

Homework 1 solution

Exercise 1

- a) The following tables provide basic descriptive statistics for maximum **breadth (mb)**.

For **maximum breadth (mb)**, we see that the mean and median values are about 134mm. The skewness is -.0287 indicating there is no noticeable skew to the maximum breadth values, so the distribution is fairly symmetric. The standard deviation is 4.89, so if the data is roughly normal, we would expect about 95% of the values to be between 124mm and 144mm. The entire span from largest to smallest value is 29 mm.

Variable: mb

Moments			
N	150	Sum Weights	150
Mean	133.973333	Sum Observations	20096
Std Deviation	4.89067963	Variance	23.9187472
Skewness	-0.0286604	Kurtosis	0.30950566
Uncorrected SS	2695892	Corrected SS	3563.89333
Coeff Variation	3.65048738	Std Error Mean	0.39932232

Basic Statistical Measures			
Location		Variability	
Mean	133.9733	Std Deviation	4.89068
Median	134.0000	Variance	23.91875
Mode	131.0000	Range	29.00000
		Interquartile Range	6.00000

Note: The mode displayed is the smallest of 2 modes with a count of 15.

- b) The next step is to analyze data by epoch group starting with **epoch 1**.

Maximum breadths (mb) in **epoch 1** have a mean of 131.4 and median of 131. The standard deviation is 5.13 and the skewness is -.18. The range of the values is 22. While the magnitude of the skewness is slightly larger and still negative, it is relatively small, so the distribution is still fairly symmetric. The values are pretty similar to those in the overall sample, though the mean and median are slightly smaller and spread (standard deviation) is slightly larger in **epoch 1**.

Variable: mb
epoch=1

Moments			
N	30	Sum Weights	30
Mean	131.366667	Sum Observations	3941
Std Deviation	5.12924901	Variance	26.3091954
Skewness	-0.1844581	Kurtosis	-0.1405281
Uncorrected SS	518479	Corrected SS	762.966667
Coeff Variation	3.90452855	Std Error Mean	0.93646846

Basic Statistical Measures			
Location		Variability	
Mean	131.3667	Std Deviation	5.12925
Median	131.0000	Variance	26.30920
Mode	131.0000	Range	22.00000
		Interquartile Range	7.00000

For **epoch 3**, the mean is around 134.5mm and the median is 136mm for **maximum breadths (mb)**, with a standard deviation of 3.48 and a large negative skew of -0.73 which indicates that the distribution shows a long tail on the left. It will be more reasonable to use robust statistics like median and interquartile range than mean to describe the central tendency of the data. The **maximal breadths (mb)** in **epoch 3** are slightly larger than those in **epoch 1** with smaller range and smaller spread. A more significant negative skewness is observed in **epoch 3** compared with **epoch 1**.

Variable: mb
epoch=3

Moments			
N	30	Sum Weights	30
Mean	134.466667	Sum Observations	4034
Std Deviation	3.481313	Variance	12.1195402
Skewness	-0.7312664	Kurtosis	-0.2725472
Uncorrected SS	542790	Corrected SS	351.466667
Coeff Variation	2.58897844	Std Error Mean	0.63559789

Basic Statistical Measures			
Location		Variability	
Mean	134.4667	Std Deviation	3.48131
Median	136.0000	Variance	12.11954
Mode	136.0000	Range	14.00000
		Interquartile Range	5.00000

Turning now to epoch 5 for the **maximal breadths (mb)**, we see that the mean of 136.2mm and the median of 137mm are slightly larger than in **epoch 1** and **epoch3**. The standard deviation is larger with a value of 5.35, and the skewness is negative though small with a value of -.13 and indicates no noteworthy asymmetry to the data.

Overall, **maximal breadths (mb)** in later epoch are larger with similar spread and range. Skewness are negative for all three epochs, though only of concern in epoch 3.

Variable: mb
epoch=5

Moments			
N	30	Sum Weights	30
Mean	136.166667	Sum Observations	4085
Std Deviation	5.35036791	Variance	28.6264368
Skewness	-0.1307951	Kurtosis	-0.3533039
Uncorrected SS	557071	Corrected SS	830.166667
Coeff Variation	3.92927876	Std Error Mean	0.97683907

Basic Statistical Measures			
Location		Variability	
Mean	136.1667	Std Deviation	5.35037
Median	137.0000	Variance	28.62644
Mode	137.0000	Range	21.00000
		Interquartile Range	7.00000

Exercise 2

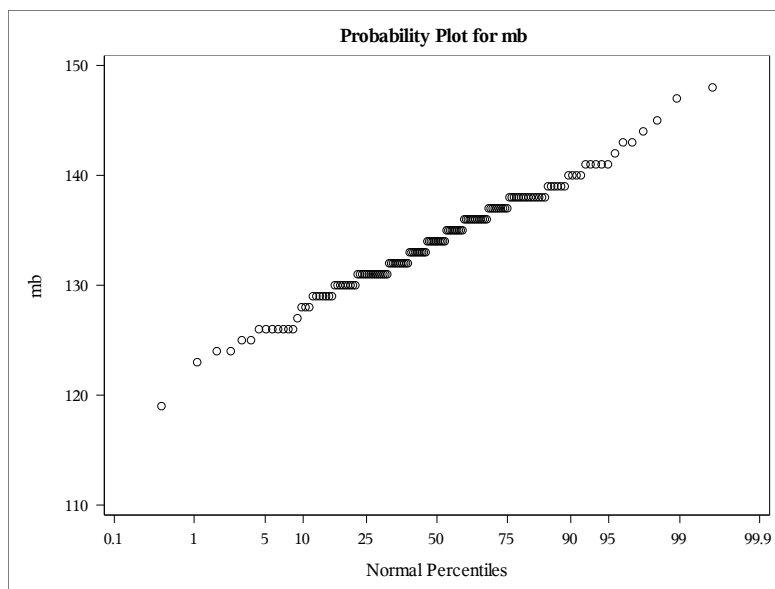
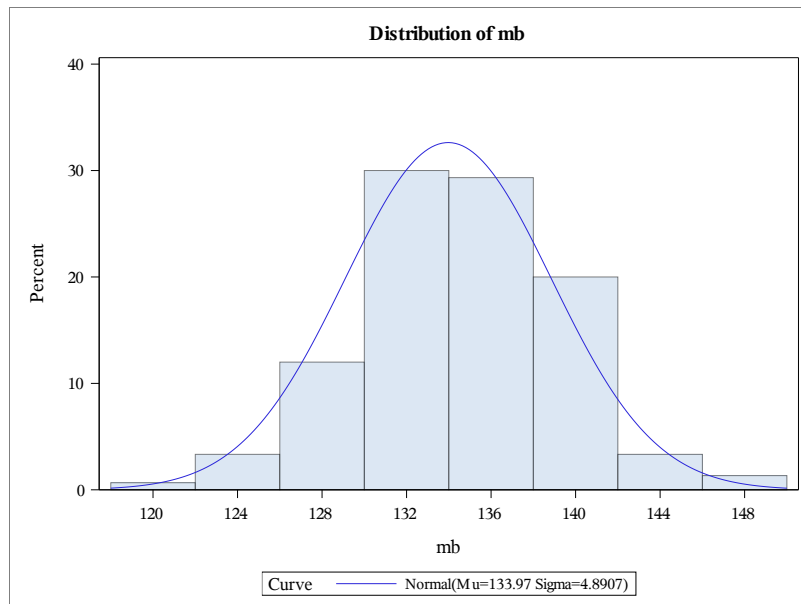
- a) To visually and quantitatively check an assumption of normality for the **maximum breadth (mb)**, we can look at a histogram, probability plot and tests for normality. The histogram looks very bell-shaped. The probability plot looks pretty straight. Visually, there are no concerns about an assumption of normality.

Looking at the normality tests, the Shapiro-Wilk test which is designed for normal distribution fails to reject the assumption of normality at a .05 level. Cramer-von Mises and Anderson-Darling have p-values of 0.18 and 0.19, indicating that they would not reject normality at a .05 level and Kolmogorov-Smirnov test has p-value greater than 0.15 which shows that we do not

have strong evidence to reject normality. We can conclude that the **maximum breadth (mb)** is reasonably close to normally distributed. We can use t tests which assume normally distributed data.

Variable: mb

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.991334	Pr < W	0.4925
Kolmogorov-Smirnov	D	0.060707	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.085139	Pr > W-Sq	0.1828
Anderson-Darling	A-Sq	0.524942	Pr > A-Sq	0.1867

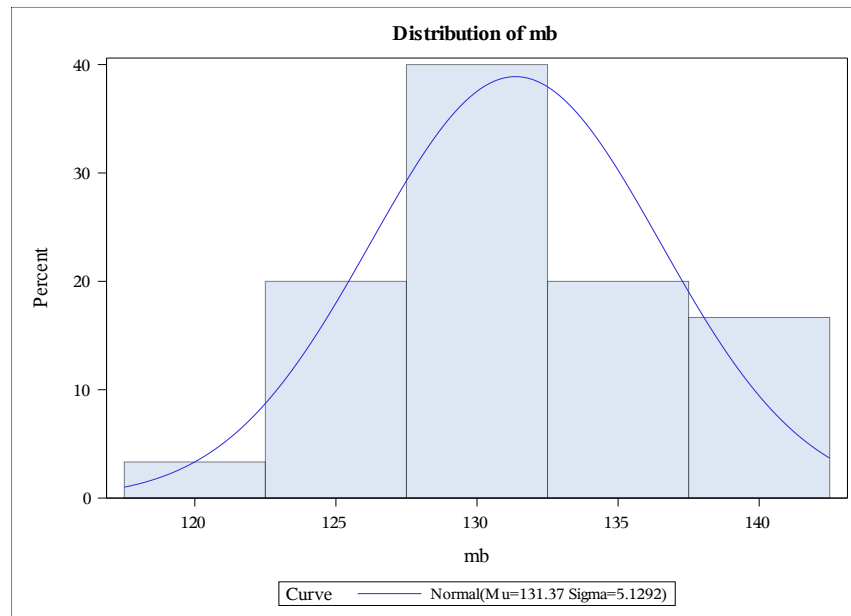


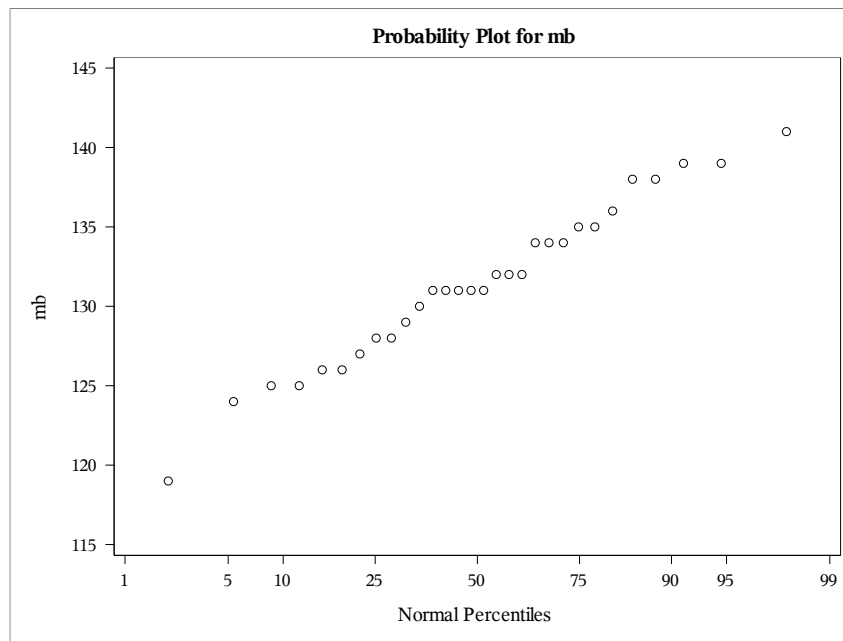
- b) Next, we want to analyze **epoch 1** for the **maximum breadth (mb)**. The histogram has a slightly higher right tail with a straight probability plot. All four tests have p-values greater than 0.05 which indicates that we do not have strong evidence to reject the normality assumption. We can use t test to test the mean in later analysis.

Variable: mb
epoch=1

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.981356	Pr < W	0.8603
Kolmogorov-Smirnov	D	0.104839	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.034956	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.222261	Pr > A-Sq	>0.2500

epoch=1



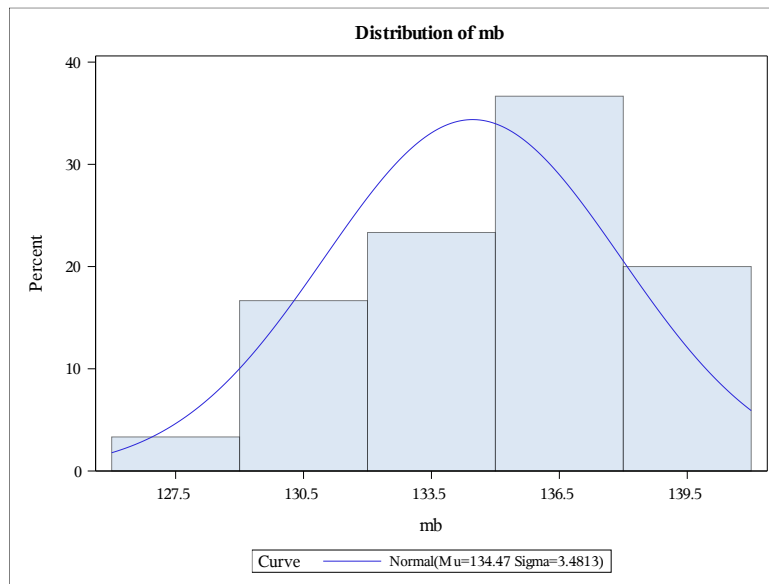


Next, we want to analyze **epoch 3** for the **maximum breadth (mb)**. The histogram has a tall right tail with a spike around 136mm. The corresponding probability plot is not straight. All four tests have p-values less than 0.05 which indicates that we reject the normality assumption at .05 level. We should choose nonparametric test for later analysis for **epoch 3**.

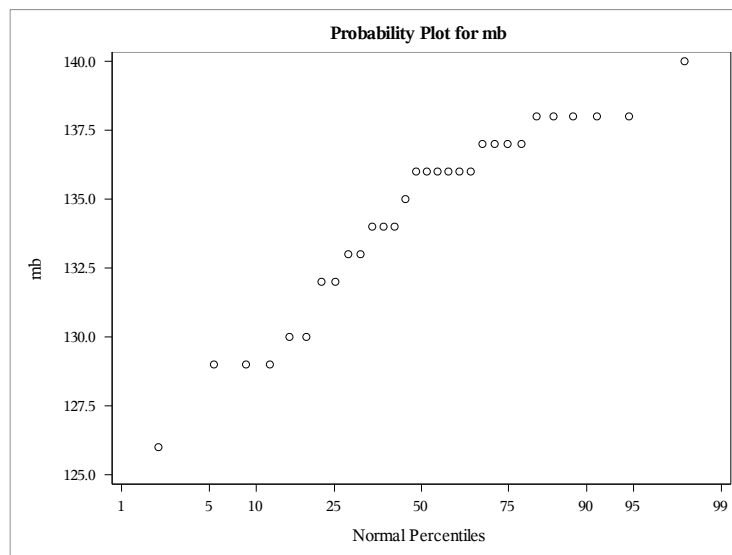
Variable: mb
epoch=3

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.922623	Pr < W	0.0314
Kolmogorov-Smirnov	D	0.203527	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.160702	Pr > W-Sq	0.0173
Anderson-Darling	A-Sq	0.940053	Pr > A-Sq	0.0164

epoch=3



epoch=3



Exercise 3

- a) Because in exercise 2 we showed that the overall **maximum breadth (mb)** is reasonably close to a normal distribution, we can use t test for testing whether the mean **maximum breadth (mb)** for the population this sample came from is significantly less than 136mm. The corresponding p-value for the t test is much less than the significance level .05 and we should reject the null hypothesis and conclude that the mean **maximum breadth (mb)** is significantly less than 136mm.

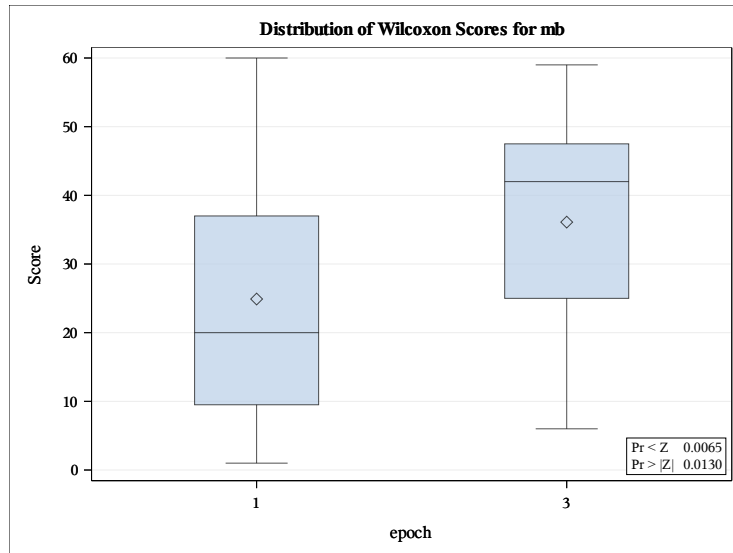
Variable: mb

DF	t Value	Pr < t
149	-5.08	<.0001

- b) The **maximum breadth (mb)** from **epoch 3** did not follow a normal distribution, so we should use a nonparametric test for significant difference between the distributions' values. Specifically, we want to use Wilcoxon two-sample test to see whether **epoch 1** had significantly greater values than **epoch 3**. The one-sided p-value in the table is for the other alternative, so we will need to use 1 minus that p-value for the question of interest. The p-value for the alternative that epoch 1 had greater values in general is about $1 - .016 = 0.984$. This is much greater than .05, so we cannot conclude that epoch 1 tended to have larger values. In fact, the one-sided test suggests that **epoch 1** actually tended to have significantly smaller maximum breadths (**mb**) than **epoch 3**.

Wilcoxon Scores (Rank Sums) for Variable mb Classified by Variable epoch					
epoch	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	30	747.0	915.0	67.428870	24.90
3	30	1083.0	915.0	67.428870	36.10
Average scores were used for ties.					

Wilcoxon Two-Sample Test					
Statistic	Z	Pr < Z	Pr > Z	t Approximation	
				Pr < Z	Pr > Z
747.0000	-2.4841	0.0065	0.0130	0.0079	0.0158
Z includes a continuity correction of 0.5.					



Exercise 4

- a) There are no extreme values in the variables we have looked at, so Pearson correlation will be fine here.

Pearson Correlation Coefficients, N = 150 Prob > r under H0: Rho=0				
	mb	bh	bl	nh
mb	1.00000	-0.06190 0.4517	-0.15697 0.0551	0.18255 0.0254
bh	-0.06190 0.4517	1.00000	0.26435 0.0011	0.14675 0.0731
bl	-0.15697 0.0551	0.26435 0.0011	1.00000	-0.00638 0.9382
nh	0.18255 0.0254	0.14675 0.0731	-0.00638 0.9382	1.00000

Looking at the correlation for the four skull measurements, there are two statistically significant correlations: the correlation between **basialveolar length** and **basibregmatic height**, and the correlation between **maximum breadth** and **nasal height**. In both cases, the correlations are positive and relatively small with values of .264 and .183, respectively. While statistically significant, there may be some question as to whether these correlations are practically significant. We can conclude that there is a slight tendency of skulls with larger **basialveolar length** to have larger **basibregmatic height** and for skulls with larger **maximum breadth** to have larger **nasal height**.

b) The following shows the correlation analysis by epoch for **epoch 1**.

There is only one statistically significant correlation: the correlation between **maximum breadth** and **nasal height** which is positive. From the magnitude of 0.51, we can conclude that there is a moderate tendency of skulls with larger **maximum breadth** to have larger **nasal height** for **epoch 1**.

epoch=1

Pearson Correlation Coefficients, N = 30 Prob > r under H0: Rho=0				
	mb	bh	bl	nh
mb	1.00000 0.00000	0.18112 0.3382	0.01504 0.9371	0.51120 0.0039
bh	0.18112 0.3382	1.00000	-0.03016 0.8743	0.03183 0.8674
bl	0.01504 0.9371	-0.03016 0.8743	1.00000	-0.11804 0.5344
nh	0.51120 0.0039	0.03183 0.8674	-0.11804 0.5344	1.00000

The correlation analysis by epoch for **epoch 5** follows.

There is only one statistically significant correlation: the correlation between **basialveolar length** and **basibregmatic height** which is 0.47. We can conclude that there is a moderate tendency of skulls with larger **basialveolar length** to have larger **basibregmatic height** for **epoch 5**. Epoch 1 and epoch 5 do not share any similar significant correlations.

epoch=5

Pearson Correlation Coefficients, N = 30 Prob > r under H0: Rho=0				
	mb	bh	bl	nh
mb	1.00000 0.9638	-0.00864 0.9638	-0.06946 0.7153	-0.10024 0.5982
bh	-0.00864 0.9638	1.00000	0.46641 0.0094	0.11628 0.5406
bl	-0.06946 0.7153	0.46641 0.0094	1.00000	0.02109 0.9119
nh	-0.10024 0.5982	0.11628 0.5406	0.02109 0.9119	1.00000

Comparing with the overall sample, we can see that the **maximum breadth** and **nasal height** relationship is stronger for epoch 1 and the **basialveolar length** and **basibregmatic height** relationship is strong for epoch 5 than in the broader sample.