

Assignment 2: Inference about mean vectors

Group 12

*Dávid Hrabovszki (davhr856), Laura Julia Melis (lauju103), Spyridon Dimitriadis (spydi472),
Vasileia Kampouraki (vaska979)*

30/11/2019

Question 1: Test of outliers.

Top 5 five countries ranked by Mahalanobis distance (as calculated in Lab1)

##	SAM	PNG	KORN	COK	MEX
##	35.01406	30.50725	26.16714	19.83400	14.23093

1.a.

Outlier testing with no multiple-testing correction procedure at significance level: 0.1%

From Chi-square we get that the critical value, for $\alpha = 0.001$ is 24.32189, so every distance that is greater than this critical value is considered as an outlier.

##	SAM	PNG	KORN
##	35.01406	30.50725	26.16714

Outlier testing with Bonferroni multiple-testing correction procedure (with $\alpha = 0.1\% / 54$), i.e. having 54 tests.

From Chi-square with Bonferroni correction we get that the critical value, for $\alpha = 0.001/54$ is 33.83184 so, again, every distance that is greater than this critical value is considered as an outlier.

##	SAM
##	35.01406

Using 0.1% significance levels and no correction procedure, we define 3 outliers in our dataset (SAM, PNG, KORN). Using the Bonferroni multiple-testing correction procedure, we conclude that only 1 observation is an outlier (SAM) based on the Mahalanobis distances.

According to McDonald in Handbook of Biological Statistics, the Bonferroni correction is appropriate when we want to be very careful not to get any false positives during the tests.

(<http://www.biostathandbook.com/multiplecomparisons.html>)

The correction happens at the expense of finding many false negatives, i.e. not finding outliers which actually are present. The Bonferroni approach might be useful in medical problems, but in our case, we believe that it is too conservative when it comes to classifying outliers.

The 0.1% significance level is considered to be low, as most tests are conducted at 5% level. This means that we want to be very certain that the outliers we define actually exist. Setting this level depends highly on the nature of the problem at hand, but in this case we think that a significance level of 5% would be more justifiable. This approach would result in 5 outliers.

1.b.

KORN seems like an outlier based on Mahalanobis distance, but not on Euclidean. This is because the Mahalanobis distance removes redundant information from correlated variables.

(<https://waterprogramming.wordpress.com/2018/07/23/multivariate-distances-mahalanobis-vs-euclidean/>)

With Mahalanobis distance we also use the relationships (covariances) between the variables (and not only the marginal variances as Euclidean distance does).

In the case of North Korea, running results were extreme in runtypes which have little correlation to other variables and low where the correlation was large.

Question 2: Test, confidence region and confidence intervals for a mean vector.

2.a.

The $100(1 - \alpha)\%$ confidence region for (μ_1, μ_2) of a p -dimensional distribution is the ellipse determined by all $\boldsymbol{\mu}$ such that:

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq c^2 = \frac{p(n-1)}{(n-p)} F_{p, n-p}(\alpha)$$

In our case, the 95% confidence ellipse for $\boldsymbol{\mu}$ consists of all values (μ_1, μ_2) satisfying:

$$45 \cdot [193.62 - \mu_1, 279.78 - \mu_2] \begin{bmatrix} 0.02 & -0.01 \\ -0.01 & 0.01 \end{bmatrix} \begin{bmatrix} 193.62 - \mu_1 \\ 279.78 - \mu_2 \end{bmatrix} \leq c^2$$

where $c^2 = \frac{2(45-1)}{45-2} \cdot F_{2,43}(0.05) = 2.047 \cdot 3.215 = 6.579535$.

If (λ_i, e_i) are the eigenvalue-eigenvector pairs of \mathbf{S} , then the i -th axis of the confidence ellipse has half length $\sqrt{p(n-1) \cdot F_{p, n-p}(\alpha) / (n-p)} \sqrt{\frac{\lambda_i}{n}}$ along the e_i direction.

Then, the axes of the confidence ellipse are: \mathbf{e}_i

Doing the eigen decomposition for the covariance matrix of our data, we obtain the following eigenvalues-eigenvectors:

$$\lambda_1 = 294.60898, \quad e'_1 = [0.5754, 0.8179]$$

$$\lambda_2 = 34.62637, \quad e'_2 = [-0.8179, 0.5754]$$

So, beginning at the center $\bar{\mathbf{x}}' = [193.62, 279.78]$, the axes of the 95% confidence ellipse are:

$$\begin{array}{ll} \text{major axis:} & \begin{bmatrix} 0.5753739 \\ 0.8178905 \end{bmatrix} \\ \text{minor axis:} & \begin{bmatrix} -0.8178905 \\ 0.5753739 \end{bmatrix} \end{array}$$

And the half length of each axis is:

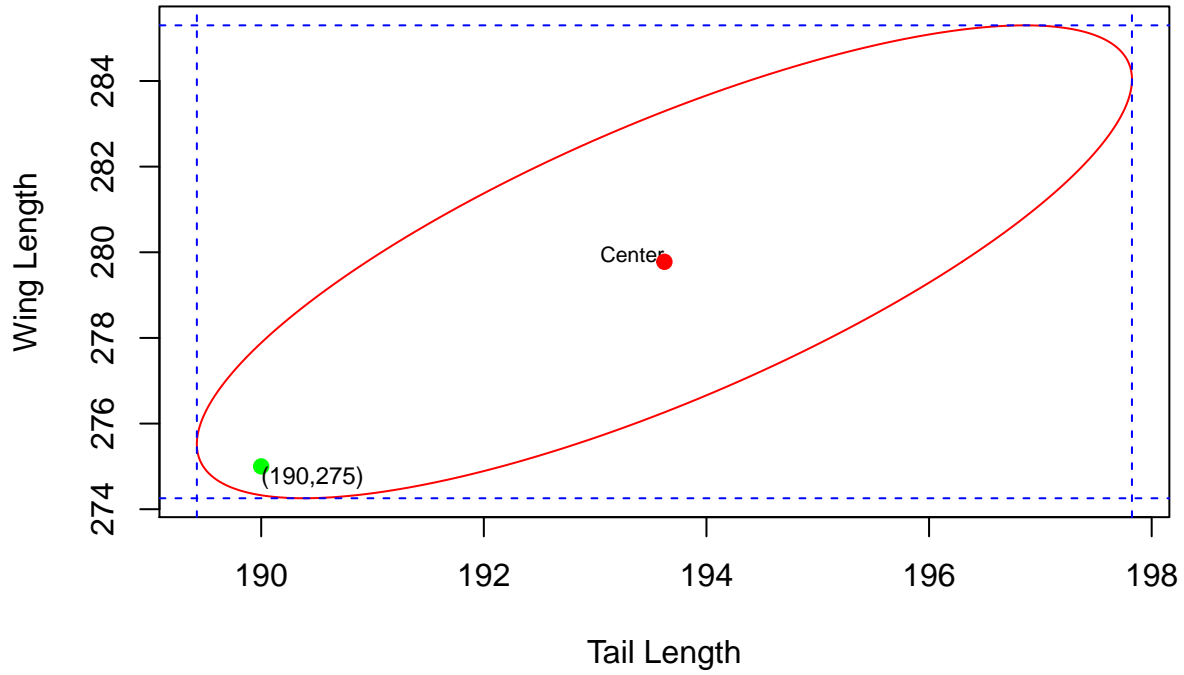
$$\text{major axis half-length:} \quad \sqrt{294.60898} \sqrt{0.1461883}$$

$$\text{minor axis half-length:} \quad \sqrt{34.62637} \sqrt{0.1461883}$$

If we plot these results, we obtain the following graph ¹

¹Sources: <https://stackoverflow.com/questions/41820683/how-to-plot-ellipse-given-a-general-equation-in-r> and <https://stackoverflow.com/questions/15915625/plotting-an-ellipse-in-matlab-given-in-matrix-form>

95% confidence ellipse for the population means



In the plot we can see the 95% confidence ellipse in red and also the confidence intervals in the dashed rectangle in blue. The green dot represents the population mean values for male hook-billed kites ($\mu' = [190, 275]$).

As we can observe, the green dot falls inside the ellipse, so we could conclude that these are in fact plausible values for the mean tail length and mean wing length for the female birds.

Also, we can confirm that $\mu' = [190, 275]$ is in the confidence region by computing the inequality explained above in this exercise:

$$45 \cdot [193.62 - 190, 279.78 - 275] \begin{bmatrix} 0.02 & -0.01 \\ -0.01 & 0.01 \end{bmatrix} \begin{bmatrix} 193.62 - 190 \\ 279.78 - 275 \end{bmatrix} \leq \frac{2(45 - 1)}{45 - 2} \cdot F_{2,45-2}(0.05)$$

$$5.54313 \leq 6.578471$$

2.b.

T^2 simultaneous confidence intervals:

$$a'\bar{x} - c\sqrt{\frac{a'Sa}{n}} \leq a'\mu \leq a'\bar{x} + c\sqrt{\frac{a'Sa}{n}}$$

where $c = \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} = \sqrt{6.578471} = 2.564853$.

So, the $100(1 - \alpha)\%$ simultaneous 95% T^2 -intervals are

$$\text{For } \mu_1 : \quad \bar{x}_1 \pm c\sqrt{\frac{s_{11}}{n}} = 193.62 \pm 2.565 \cdot 1.638$$

$$\Rightarrow \quad \boxed{189.4217 \leq \mu_1 \leq 197.8227}$$

$$\text{For } \mu_2 : \quad \bar{x}_2 \pm c\sqrt{\frac{s_{22}}{n}} = 279.78 \pm 2.565 \cdot 2.153$$

$$\Rightarrow \quad \boxed{274.2564 \leq \mu_2 \leq 285.2992}$$

Bonferroni $100(1 - \alpha)\%$ confidence intervals:

$$\bar{x}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}}, \text{ for } i = 1, 2, \dots, p$$

where m is the number of tests being carried out and $t_{n-1}\left(\frac{\alpha}{2p}\right) = t_{44}\left(\frac{0.05}{4}\right) = 2.32$.

So, the 95% Bonferroni intervals for the two population means are:

$$\text{For } \mu_1 : \quad \bar{x}_1 \pm t_{44}\left(\frac{0.05}{4}\right)\sqrt{\frac{s_{11}}{n}} = 193.62 \pm 2.3207 \cdot 1.638$$

$$\Rightarrow \quad \boxed{189.8216 \leq \mu_1 \leq 197.4229}$$

$$\text{For } \mu_2 : \quad \bar{x}_2 \pm t_{44}\left(\frac{0.05}{4}\right)\sqrt{\frac{s_{22}}{n}} = 279.78 \pm 2.3207 \cdot 1.638$$

$$\Rightarrow \quad \boxed{274.7819 \leq \mu_2 \leq 284.7736}$$

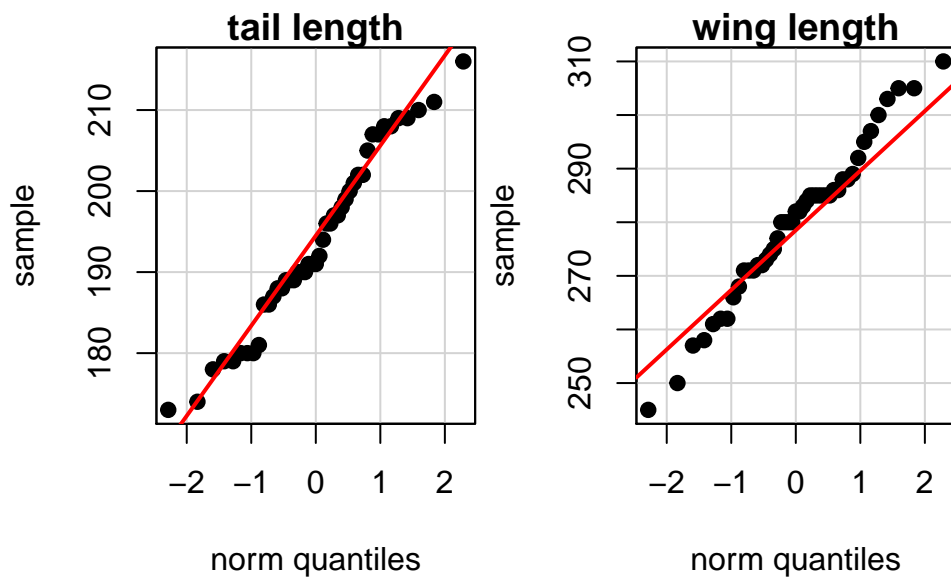
Comparison and advantage of T^2 -intervals over Bonferroni.

The simultaneous T^2 -intervals are a bit wider than the Bonferroni confidence interval in both variables, the Bonferroni interval falls within the T^2 -interval. Bonferroni correction tries to ensure that the probability of declaring even one false positive is no more than, e.g., 5%. The Bonferroni correction declares as significant (rejects the null) any hypothesis where the p-value is $\leq 0.05/2m$.

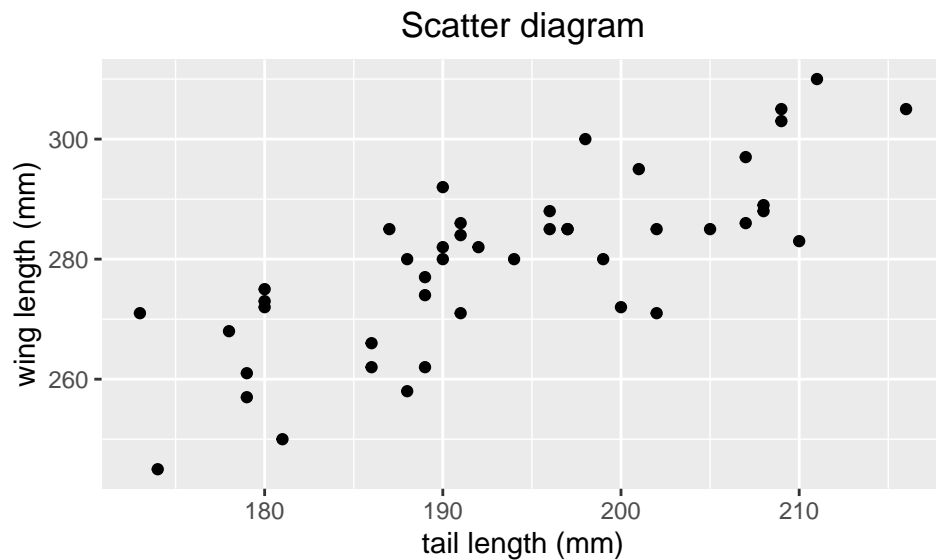
If we are interested only in the component means, the Bonferroni intervals provide more precise estimates than the T^2 -intervals.

2.c.

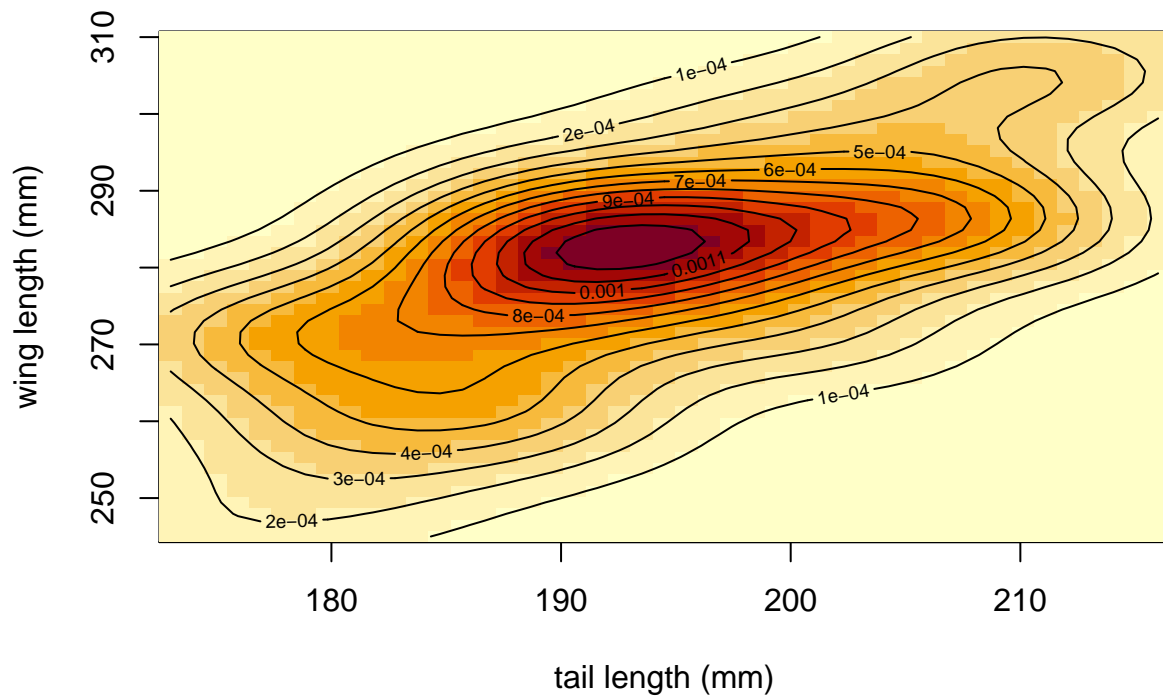
We will explore if the data is plausibly coming from a bivariate Normal distribution analyzing Q-Q plots of each variable as well as the scatterplot of the observations and the sample contour:



We could say that the univariate normal distribution is not a viable normal model for each variable separately.



The observations in the scatterplot should be distributed forming an ellipse and, as we can observe, the dots have a wide spread.



From the contour plot of the sample, the curves don't have an ellipse shape.

Summarizing the three graphs, we could say that the bivariate normal distribution is not a viable population model.

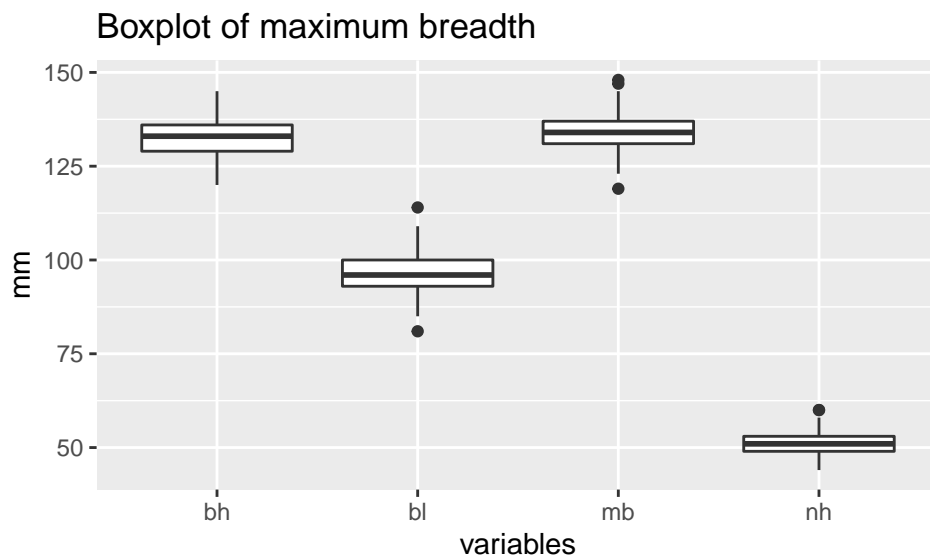
Question 3: Comparison of mean vectors (one-way MANOVA)

3.a.

First we construct a data frame for the descriptive statistics to obtain a general image of the data.

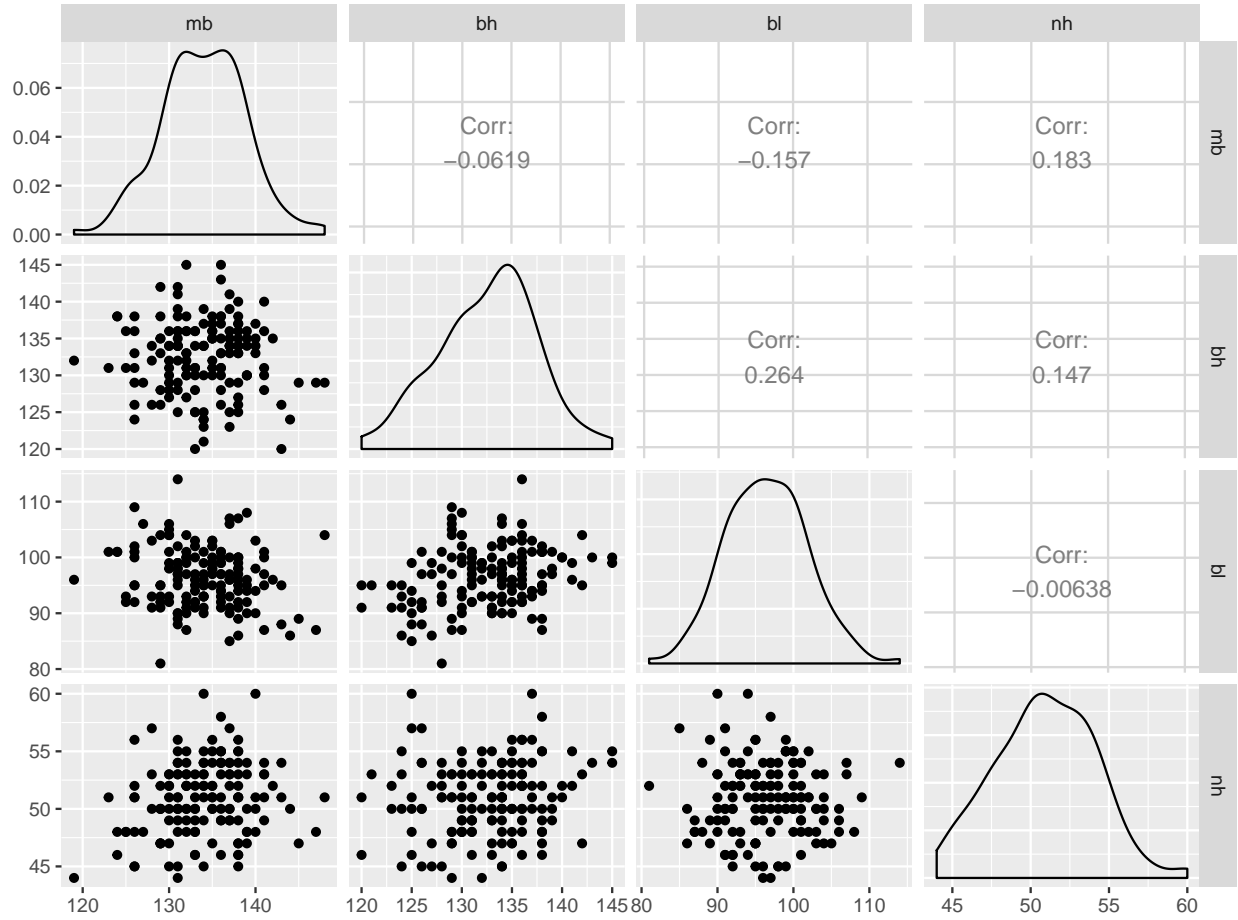
##	mb	bh	bl	nh
## mean	133.973	132.547	96.460	50.933
## median	134.000	133.000	96.000	51.000
## min	119.000	120.000	81.000	44.000
## max	148.000	145.000	114.000	60.000
## range	29.000	25.000	33.000	16.000
## sd	4.891	4.939	5.378	3.208
## skewness	-0.028	-0.174	0.138	0.082
## kurtosis	0.216	-0.149	0.143	-0.207

Boxplot



We use boxplots to examine our data. We can observe three outlier for maximum breadth and for basialiveolar length we observe two outliers. Finally, for nasal height there is one outlier.

Looking at the boxplot, we get a general image about the standard deviation which is indicated by the height of the box. For nasal height the standard deviation is the lowest.



We observe from the pairplot that the variables are weakly correlated.

3.b.

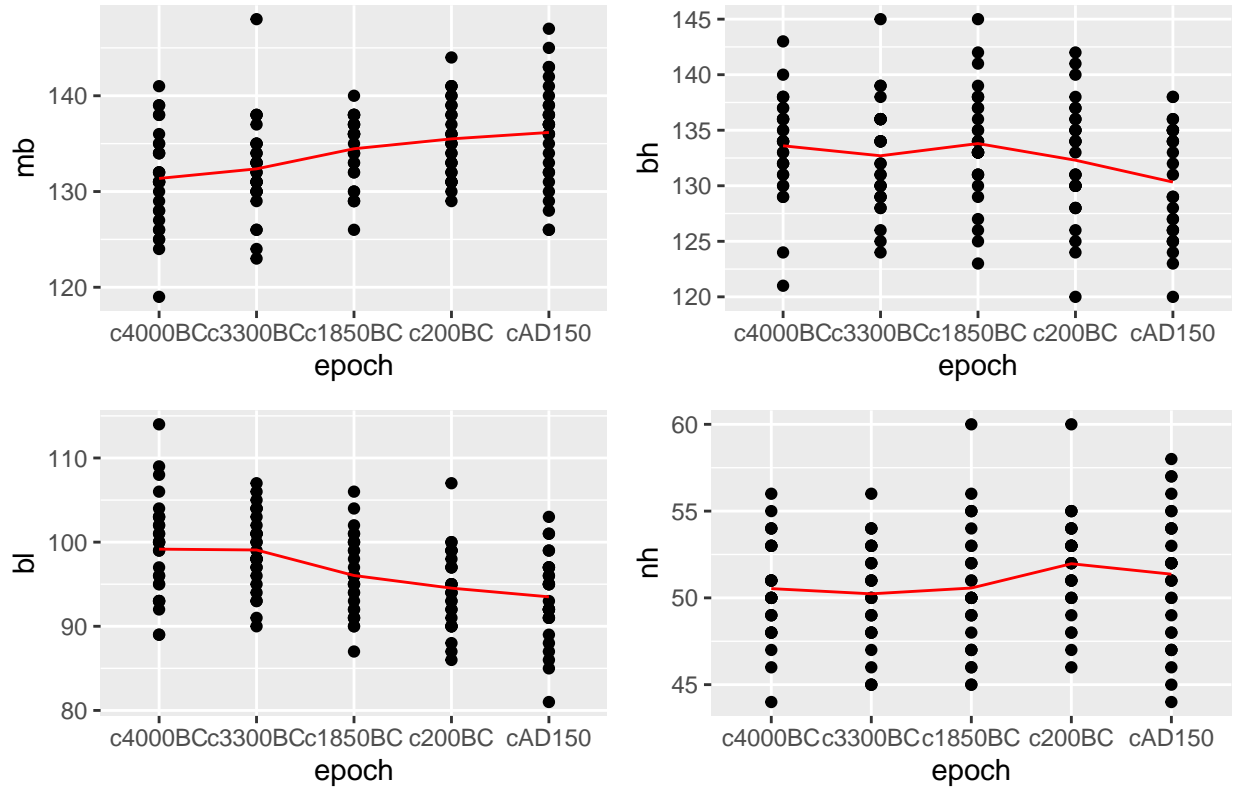
Here is the data frame with the means of each variable in each epoch.

```
##           mb           bh           bl           nh
## c4000BC_mean 131.3667 133.6000 99.16667 50.53333
## c3300BC_mean 132.3667 132.7000 99.06667 50.23333
## c1850BC_mean 134.4667 133.8000 96.03333 50.56667
## c200BC_mean  135.5000 132.3000 94.53333 51.96667
## cAD150_mean  136.1667 130.3333 93.50000 51.36667
```

Looking at the data frame we see for the maximum breadth that the mean was increasing slowly but steadily through the epochs. For basibregmatic height we can't say that there are significant changes but it seems as the mean was decreasing through the epochs, except in c1850BC when there was an increase. Basialveolar length was the characteristic that its mean changed the most. There was a quite significant decrease of the mean, compared at least to the other characteristics. Finally, nasal height's mean remained almost stable.

Above, we plot the means to obtain a better image of the changes.

Plot of Means



Now we can observe better the differences of means between the epochs, so we will continue by computing and reporting simultaneous confidence intervals.

3.c.

simultaneous confidence intervals:

$$a'\bar{x} - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{a'Sa}{n}} \leq a'\mu \leq a'\bar{x} + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{a'Sa}{n}}$$

The confidence interval for mb is [132.7147 , 135.232]

The confidence interval for bl is [131.2755 , 133.8178]

The confidence interval for bh is [95.076 , 97.844]

The confidence interval for nl is [50.10776 , 51.7589]

MANOVA

MANOVA (multivariate analysis of variance) is a type of multivariate analysis used to analyze data that involves more than one dependent variable at a time. MANOVA allows us to test hypotheses regarding the effect of one or more independent variables on two or more dependent variables. We assume that the assumptions of manova apply in our case.

```
##           Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## epoch      4           0.48182    4.231   16   562 8.278e-08 ***
## Residuals 145
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We want to check if the means differ between the epochs. Thus, first we used the function manova and looking at the results we get $p\text{-value} = 8.278e-08 < 0.05$, which means that we reject the null Hypothesis and we conclude that there are significant differences between the means. However, we don't know in which variables those differences occur.

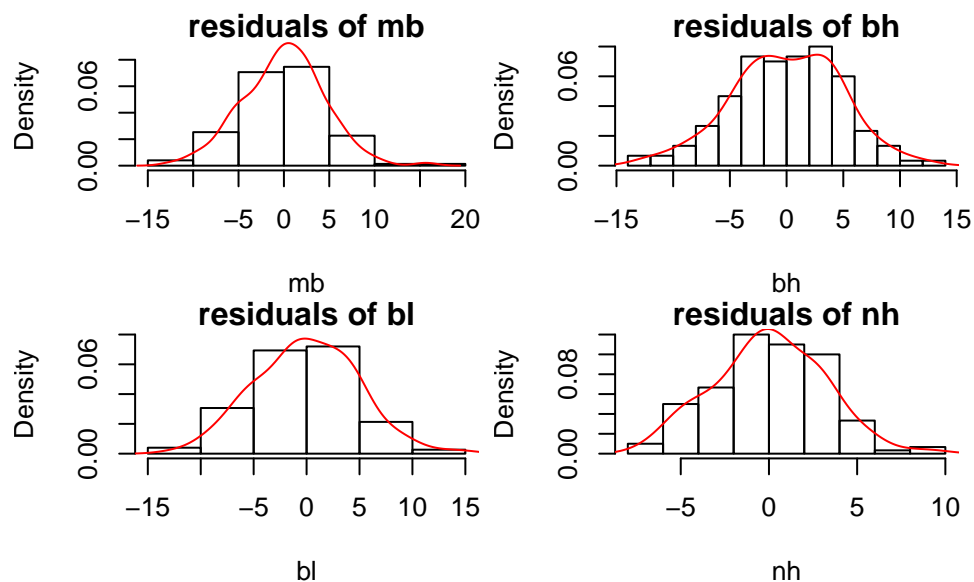
For this reason we use anova to test each one variable separately.

```
## Response mb :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## epoch           4   502.83  125.707   5.9546 0.0001826 ***
## Residuals      145  3061.07   21.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response bh :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## epoch           4   229.9   57.477   2.4474 0.04897 *
## Residuals      145  3405.3   23.485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response bl :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## epoch           4   803.3  200.823   8.3057 4.636e-06 ***
## Residuals      145  3506.0   24.179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response nh :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## epoch           4    61.2   15.300   1.507 0.2032
## Residuals      145  1472.1   10.153
```

As we see from the results, only nasal height's p-value is greater than 0.5 ($p\text{-value} = 0.2032$) and all the other variables have significance differences in their means between the epochs.

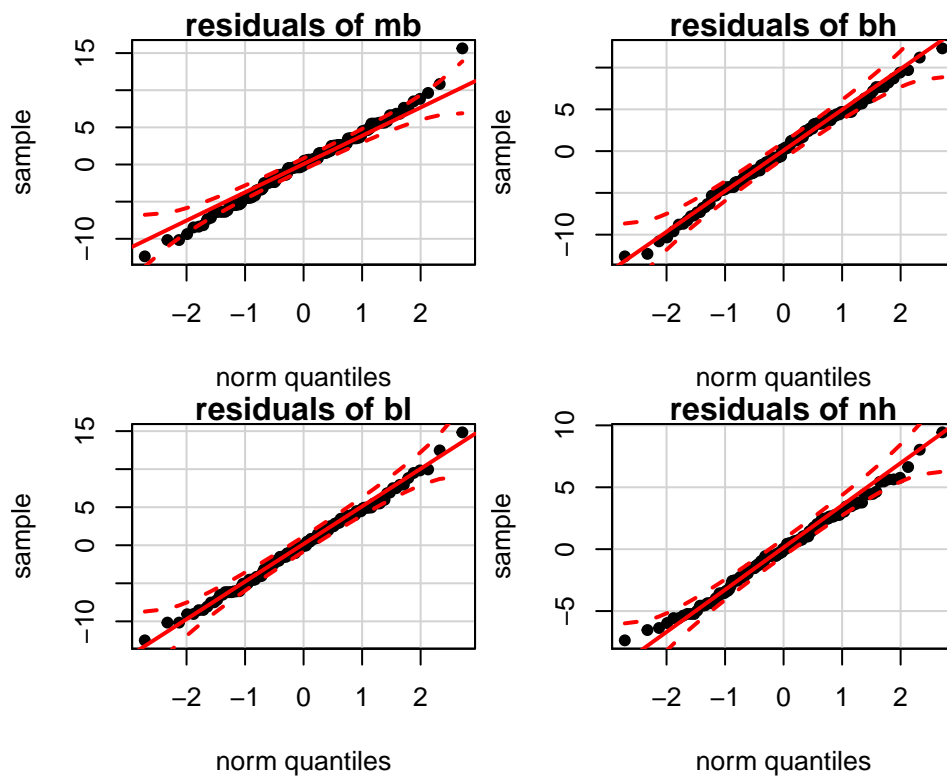
However, that was expected as we have already come to this conclusion from question (b) where we saw that nasal height's mean was the only one to remain roughly stable.

Inspection of the residuals whether they have mean 0 and if they deviate from normality (graphically).



From the histograms of the residuals we can say that all the variables' residuals have mean 0. Moreover, they all look normal distributed except of bh which looks a bit suspicious.

Q-Q plots



From Q-Q plots we observe that all residuals look normal with 95% significance level.

Appendix

```
knitr::opts_chunk$set(echo = TRUE, message = F, warning = F, error = F)
# NEEDED LIBRARIES
library(ggplot2)
library(tidyr)
library(gridExtra)
library(grid)
library(car)
library(heplots)
library(MASS)
library(fmsb)
library(GGally)

RNGversion('3.5.1')
#####--QUESTION 1--#####
#read data
data1 <- read.table("T1-9.dat", stringsAsFactors = FALSE)
colnames(data1) <- c("Country", "100m", "200m", "400m",
                    "800m", "1500m", "3000m", "Marathon")

#Mahalanobis distance ( mahdist )

#center the raw data by the means
resid = sapply(data1[,-1],function(x)x-mean(x))

C = cov(data1[,-1])
mahdist = resid%%solve(C)%*%t(resid)

mahdist = diag(mahdist)
names(mahdist) = data1$Country
head(sort(mahdist, decreasing=TRUE),5)
alpha <- 0.001

chi_value <- qchisq(1-alpha, df = 7)
mahdist_outliers <- mahdist[mahdist>chi_value]
sort(mahdist_outliers, decreasing=TRUE)
chi_value_bonferroni <- qchisq(1-alpha/length(mahdist), df = 7)
mahdist_outliers_bonferroni <- mahdist[mahdist>chi_value_bonferroni]
sort(mahdist_outliers_bonferroni, decreasing=TRUE)
#####--QUESTION 2--#####
# Importing and modifying data file
df = read.table("T5-12.dat")
colnames(df) <- c("Tail.length", "Wing.length")
# Initial values
mu <- c(190,275)
xbar <- as.numeric(colMeans(df))
p <- 2
n <- nrow(df)
alpha <- 0.05
S <- cov(df)
# Axes of the confidence ellipse
c2 <- ((p*(n-1))/(n-p))*qf(1-alpha,df1=p,df2=n-p)
```

```

lambdas <- eigen(S)$values
e <-eigen(S)$vector
#PLOT
f_value <- qf(1-alpha,df1=p,df2=n-p)

# Points of the ellipse
theta <- seq(0, 2*pi, length = 1000)
r <- sqrt((n-1)*p/(n-p)*f_value/n)
v <- rbind(r*cos(theta), r*sin(theta))
z <- backsolve(chol(solve(S)), v)+xbar

# Confidence intervals (rectangle)
c <- sqrt(c2)
low1 <- xbar[1] - c*sqrt(S[1,1]/n)
upp1 <- xbar[1] + c*sqrt(S[1,1]/n)
low2 <- xbar[2] - c*sqrt(S[2,2]/n)
upp2 <- xbar[2] + c*sqrt(S[2,2]/n)

# Plot of the ellipse
{plot(t(z), type="l", xlab="Tail Length", ylab="Wing Length",
      main="95% confidence ellipse for the population means", col="red")
text(xbar[1], xbar[2], "Center", cex=0.7, adj = c(1,0))
points(xbar[1], xbar[2], pch=19,col="red")
points(mu[1], mu[2], pch=19,col="green")
text(mu[1], mu[2], "(190,275)", cex=0.75, adj = c(0,1))
abline(v=low1, col="blue", lty=2)
abline(v=upp1, col="blue", lty=2)
abline(h=low2, col="blue", lty=2)
abline(h=upp2, col="blue", lty=2)
}
right_side<- n*(xbar-mu)%*%solve(S)%*%(xbar-mu)
result <- right_side <= c2 #TRUE
CI1 <- c(low1, upp1)
CI2 <- c(low2, upp2)
t <- qt(1-alpha/(2*p),df=n-1)

low1 <- xbar[1] - t*sqrt(S[1,1]/n)
upp1 <- xbar[1] + t*sqrt(S[1,1]/n)
low2 <- xbar[2] - t*sqrt(S[2,2]/n)
upp2 <- xbar[2] + t*sqrt(S[2,2]/n)

bonf1 <- c(low1, upp1)
bonf2 <-c(low2, upp2)
par(mfrow=c(1,2), mar=c(4,4,1,0)+0.1)
qqPlot(df$Tail.length,ylab="sample",main = "tail length",envelope=FALSE,
       col.lines="red",pch=19,id=F)

qqPlot(df$Wing.length,ylab="sample",main = "wing length",envelope=FALSE,
       col.lines="red",pch=19,id=F)
ggplot(df, aes(x = Tail.length, y = Wing.length)) +
  geom_point() +
  ggtitle("Scatter diagram")+
  labs(x="tail length (mm)",y="wing length (mm)") +

```

```

  theme(plot.title = element_text(hjust = 0.5)) #adjust the title to the centre
#library(MASS)
bivn.kde <- MASS::kde2d(df$Tail.length, df$Wing.length, n = 45) #MASS package
# Contour plot overlayed on heat map image of results
image(bivn.kde, xlab="tail length (mm)", ylab="wing length (mm)")
contour(bivn.kde, add = TRUE) # from base graphics package
#####--QUESTION 3--#####
#library(heplots)
data = Skulls

statsdf = data.frame(cbind(mean=apply(data[-1], mean),
                             median=apply(data[-1], median),
                             min=apply(data[-1], min),
                             max=apply(data[-1], max),
                             range=apply(data[-1], function(x)max(x)-min(x)),
                             sd=apply(data[-1], sd),
                             skewness=apply(data[-1], timeDate::skewness),
                             kurtosis=apply(data[-1], timeDate::kurtosis)
                             ))

round(as.data.frame(t(statsdf)),3)
#statsdf
df = gather(data[,c(2,3,4,5)], "variables", "mm")

boxplot1 <- ggplot(data = df, mapping = aes(x = variables, y = mm)) +
  ggtitle("Boxplot of maximum breadth") +
  geom_boxplot()

boxplot1
#library(GGally)
ggpairs(data[, -1], progress=F)
c4000BC_mean <- sapply(data[1:30, 2:5], mean)

c3300BC_mean <- sapply(data[31:60, 2:5], mean)

c1850BC_mean <- sapply(data[61:90, 2:5], mean)

c200BC_mean <- sapply(data[91:120, 2:5], mean)

cAD150_mean <- sapply(data[121:150, 2:5], mean)

means <- as.data.frame(rbind(c4000BC_mean, c3300BC_mean, c1850BC_mean,
                             c200BC_mean, cAD150_mean))

means
#plots for mean

mplot1 <- ggplot(data, aes(x=epoch, y=mb)) + geom_point() +
  stat_summary(aes(y = mb, group=1), fun.y=mean, colour="red", geom="line", group=1)

mplot2 <- ggplot(data, aes(x=epoch, y=bh)) + geom_point() +
  stat_summary(aes(y = bh, group=1), fun.y=mean, colour="red", geom="line", group=1)

```

```

mplot3 <- ggplot(data, aes(x=epoch, y=bl)) + geom_point() +
  stat_summary(aes(y = bl,group=1), fun.y=mean, colour="red", geom="line",group=1)

mplot4 <- ggplot(data, aes(x=epoch, y=nh)) + geom_point() +
  stat_summary(aes(y = nh,group=1), fun.y=mean, colour="red", geom="line",group=1)

grid.arrange(mplot1,mplot2,mplot3,mplot4,top = textGrob("Plot of Means",
  gp=gpar(fontsize=18,font=3)))

n <- nrow(data)
p <- 4
alpha <- 0.05
xbar <- as.numeric(colMeans(data[-1]))
S <- cov(data[-1])

c2 <- ((p*(n-1))/(n-p))*qf(1-alpha,df1=p,df2=n-p)

# Confidence intervals (rectangle)
c <- sqrt(c2)
low1 <- xbar[1] - c*sqrt(S[1,1]/n)
upp1 <- xbar[1] + c*sqrt(S[1,1]/n)
low2 <- xbar[2] - c*sqrt(S[2,2]/n)
upp2 <- xbar[2] + c*sqrt(S[2,2]/n)

low3 <- xbar[3] - c*sqrt(S[3,3]/n)
upp3 <- xbar[3] + c*sqrt(S[3,3]/n)
low4 <- xbar[4] - c*sqrt(S[4,4]/n)
upp4 <- xbar[4] + c*sqrt(S[4,4]/n)

cat("The confidence interval for mb is [",low1,",",upp1,"]\n")
cat("The confidence interval for bl is [",low2,",",upp2,"]\n")
cat("The confidence interval for bh is [",low3,",",upp3,"]\n")
cat("The confidence interval for nl is [",low4,",",upp4,"]\n")
#Manova
dfMANOVA = manova(cbind(mb,bh,bl,nh)~epoch, data=data)
summary(dfMANOVA, test="Hotelling-Lawley")
#after seeing that Ho has been rejected
#we use aov to do testing in individual responses
summary.aov(dfMANOVA, test="Hotelling-Lawley")
par(mfrow=c(2,2),mar=c(4, 3.8, 1, 0.5)) #(bottom, left, top, right)
dens = density(dfMANOVA$residuals[,1])
hist(dfMANOVA$residuals[,1], freq=F, xlab="mb",main="residuals of mb", ylim=c(0,0.09))
lines(dens,col="red")

dens = density(dfMANOVA$residuals[,2])
hist(dfMANOVA$residuals[,2], freq=F, xlab="bh",main="residuals of bh")
lines(dens,col="red")

dens = density(dfMANOVA$residuals[,3])
hist(dfMANOVA$residuals[,3], freq=F, xlab="bl",main="residuals of bl", ylim=c(0,0.08))
lines(dens,col="red") #prob density

dens = density(dfMANOVA$residuals[,4])

```



```

hist(dfMANOVA$residuals[,4], freq=F, xlab="nh",main="residuals of nh")
lines(dens,col="red")
#dfMANOVA$residuals[,2] , ylim=c(0,0.12)
par(mfrow=c(2,2),mar=c(4, 3.8, 1, 0.5)) #(bottom, left, top, right)
qqPlot(dfMANOVA$residuals[,1],ylab="sample",main = "residuals of mb",envelope=.95,
       col.lines="red",pch=19,id=F)
qqPlot(dfMANOVA$residuals[,2],ylab="sample",main = "residuals of bh",envelope=.95,
       col.lines="red",pch=19,id=F)
qqPlot(dfMANOVA$residuals[,3],ylab="sample",main = "residuals of bl",envelope=.95,
       col.lines="red",pch=19,id=F)

qqPlot(dfMANOVA$residuals[,4],ylab="sample",main = "residuals of nh",envelope=.95,
       col.lines="red",pch=19,id=F)

```