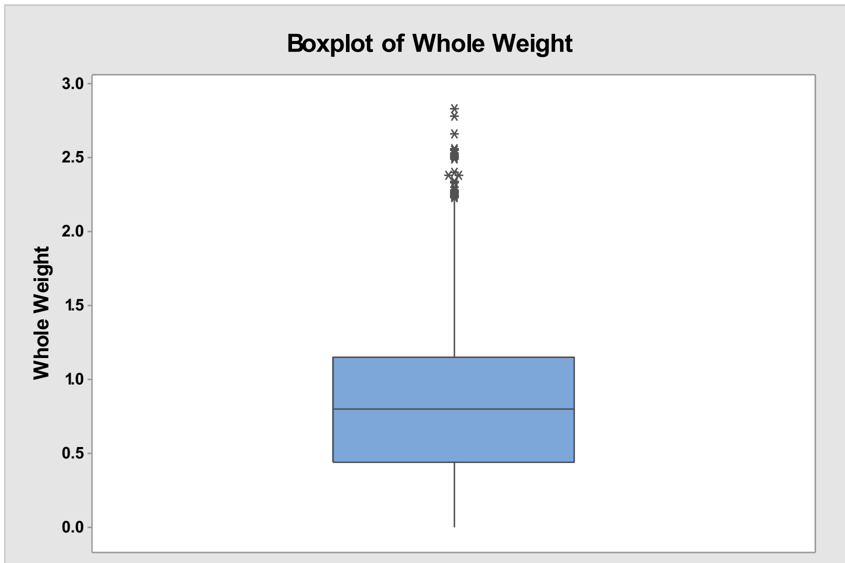
Caleb McWhorter

Minitab HW #1



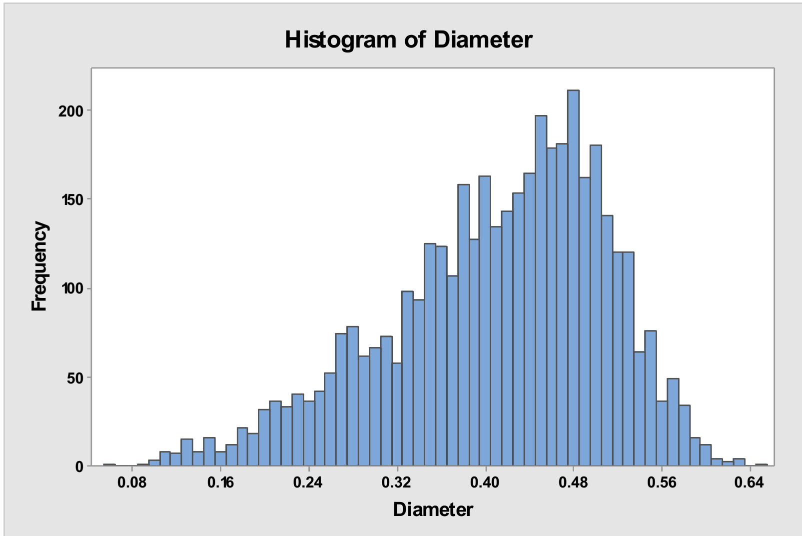
1. The boxplot of 'whole weight' is given above. Clearly, looking at the distribution shows that it is right skewed. The distribution as whole is not normally distributed. The outliers are marked by stars on the boxplot. While there are many outliers on the right, there are no outliers on the left.

Descriptive Statistics: Whole Weight

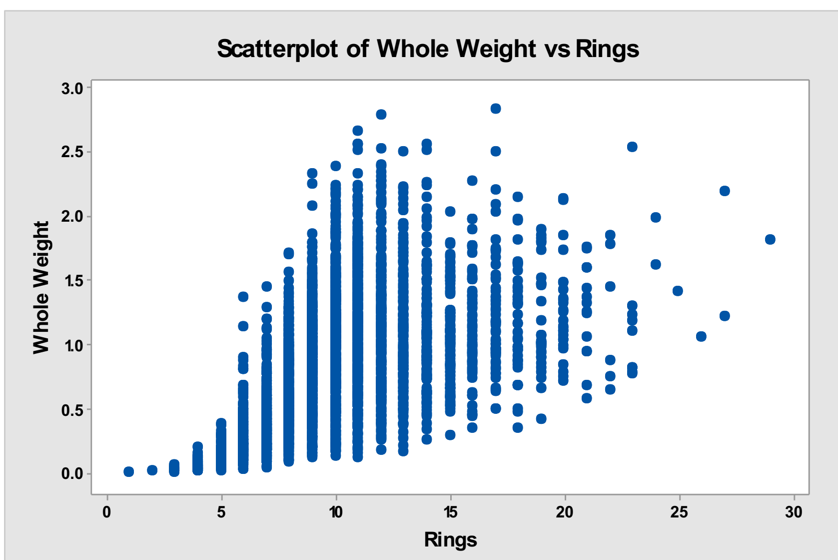
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Whole Weight	4177	0	0.82874	0.00759	0.49039	0.00200	0.44150	0.79950	1.15350

Variable	Maximum	Range	IQR
Whole Weight	2.82550	2.82350	0.71200

2. Above are the basic statistics for the variable 'whole weight'. The 5-number summary (min, Q1, Median, Q3, Max) are clearly visible, as are the mean and standard deviation. The N indicates that 4,177 abalone were examined, and N* indicates none had missing values. Also shown are the range for this variable and the IQR, added by using the 'Statistics' option.



3. Above is the histogram for the variable 'diameter'. The variable does seem 'a bit' normally distributed, albeit clearly left skewed. However, notice the sharp peaks along the distribution. It may have the correct 'shape' (with a skew) but I would not count tests requiring normality to work well until the number in the sample size is larger.

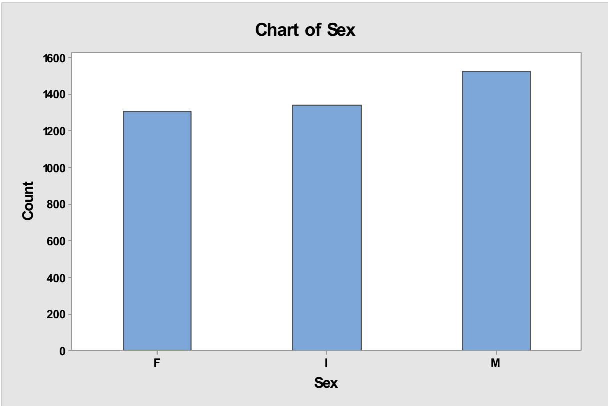


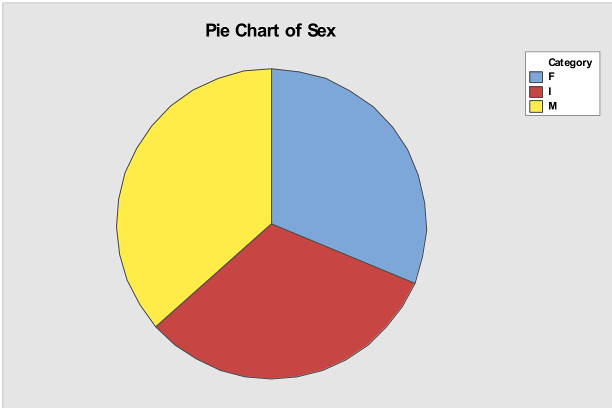
4. The scatterplot is given above. Clearly, the greater of number of rings, the greater the weight. This should make sense if rings can be used to count the age of the abalone. However, there does not seem to be a good linear fit. So while there is a relationship between the two, it is probably not a linear relationship.

Regression Analysis: Whole Weight versus Rings

```
The regression equation is
Whole Weight = 0.01227 + 0.08219 Rings


S = 0.412670   R-Sq = 29.2%   R-Sq(adj) = 29.2%


Analysis of Variance

Source        DF       SS       MS        F      P
Regression     1   293.26  293.262  1722.07  0.000
Error       4175   710.99    0.170
Total       4176  1004.25
```

5. We see the regression for the two variables above. The R^2 value is only 29.2%, meaning the data is only '29.2% linear', which confirms our discussion from the previous part – that the data is not very linear. However, if one were to fit a line to the scatterplot, the line fitting the data best would be Whole Weight = 0.01227 + 0.08219 Rings, i.e. y= 0.01227 + 0.08219 x.





6. Above are a bar chart and pie graph of the variable 'sex'. Notice from each, we can see that approximately the same number of female, male, and infant abalone were examined (in the bar chart because they are all roughly the same height and in the pie chart because they all roughly have the same area). Using Stat >> Tables >> Tally Individual Variables, we find there were 4,177 total abalones: 1, 307 females, 1,342 infants, and 1,528 males.

One-Sample Z: Height

Test of $\mu = 0.1385$ vs $\neq 0.1385$
The assumed standard deviation = 0.04183

Variable	N	Mean	StDev	SE Mean	99% CI	Z	P
Height	4177	0.139516	0.041827	0.000647	(0.137849, 0.141184)	1.57	0.116

One-Sample Z: Height

Test of $\mu = 0.1385$ vs > 0.1385
The assumed standard deviation = 0.04183

Variable	N	Mean	StDev	SE Mean	99% Lower Bound	Z	P
Height	4177	0.139516	0.041827	0.000647	0.138011	1.57	0.058

7. The first table above (using a two-sided) was used to construct a 99% confidence interval for the variable 'height'. So we are 99% certain that the true average height of the abalone is between 0.1378mm to 0.1412mm. The second 'One-Sample Z: Height' was used to test the hypothesis using $\alpha = 0.01$. Notice that the test statistic is 1.57 and we have a p-value of $p = 0.058$. Since p is not less than α, we fail to reject the null hypothesis that $\mu = 0.1385$.

Tally for Discrete Variables: Rings

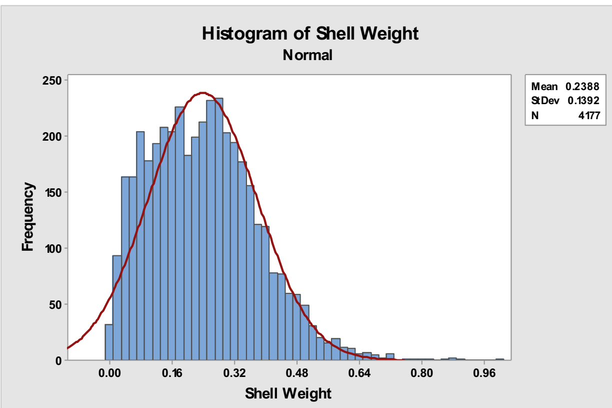
Rings	Count
1	1
2	1
3	15
4	57
5	115
6	259
7	391
8	568
9	689
10	634
11	487
12	267
13	203
14	126
15	103
16	67
17	58
18	42
19	32
20	26
21	14
22	6
23	9
24	2
25	1
26	1

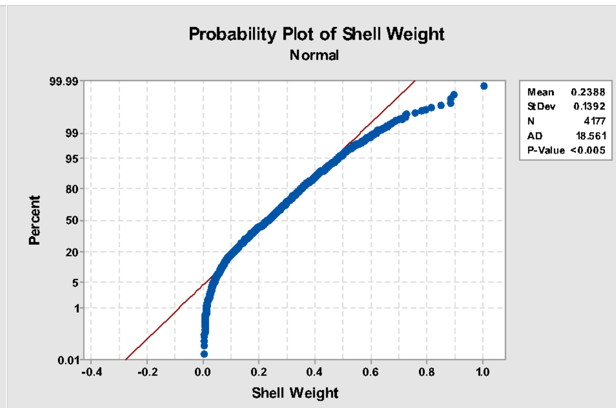
```
27        2
29        1
N=     4177
```

Test and CI for One Proportion

```
Sample     X      N  Sample p          99% CI
1        267   4177  0.063921  (0.054555, 0.074297)
```

8. A tally for the variable ‘rings’ is given above. Notice that there were 267 abalone with 12 rings. It also tells us there were 4,177 abalone. Using 267 as the number of events and 4,177 as the number of trials, we obtain a 99% confidence interval of (0.0546, 0.0742), i.e. we are 99% certain that the proportion of abalone possessing 12 rings is between 5.46% and 7.42%. Since 4% is not in this interval, it does not seem likely that 4% of abalone have 12 rings.





9. The histogram and normality plot for ‘shell weight’ is given above. In both cases, the data has the right ‘shape’ but clearly is not normal – the peaks on the histogram are off the curve and it is missing portions on the left side, while on the normality plot there are clearly issues at both ends of the distribution. Of course, the Central Limit Theorem (CLT) states that if we take large enough samples from this data, the sampling distribution will be normally distributed.