Analysis of Egyptian Skull Data

Ramit Nandi, Utsyo Chakraborty, Sandeepan Sarkar

2023-03-29

Contents

1 The data - possible research questions

The data consists of *measurements (in mm)* made on male Egyptian skulls from three time periods, viz. 4000 BC (period 1 - the early predynastic period), 3300 BC (period 2 - the late predynastic period) and 1850 BC (period 3 - the 12th and 13th dynasties). Anthropocentric measurements have been made on four physical parameters of the skulls. They are:

- **Maximum breadth** (mb)
- **Basibregmatic height** (bh)
- **Basialveolar length** (bl)
- **Nasal height** (nh)

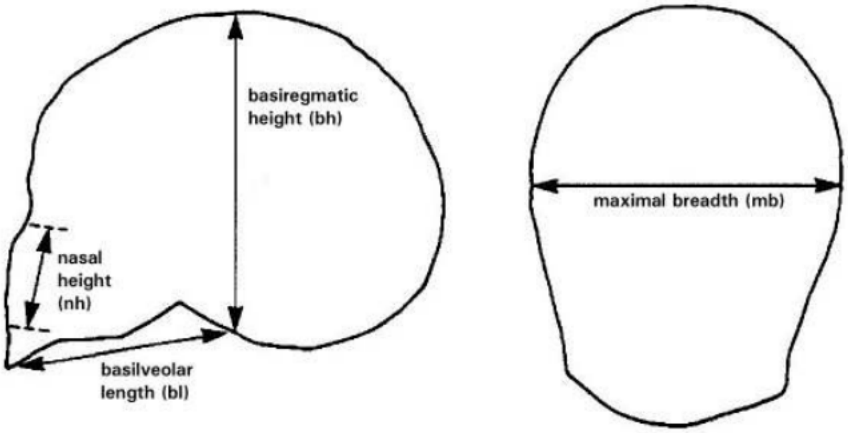


Figure 9: Diagram of the skull measurements

Here is a glimpse of the data.

```
##    mb  bh  bl nh period
## 1 131 138  89 49 4000BC
## 2 125 131  92 48 4000BC
## 3 131 132  99 50 4000BC
## 4 119 132  96 44 4000BC
## 5 136 143 100 54 4000BC
## 6 138 137  89 56 4000BC
```

Before we start analyzing this data, the following research questions come up:

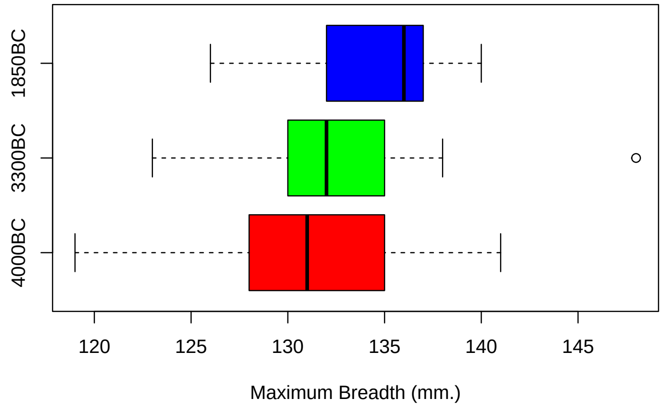
1. How are the four skull measurements *related*?
2. Are there any *significant differences in the skull sizes among time periods*?
3. Do they *show any changes with time*?
4. Can we *reduce the dimensionality* of the data in hand?
5. Given *new* skull measurements, can we *predict the time period* to which the skull might belong to?

2 Exploratory Data Analysis

2.1 Univariate Analysis

First, we look at univariate boxplots and try to detect patterns and outliers in the individual variables.

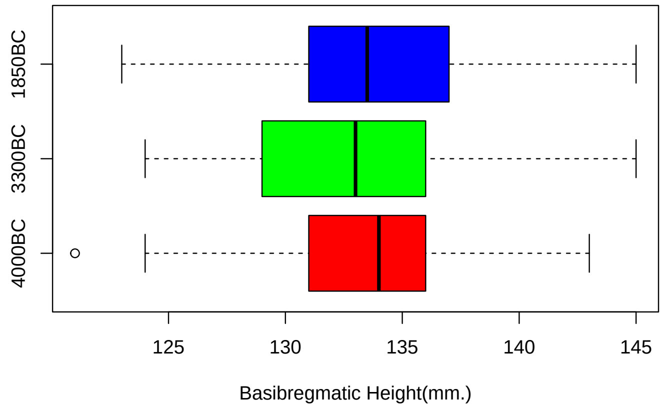
1. *Maximum Breadth*



Comments:

- Besides being negatively skewed, the average maximum breadth for period 1850 BC is larger compared to those in 4000 BC and 3300 BC.
- There is a *significant* outlier present (4th observation) in 3300 BC.

2. *Basibregmatic Height*



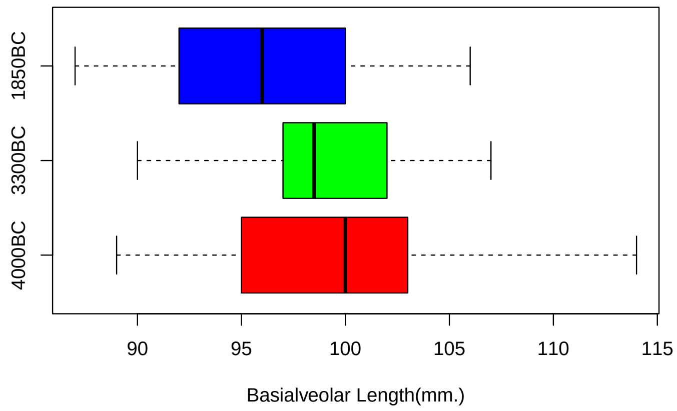
Comments:

- There does no seem to be much difference between the average values of basibregmatic height among the three periods.

- The heights for period 2 are more dispersed compared to the other two.

- There is an outlier present in 4000 BC.

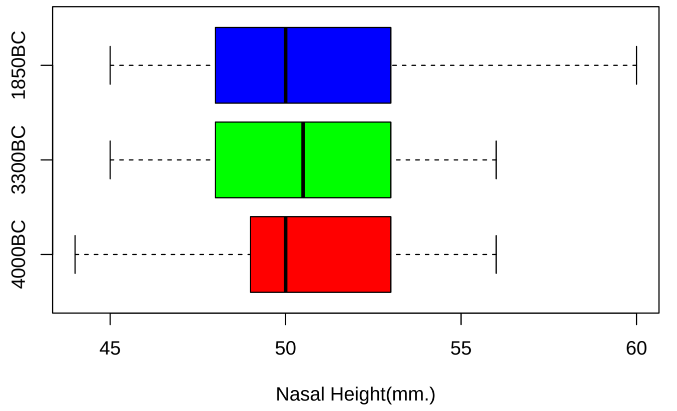
3. *Basialveolar Length*



Comments:

- The average basialveolar length seems to be decreasing over the three periods.

4. *Nasal Height*



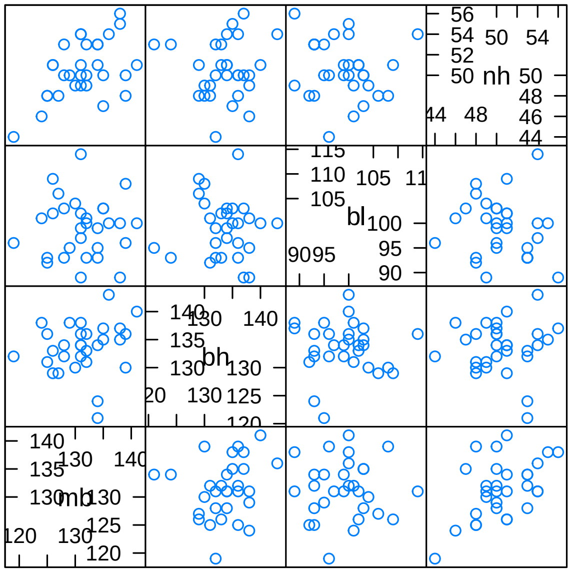
Comments:

- There does no seem to be much difference between the average values of nasal height among the three periods.
- The distribution of nasal height seem to be very similar for the periods of 3300 BC and 1850 BC.

2.2 Bivariate Analysis

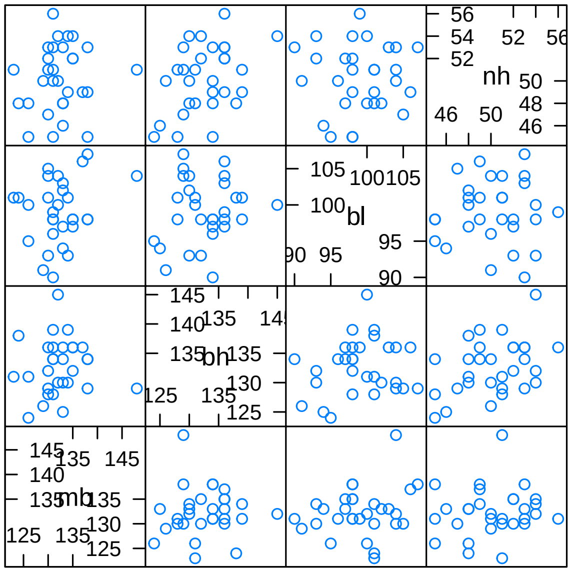
We now examine the bivariate relationships among the four variables, taken two at a time. This can be visualized using a **scatterplot matrix**. The period wise scatterplot matrices have been displayed below.

4000BC

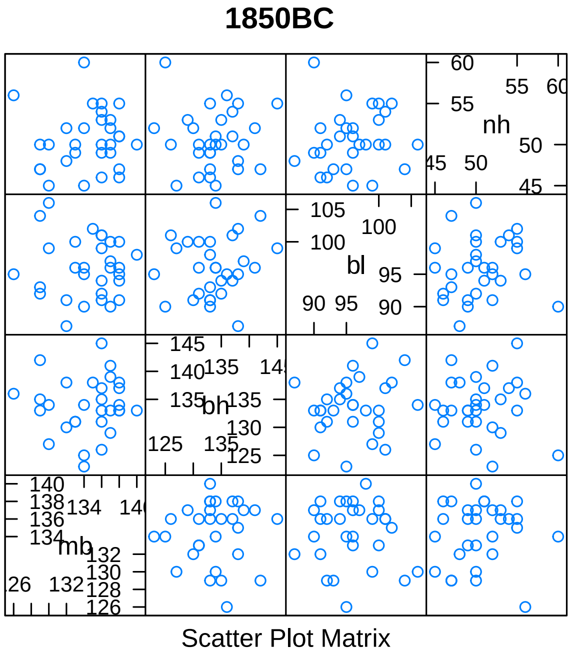


Scatter Plot Matrix

3300BC



Scatter Plot Matrix

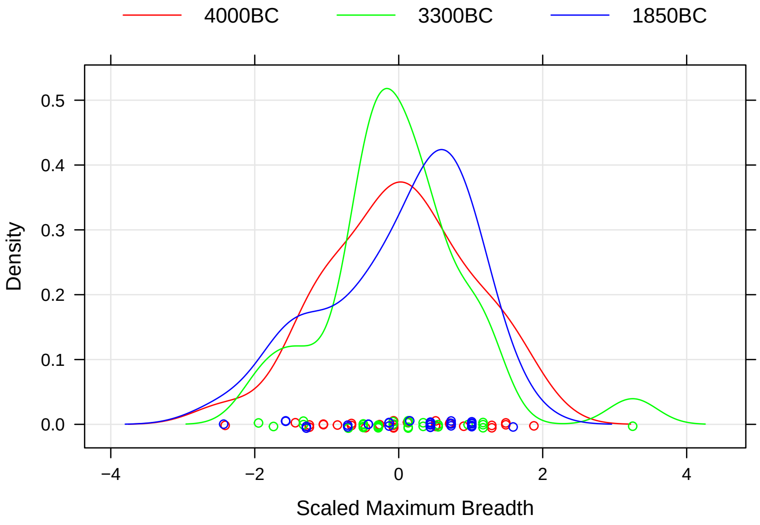


Comment: From the three scatterplot matrices, there does not seem to be any significant correlation between the variables used in the analysis.

2.3 Checking univariate normality

We now check if the measurements of each of the four variables are normally distributed period-wise. **Density plots** have been used for graphical visualization. We also use the **Shapiro-Wilk's test** to test normality. The Shapiro-Wilk's testing procedure tests the null hypothesis of normality against a non-normal alternative.

1. *Maximum Breadth*



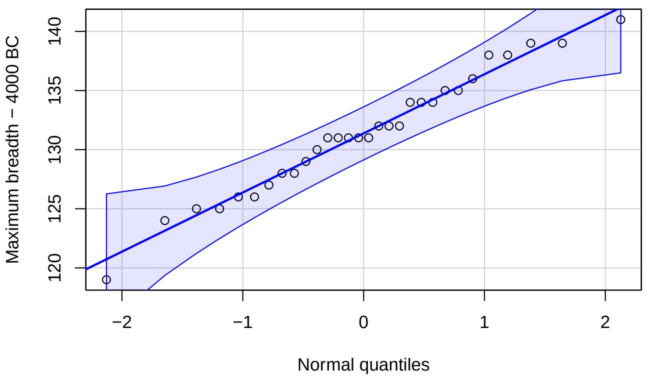
Comment:

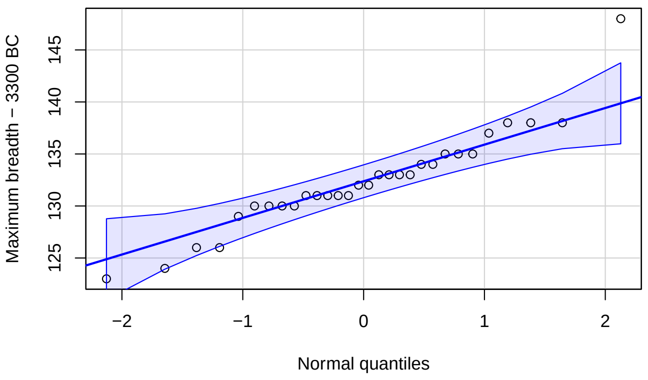
- The above density plot does not reveal sufficient evidence against normality.
- We notice that the scaled distribution of maximum breadth is negatively skewed for the period 1850 BC. This causes a separation of this period from the other two.

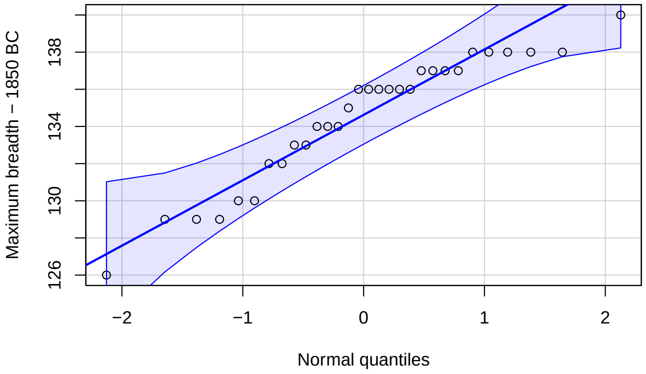
The p-values of the Shapiro Wilk's test has been tabulated below:

- 4000 BC: 0.8603
- 3300 BC: 0.04029
- 1800 BC: 0.03139

If the p-value of the test is >0.05, then we have a sufficient evidence to suggest that the univariate distribution is not significantly different from a univariate normal distribution (at 5% level of significance). However we should **not** rule out normality straight away on the basis of p-values alone, as the Shapiro Wilk's is highly sensitive to fluctuations which may have arisen due to chance causes. Inspecting **QQ (quantile-quantile) plots (with confidence bands)** of the variable might be a better way of checking for normality.

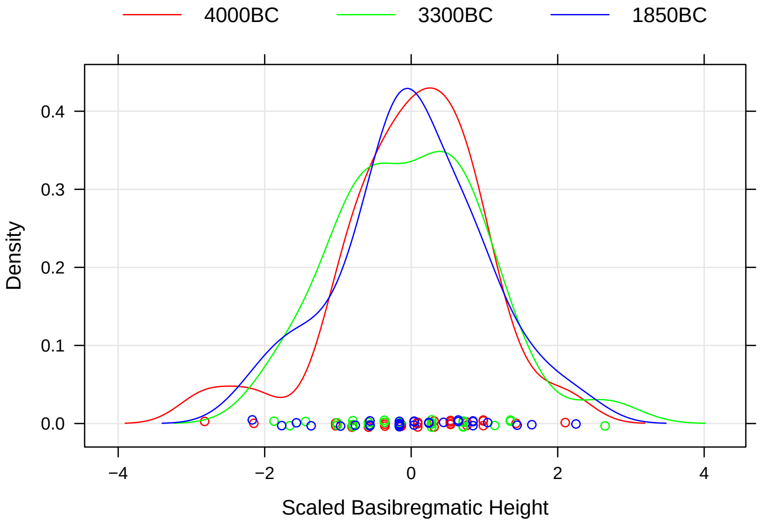






Comment: There does not seem to be sufficient evidence against normality.

2. *Basibregmatic Height*



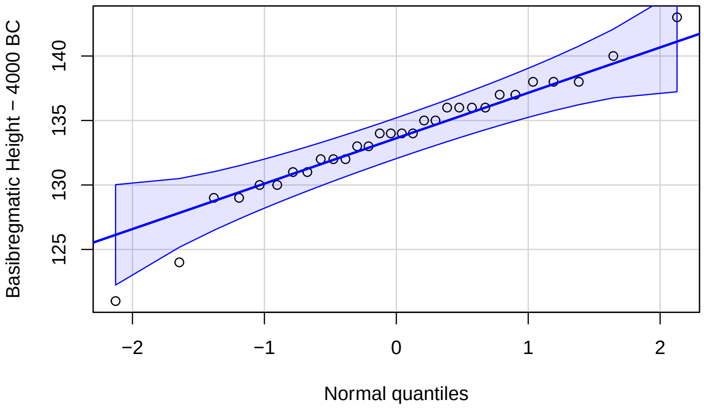
Comment:

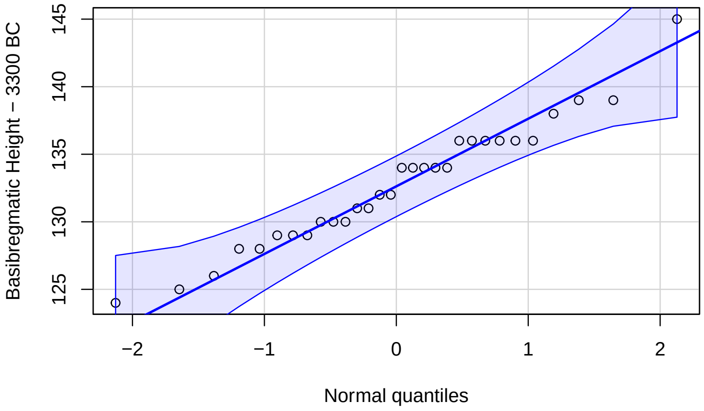
- The above density plot does not reveal sufficient evidence against normality.

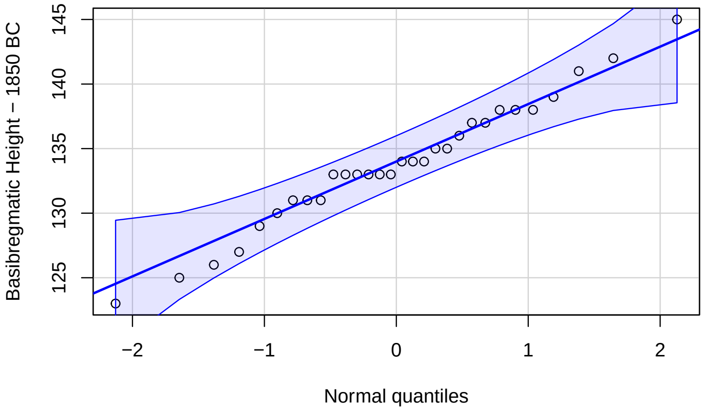
The p-values of the Shapiro Wilk's test has been tabulated below:

- 4000 BC: 0.2536
- 3300 BC: 0.5799
- 1800 BC: 0.9031

The QQ plots (with confidence bands) for the three periods are given below:

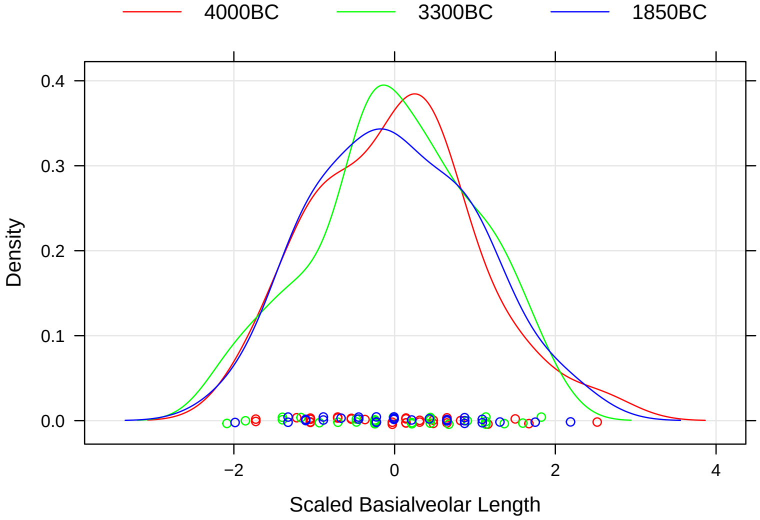






Comment: There does not seem to be sufficient evidence against normality.

3. *Basialveolar Length*



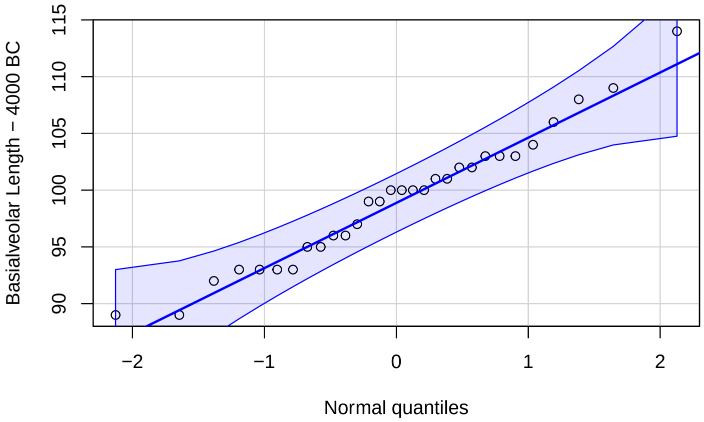
Comment:

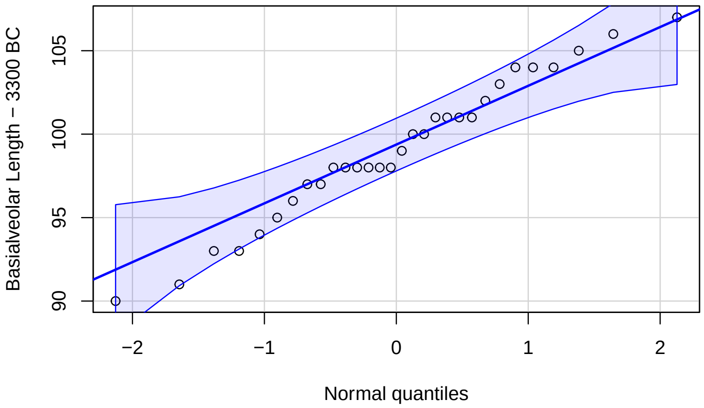
- The above density plot does not reveal sufficient evidence against normality.

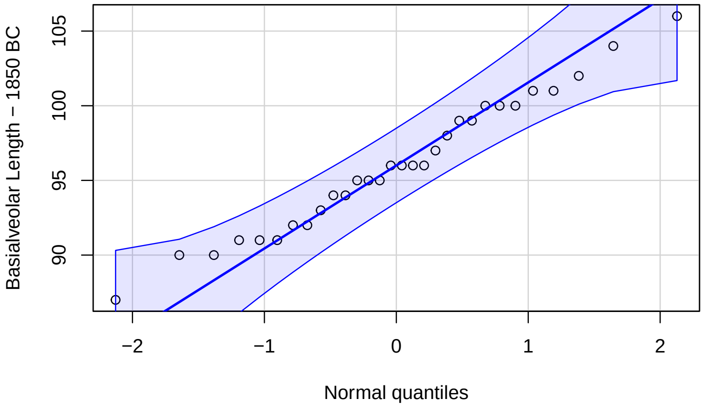
The p-values of the Shapiro Wilk's test has been tabulated below:

- 4000 BC: 0.6282
- 3300 BC: 0.7495
- 1800 BC: 0.8849

The QQ plots (with confidence bands) for the three periods are given below:

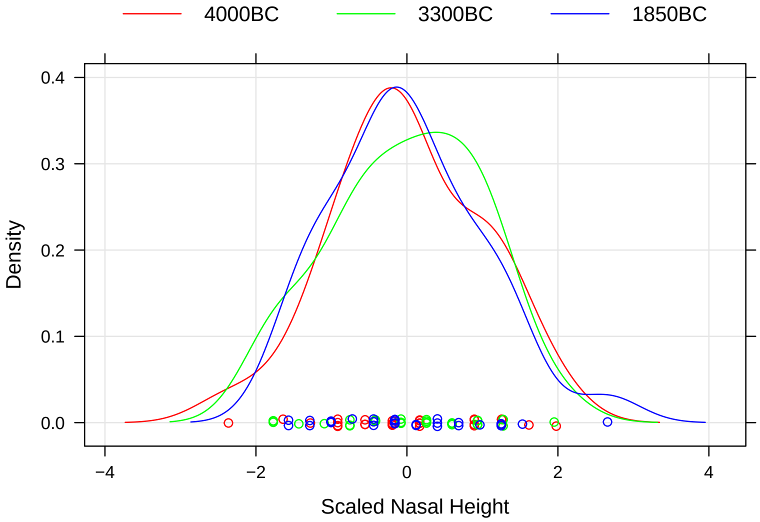






Comment: There does not seem to be sufficient evidence against normality.

4. *Nasal Height*



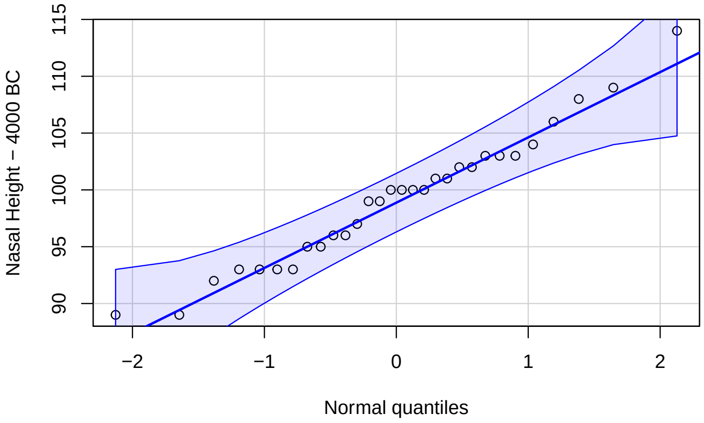
Comment:

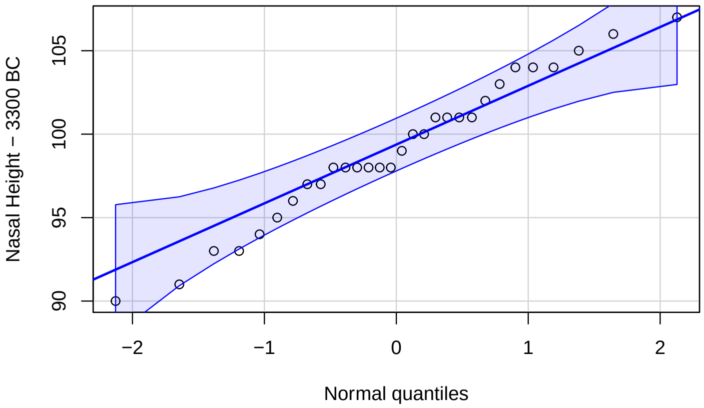
- The above density plot does not reveal sufficient evidence against normality.
- The scaled distribution of nasal height for 3300 BC is negatively skewed.

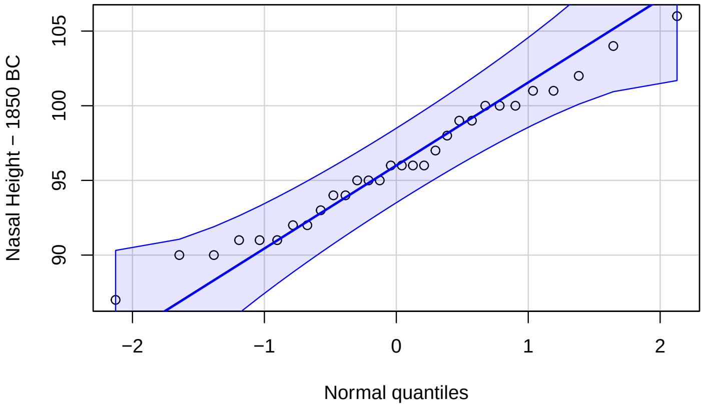
The p-values of the Shapiro Wilk's test has been tabulated below:

- 4000 BC: 0.6772
- 3300 BC: 0.3762
- 1800 BC: 0.3881

The QQ plots (with confidence bands) for the three periods are given below:







Comment: There does not seem to be sufficient evidence against normality.

A general observation from the p values of Shapiro Wilk's test: We fail to reject normality at 1% level of significance for any of the cases.

2.4 Checking Multivariate Normality

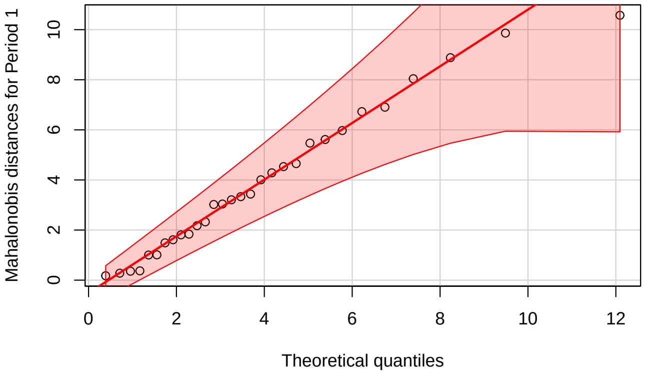
We use **Multivariate Shapiro-Wilk's test** to test for group-wise multivariate normality. However as before, this test is highly sensitive to random fluctuations which may have arisen due to chance causes. Thus we validate our observations by observing **Chi-square plots** (also known as *gamma plots*). In these diagrams, we plot the period wise *Mahalonobis distance* against the theoretical quantiles of the χ^2 distribution with degrees of freedom being equal to the number of variables.

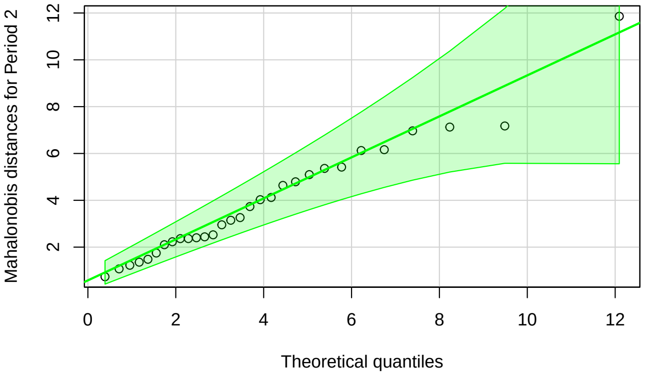
The period wise Multivariate Shapiro Wilk's test gave us the following p-values:

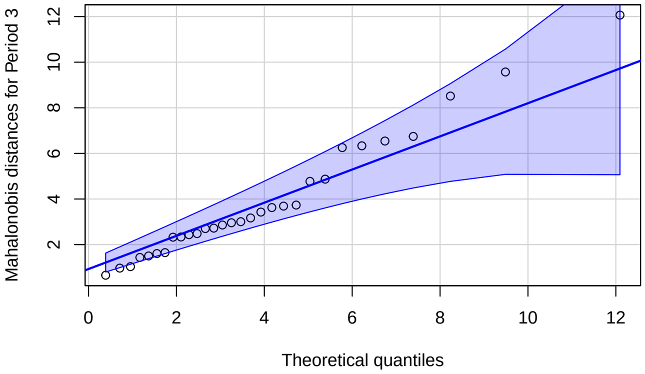
- Period 1: p-value = 0.05208
- Period 2: p-value = 0.01798

- Period 3: p-value = 0.01038

We now inspect the Chi-square plots as shown below:





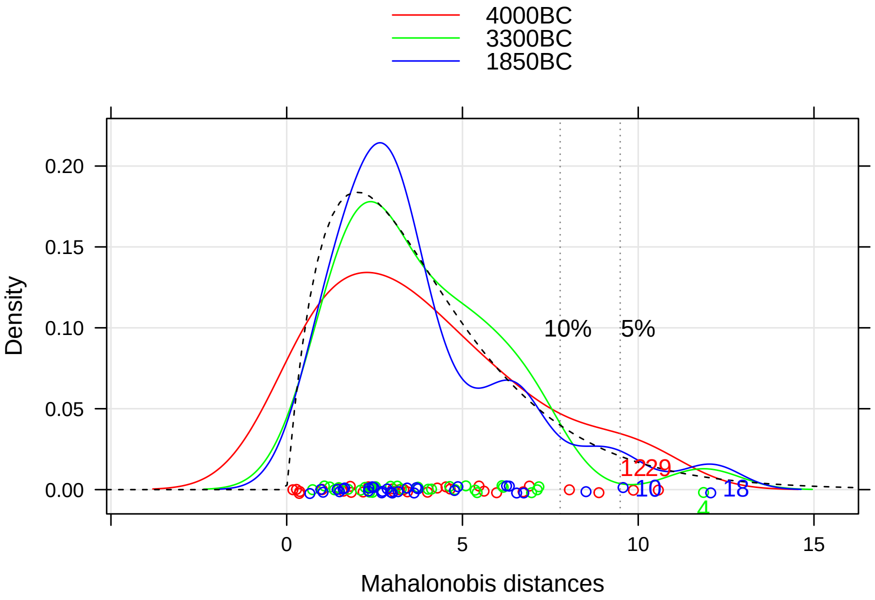


Comment: Based on the data and Chi-square plots, we do not have sufficient evidence to reject multivariate normality. Hence from this point forward, we assume that the three populations corresponding to the three time periods are *Multivariate Normal* with some mean vector and dispersion matrix.

2.5 Checking presence of outliers

We may be interested in checking if there are any outliers in our data. This because outliers may affect any future analyses we might perform on the data (especially MANOVA, which we intend to perform). To detect this, we use *Mahalonobis distance* as our metric.

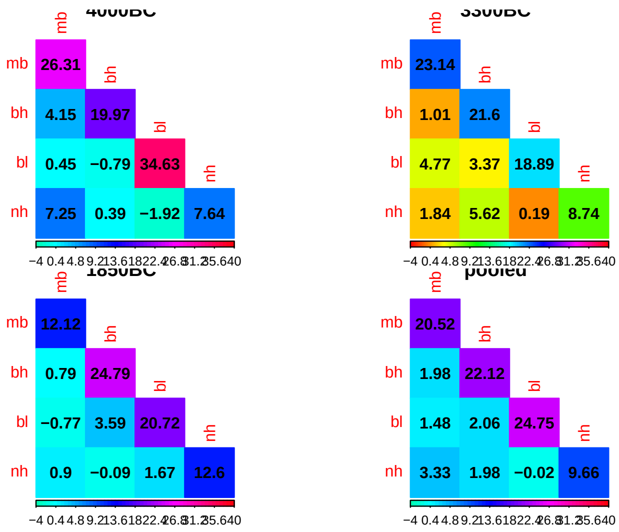
The density plot of the Mahalonobis distance for the three periods, with cut off points (corresponding to the χ^2 distribution with 4 degrees of freedom) at 10% and 5% level of significance have been plotted below.



Comments:

- There are not many outliers in the data
- Compared to other observations in 3300 BC, the 4th observation is quite different. This was also indicated in the univariate boxplot corresponding to *maximum breadth*).

2.6 Homogeneity of Dispersion Matrices



The above diagram graphically denotes the dispersion matrices for the three populations (periods). A critical assumption to carry out MANOVA is that the dispersion matrices corresponding to the three populations need to be **homogeneous**, i.e., equal. We perform **Box's M Test** to check if the dispersion matrices corresponding to the three populations (time periods) are homogeneous. Since we have already accepted the hypothesis that our populations are distributed with multivariate normal distributions, we can trust the conclusion of Box's M Test.

The p-value corresponding to the Box's M Test performed on our data is 0.3943. We thus have sufficient evidence to conclude that the dispersion matrices corresponding to three populations are equal.

3 MANOVA and further analysis

We try to answer the following question: **'is there any significant difference in the measurements of skulls among the three periods?** This can be answered using MANOVA (Multivariate Analysis of Variance).

MANOVA, which is an extension of univariate ANOVA requires the following assumptions:

- Independent Random Sampling from the population
- Dependent variables are multivariate normally distributed within each group of the independent variables (which are categorical).
- The population covariance matrices of each group are equal (which is an extension of the homoscedasticity assumption of univariate ANOVA).
- Dependent variables are not correlated.

In our exploratory data analysis we have already verified that the above assumptions hold. We thus proceed with our MANOVA.

3.1 Model and Testing Problem

- Dependent variables: Maximum breadth (mb), Basibregmatic height (bh),Basialveolar length (bl), Nasal Height (nh) - **Continuous** in nature
- Independent variables: Period - **Categorical** in nature

Our model would thus be: $\underline{y}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{\epsilon}_{ij}$

where:

- $\underline{y}_{ij}$: Vector of variables for the j^{th} skull in the i^{th} period.
- $\underline{\mu}$: General effect of variation in the skull measurements.
- $\underline{\alpha}_j$: Additional effect of the i^{th} period.
- $\underline{\epsilon}_{ij}$: Error term

We want to test the hypothesis $H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \underline{\mu}_3$ against H_1 : not all $\underline{\mu}_i$'s are equal, i.e., at least one pair of periods is different on at least one variable.

Using the usual test statistic and critical region, the MANOVA results are displayed below:

```
##           Df  Pillai approx F num Df den Df Pr(>F)
## period     2 0.17221   2.0021      8    170 0.0489 *
## Residuals 87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At 5% level of significance, we can reject our null hypothesis. Thus in light of the data, there is a significant difference of means between at least one pair of periods. We now perform **paired comparison tests** to check which pair causes the difference.

3.2 Paired Comparison tests

1. For pair 1: Period 1 (4000 BC) vs. Period 2 (3300 BC)

```
##           Df   Pillai approx F num Df den Df Pr(>F)
## period     1 0.027674  0.39135      4     55 0.8139
## Residuals 58
```

2. For pair 2: Period 2 (3300 BC) vs. Period 2 (1850 BC)

```
##           Df  Pillai approx F num Df den Df  Pr(>F)
## period     1 0.18976   3.2203      4     55 0.01908 *
## Residuals 58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

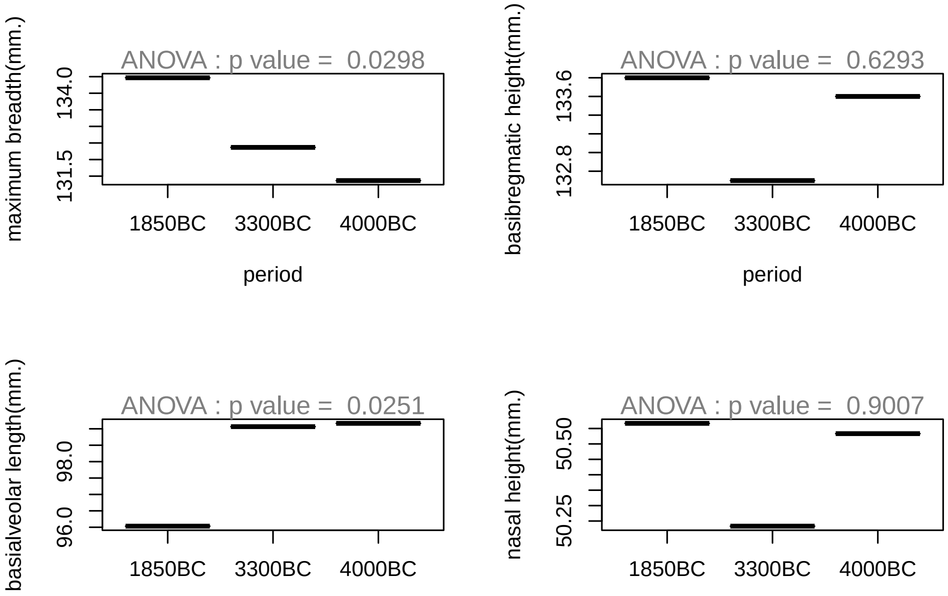
3. For pair 3: Period 1 (4000 BC) vs. Period 3 (1850 BC)

```
##           Df  Pillai approx F num Df den Df  Pr(>F)
## period     1 0.18757   3.1744      4     55 0.02035 *
## Residuals 58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In order to keep the overall size of the test 5%, we equi-allocate level (5/3)% to each of the paired comparison test. Note that at this level, we cannot reject any of the three hypotheses. However the high p-value (0.8139) for the first paired comparison suggests that the mean vectors for the periods 4000 BC and 3300 BC are **very similar**.

3.3 Component-wise ANOVA

We inspect the component-wise means for each period.



Comment:

- From the above plots, it is reasonably clear that there is significant difference of means for Maximum breadth and Basialveolar length.
- There seems to be an *increasing* trend in the maximum breadth and *decreasing* trend in basialveolar length with the passage of time.

To verify what we suspect, we use **Jonckheere-Terpstra** test. The Jonckheere-Terpstra test is a *rank-based nonparametric test* that can be used to determine if there is a *statistically significant trend* between an **ordinal/categorical independent variable** and a **continuous or ordinal dependent variable**. [This test is analogous to the **Kruskal Wallis test**, but with ordered alternatives.More information on this test and how it is carried out can be found here: https://en.wikipedia.org/wiki/Jonckheere%27s_trend_test]

We apply this test to 'mb' and 'bl' respectively. The results have been tabulated below:

Variable	Null	Alternative	p-values
mb	Equal	Increasing	0.04274
bl	Equal	Decreasing	0.01182

Thus, our trend hypotheses have been verified.

4 Principal Component Analysis

After performing MANOVA, we would like to know if we can reduce the dimensionality of our data by getting to know which variables explain its variability in a better, more efficient way. To get to know this information, we perform **Principal Component Analysis**.

4.1 Results of PCA

The results of PCA (after appropriately scaling the covariates) are displayed below:

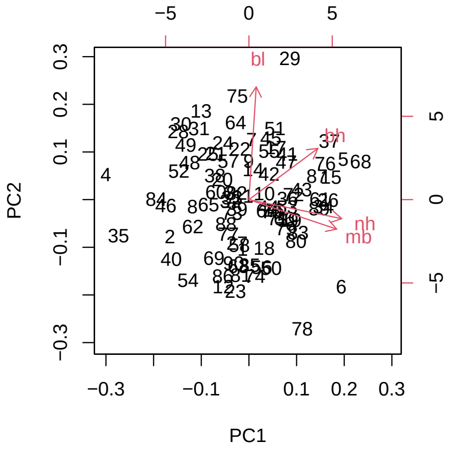
```
## Standard deviations (1, .., p=4):
## [1] 1.1493294 1.0175312 0.9376977 0.8742970
##
## Rotation (n x k) = (4 x 4):
##           PC1        PC2        PC3         PC4
## mb 0.60371980 -0.2291202  0.4015625 -0.64944124
## bh 0.47444000  0.3981122 -0.7658352 -0.17294464
## bl 0.05017241  0.8758528  0.4787798  0.03368199
## nh 0.63867974 -0.1479608  0.1517037  0.73971735
```

The four principal components are:

1. $PC_1 = 0.60mb + 0.47bh + 0.05bl + 0.64nh$
2. $PC_2 = -0.22mb + 0.39bh + 0.88bl - 0.14nh$
3. $PC_3 = 0.40mb - 0.76bh + 0.48bl + 0.15nh$
4. $PC_4 = -0.65mb + 0.17bh + 0.03bl + 0.74nh$

4.2 Biplot

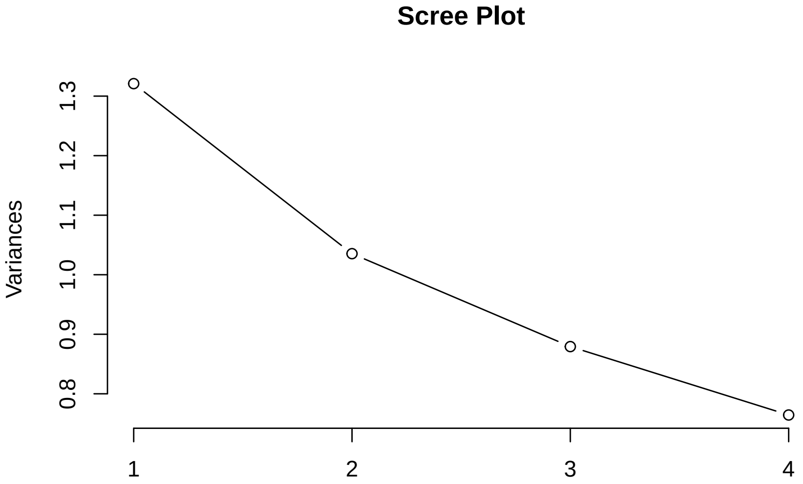
We study a **biplot** to check which variables have a major contribution to the overall variability inside the principal components.



Comment: From the biplot, it seems that mb, nh and bh (to a certain extent) contribute mainly to the first principal component. The contribution of bl is mainly to the second principal component.

4.3 Scree plot

To graphically visualize the variance of each principal component, we examine the **scree plot**.



Comment: Only 59% of the total variability in the data is explained by the first two principal components. Further still, only 81% is explained by the first three. We conventionally continue to add principal components till >90% of the variability is achieved. Since even three principal components do not yield >90% of variability, there seems to be little point in using them.

We do not use PCA in our analysis further.

5 (Linear) Discriminant Analysis and Classification

5.1 Strategy

We are now interested in trying to answer the query **"given new skull measurements, can we tell which period the skull belongs to?**

To do this, we perform the following tasks:

1. We split the data into two parts: a **training** set and a **testing set** with 80:20 ratio.
2. We apply **Discriminant Analysis** on the training set to find axes which can discriminate (or separate) the data the most based on period.

3. We then try to develop a classification rule based on our discriminant analysis and apply it to the testing set. This would enable us to assess how the classification rule performs on “new” data not used to formulate the discrimination rule.

5.2 LDA on Training Data

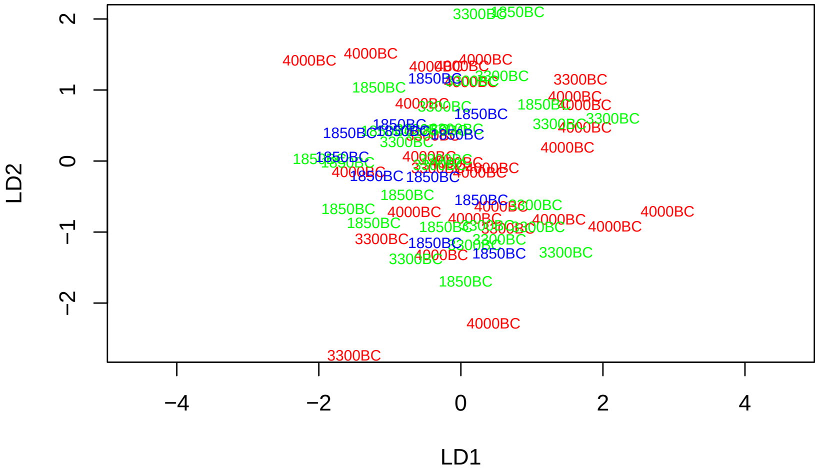
Since we want to randomly split the data in hand into two sets, we do so in a *stratified manner*. This would ensure that all groups are equally represented in both the sets and thus we can minimize bias while discriminating and classifying.

Since Box’s M Test ensures that the (population) dispersion matrices of the three periods are homogeneous, we can use **Linear Discriminant Analysis**. We work with the special case that the prior probabilities of each of the three groups are same (=0.33) and that each group incurs equal costs of misclassification. Performing LDA under these assumptions gives us the following results:

```
## Call:
## lda(period ~ . - period, data = data1)
##
## Prior probabilities of groups:
##    4000BC    3300BC    1850BC
## 0.3333333 0.3333333 0.3333333
##
## Group means:
##              mb       bh       bl       nh
## 4000BC 131.2500 133.2083 98.75000 50.33333
## 3300BC 132.1250 132.4167 98.87500 50.08333
## 1850BC 134.5417 133.2500 96.16667 50.66667
##
## Coefficients of linear discriminants:
##            LD1         LD2
## mb -0.16590528 -0.11182137
## bh -0.02456190  0.14499203
## bl  0.15242099 -0.08567125
## nh  0.03887379  0.11879371
##
## Proportion of trace:
##    LD1    LD2
## 0.9403 0.0597
```

Studying the above output, we can see that 94% of the separation in the training data has been achieved by LD1 alone. Inspecting the coefficients of LD1, we see that ‘mb’ and ‘bl’ have a major contribution. Also notice that they have opposite signs.

In order to graphically study the extent of separation done by the two linear discriminants, we look at the following plot.

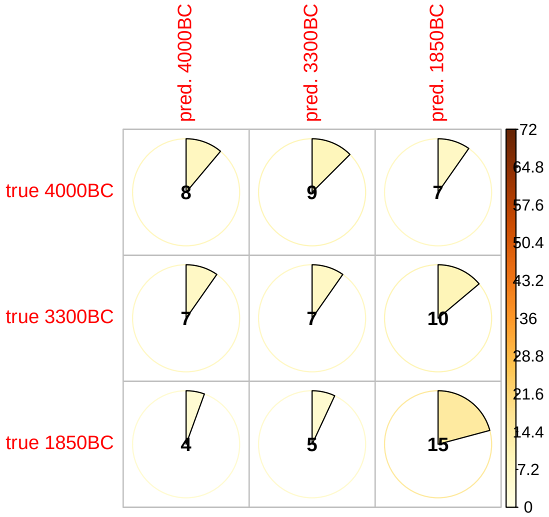


Comment: As we can see, most of the separation is done by the first linear discriminant alone. However we cannot clearly separate any of three groups, which is not a good sign.

We concretely assess the performance of our discrimination rule on the testing data set.

5.3 Performance

First we see how our discrimination rule performs discrimination on the training data set.

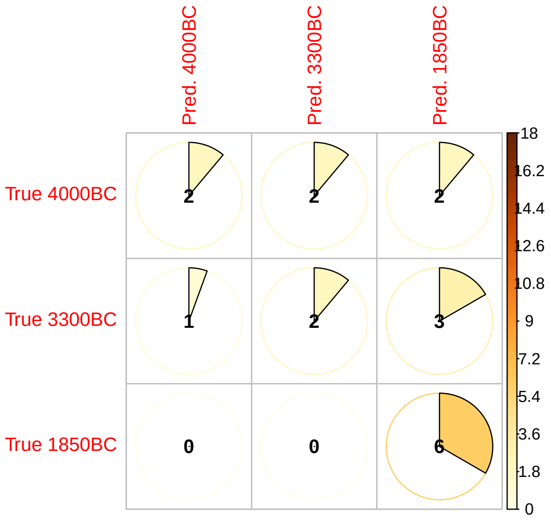


Comment:

- The discrimination is best for 1850 BC (Period 3)
- The misclassification rate (1- total % of correct classifications) is 58.33%

We now classify the 'unseen' observations of the testing data into periods using the discrimination rule we have devised from the training data.

The results have been displayed below:



Comment:

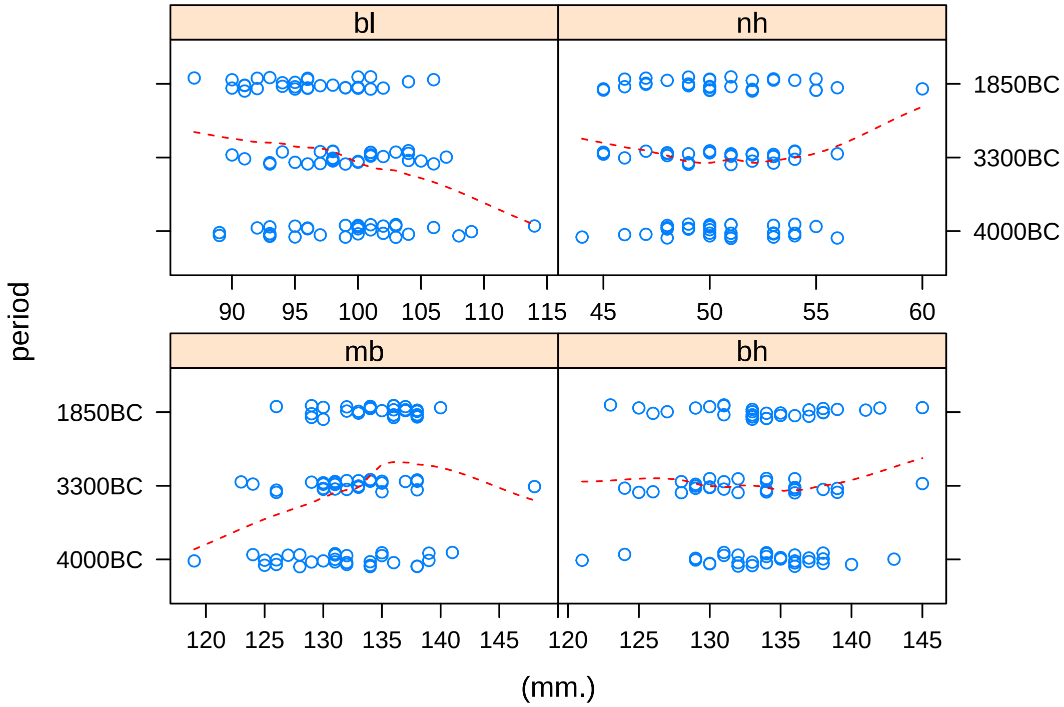
- We see that the most number of correct classifications occur for 1850 BC (Period 3). In fact, there are no misclassifications for this period.
- The misclassification occurs most frequently for 3300 BC (Period 2).
- The misclassification rate (1- total % of correct classifications) is 44.44%

6 Multinomial Logistic Regression

6.1 Motivation

Multinomial logistic regression is a classification method that generalizes logistic regression to multiclass problems, i.e. with more than two possible discrete outcomes. It is a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables, which may be real-valued, binary-valued, categorical-valued etc.

In our case, the conditional distribution of the response (dependent variable) given the covariate information is $Multinomial(n = 3, \mu_1(x), \mu_2(x), \mu_3(x))$, and our independent variables are continuous. Hence, we may try multinomial logistic regression.



We observe that the fitted LOESS curves could also suggest the use of multinomial logistic regression.

6.2 Model, Prediction and Performance

We fit a multinomial logistic regression model to the training data and check how well it performs discrimination on this dataset.

We fix 1850 BC (period 3) as the *reference level* and perform the procedure. The results of the fitting are displayed below.

```
##             log(mu[,1]/mu[,3]) log(mu[,2]/mu[,3])
## (Intercept)         8.95549873        8.382332602
## mb                 -0.18027237       -0.132133022
## bh                 -0.01023755       -0.043738364
## bl                  0.14395531        0.150142889
## nh                  0.04679874        0.008397481
```

Comment:

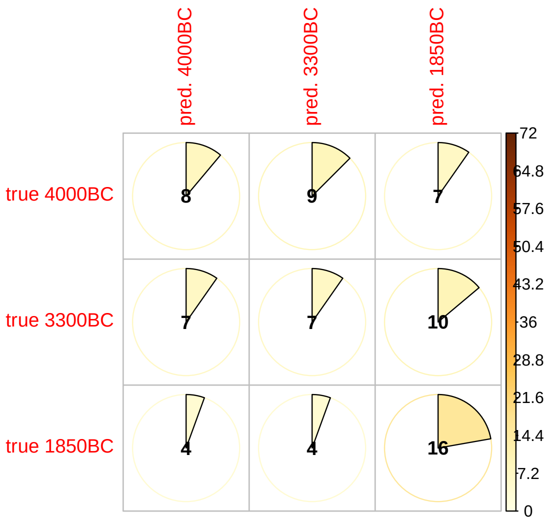
- Studying the coefficients of 'mb' and 'bl', we suspect that these two variables have an important role to play in the model. This also is in line with what we suspected from our earlier plots.

We peform the prediction as follows.

- If $log(\hat{p_1}/\hat{p_3}) < 0$ and $log(\hat{p_2}/\hat{p_3}) < 0$, then $\hat{p_3} > max(\hat{p_1}, \hat{p_2})$, and we classify the observation in Period 3.
- If $log(\hat{p_1}/\hat{p_3}) > 0$ and $log(\hat{p_2}/\hat{p_3}) < 0$, then $\hat{p_1} > \hat{p_3} > \hat{p_2}$, and we classify the observation in Period 1.
- If $log(\hat{p_1}/\hat{p_3}) < 0$ and $log(\hat{p_2}/\hat{p_3}) > 0$, then $\hat{p_2} > \hat{p_3} > \hat{p_1}$, and we classify the observation in Period 2.
- If $log(\hat{p_1}/\hat{p_3}) > 0$ and $log(\hat{p_2}/\hat{p_3}) > 0$, then if $log(\hat{p_1}/\hat{p_3}) > log(\hat{p_2}/\hat{p_3})$, then we classify the observation in Period 1, otherwise Period 2.

(Note that the *predict* function in R gives us the values $log(\hat{p_1}/\hat{p_3})$ and $log(\hat{p_2}/\hat{p_3})$ respectively.)

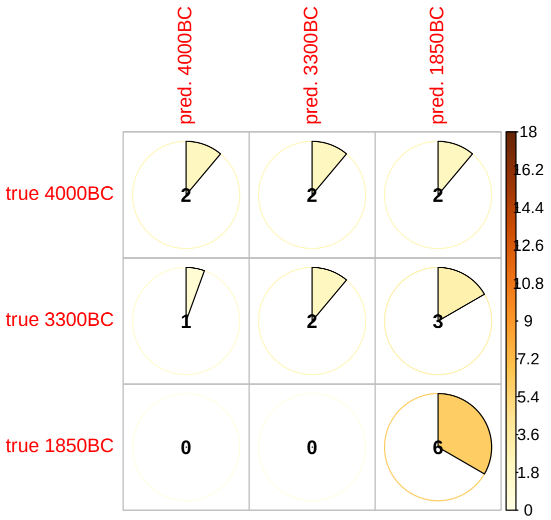
We now check how the model discriminates on the training data. The results are displayed below.



Comment:

- The misclassification rate is 56.94%

We now check how the model "classifies" observations in the testing data set. Using the algorithm for classification as above, the results have been displayed below.



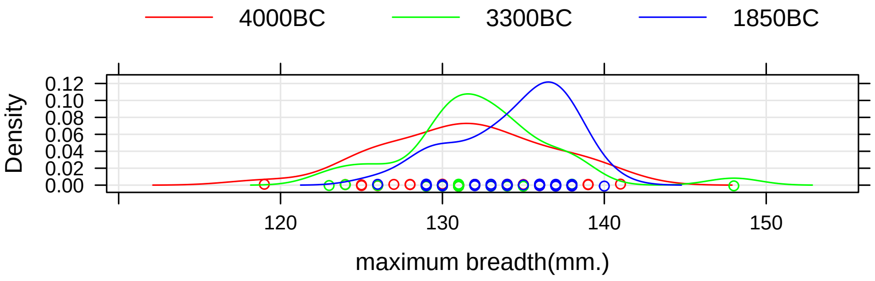
Comment:

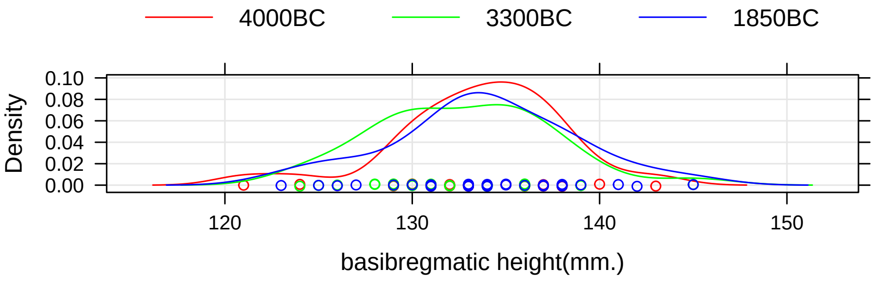
- The misclassification rate is 44.44%
- This suggests no improvement whatsoever over LDA.

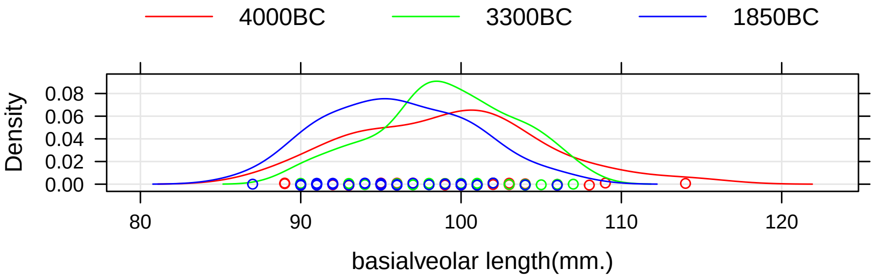
7 Conclusion

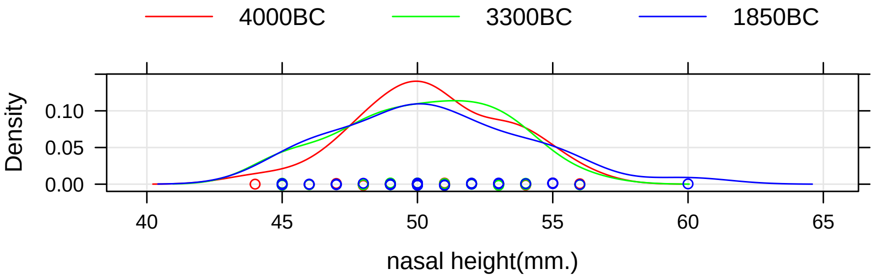
7.1 Why do we get such a performance?

We notice that LDA and Multinomial Logistic Regression give us the same misclasssifcation rate, and hence we cannot prefer to use one over the other for prediction purposes. A plausible reason for this could be understood by looking at the density plots below.









Comments:

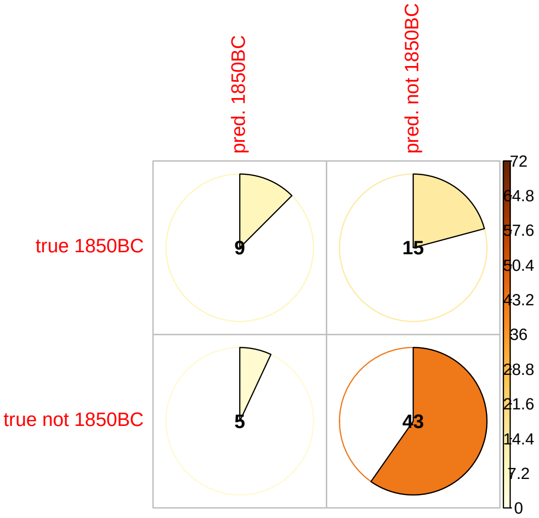
- It is clear from the above plots that the data is overlapping to a great extent (for ‘bh’ and ‘nh’) to clearly separate out and distinguish effects among periods.
- For ‘mb’ and ‘bl’, notice that only for period 3 (i.e., 1850 BC) are the observations quite different from periods 1 and 2. This is why the number of misclassifications for LDA and Multinomial Logistic Regression are least for this period. One reason for this might be that the time periods are not equi-spaced: there is more difference in time between periods 2 and 3 than it is between 1 and 2. As a result, the evolution in skull dimensions has not been effectively captured in the data.

7.2 Exploring some alternative approaches

1. **Applying LDA on Period 3 alone**

Motivated by our previous discussion, it would seem logical to apply LDA on on period 3 alone (due to the sufficiently clear mean separation between period 3 and the other two).

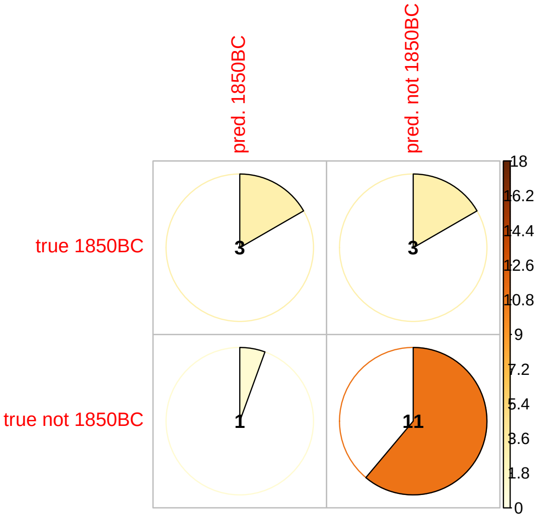
Applying LDA on the training data and assessing how the discrimination rule performs, we observe the following:



Comment:

- The misclassification rate is 27.78%
- This would imply that the discrimination rule can discriminate observations in the training data on the basis of whether they belong to Period 3 or not with a good accuracy.

Finally using this discrimination rule on the testing data and assessing the accuracy of classification, we observe the following:

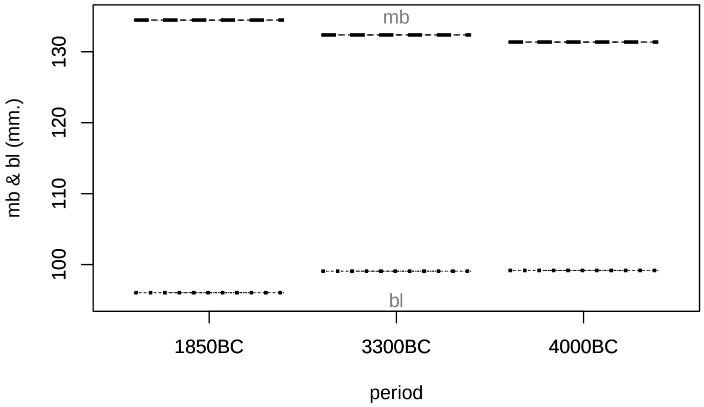


Comment:

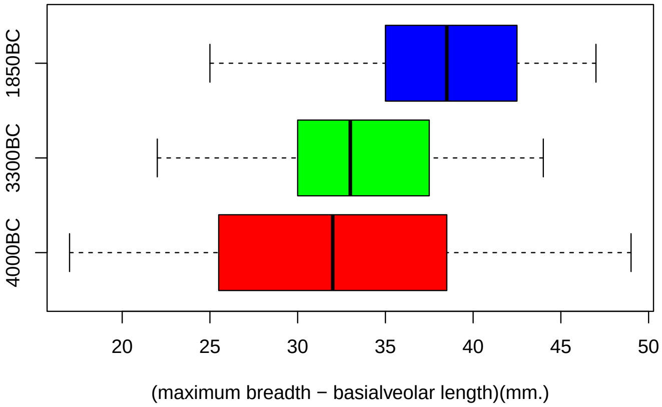
- The misclassification rate is 22.22%
- This would imply that the discrimination rule can classify observations in the testing data on the basis of whether they belong to Period 3 or not with a good accuracy.

2. **A new classifier?**

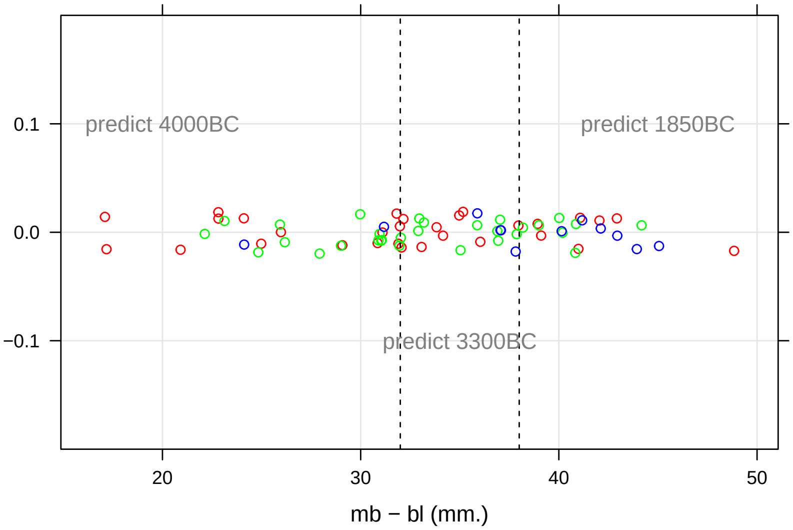
Refering back to the section of component wise ANOVA , we see that *mb* is increasing and *bl* is decreasing with time. Combining the two in a single plot, we see *mb-bl* **increases** on an average from Period 1 to 3. This could hint at the use of *mb-bl* as a classifier while performing discriminant analysis.



Period wise boxplots of *mb-bl* give us similar conclusions (i.e., *mb-bl* increases on an average with the passage of time)



Note that if we plot the values of *mb-bl* for the three populations on a single plot, the result would look like this:

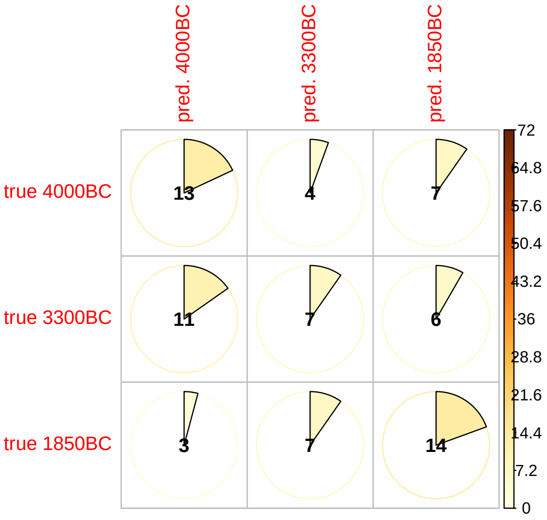


The above plot (complemented by the fact that *mb-bl* shows an increasing trend) could suggest the following rule for classification/discrimination:

- If the value of *mb-bl* for the i^{th} observation is **lesser than (or equal to)** the 33.33% quantile, then we classify the observation into **Period 1**, i.e., 4000 BC.
- Else, if the value of *mb-bl* for the i^{th} observation is **greater than the 33.33% quantile** but **lesser than (or equal to) the 66.67% quantile**, we classify the observation into Period 2, i.e., 3300 BC.
- Else, we classify the observation into **Period 3**, i.e., 1850 BC.

(Note that we equally split the total variation into three equal parts. The divisions could have been made unequally as well. However this might have required some proper justification.)

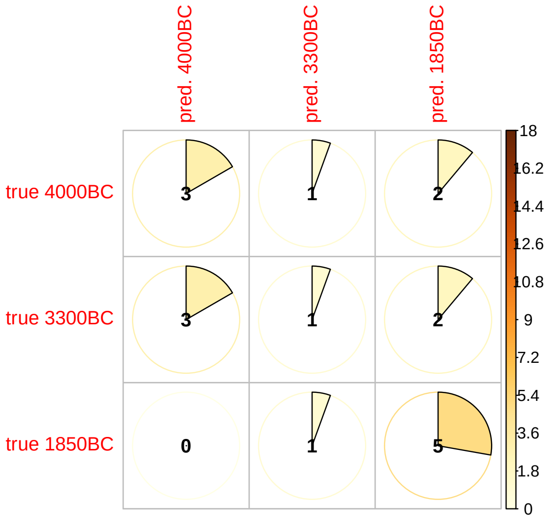
Implementing this discrimination/classification rule on the *training data*, we get the following results:



Comment:

- The misclassification rate is 52.78% approximately.

Using this discrimination rule on the *testing data* and assessing the accuracy of classification, we observe the following:



Comment:

- The misclassification rate is 50%

3. **Using only *mb* and *bl* in Multinomial Logistic Regression**

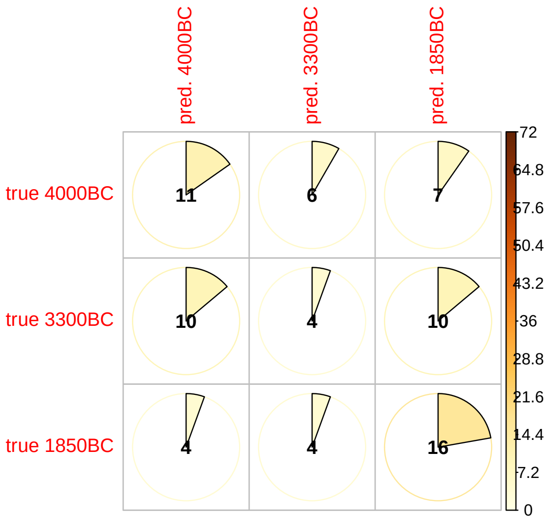
Motivated by our conclusion from our multinomial logistic regression analysis, we try using only *mb* and *bl* as covariates.

We fit a multinomial logistic regression model to the training data and check how well it performs discrimination on this dataset.We fix 1850 BC (period 3) as the *reference level* and perform the procedure. The results of the fitting are displayed below.

```
##             log(mu[,1]/mu[,3]) log(mu[,2]/mu[,3])
## (Intercept)          9.6560367          3.7865015
## mb                  -0.1733679         -0.1337963
## bl                   0.1376550          0.1443183
```

We perform discrimination (and classification) using a similar algorithm as proposed before.

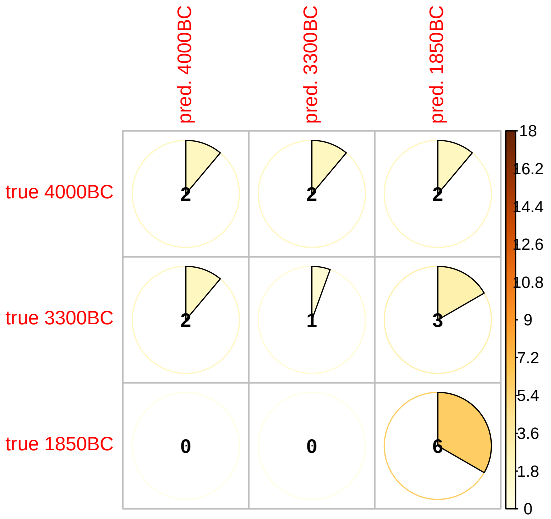
We now check how the model discriminates on the training data. The results are displayed below.



Comment:

- The misclassification rate is 52.78%

Finally using this discrimination rule on the testing data and assessing the accuracy of classification, we observe the following:



Comment:

- The misclassification rate is 50%
- This does not suggest any improvement compared to the previous multinomial logistic regression model.