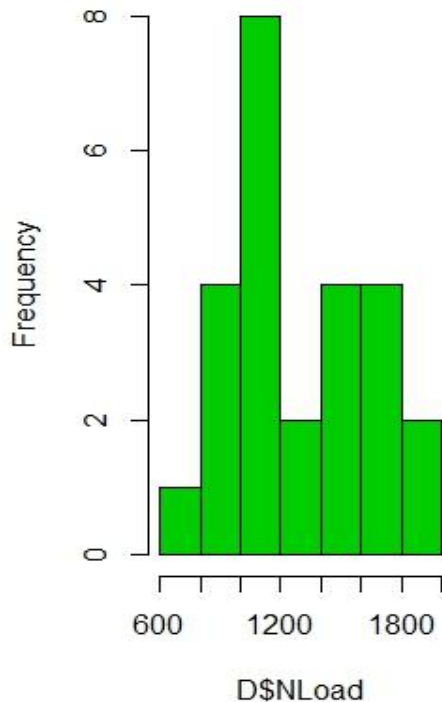Ryan Kladar (s141094)
Course 02402 – Introductory Statistics
Project 1 – Skive Fjord Nitrogen Load Analysis
20 October 2014

1. Descriptive Analysis

### Histogram of D$NLoad



a) The density of the nitrogen loads is shown to the left using a histogram. We can see that while the plot is somewhat symmetrical, the density skews to the left, indicating that the density is positive. The frequency of loads (mode) below the mean (1361.9), is greater than the frequency of loads above the mean. This can be also illustrated with the default density plot function in R of a data set (below).

### density.default(x = D$NLoac
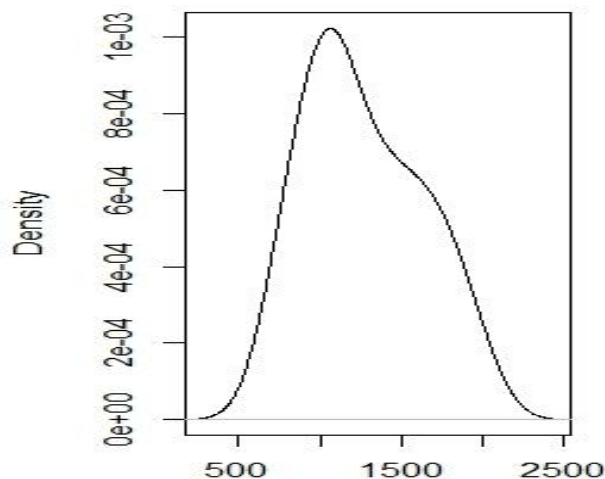


The density of nitrogen loads shows the positive skew of the data. The peak is to the left of the mean. The data is not quite normally distributed, but shows a correlation with the normal distribution.
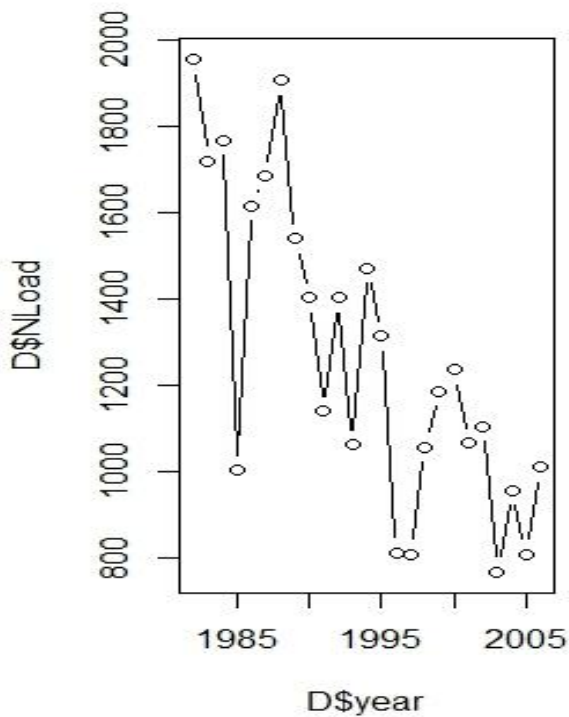
b) Shown are the time series plots for the nitrogen loads vs. the year they were measured. First, the basic time series (left), and second, the time series segmented by VMP period (bottom left). It is clear that nitrogen loads into the Skive fjord has been steadily decreasing over the 25 year time period. Though 1985 saw a relatively low load for that era (average greater than 1500), the remaining series decreases rather predictably.
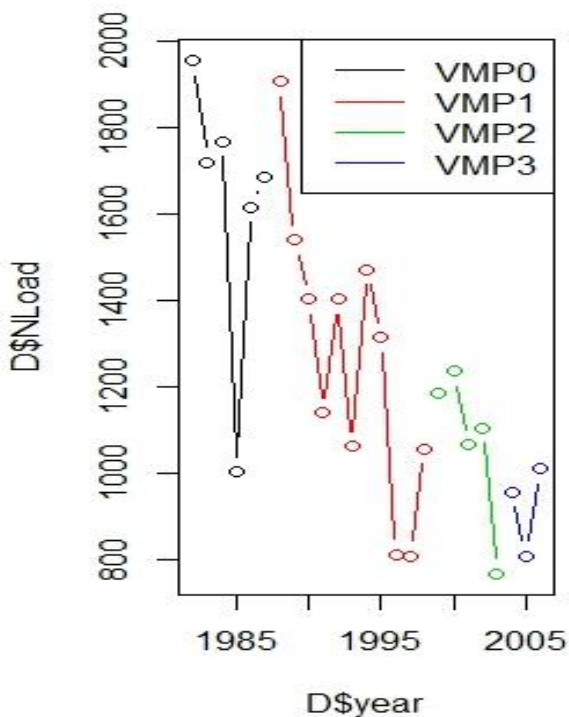


In the time series segmented by VMP period, the results must be looked at carefully. The average of each period appears to be less than the period prior. Each period is also shorter than the period prior, if VMP0 is taken to extend backward in time off the graph, before VMP was implemented. The nitrogen loads also never increase after a new VMP period is started. After a new VMP is adopted, nitrogen loads either remain static or decrease, which suggests emitters are following the rules of each new VMP. After the 3rd and final VMP, nitrogen loads remain quite low. This is some qualitative evidence of the success of the VMP programs.

**VMP = 0**

**VMP = 1**

**VMP = 2**

**VMP = 3**
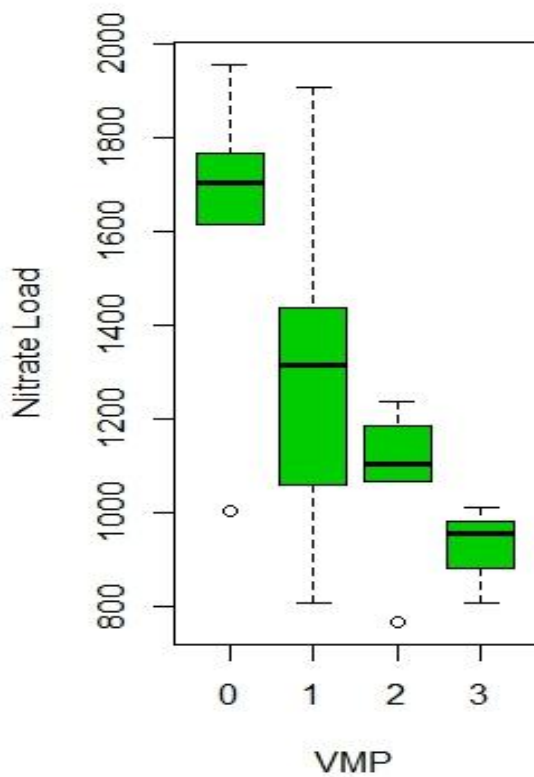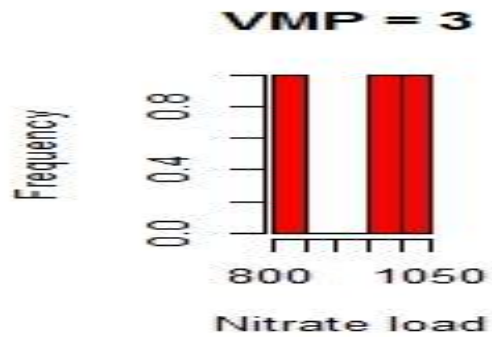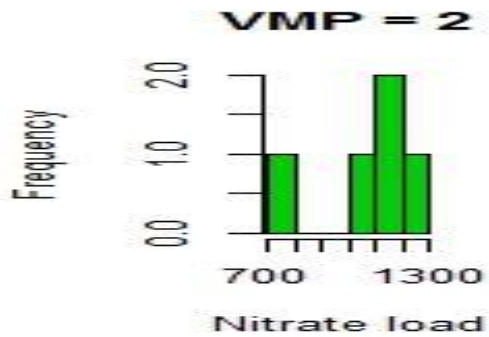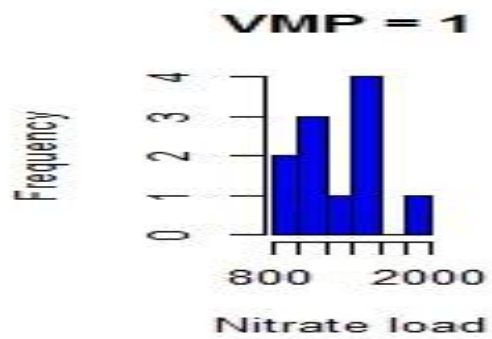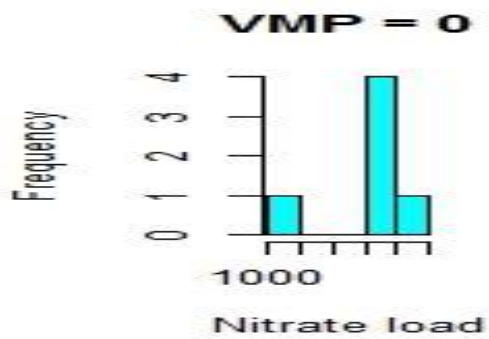


c) The empirical density of VMP0 is skewed to the right (negative). The empirical density of VMP1 is more evenly distributed, but is still skewed to the left (positive). The empirical density of VMP2 is skewed considerable to the right (negative) and the empirical density of VMP3 is not evenly distributed, skewed to the right (negative). Both the mean and median decrease for all periods, but the spread, or variances, decrease with time. The least varied periods appear to be VMP2 and VMP3, with the most varied periods as VMP0 and VMP1.

| d) | Number Of obs. $N$ | Sample Mean $(x)$ | Sample Variance $(s^2)$ | Standard Deviation $(s)$ | Lower Quartile $(Q_1)$ | Median $(Q_2)$ | Upper Quartile $(Q_3)$ |
|---|---|---|---|---|---|---|---|
| VMP0 1982-1987 | 6 | 1625 | 105522.2 | 324.84 | 1633 | 1702 | 1756 |
| VMP1 1988-1998 | 11 | 1265 | 107970.2 | 328.59 | 1060 | 1314 | 1437 |
| VMP2 1999-2003 | 5 | 1073 | 33428.26 | 182.83 | 1066 | 1106 | 1187 |
| VMP3 2004-2006 | 3 | 926.2 | 10828.9 | 104.1 | 883.6 | 956.9 | 984.1 |

e) In the table we can see the clear, regular decrease in the means of each period. We can also easily see the difference in periods one and two compared to periods three and four in terms of variance and standard deviation. Period 1 and 2 are characterized by large variances and deviations, meaning high unpredictability in year-to-year nitrate loads, not easily seen in the boxplot. One thing the VMP2 accomplished was the reduction in variability of nitrate loads, which, while not affecting the total load over the period considerably, does make year-to-year management of the ecosystem easier. The boxplot shows the decrease in variability, but only qualitatively. The table quantifies the reduction in variance. The prediction for next year's load can be much more precise. VMP3 also decreased average load and decreased variability even further, another important step to nutrient level regulation attainment.

2. Confidence Intervals and Hypothesis Tests

| f) | Lower Limit of CI | Upper Limit of CI |
|---|---|---|
| VMP0 1982-1987 | 1283.65 | 1965.45 |
| VMP1 1988-1998 | 1044.7 | 1486.2 |
| VMP2 1999-2003 | 845.88 | 1299.91 |
| VMP3 2004-2006 | 667.65 | 1184.67 |

$$\bar{x} \pm t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

This equation gives a confidence interval around a sample mean. For the VMP0:

$$1625 \pm t_{0.975} \cdot \frac{324.84}{\sqrt{6}} = [1238.65, 1965.45]$$

g) All of the confidence bands overlap with the preceding and succeeding periods. In fact, the only two intervals that do not overlap are the VMP0 and VMP3 periods. What is surprising is that the band widths do not change drastically. While the confidence band varies by about 700 for VMP0, the other three periods' intervals are all about 400. The interval does not decrease very drastically with the mean. While the mean drops, the intervals stay static.

h) For a null- and alternative hypothesis impact on the Nitrate load into Skive fjord:

$$H_0 : \mu_{vmp0} = \mu_{vmp3}$$
$$H_1 : \mu_{vmp0} \neq \mu_{vmp3}$$

That is to say that up the mean of VMP0 is not statistically significantly different than the mean of VMP3, we will accept the null hypothesis, and the VMPs will have not had an effect. If the mean of VMP0 is significantly different than the mean of VMP3, based on an arbitrary significance factor α of our choosing (typical is 0.05, or 5% chance that the differences between the two means is due to random variation). This type of hypothesis testing is known as "non-directional" testing. This means that we do not care if the difference between the means is positive or negative, just whether or not it exists. In this case it may also be useful to perform directional hypothesis testing, specifically negative direction testing, to see if the VMPs have lowered the nitrate load. It would be very problematic if the regulation actually *increased* the nitrate load.

The formula for calculating the test statistic needed in this hypothesis test is as follows:

$$t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Where $\mu = \mu_0$ under the null hypothesis, $H_0$.

To use this observed t value to find statistical significance, we must find a P-value to compare to our significance factor α:

$$P_{value} = 2 \cdot P(T > |t_{obs}|)$$

Where T follows a t-distribution of (n-1) degrees of freedom.

In R, the function t.test will give a P value for a given data set. For the hypothesis stated above, the P value is 0.003282, or 0.3%. It is extremely unlikely that the change in mean from VMP0 to VMP3 is due to random variation in the data, and we can confidently assume, with the standard significance threshold α = 5%, the alternative hypothesis. The VMPs are effective in changing the mean nitrate load into the fjord.

If we perform directional hypothesis testing, that is determining the direction of significant change, we can use this hypothesis:

$$H_0 : \mu_{vmp0} \leq \mu_{vmp3}$$
$$H_1 : \mu_{vmp0} > \mu_{vmp3}$$

This hypothesis is used because we are interested in determining if the VMP decreased the nitrate load, which is their goal. If the VMPs increased the nitrate load, or the load remained the same, we accept the null hypothesis. It is determined that the change is indeed negative, and the likelihood of change being due to random chance is very small. The VMPs are effective in decreasing nitrate load.

i) While the overall change in nitrate load is a decrease likely due to the VMPs, we haven't determined that any one VMP is responsible, or if any were counterproductive to our goal of decreasing nitrate loads. We establish these three hypothesis tests:

1) $H_0 : \mu_{vmp0} = \mu_{vmp1}$    $H_1 : \mu_{vmp0} \neq \mu_{vmp1}$

2) $H_0 : \mu_{vmp1} = \mu_{vmp2}$    $H_1 : \mu_{vmp1} \neq \mu_{vmp2}$

3) $H_0 : \mu_{vmp2} = \mu_{vmp3}$    $H_1 : \mu_{vmp2} \neq \mu_{vmp3}$

We test each, using R's t.test function.

1) P-value = 0.00465
2) P-value = 0.0781
3) P-value = 0.1346

Using a significance threshold of 0.05 (5%), we see that while VMP1 had a statistically significant decrease on the nitrate loads over VMP0, the others, VMP2 over VMP1 and VMP3 over VMP2, did not meet our criteria for statistical significance. They do give weak evidence against the null hypothesis, but could also be due to random variation in the data. When all VMPs are combined, we can reject the null hypothesis, but, for certain individual VMPs, we cannot.

Conclusion)

The 3 VMPs enacted over the period of 25 years have impacted the nitrate loading in Skive Fjord and decreased it consistently over time. Observing the graphs in the statistical description section show qualitatively the decrease of nitrate loads, and would suggest the VMPs are a driver of this change. Each VMP era has a lesser mean than the era before, and the variance decreases across the period, giving us confidence in our qualitative assumptions. The hypothesis testing and confidence intervals show that while the graphs seem like enough evidence, our assumptions can sometimes be wrong. Most confidence intervals overlap with the intervals in the preceding era, and some overlap the mean of the preceding era. (This is an easy way to test significance using confidence intervals and critical values, P-value testing is also correct.) This means that the likelihood of decreasing loads due to random chance is greater. However, we still maintain weak evidence against the null hypothesis. If more years or data points were given in each VMP period, we would have more degrees of freedom and thus more confidence in our testing.

SEE ACCOMPANYING .R FILE FOR R CODE USED IN THIS PROJECT.