Multivariate Analysis of Egyptian Skulls

Terry Situ

Introduction

We analyzed a data set on the measurements performed on 150 male Egyptian skulls over 5 time periods. Period 1 is 4000 B.C.; period 2 is 3300 B. C.; period 3 is 1850 B. C.; period 4 is 200 B.C.; and period 5 is 150 A.D. The variables consisted of the maximum breadth of skull in mm, basibregmatic height of the skull in mm, basialveolar length of skull in mm, and also nasal height of the skull. More information on these attributes of the skull can be found in the appendix. Researchers suggested that a change in skull size over time is evidence of interbreeding of a resident population with immigrant populations.

The purpose of our study was to analyze whether or not there was a significant change in the measurements of the skulls over time. We perform a multivariate analysis using MANOVA on this data set. We check out normality assumptions, perform hypothesis tests on the differences in the means of the measurements between time periods and construct confidence intervals on these difference of means, principle component analysis, and outlier detection analysis.

Verifying Assumptions

We first verify assumptions of equal covariance matrices for the data on the skulls within each time period. We perform Box's test for equality of covariance matrices. We calculate a C value of 45.667 and compare it to Box's chi squared approximation. At an alpha of .05, we fail to reject our null hypothesis and thus conclude that we can assume they have equal covariance matrices.

We plot the scatter plots and density plots in order to check our assumptions of normality.

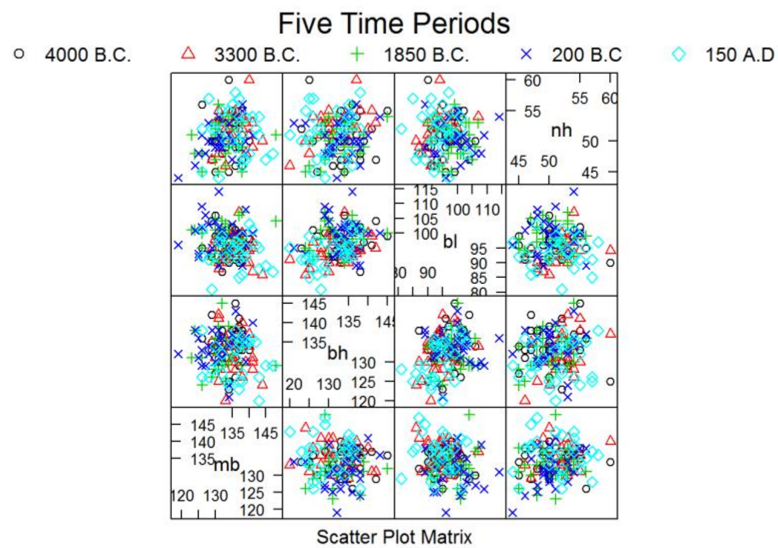


Figure 1: Scatter plot matrix

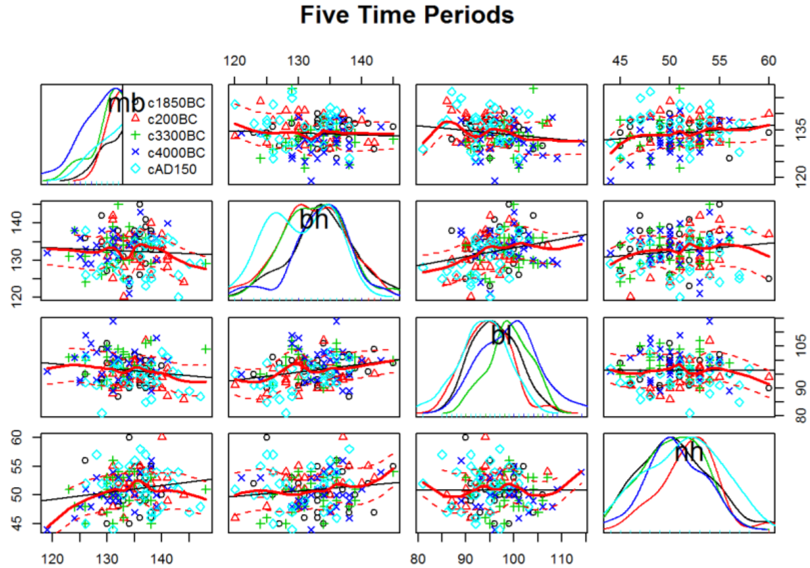


Figure 2: Density plot matrix

Because the scatterplots are roughly elliptical in shape and the density plots have the shape of a Gaussian kernel, our assumptions of bivariate normality are not violated. We also perform the Anderson-Darling test for normality and the Shapiro Wilk's Normality Test for each time period, and we plot the QQ-plots for each variable from each time interval. Refer to appendix, sections 4-8. A few of these normality checks fail but overall the QQ-plots are well behaved as the data generally falls along the line. So with all these results in mind, we conclude that our data is normally distributed.

Outlier Detection

Using the LOF algorithm to detect outliers on the entire data set, we found the top 5 outliers to be observations 147, 29, 34, 63, and 131. When detecting the multivariate outliers on the entire dataset based on Mahalanobis distance and the adjusted Mahalanobis's distance, we also detect 5 outliers. When detecting outliers within each time period by using box plots, we detect 3 outliers. Refer to appendix, section 3. We decide not to exclude the outliers because we have no reason to believe any measurements were taken improperly.

Descriptive Statistics

The descriptive statistics for the data set are as follows:

Table 1: Matrix of Mean Vectors: x_bar_i corresponds to the i^{th} time period

```
##        x_bar_1   x_bar_2   x_bar_3   x_bar_4   x_bar_5
## mb 131.36667 132.36667 134.46667 135.50000 136.16667
## bh 133.60000 132.70000 133.80000 132.30000 130.33333
## bl  99.16667  99.06667  96.03333  94.53333  93.50000
```

```
## nh  50.53333  50.23333  50.56667  51.96667  51.36667
```

Table 2: Covariance Matrix for the data: S_i corresponds to the covariance matrix for the i^{th} time period

```
s1
##           mb         bh          bl          nh
## mb 26.309195  4.1517241  0.4540230  7.2459770
## bh  4.151724 19.9724138 -0.7931034  0.3931034
## bl  0.454023 -0.7931034 34.6264368 -1.9195402
## nh  7.245977  0.3931034 -1.9195402  7.6367816
s2
##           mb        bh          bl        nh
## mb 23.136782  1.010345  4.7678161 1.8425287
## bh  1.010345 21.596552  3.3655172 5.6241379
## bl  4.767816  3.365517 18.8919540 0.1908046
## nh  1.842529  5.624138  0.1908046 8.7367816
s3
##            mb          bh          bl          nh
## mb 12.1195402  0.78620690 -0.7747126  0.89885057
## bh  0.7862069 24.78620690  3.5931034 -0.08965517
## bl -0.7747126  3.59310345 20.7229885  1.67011494
## nh  0.8988506 -0.08965517  1.6701149 12.59885057
s4
##           mb        bh        bl       nh
## mb 15.362069 -5.534483 -2.172414 2.051724
## bh -5.534483 26.355172  8.110345 6.148276
## bl -2.172414  8.110345 21.085057 5.328736
## nh  2.051724  6.148276  5.328736 7.964368
s5
##            mb         bh          bl          nh
## mb 28.6264368 -0.2298851 -1.8793103 -1.9942529
## bh -0.2298851 24.7126437 11.7241379  2.1494253
## bl -1.8793103 11.7241379 25.5689655  0.3965517
## nh -1.9942529  2.1494253  0.3965517 13.8264368
```

Table 3: Correlation Matrix for the Data: $corr_i$ corresponds to the correlation matrix of the i^{th} population

```
corr1
```

```
##           mb          bh          bl          nh
## mb 1.0000000  0.18111709  0.01504250  0.51119663
## bh 0.1811171  1.00000000 -0.03015856  0.03182998
## bl 0.0150425 -0.03015856  1.00000000 -0.11804243
## nh 0.5111966  0.03182998 -0.11804243  1.00000000
```

```
corr2
```

```
##           mb        bh         bl         nh
## mb 1.0000000 0.0451987 0.22804974 0.12959464
## bh 0.0451987 1.0000000 0.16661771 0.40943788
## bl 0.2280497 0.1666177 1.00000000 0.01485165
## nh 0.1295946 0.4094379 0.01485165 1.00000000
```

```
corr3
```

```
##             mb           bh          bl           nh
## mb  1.00000000  0.045361639 -0.04888452  0.072740949
## bh  0.04536164  1.000000000  0.15853987 -0.005073467
## bl -0.04888452  0.158539868  1.00000000  0.103360525
## nh  0.07274095 -0.005073467  0.10336052  1.000000000
```

```
corr4
```

```
##            mb         bh         bl        nh
## mb  1.0000000 -0.2750548 -0.1207064 0.1854891
## bh -0.2750548  1.0000000  0.3440480 0.4243701
## bl -0.1207064  0.3440480  1.0000000 0.4112076
## nh  0.1854891  0.4243701  0.4112076 1.0000000
```

```
corr5
```

```
##              mb           bh          bl          nh
## mb  1.000000000 -0.008643059 -0.06946376 -0.10024009
## bh -0.008643059  1.000000000  0.46640665  0.11628069
## bl -0.069463759  0.466406653  1.00000000  0.02109056
## nh -0.100240089  0.116280694  0.02109056  1.00000000
```

Results

We perform a MANOVA analysis. Thus we reject the null hypothesis that all of the mean vectors are the same. In order to identify which of the mean vectors are different, we inspect the Scheffe's 95% confidence intervals. We opted to use the Scheffe's confidence intervals over Bonferroni's after comparing $\sqrt{C_\alpha}$ with Bonferroni's T value. Scheffe's multiplier was smaller than Bonferroni's multiplier.

Table 4: MANOVA test results

```
##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##             mb          bh         bl        nh
## mb 3061.066667    5.333333   11.46667  291.3000
## bh    5.333333 3405.266667  754.00000  412.5333
## bl   11.466667  754.000000 3505.96667  164.3333
## nh  291.300000  412.533333  164.33333 1472.1333
##
## ------------------------------------------
##
## Term: epoch
##
## Sum of squares and products for the hypothesis:
##           mb         bh        bl          nh
## mb  502.8267 -228.14667 -626.6267  135.43333
## bh -228.1467  229.90667  292.2800  -66.06667
## bl -626.6267  292.28000  803.2933 -180.73333
## nh  135.4333  -66.06667 -180.7333   61.20000
##
## Multivariate Tests: epoch
##                  Df test stat  approx F num Df   den Df     Pr(>F)

## Pillai            4 0.3533056  3.512037     16 580.0000 4.6753e-06 *
**
## Wilks             4 0.6635858  3.900928     16 434.4548 7.0102e-07 *
**
## Hotelling-Lawley  4 0.4818191  4.230974     16 562.0000 8.2782e-08 *
**
```

```
## Roy                  4 0.4250954 15.409707      4 145.0000 1.5883e-10 *
**
```

Table 5: Scheffe's 95% Confidence Intervals

	Lower	Upper
4000 B.C. - 3300 B.C.		
mb	-4.739220	2.739220
bh	-3.043848	4.843848
bl	-3.901737	4.101737
nh	-2.293094	2.893094
4000 B.C. - 1850 B.C.		
mb	-6.8392206	0.6392203
bh	-4.1438484	3.7438484
bl	-0.8684037	7.1350704
nh	-2.6264271	2.5597605
4000 B.C. - 200 B.C.		
mb	-7.8725536	-0.3941131
bh	-2.6438484	5.2438484
bl	0.6315963	8.6350704
nh	-4.0264271	1.1597605
4000 B.C. - 150 A.D.		
mb	-8.5392203	-1.060780
bh	-0.6771818	7.210515
bl	1.6649296	9.668404
nh	-3.4264271	1.759760
3300 B.C. - 1850 B.C.		
mb	-5.8392203	1.639220
bh	-5.0438484	2.843848
bl	-0.9684037	7.035070
nh	-2.9264271	2.259760
3300 B.C. - 200 B.C.		
mb	-6.8725536	0.6058869
bh	-3.5438484	4.3438484
bl	0.5315963	8.5350704
nh	-4.3264271	0.8597605
1850 B.C. - 200 B.C.		
mb	-4.772554	2.705887
bh	-2.443848	5.443848
bl	-2.501737	5.501737
nh	-3.993094	1.193094
1850 B.C. - 150 A.D.		
mb	-5.4392203	2.039220
bh	-0.4771818	7.410515
bl	-1.4684037	6.535070

nh	-3.3930938	1.793094
200 B.C. – 150 A.D.		
mb	-4.405887	3.072554
bh	-1.977182	5.910515
bl	-2.968404	5.035070
nh	-1.993094	3.193094

After observing the confidence intervals for 4000 B.C. and 200 B.C. we conclude that the maximum breadth of the skull for Egyptians grew over that period of time and the nasal length shrunk.

Further Analysis

We performed a Principal Components analysis to see if it was possible to reduce the number of dimensions. We find that the first 3 principal components account for approximately 90% of the total variation. We decide that simply removing one dimension isn't significant enough to warrant further analysis.

```
Importance of components:
                          PC1    PC2    PC3    PC4
Standard deviation     5.9646 4.8717 4.3540 3.0431
Proportion of Variance 0.4064 0.2712 0.2166 0.1058
Cumulative Proportion  0.4064 0.6776 0.8942 1.0000
```

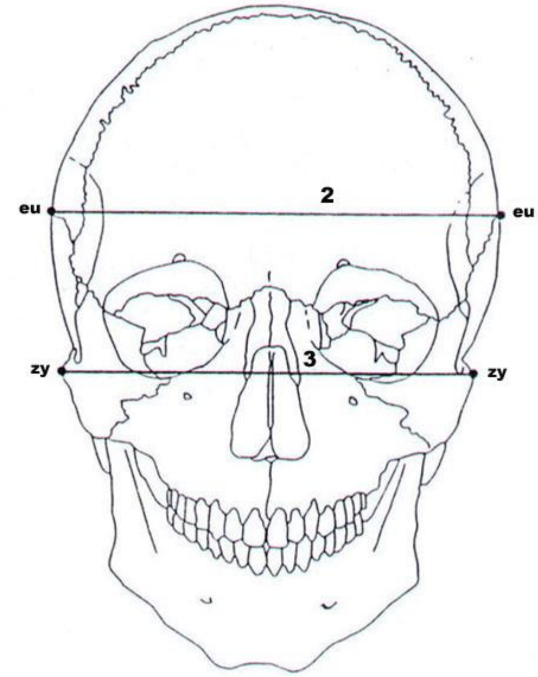
Reference

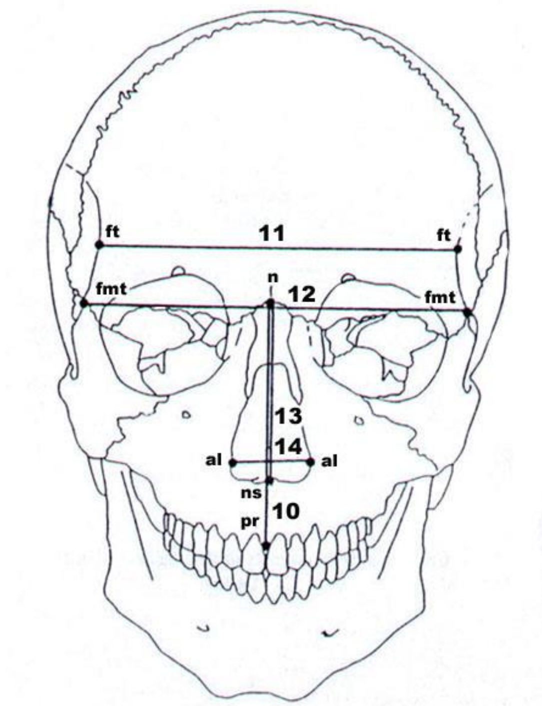
http://www.redwoods.edu/instruct/agarwin/anth_6_measurements.html

Appendix

Section 1: Definition of Skull Terminology

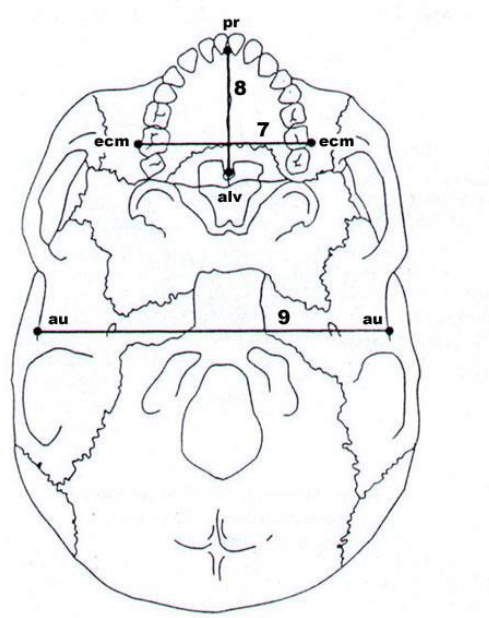
2→Maximum Cranial Breadth (eu-eu): maximum width of skull perpendicular to midsagittal plane wherever it is located, with the exception of the inferior temporal lines and the area immediately surrounding them.

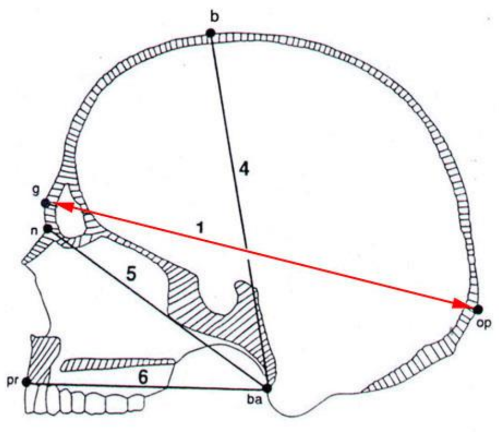




13→Nasal height of skull: direct distance from nasion (n) to the midpoint of a line connecting the lowest points of the inferior margin of the nasal notches

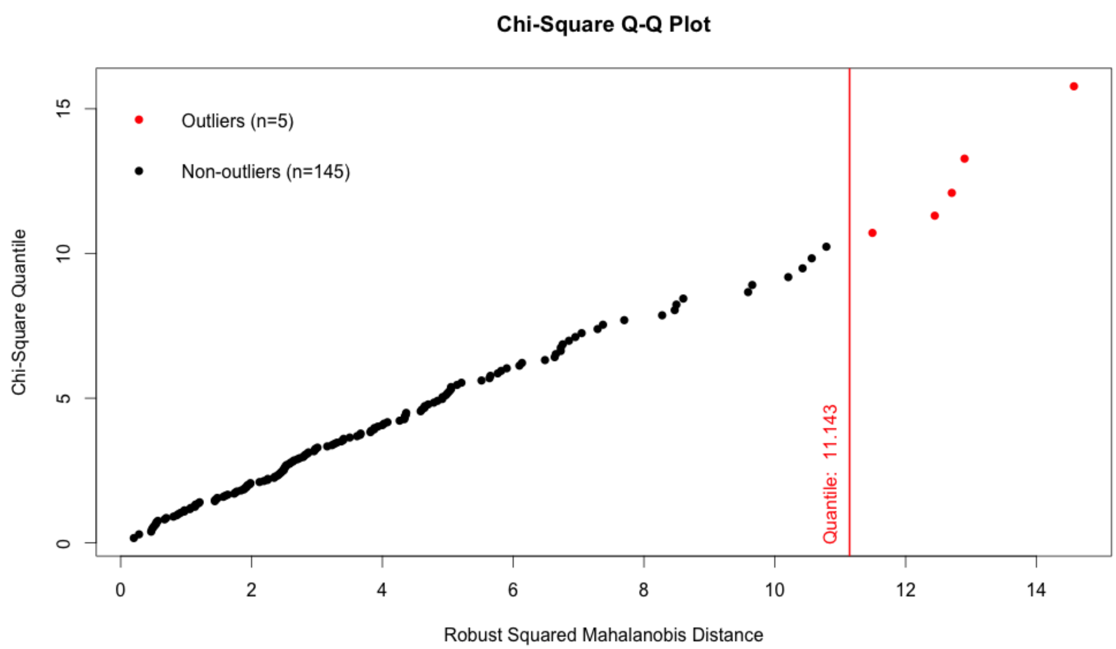
8 → Basialveolar length: direct distance from prosthion (pr) to alveolon (alv). Instrument: spreading or sliding caliper. Comment: Sliding caliper applicable only if incisor teeth have been lost. Position skull with basilar portion facing up.



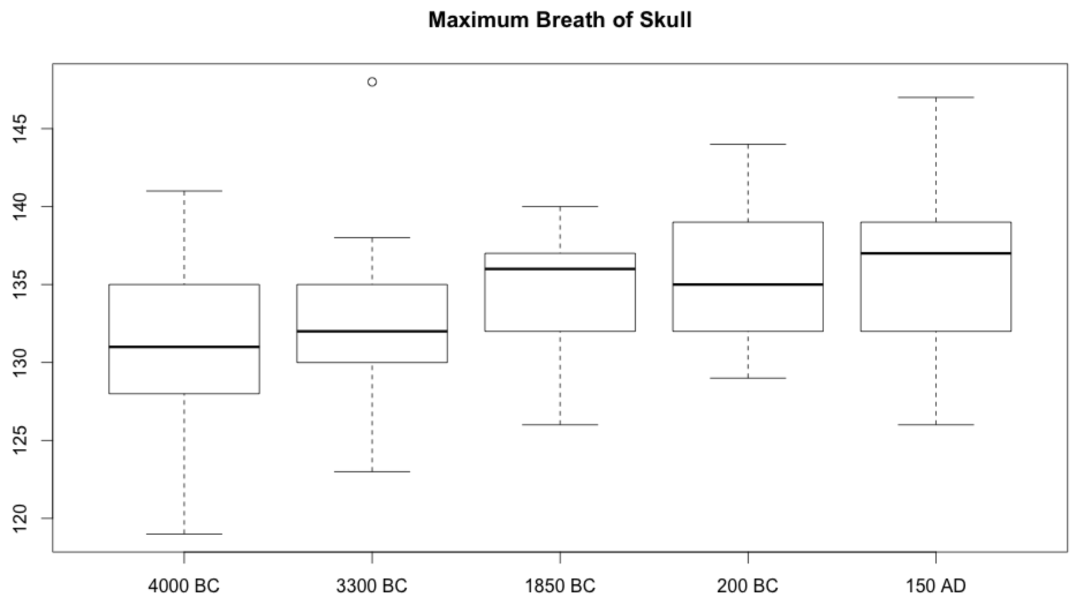


4 → Basibregmatic height of skull: direct distance from the lowest point on the anterior margin of foramen magnum (ba), to bregma (b)

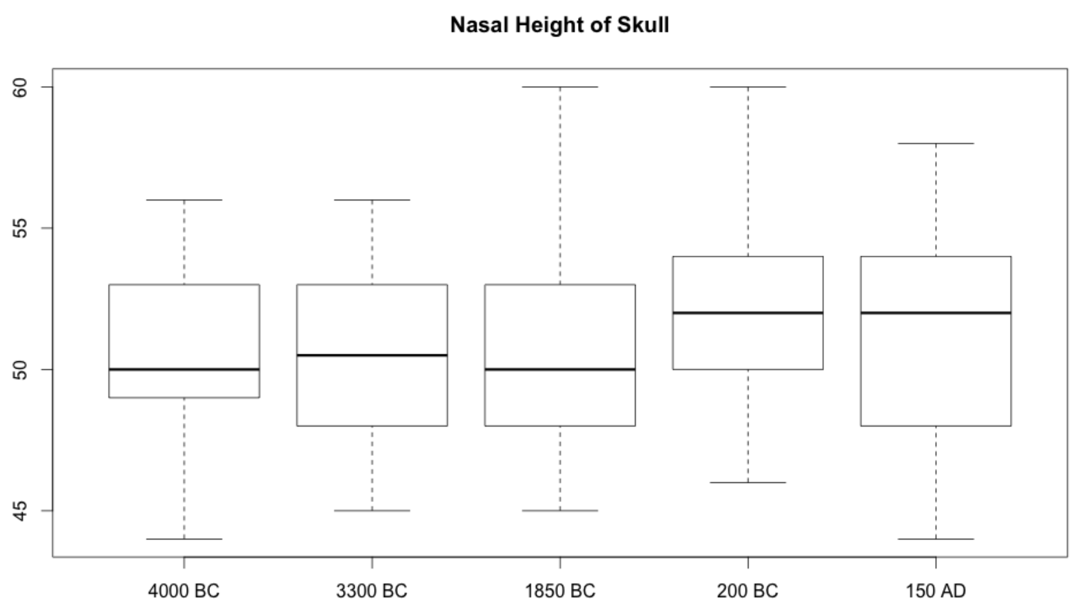
Section 2: Outlier Plots



Section 3: Boxplots

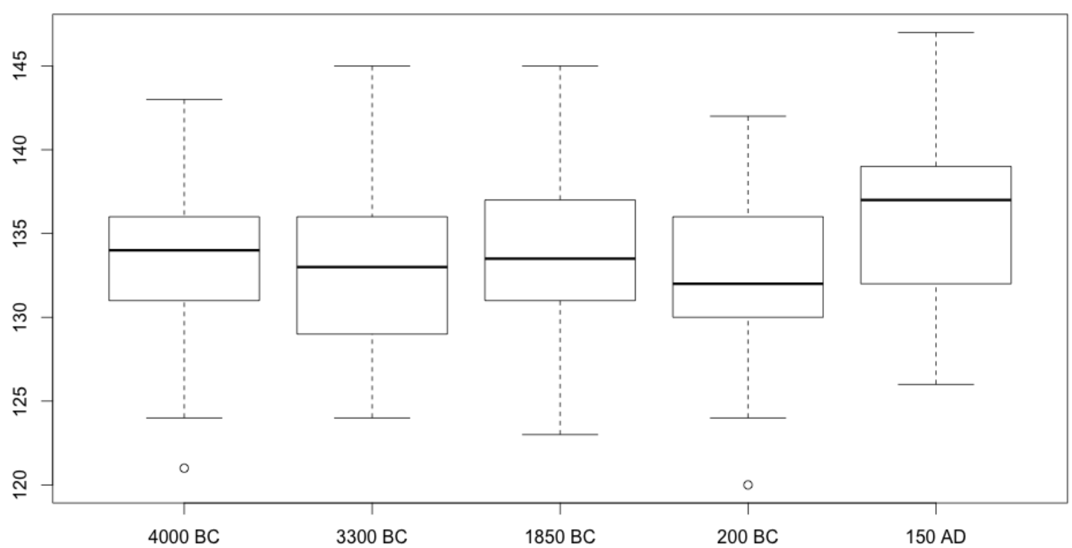


Maximum Breath of Skull
120
125
130
135
140
145
4000 BC
3300 BC
1850 BC
200 BC
150 AD

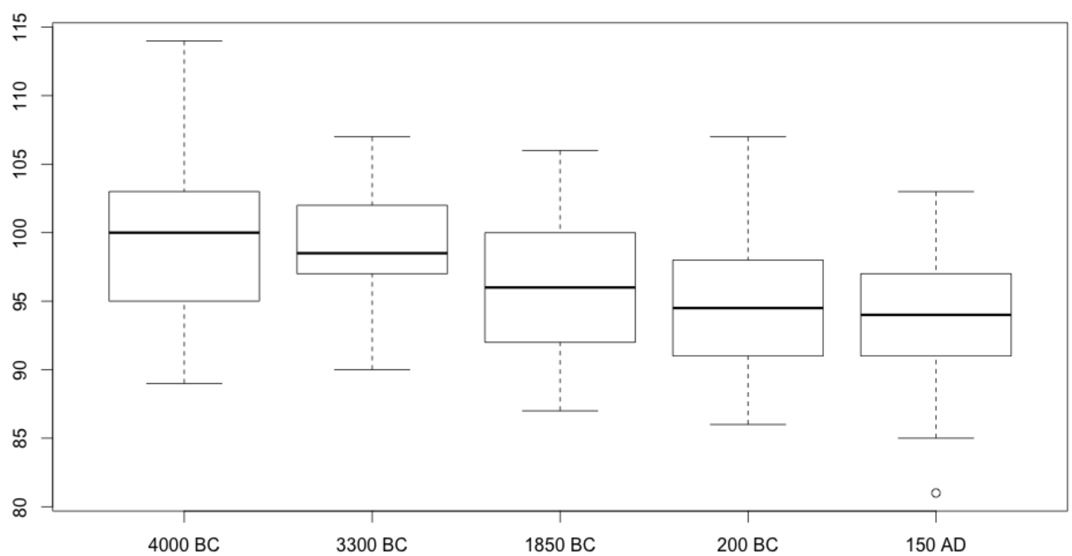


Nasal Height of Skull
45
50
55
60
4000 BC
3300 BC
1850 BC
200 BC
150 AD

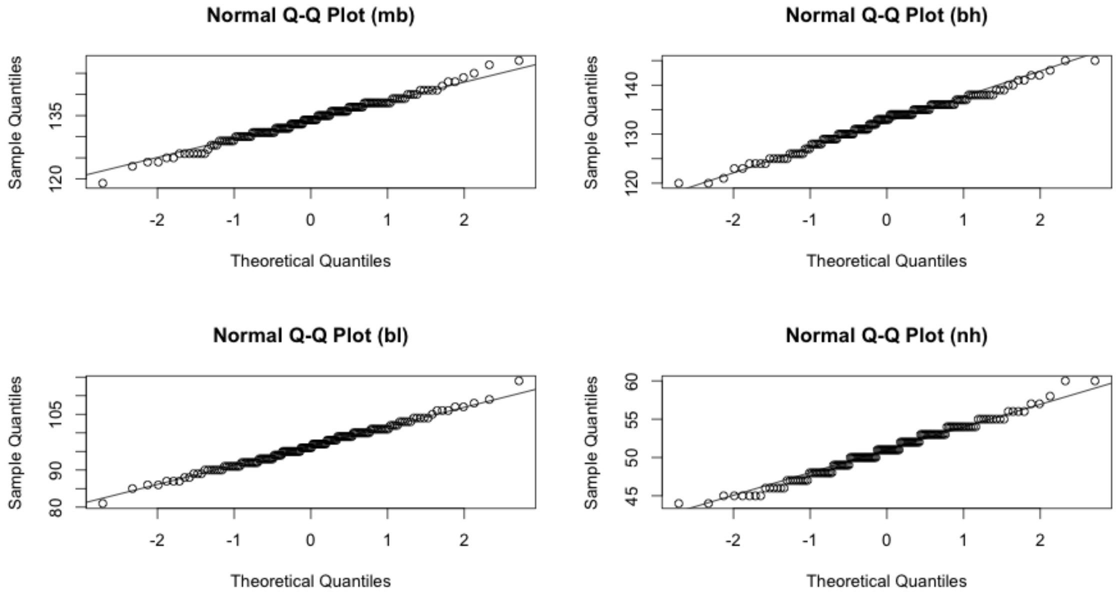
Basibregmatic Height of Skull



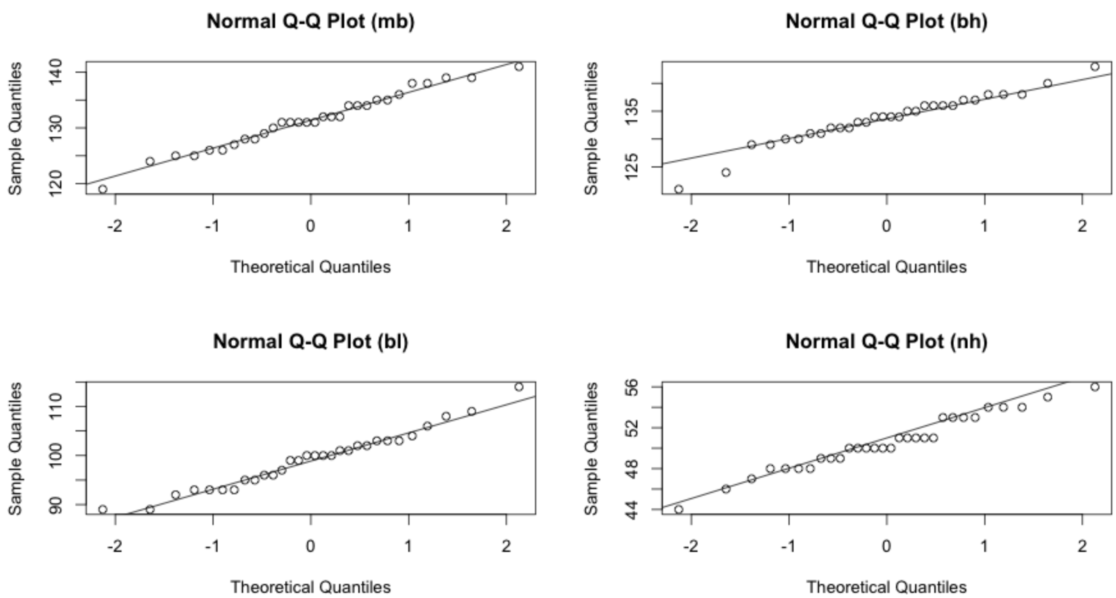
Basivaleor Length of Skull



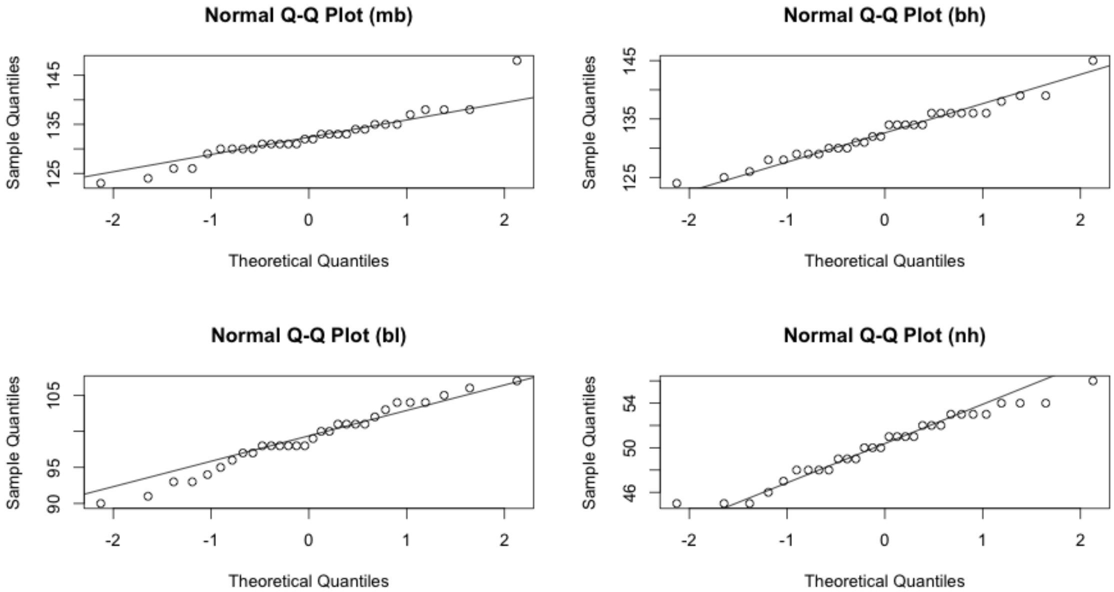
Section 4: QQ Plot of Data



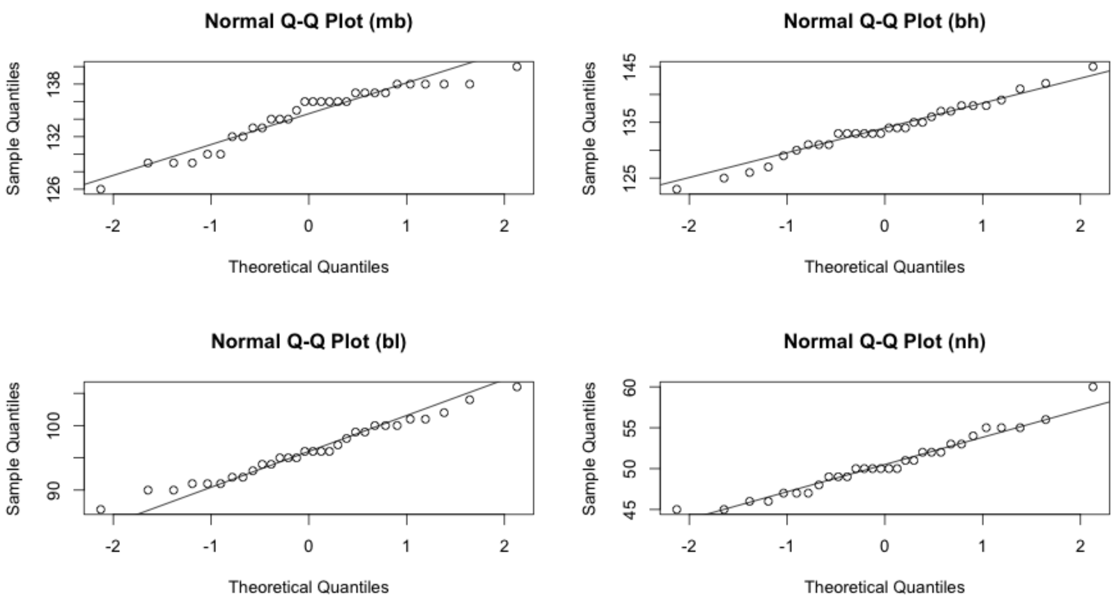
Section 5: QQ Plot of Data of Time Period 1



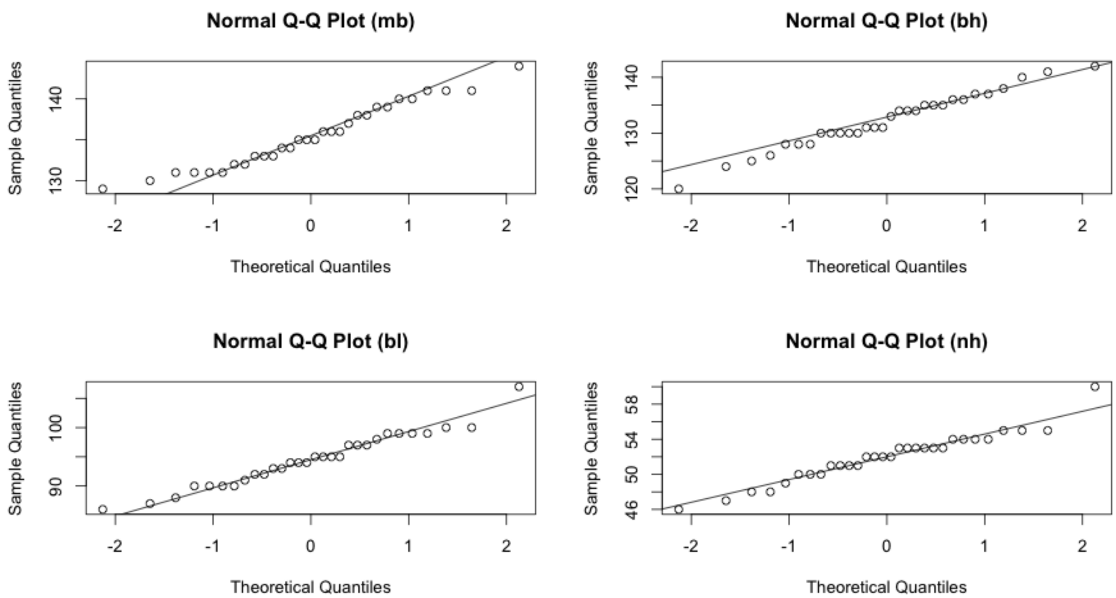
Section 6: QQ Plot of Data of Time Period 2



Section 7: QQ Plot of Data of Time Period 3



Section 8: QQ Plot of Data of Time Period 4



Section 9: QQ Plot of Data of Time Period 5