MANOVA - Egyptian Skull Size

Greg Johnson

We have data on 90 Egyptian skulls, specifically on various measurements and the period they're thought to be from. We want to compare the skull measurements between periods. Our analysis of choice is the one-way MANOVA

Let's check our MANOVA assumptions.

The first assumption is that each group is a random sample from its respective population and that each group is independent of each other. This assumption is usually fulfilled in the design of the experiment and isn't tested for after the fact.

The second assumption is that all populations have a common covariance matrix Σ. We can test this with Box's M Test.

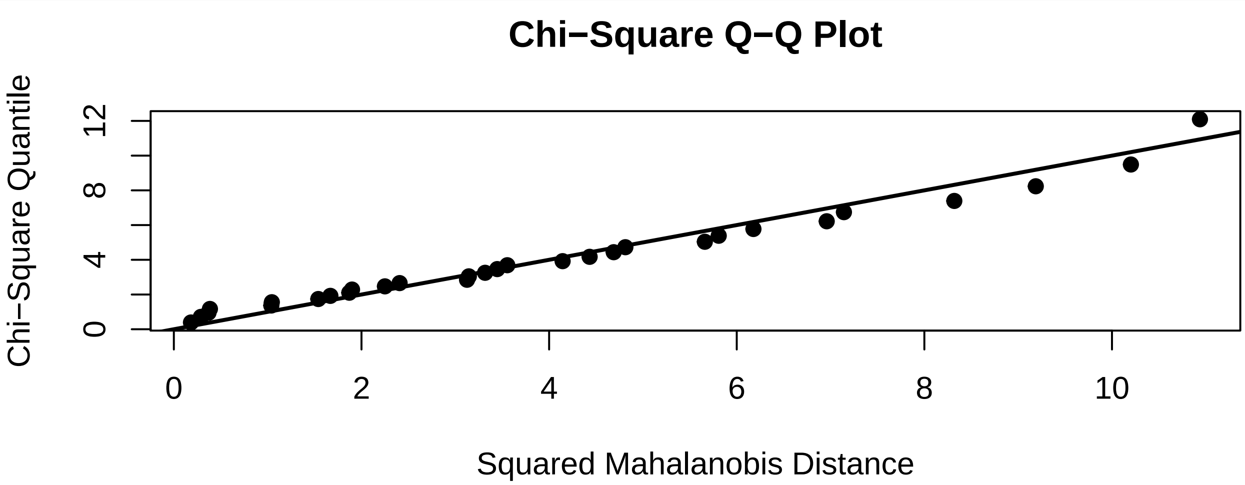
```
boxM(skull[,1:4],skull[,5])
```

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  skull[, 1:4]
## Chi-Sq (approx.) = 21.048, df = 20, p-value = 0.3943
```

So we fail to reject the null hypothesis that the groups have a common covariance matrix. Our assumption isn't violated.

The third assumption asserts multivariate normality for each population. We can check this with Mardia's multivariate skew and kurtosis statistics as well as a chi-square QQ plot:

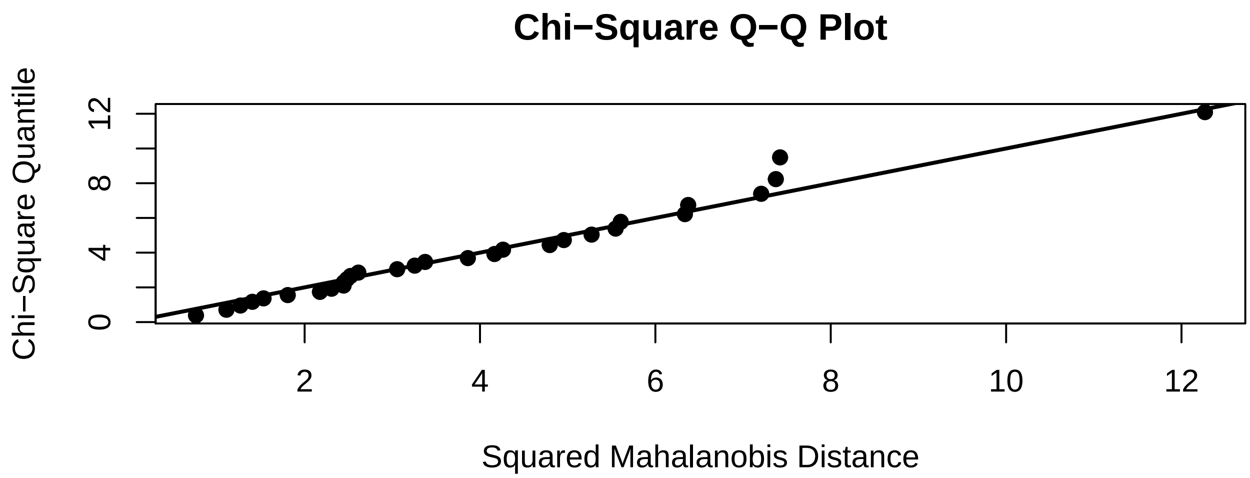
```
mardiaTest(skull[1:30,1:4],qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : skull[1:30, 1:4]
##
##    g1p            : 5.582262
##    chi.skew       : 27.91131
##    p.value.skew   : 0.1115155
##
```

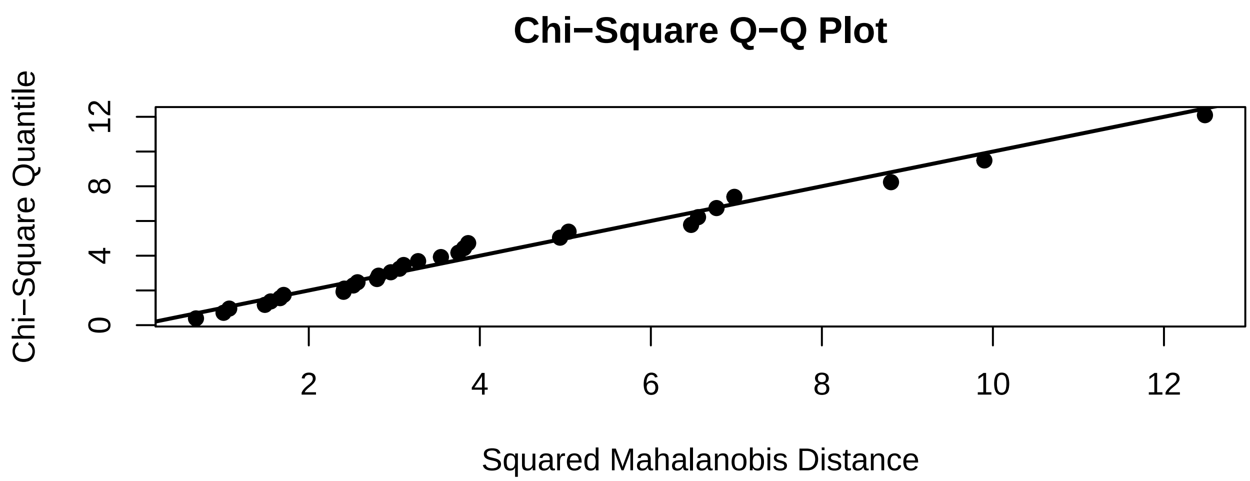
```
##    g2p            : 24.78416
##    z.kurtosis     : 0.3099655
##    p.value.kurt   : 0.7565872
##
##    chi.small.skew : 31.93878
##    p.value.small  : 0.043955
##
##    Result         : Data are multivariate normal.
## ---------------------------------------
```

```
mardiaTest(skull[31:60,1:4],qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : skull[31:60, 1:4]
##
##    g1p            : 3.37514
##    chi.skew       : 16.8757
##    p.value.skew   : 0.6610296
##
##    g2p            : 22.14362
##    z.kurtosis     : -0.7337989
##    p.value.kurt   : 0.4630713
##
##    chi.small.skew : 19.31078
##    p.value.small  : 0.5017151
##
##    Result         : Data are multivariate normal.
## ---------------------------------------
```

```
mardiaTest(skull[61:90,1:4],qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : skull[61:90, 1:4]
##
##    g1p              : 4.120737
##    chi.skew         : 20.60369
##    p.value.skew     : 0.42078
##
##    g2p              : 23.60906
##    z.kurtosis       : -0.1545345
##    p.value.kurt     : 0.8771883
##
##    chi.small.skew : 23.5767
##    p.value.small  : 0.2613716
##
##    Result          : Data are multivariate normal.
## ---------------------------------------
```

Looks like each sample group is multivariate normal! All of our assumptions are fulfilled for going forward with the MANOVA.

```
fit <- manova(skull[,1:4] ~ skull[,5])
summary(fit, test="Wilks")
```

```
##             Df   Wilks approx F num Df den Df  Pr(>F)
## skull[, 5]   1 0.86076   3.4374      4     85 0.01182 *
## Residuals   88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
xbar<-apply(skull[,1:4],2,mean)
xbar1<-apply(skull[1:30,1:4],2,mean)
xbar2<-apply(skull[31:60,1:4],2,mean)
xbar3<-apply(skull[61:90,1:4],2,mean)

B<-matrix(0,4,4)
for(g in 1:3){
  temp<-get(paste("xbar",g,sep=""))
  B<-B+30*(temp-xbar)%*%t(temp-xbar)
}
```

```r
W<-matrix(0,4,4)
for(i in 1:nrow(skull)){
  temp<-skull[i,1:4]
  if(i %in% 1:30){tempbar<-xbar1}
  if(i %in% 31:60){tempbar<-xbar2}
  if(i %in% 61:90){tempbar<-xbar3}
  W<-W+(temp-tempbar)%*%t(temp-tempbar)
}

Tot<-matrix(0,4,4)
for(i in 1:nrow(skull)){
  temp<-skull[i,1:4]
  Tot<-Tot+(temp-xbar)%*%t(temp-xbar)
}
```

Using Wilks' Lambda and an alpha of 0.05, our MANOVA tells us that the three time periods differ in their mean vector of skull measurements. As a follow up, we can construct 95% simultaneous confidence intervals to determine which mean components differ between the time periods. Our formula takes the form:

$$\bar{x}_{ki} - \bar{x}_{li} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_l}\right)}$$

```r
xbar1-xbar2-qt(.05/24,87,lower.tail=FALSE)*sqrt(diag(W)/(87*15))
```

```
## maxbreath basheight baslength nasheight
## -4.442312 -2.673706 -3.680110 -2.061423
```

```r
xbar1-xbar2+qt(.05/24,87,lower.tail=FALSE)*sqrt(diag(W)/(87*15))
```

```
## maxbreath basheight baslength nasheight
##  2.442312  4.473706  3.880110  2.661423
```

```r
xbar1-xbar3-qt(.05/24,87,lower.tail=FALSE)*sqrt(diag(W)/(87*15))
```

```
##  maxbreath  basheight  baslength  nasheight
## -6.5423115 -3.7737058 -0.6467765 -2.3947561
```

```r
xbar1-xbar3+qt(.05/24,87,lower.tail=FALSE)*sqrt(diag(W)/(87*15))
```

```
## maxbreath basheight baslength nasheight
## 0.3423115 3.3737058 6.9134432 2.3280894
```

```r
xbar2-xbar3-qt(.05/24,87,lower.tail=FALSE)*sqrt(diag(W)/(87*15))
```

```
##  maxbreath  basheight  baslength  nasheight
## -5.5423115 -4.6737058 -0.7467765 -2.6947561
```

```r
xbar2-xbar3+qt(.05/24,87,lower.tail=FALSE)*sqrt(diag(W)/(87*15))
```

```
## maxbreath basheight baslength nasheight
##  1.342312  2.473706  6.813443  2.028089
```

For BasBrth:

$$\tau_{11} - \tau_{21} : (-4.44, 2.44)$$

$$\tau_{11} - \tau_{31} : (-6.54, 0.34)$$

$$\tau_{21} - \tau_{31} : (-5.54, 1.34)$$

For BasHght:

$$\tau_{12} - \tau_{22} : (-2.67, 4.47)$$
$$\tau_{12} - \tau_{32} : (-3.77, 3.37)$$
$$\tau_{22} - \tau_{32} : (-4.67, 2.47)$$

For BasLgth:

$$\tau_{13} - \tau_{23} : (-3.68, 3.88)$$
$$\tau_{13} - \tau_{33} : (-0.64, 6.92)$$
$$\tau_{23} - \tau_{33} : (-0.75, 6.81)$$

For NasHgth

$$\tau_{14} - \tau_{24} : (-2.06, 2.66)$$
$$\tau_{14} - \tau_{34} : (-2.39, 2.33)$$
$$\tau_{24} - \tau_{34} : (-2.69, 2.03)$$

All the simultaneous intervals include zero which conflicts with our MANOVA results. Perhaps there's more power in running four univariate ANOVA's on each of the four outcome variables to see where the differences lie that made our MANOVA significant.

```
anova1<-anova(lm(skull[,1]~skull[,5]))
anova1
```

```
## Analysis of Variance Table
##
## Response: skull[, 1]
##            Df  Sum Sq Mean Sq F value   Pr(>F)
## skull[, 5]  1  144.15 144.150   7.081 0.009259 **
## Residuals  88 1791.45  20.357
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova2<-anova(lm(skull[,2]~skull[,5]))
anova2
```

```
## Analysis of Variance Table
##
## Response: skull[, 2]
##            Df Sum Sq Mean Sq F value Pr(>F)
## skull[, 5]  1    0.6   0.600  0.0272 0.8695
## Residuals  88 1944.3  22.094
```

```
anova3<-anova(lm(skull[,3]~skull[,5]))
anova3
```

```
## Analysis of Variance Table
##
## Response: skull[, 3]
##            Df  Sum Sq Mean Sq F value  Pr(>F)
## skull[, 5]  1  147.27 147.267  5.9013 0.01716 *
## Residuals  88 2196.02  24.955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova4<-anova(lm(skull[,4]~skull[,5]))
anova4
```

```
## Analysis of Variance Table
##
## Response: skull[, 4]
##            Df Sum Sq Mean Sq F value Pr(>F)
## skull[, 5]  1   0.02  0.0167  0.0017 0.9668
## Residuals  88 842.21  9.5705
```

So it looks like there are significant differences in the means of MaxBrth and BasLgth between the periods and this is what's giving us the MANOVA results.