Exam1

Michael Pena

2024-07-18

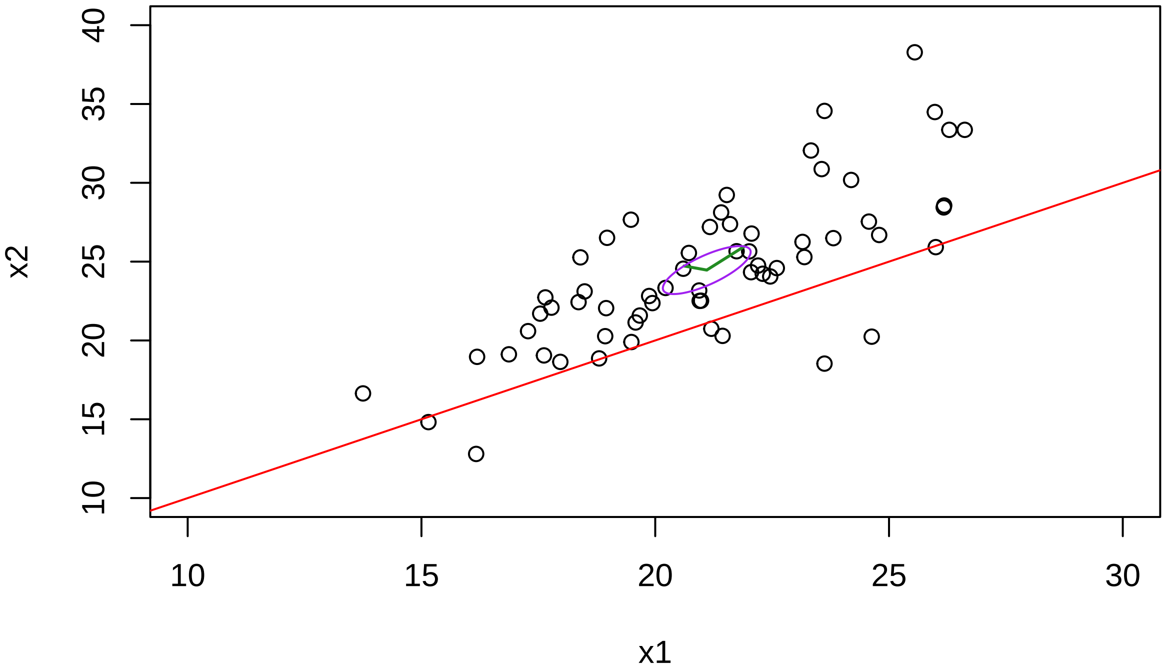
```
attach(data)
x1 = female
x2 = male
n = length(x1)
```

(a).

$$H_0 : \begin{bmatrix}\mu_{male}\\ \mu_{female}\end{bmatrix} = \begin{bmatrix}\mu_{female}\\ \mu_{male}\end{bmatrix} H_A : \begin{bmatrix}\mu_{male}\\ \mu_{female}\end{bmatrix} \neq \begin{bmatrix}\mu_{female}\\ \mu_{male}\end{bmatrix}$$

(b).

```
#plot data and line
plot(x1,x2, xlim = c(10,30) , ylim = c(10,40))
abline(c(0,0),c(1,1), col = "red")
# plot eigenvectors
x.bar = apply(data,2,mean)
S.x.bar = matrix(var(data),ncol=2)/n
eig = eigen(S.x.bar)
lam1 = eig$'values'[1]
lam2 = eig$'values'[2]
v1 = eig$vectors[,1]
v2 = eig$vectors[,2]
dist = qchisq(.95,2)
a1 = v1*sqrt(lam1)*sqrt(dist)
a2 = v2*sqrt(lam2)*sqrt(dist)
lines(c(x.bar[1],x.bar[1]+a1[1]),c(x.bar[2],x.bar[2]+a1[2]),lwd=1.5,col="forestgreen")
lines(c(x.bar[1],x.bar[1]+a2[1]),c(x.bar[2],x.bar[2]+a2[2]),lwd=1.5,col="forestgreen")
# put in ellipse
library(plotrix)
d1 = sqrt(sum(a1*a1))
d2 = sqrt(sum(a2*a2))
theta = acos(v1[1])
draw.ellipse(x=x.bar[1],y=x.bar[2],a=d1,b=d2,angle=theta,deg=F, border = "purple")
```



It is very clear that the confidence ellipse does not ever cross the hypothese line. We can probably assume we are going to reject null because of this.

(c).

```
# making confidence interval
n = length(x1)
t = qt(.975,n-1)
s1 = sd(x1) ; s2 = sd(x2)
sprintf("95 percent confidence interval for females: ( %f , %f )",mean(x1)-t*s1/sqrt(n),mean(x1)+t*s1/s
```

```
## [1] "95 percent confidence interval for females: ( 20.334997 , 21.869197 )"
```

```
sprintf("95 percent confidence interval for males: ( %f , %f )",mean(x2)-t*s2/sqrt(n),mean(x2)+t*s2/sqr
```

```
## [1] "95 percent confidence interval for males: ( 23.221276 , 25.709047 )"
```

(d).

going to use form. from notes

$$\bar{x}_j \pm \sqrt{\frac{p(n-1)s_j^2}{n(n-p)} \cdot F^{+}_{p,n-p}(\alpha)}$$

```
confInt <- function(data,alpha = .05){
  X = data
  p = dim(X)[2]
  n = dim(X)[1]
  a = alpha
  s = apply(X,2,sd)
  xbar = apply(X,2,mean)
  CONF <- data.frame()
  for(j in 1:p){
    num = p*(n-1)*(s[j])^2*qf(1-a,n-1,n-1)
    den = n*(n-p)
    CONF[j,1] = paste0("x",j)
    CONF[j,2] = xbar[j] - sqrt(num/den)
    CONF[j,3] = xbar[j] + sqrt(num/den)
  }

  names(CONF) = c("var","upper","lower")
 return("confidence intervals" = CONF)
}
confInt(data,.05)
```

```
##   var    upper    lower
## 1  x1 20.42572 21.77847
## 2  x2 23.36839 25.56193
```

(e).

```
library(mvtnorm)
X = as.matrix(data)
mu.mle = as.matrix(apply(X,2,mean))
sigma.mle  = matrix(0,2,2)
for(i in 1:n){sigma.mle = sigma.mle + (X[i,] - mu.mle)%*%t(X[i,]-mu.mle)}
sigma.mle = sigma.mle/n
# Lambda Ratio for Ranjosh
Se = matrix(c(12,8,8,12), nrow = 2, ncol = 2, byrow = T)
Me = c(22.5,24.5)
# Lambda Ratio for Ethan
Sr = matrix(c(9,8,8,16), nrow = 2, ncol = 2, byrow = T)
Mr = c(21.5,26)
Eth.Rat = prod(dmvnorm(X,Me,Se))/prod(dmvnorm(X,mu.mle,sigma.mle))
Ran.Rat = prod(dmvnorm(X,Mr,Sr))/prod(dmvnorm(X,mu.mle,sigma.mle))
# print
Ran.Rat;Eth.Rat
```

```
## [1] 0.0001009966
```

```
## [1] 4.596644e-12
```

Looks like the Likelihood Ratio value of Ranjosh's parameters yeild as higher ratio. This leads me to think that Ranjosh had a better estimate. You could say Ranjosh is more “right”.

(f).

```
Sn.r = matrix(0,2,2)
Sn.e = Sn.r
for(i in 1:n){Sn.r = Sn.r + (X[i,] - Mr)%*%t(X[i,]-Mr)}
for(i in 1:n){Sn.e = Sn.e + (X[i,] - Me)%*%t(X[i,]-Me)}
Ran.Rat = prod(dmvnorm(X,Mr,Sn.r/n))/prod(dmvnorm(X,mu.mle,sigma.mle))
Eth.Rat = prod(dmvnorm(X,Me,Sn.e/n))/prod(dmvnorm(X,mu.mle,sigma.mle))
Ran.Rat;Eth.Rat
```

```
## [1] 0.02555809
```

```
## [1] 4.014658e-06
```

Not assuming covariance matrices and basing them off the means does cause the Lambda ratios to get bigger. Still, Ranjosh has a better case for his mean vector as the likelihood ratio remains the bigger of the two.

(g).

```r
line = cbind(seq(20,23,by=0.001),seq(20,23,by=0.001))
mhds = rep(0,length(line[,1]))
for(i in 1:length(line[,1])){
  mhds[i] = mahalanobis(line[i,], center = x.bar, cov = S.x.bar)
}
line[which(mhds == min(mhds)),]
```

```
## [1] 20.469 20.469
```

```r
# code from example
Sn  = matrix(0,2,2)
for(i in 1:n){Sn = Sn + (X[i,] - x.bar)%*%t(X[i,]-x.bar)}
mu.0 = line[which(mhds == min(mhds)),]
A = Sn/(n*(n-1))
T2 = t(x.bar - mu.0)%*%solve(A)%*%(x.bar - mu.0)
p = 2
val = (n-p)*T2/((n-1)*p)
# getting p-values
pval.cs = 1-pchisq(T2,p)
pval.F = 1 - pf(val,p,n-p)
# return vals
pval.cs;pval.F
```

```
##              [,1]
## [1,] 7.993606e-15
```

```
##              [,1]
## [1,] 3.607821e-10
```

These numbers are practically zero. If we go with 99% confidence, both p-values are in the reject null region. So both F-score and Chi-Squared score will lead to the same conclusion. Therefore, with 99% confidence, there is sufficient evidence to reject the claim that non-management males and females have the same average salary.