

Lab 2 732A97 Multivariate Statistical Methods

Raymond Sseguya

2019-11-28

Inference about mean vectors

Question 1: Test of outliers

a) The Mahalanobis distance is approximately chi-square distributed ...

```
trackrcs <- read.table("T1-9.dat",
  col.names = c("countries", "x100m", "x200m",
    "x400m", "x800m", "x1500m", "x3000m", "marathon"))

trackrcs2 <- (trackrcs)[,-1]
rownames(trackrcs2) <- trackrcs[,1]

C <- cov((trackrcs)[,-1])
x_bar = apply(trackrcs2,1,mean)
d0 = as.matrix(trackrcs2-x_bar)
deviation = sqrt( d0%*%t(d0) )

d_sq_m <- d0%*%solve(C)%*%t(d0)
diagonal_vector3 <- diag(d_sq_m)
deviation_countries3 <-
  cbind.data.frame(countries = as.vector(trackrcs[,1]),diagonal_vector3)
deviation_countries_ordered3 <-
  deviation_countries3[order(-deviation_countries3$diagonal_vector3), ]
# deviation_countries_ordered3[1:5,]

chisq.test(deviation_countries_ordered3$diagonal_vector3,
  p=rep(0.1, nrow(deviation_countries_ordered3)), rescale.p = TRUE )

##
## Chi-squared test for given probabilities
##
## data: deviation_countries_ordered3$diagonal_vector3
## X-squared = 2034800, df = 53, p-value < 2.2e-16

named_Mahalanobis <- as.vector(deviation_countries_ordered3[,2])
names(named_Mahalanobis) <- rownames(deviation_countries_ordered3)
ch_s <- combn(x=named_Mahalanobis, m=2,
  FUN = function(c){
    sg <- 0.1/100
    pv <- chisq.test(x=c, p=rep(sg,2), rescale.p = TRUE)$p.value
    pvname <- paste0(names(c)[1]," ",names(c)[2])
    assign(pvname, pv)
```

```
    return(list(pv, pvname))
  })

## not outliers
print(unlist(ch_s[2, which(ch_s[1,] > sg)]))
```

```
## [1] "SIN MYA" "MYA TPE" "IND ISR" "COL KORN" "KORN AUT" "TUR ARG"
## [7] "TUR GRE" "ARG GRE" "NZL BRA" "CAN FIN" "ESP FRA" "ESP POR"
## [13] "ESP NED" "FRA POR" "FRA NED" "POR NED" "NED ITA" "NED BEL"
## [19] "ITA BEL" "NOR MEX" "MEX IRL" "MEX CZE" "IRL CZE" "CZE POL"
## [25] "POL JPN" "JPN ROM" "USA GER" "USA CHN" "GER CHN"
```