$

We can see that we rejected H_0 100 times, then:

$$\hat{\pi} = 1$$

```
mu<-c(0,0,0)
sig<-matrix(c(2,0,1,0,2,0,1,0,2),3,3)
power<-0
n<-200
p<-3
chiCritical<-qchisq(0.95,p*(p-1)/2)
for (i in 1:100) {
  data<-mvrnorm(n,mu,sig)
  R<-cor(data)
  statisti<--n*log(det(R))
  if(statisti>chiCritical)
    power<-power+1
}
(power<-power/100)
```

3

a. given n, $x_1, x_2, ..., x_n \sim N_2(\mu, \sigma^2 I)$ i.i.d.

$$\begin{cases} H_0 : \mu = (0,0)^T \\ H_1 : \text{otherwise} \end{cases}$$

We know that the MLE for the expectation without constraints is: $\bar{X}$ so:

$$L(\mu_1) = (2\pi * |\sigma^2 I|)^{-n/2} * \exp^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^T \Sigma^{-1}(x_i - \bar{x})}$$

and:

$$L(\mu_0) = (2\pi * |\sigma^2 I|)^{-n/2} * \exp^{-\frac{1}{2}\sum_{i=1}^{n}(x_i)^T \Sigma^{-1}(x_i)}$$

Therefore, the generalize likelihood ratio is (when σ is known):

$$\frac{L(\mu_1)}{L(\mu_0)} = \exp^{\frac{1}{2}\sum_{i=1}^{n}[(x_i)^T \frac{1}{\sigma^2} I(x_i) - (x_i - \bar{x})^T \frac{1}{\sigma^2} I(x_i - \bar{x})]}$$

$$2log(\frac{L(\mu_1)}{L(\mu_0)}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}[(x_i)^T(x_i) - (x_i - \bar{x})^T(x_i - \bar{x})]$$

Because $x_i = (x_{1i}, x_{2i})^T, \bar{x} = (\bar{x_1}, \bar{x_2})^T$ we get:

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}[2x_{1i}\bar{x_1} - \bar{x_1}^2 + 2x_{2i}\bar{x_2} - \bar{x_2}^2] = \frac{1}{\sigma^2}\sum_{i=1}^{n}\bar{x_1}(x_{1i} - \bar{x_1}) + \bar{x_1} * x_{1i} + \bar{x_2}(x_{2i} - \bar{x_2}) + \bar{x_2} * x_{2i}$$

We notice that:

$$(1)\sum_{i=1}^{n}(x_{ji} - \bar{x_j}) = 0 \quad \forall j \qquad (2)\sum_{i=1}^{n}x_i = n\bar{x}$$

so:

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}\bar{x_1} * x_{1i} + \bar{x_2} * x_{2i} = \frac{n(\bar{x_1}^2 + \bar{x_2}^2)}{\sigma^2} = (\frac{\sqrt{n}\bar{x_1}}{\sigma})^2 + (\frac{\sqrt{n}\bar{x_2}}{\sigma})^2$$

by (1) and (2).

Under H_0 $\frac{\sqrt{n}\bar{x_1}}{\sigma}, \frac{\sqrt{n}\bar{x_2}}{\sigma} \sim N(0,1)$ and by definition sum of squared random variables standard normal distribution is distributed χ^2 and that's what we asked to show.

b. We will estimate the power in the following point:

$$\mu = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The process is the same. We chose to do 50,000 replications. In each replication we check if the statistic is larger then the critic value:

$$\chi^2_{2,0.95} = 5.9914$$

and the power is:

$$\hat{\pi} = 1$$

```
n<-44
p<-2
mu<-c(1,1)
sig<-diag(2)
chiCritical<-qchisq(0.95,2)
power<-0
for (i in 1:50000) {
 data<-mvrnorm(n,mu,sig)
 xavg<-colMeans(data)
 statisti<-n*t(xavg)%*%xavg
 if(statisti>chiCritical)
   power<-power+1
}
(power<-power/50000)
```

c. Now σ^2 is unknown. The solution to this problem is like in previous courses- to estimate the variance. First of all we will find the MLE of the variance:

$$f_x(x) = (2\pi)^{-1}|\sigma^2 I|^{-\frac{1}{2}} * \exp^{-\frac{1}{2}(x_i-\mu)^T(\sigma^2 I)^{-1}(x_i-\mu)} = (2\pi\sigma^2)^{-1} * \exp^{-\frac{1}{2\sigma^2}(x_i-\mu)^T(x_i-\mu)}$$

and the log likelihood function according to n samples is:

$$log(L(\mu,\sigma^2)) = log((2\pi\sigma^2)^{-n} * \exp^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^T(x_i-\mu)}) =$$

$$= -nlog(2\pi) + -nlog(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^T(x_i-\mu)$$

We will derivative according to σ^2 and equal it to 0:

$$\frac{\partial l(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{\sigma^2} + \frac{\sum_{i=1}^n (x_i-\mu)^T(x_i-\mu)}{2(\sigma^2)^2} = 0 \Rightarrow \sigma^2 = \frac{\sum_{i=1}^n (x_i-\mu)^T(x_i-\mu)}{2n}$$

Under H_0 $\mu = (0,0)^T$ so:

$$\sigma_0^2 = \frac{\sum_{i=1}^n x_i^T x_i}{2n}$$

Under H_1 $\mu \neq (0,0)^T$ and we know that the MLE for μ is $\bar{x}$ so:

$$\sigma_1^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})}{2n}$$

The likelihood functions:

$$L(\theta_1) = (2\pi\sigma_1^2)^{-n} * \exp^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^n (x_i-\bar{x})^T(x_i-\bar{x})} =$$

$$= (2\pi)^{-n} (\frac{\sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})}{2n})^{-n} * \exp^{-n}$$

$$L(\theta_0) = (2\pi\sigma_0^2)^{-n} * \exp^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i)^T(x_i)} =$$

$$= (2\pi)^{-n} (\frac{\sum_{i=1}^n x_i^T x_i}{2n})^{-n} * \exp^{-n}$$

And the generalize likelihood ratio is:

$$\frac{L(\theta_1)}{L(\theta_0)} = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})}{\sum_{i=1}^n x_i^T x_i} \right)^{-n} = \left(1 - \frac{n(\bar{x_1}^2 + \bar{x_2}^2)}{\sum_{i=1}^n x_{1i}^2 + x_{2i}^2} \right)^{-n}$$

According to Wilks:

$$2log(\frac{L(\theta_1)}{L(\theta_0)}) = -2nlog\left(\frac{\sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})}{\sum_{i=1}^n x_i^T x_i} \right) =$$

$$= -2nlog\left(1 - \frac{n(\bar{x_1}^2 + \bar{x_2}^2)}{\sum_{i=1}^n x_{1i}^2 + x_{2i}^2} \right) \underset{app.}{\sim} \chi_2^2$$

So the statistic will be:

$$\chi_{st}^2 = -2nlog\left(1 - \frac{n(\bar{x_1}^2 + \bar{x_2}^2)}{\sum_{i=1}^n x_{1i}^2 + x_{2i}^2} \right)$$

d. We will estimate the power in the following point, now the variance is unknown so the statistic will be like in the pervious task:

$$\mu = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The process is the same. We chose to do 50,000 replications. In each replication we check if the statistic is larger then the critic value:

$$\chi^2_{2,0.95} = 5.9914$$

and the power is:

$$\hat{\pi} = 1$$

```
n<-44
p<-2
mu<-c(1,1)
sig<-diag(2)
chiCritical<-qchisq(0.95,2)
power<-0
for (i in 1:50000) {
  data<-mvrnorm(n,mu,sig)
  xavg<-colMeans(data)
  statisti<--2*n*log(1- n*(sum(xavg^2))/sum(data^2))
  if(statisti>chiCritical)
    power<-power+1
}
(power<-power/50000)
```