The Arcsine Transformation: Has the time come for retirement?

Emily Wilson, Melanie Underwood,
Olivia Puckrin, Karla Letto,
Rebecca Doyle, Holly Caravan,
Stacey Camus, Kate Bassett

Abstract

The merits of using the arcsine transformation prior to analyzing proportion data is being questioned in the published literature. While arcsine transformation stabilizes variance and normalizes proportional data, there are several reasons why this method can be problematic. An alternative analysis proposed to address the problems with normality of proportion data is the Generalized Linear Model logistic regression analysis. We compared the frequency of use of arcsine through time in ten leading biological journals. We tested the effectiveness of both arcsine transformation and logistic regression in making the residuals meet the assumptions of normality, homogeneity and independence by noting changes in the residual plots and changes in the p-value and significance decision compared to the linear regression on untransformed data using 40 data sets from the published literature. In the leading biological journals there is an obvious trend of an increased use of arcsine transformation on percentage data starting around the 1970s. Logistic regression was able to improve the residuals' normality, homogeneity and independence more often than arcsine. The arcsine transformation increased and decreased p values at almost the same rate. In comparison, logistic regression increased the p-value in 86% of the data sets, often resulting in a change in significance. The results suggest that logistic regression should be used as an alternative to the arcsine transformation in biological analysis.

Keywords: arcsine, logistic regression, Generalized Linear Model, proportion, voodoo statistic

Introduction

The use of arcsine transformation, also known as inverse transformation (Rao 1998) or angular transformation (Snedecor and Cochran 1989), has been open for debate as to the usefulness in analysis of proportion data that tends to be skewed when the distribution is not normal.

Where is the literature behind arcsine transformation recommendation coming from? The most frequently cited source is Sokal and Rolf (1981), cited by 675 articles. The rational provided by Sokal and Rolf (1981) for using arcsine transformation on proportion and percentage data was arrived at due to it's ability to eliminate the function that ties variance to the mean, by the "stretching out of both tails and compressing the middle" (needs a reference). This would address the concerns of Snedecor and Cochran (1989), who point out that plots of data near 0 and 1, will be scattered and result in high variance. Data can then be considered the proportion of success, which gets around the difficulties of non-normal data (Rao 1998). Along the same rational, Warton and Hui (in press) speak of the motivation of researchers to use arcsine transformation in order to "stabilize the variance of binomial". Additional advocates of arcsine transformation, Kaplan et al. (1975) provide research specific rational for its use. In the study of sclerotic glomerili in human kidneys, large proportions are considered to be the same, as seen on average within 200 kidneys. Use of arcsine transformation is a convenient way of carrying out unweighted linear regressions, which they found beneficial in the analysis.

Although arcsine transformation is a useful tool in stabilizing variances and normalizing proportional data, there are several reasons why this method can be problematic. The equalization of variance in proportional data when using arcsine transformations requires the numbers of trials to be equal for each data point, while the efficacy of arcsine transformation in normalizing proportional data is dependent on sample size, n, and doesn't perform well at extreme ends of the distribution (Worton and Hui in press, Hardy 2002). Another argument against arcsine transformation is that it does not confine proportional data between 0 and 1, resulting in the extrapolation of proportional values that aren't biologically sensible (Hardy 2002). In an example provided by Hardy (2002), the arcsine transformation of the relationship between sex ratio data and distance from a pollutant predicted a sex ratio greater than 1 for males as the distance from the pollutant increased (Hardy, 2002).

An alternative to arcsine transformation that is becoming more prevalent in today's biological analyses is the logistic regression, an analytical method which is designed to deal with proportional data (Steel and Torrie, 1997). Logistic regression allows for binomially distributed proportional data, unlike arcsine transformation that attempts to normalize the data (Worton and Hui in press). The logit link function used in logistic regression provides a more biologically relevant analysis, where the proportional data never falls outside of 0 and 1 (Worton and Hui in press). This link function also can deal with unbalanced data, whereas the arcsine transformation can only effectively equalize variance in proportional data when data points have an equal number of trials (Jaeger 2008, Worton and Hui in press). Also, logistic regression produces easily interpretable and biologically relevant coefficients, unlike the arcsine transformation (Worton and Hui in press).

Although logistic regression seems like the better alternative, arcsine transformation is still a widely used method in scientific studies (See Figure 1). Is logistic regression really an advantageous method over arcsine transformation? This document addresses the effectiveness of arcsine transformation by comparing output of untransformed with transformed proportional data to the output of the untransformed proportional data using logistic regression.

Materials and Methods

In order to determine the frequency of arcsine transformation use in leading journals, we searched for the number of times arcsin(e), arc sin(e), and arc-sin(e) was mentioned in influential journals such as Ecology, Science, and International Committee for the Exploration of the Sea. We graphed these results as publication per year from 1930 to 2010.

In order to evaluate the efficacy of the arcsine transform in addressing violation of assumptions for the GLM, we each searched the literature for five data sets (n total = 40) of proportions as the response variable and mostly choose tests that had at least ten data points. We used the statistical packages Minitab, SPLUS and R to analyze the data sets. We transformed the proportion data using the squareroot arcsine transformation (Sokal and Rohlf 1995). We ran a General Linear Model on the original proportion data, and on the transformed proportion data, and a Generalized Linear Model on the original data, using a binomial distribution with logit link.

From these analyses, we generated three plots per model to test the assumptions for residuals: the QQplot testing normality, the Residuals vs. Fitted values to test for homogeneity, and the Residuals vs. Residual Lag plot to test for independence. We constructed ANOVA tables for GLM, and ANODEV tables for GzLM. We compared the plots for improvements in meeting assumptions, and the p-values in regards to the change in decision and loss or gain in power, for GLM proportion vs. GLM Arcsine Transformed, and the GLM proportion vs. GzLM proportion. Finally, we constructed a summary table to examine the results of these comparisons across all data sets tested.

Results

There is an obvious trend of an increased use of arcsine transformation on percentage data starting around the 1970s. *Animal Behaviour* and *Ecology* showed the highest frequency use of arcsine transformation with over 50 publications per year. *The Journal of Wildlife Management*, *The American Naturalist*, *Heredity* and *Agriculture, Ecosystem and Environments* all showed increased use of arcsine starting in the late 1990s to early 2000s but the amount of publications using this transformation is now currently declining. *Nature* and *Annual Review of Physiology* showed the lowest use of the arcsine transformation. The *ICES Journal of Marine Science* started using transformations in the early 2000s and is still increasing (Fig. 1).

The arcsine transformation had less of an effect on the three residual plots analyzed than the logistic regression. When both the arcsine transformation and logistic regression altered the residuals plots the residual's normality, homogeneity and independence increased more often than they decreased. However, this difference was greater for the logistic regression than the arcsine transformation (Table 1).

Table 1: A comparison of the change in normality, homogeneity and independence for both the linear model with an arcsine transformation and the Generalized Linear Model logistic regression with the linear model without any transformations. Values are out of 40 different data sets.

	Normality			**Homogeneity**			**Independence**		
	Better (%)	Worse (%)	No Change (%)	Better (%)	Worse (%)	No Change (%)	Better (%)	Worse (%)	No Change (%)
Arcsine Transform	15	12.5	72.5	10	10	80	15	7.5	77.5
Logistic Regression	32.5	17.5	50	22.5	12.5	65	27.5	5	67.5

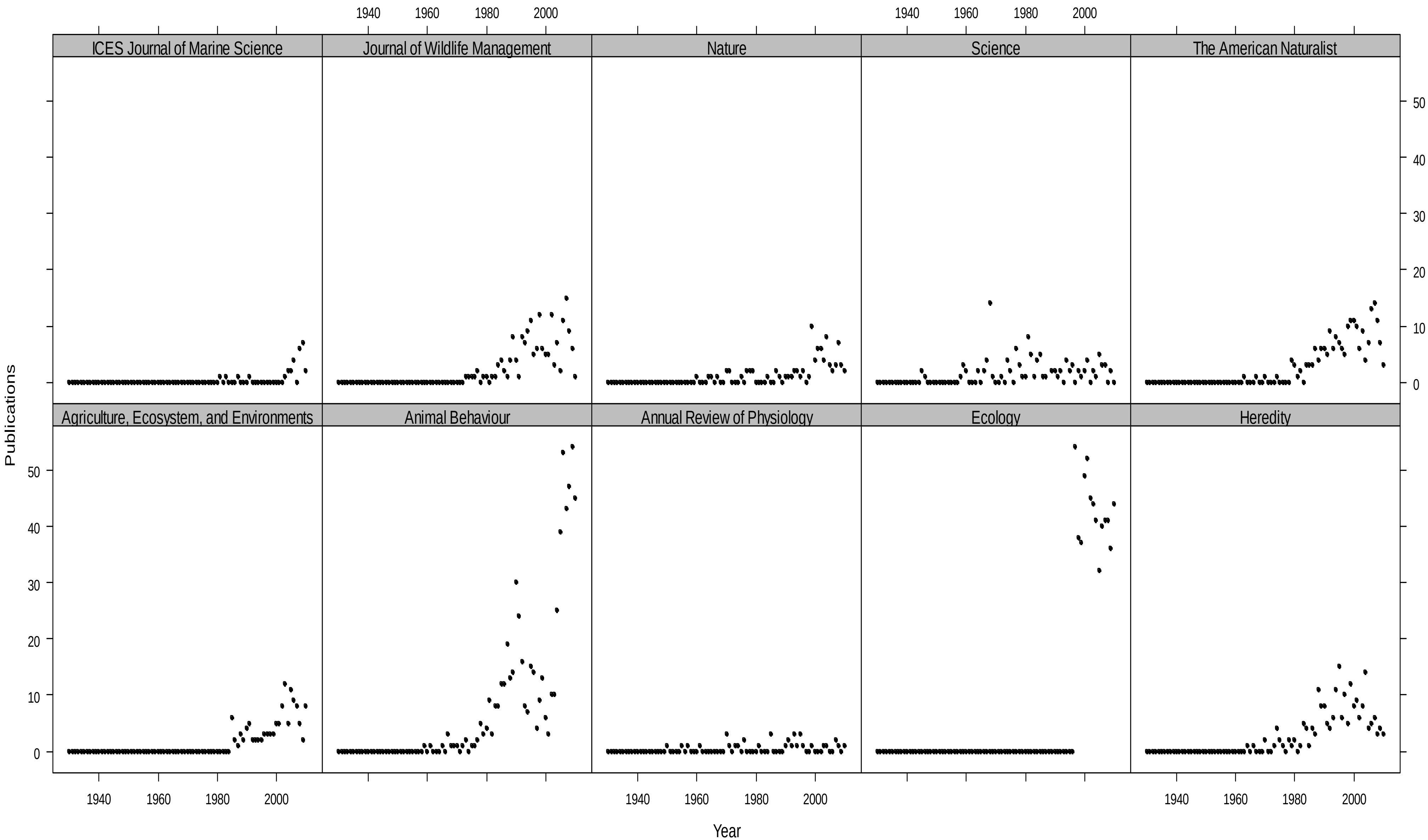


Figure 1: Trends in the use of arcsine transformations of percentage data in leading journals across varying fields of biology from 1930-2010.

In the majority of data sets both the arcsine transformation and logistic regression affected the original p-value obtained by the general linear model without any transformations. This occurred in 95.2% of the data sets for arcsine transformation and 97.6% of the data sets for logistic regression. However, the arcsine transformation increased the p-value in 50% of the data sets and decreased it in 45.2% of the data sets, while the logistic regression increased it in 85.7% of the data sets and only decreased it in 11.9% of the data sets. Therefore, the degree of the effect was different for both analyses (Table 2).

The change in p-value due to t arcsine transformation changed the final decision of significance in only 4.8% of the data sets. However, logistic regression changed the decision of significance in 33.3% of the data sets (Table 2).

Table 2: A comparison of the change in P-value and decision for both the linear model with an Arcsine transformation and the Generalized Linear Model logistic regression with the linear model without any transformations. Values are out of 40 different data sets.

	P-Value			Change in Decision	
	No Change (%)	Decrease (%)	Increase (%)	Yes (%)	No (%)
Arcsine Transform	4.8	45.2	50	4.8	95.2
Logistic Regression	2.4	11.9	85.7	33.3	66.7

Discussion

The arcsine transformation of percentage and proportion data with binomial distributions has been widely accepted for many years. Its use however, was not particularly prevalent until the 1980s, when it became increasingly common. This coincides with the publication of Biometry 2 ed. by Sokal and Rohlf in 1981 which promotes the arcsine transformation of proportional data. This particular edition has been cited by approximately 675 publications to date.

The use of arcsine transformations across different fields of research has followed a similar trend. Rarely any incidences of the use of arcsine transformations were seen before 1960, with a scattering of references to it between 1960 and 1970. In 1980 the use of arcsine transformation gained speed, increasing steadily to present day. Older, more defined journals have had relatively low occurrences of arcsine transformation. Among different fields, for example *Nature*, *Heredity*, and the *ICES Journal of Marine Science*, the highest number of publications using the arcsine transformation in any given year did not exceed ten publications before 1990. A small differences across fields is when the use if arcsine transformations shows up in publications. In the field of genetics, represented by the journal *Heredity*, articles were being published as early as the mid 1960s, while in the *ICES Journal of Marine Science*, the first reference to the use of arcsine transformations is not until 1980. There is also variation in the number of publications referencing arcsine transformations within fields. Within the field of ecology, different journals vary in the amount of publications per year with arcsine transformed data. In older ecology-based publications, there is a noticeable range in the use of arcsine transformation (ex. *TREE*, *Agriculture*, *Ecology and Environment*, and *The American Naturalist*). While *TREE* peaked in 1997 at five publications, the journal *Ecology* published 55 articles that same year. This indicates a general assumption that the arcsine transformation is the required method for dealing with percent and proportion data.

After examining the effect of arcsine transformations of percent and proportion data on residuals, it appears as though this particular transformation is not all it's cracked up to be. On a broad scale, in 85-90% of the cases, the arcsine transformation had either no effect on the residual plots, or actually made them worse. This is supported in various publications which found the same thing (Mech and Goyal 1995, Hussman et *al.* 2003,). In a few small cases the arcsine transformation actually did improve the residual plot, though relative to the number of instances in which the plots became worse, or did not change at all, this does not lend much support to the use of the transformation. Where the arcsine transformation resulted in better residuals, it did not actually effect the decision. Data sets which were improved by the arcsine transformation had particularly bad residuals; it is possible that the arcsine transformation may be useful for certain data sets, but not others.

In comparison to the arcsine transformation, we saw different results when using the logistic regression instead. Performing a logistic regression as opposed to a general linear model on untransformed data improved the quality of residuals by 50% compared to the effect of the arcsine transformation. The arcsine transformation increased and decreased p values at almost the same rate, but despite this, did not seem to influence the outcome of the decision. The logistic regression on the other hand, increased the p value 86% of the time, leading to a change in decision roughly 33% of the time. The arcsine transformation is not likely to change the decision that was arrived at using unaltered data, indicating that transformations may not be necessary. The logistic regression often resulted in significant p values from unaltered data becoming not significant after the analysis. If the data is arcsine transformed, there is the possibility of getting more significant results than if the data had been left unaltered and subjected to a logistic regress, where we see fewer significant results. This could mean that the arcsine transformation is biased towards significant p values, which may not be the case.

Based on the results of this study, and supported by Betts et al. (2007) among others (see references), we recommend that the first step always be using the original, *untransformed* data to examine the residuals. Only if the residuals do not meet the assumptions, would you consider an arcsine transformation. If the arcsine transformation improved the residuals, then it would be a suitable method. However, based on the results of this study, that is not likely to be the case. More recently, it is being suggested that the logistic regression be used as an alternative to the arcsine transformation (Jaeger 2008, Worton and Hui 2010).

References Cited

Betts et al., (2007) Uneven rates of landscape change as a source of bias in roadside wildlife surveys. *J. Wildl. Manage.* 71, 2266-2273.

Hardy, I.C.W., ed (2002) Sex *ratios: Concepts and research methods*, Cambridge University Press.

Hussman et al., (2003) Correlation patterns of Marrow fat in Rocky Mountain Elk Bones. *J. Wildl. Manage.* 67:742-746.

Jeager, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards Logit Mixed Models. *J. Mem. Lang.* 59, 434-446.

Kaplan, C. B. *et al*. (1975) Age-related incidence of sclerotic glomeruli in human kidneys. *Am J Pathol*, 80, 227–234.

Mech, L.D. and Goyal S.M. (1995) Effects of Canine Parvovirus on Gray Wolves in Minnesota. *J. Wildl. Manage.* 59(3): 565-570.

Rao, P.V. (1998). Statistical Research Methods in the Life Sciences. Brooks/Cole Publishing Company.

Sahai, H. and Ageel, M.I. (2000). The analysis of variance: fixed, random, and mixed models. Springer.

Snedecor, G.W. *et al*. (1989). Statistical Methods, (8^{th} ed) Iowa State University Press, Iowa.

Sokal, R.R. and Rohlf, J.F. 1981, Biometry: the principles and practice of statistics in biological research. 2nd ed., W. H. Freeman and Company, San Francisco, 859p.

Worton, D.I. and Hui, F.K.C., (2010) The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, In Press.

Zar, J. H. (1998) Biostatistical Analysis. (4^{th} ed) Prentice Hall.

Appendix A: Kate Bassett

Dataset 1

Reference:

Taylor, R. M., M. D., Dr. P.H., M. A. Haseeb, D.K.S.M., Dip Bact., T. H. Work, M.D., M.P.H. (1955). A Regional Reconnaissance on Yellow Fever in the Sudan. Bulletin of the World Health Organization, 12, 711-725.

Raw Data:

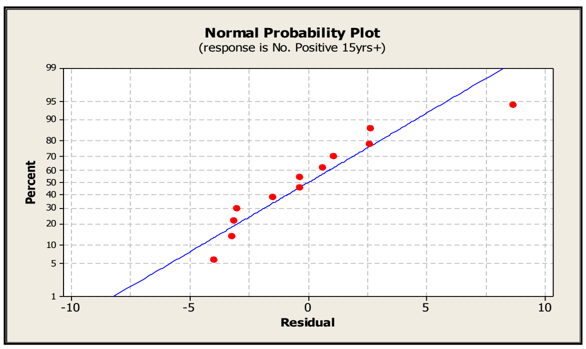
Proportion of blood samples, testing positive for Yellow Fever, affecting ages 15 years or older.

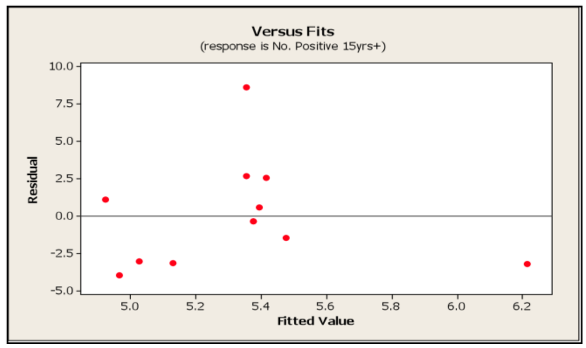
Total Blood Samples Testing for Yellow Fever	Total Positive Results	No. Positive People 15yrs+	Proportion Positive People 15yrs+	Arcsin	Degrees
80	4	3	0.75	1.05	60.00
27	2	2	1.00	1.57	90.00
22	2	2	1.00	1.57	90.00
41	16	8	0.50	0.79	45.00
19	7	1	0.14	0.39	22.21
17	6	6	1.00	1.57	90.00
39	8	5	0.63	0.91	52.24
39	6	5	0.83	1.15	65.91
38	9	8	0.89	1.23	70.53
44	6	4	0.67	0.96	54.74

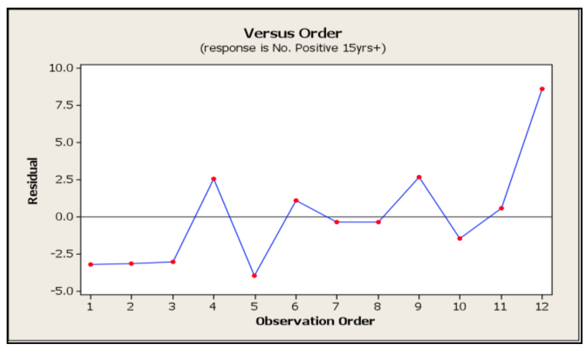
40	7	6	0.86	1.18	67.79
38	14	14	1.00	1.57	90.00

Non-Transformed Data:

LM: Linear Regression



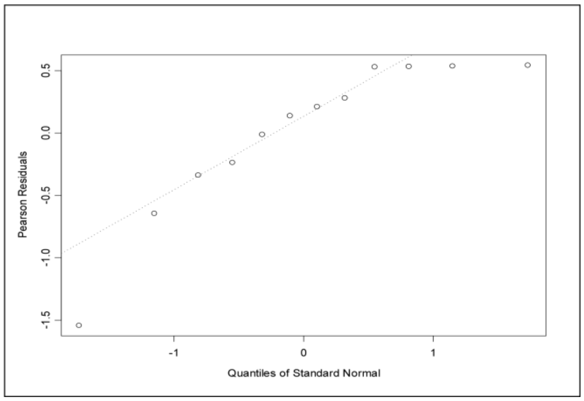


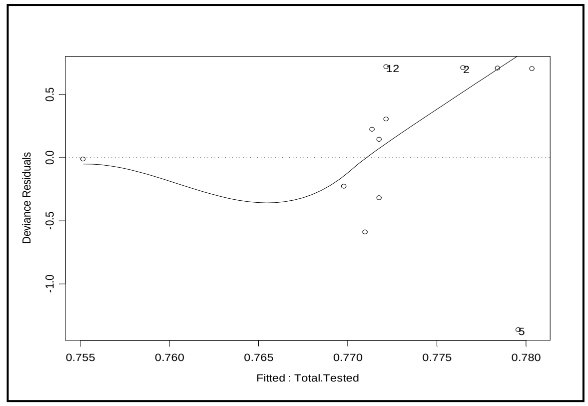


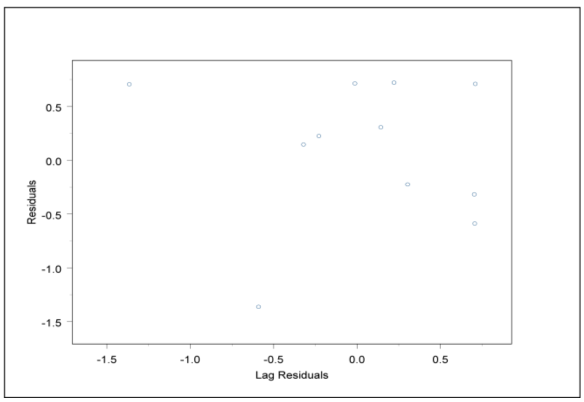
ANOVA

Source	DF	Seq SS	Adj SS	Adj MS	F	P
total tested	1	0.00055	0.00055	0.00055	0.01	0.933
Error	10	0.73506	0.73506	0.07351		
Total	11	0.73560				

GzLM: Logistic Regression





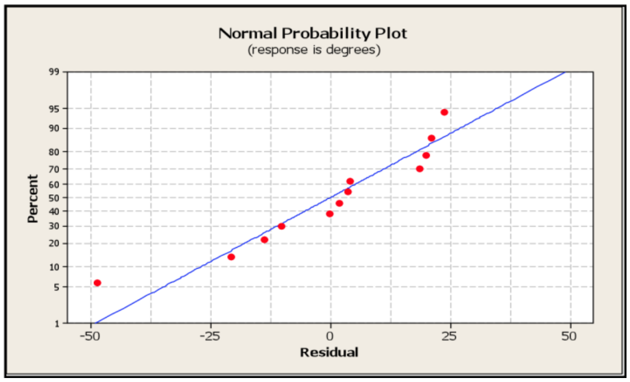


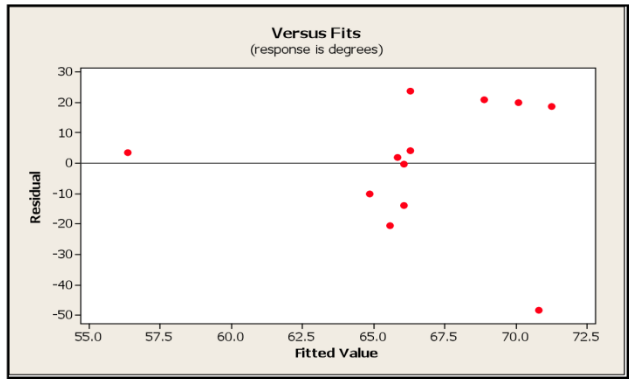
ANODEV Table:

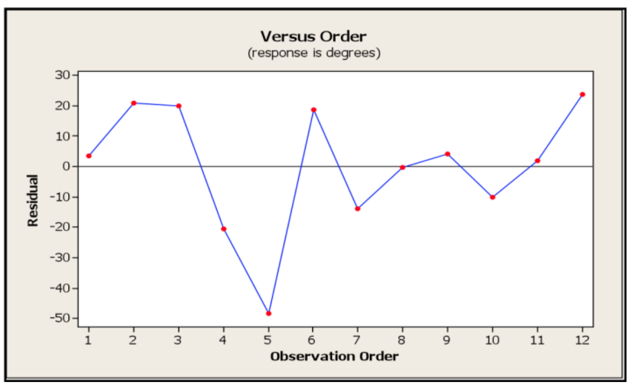
	Df	Dev	Res df	Res dev	P
Null			11	4.547819	
Total Tested	1	0.002679815	10	4.545139	0.958714407

Arcsin Transformed Data:

LM: Linear Regression







ANOVA Table:

Source	DF	SeqSS	Adj SS	AdjMS	F	P
total tested	1	167.3	167.3	167.3	0.34	0.571
Error	10	4868.0	4868.0	486.8		
Total	11	5035.3				

Dataset 2

Reference:

Bruce-Chwatt, L. J., C. Garrett-Jones and B. Weitz. (1966). Ten Years' Study (1955-64) of Host Selection by Anopheline Mosquitos. Bulletin of the World Health Organization, 35, 405-439.

Raw Data:

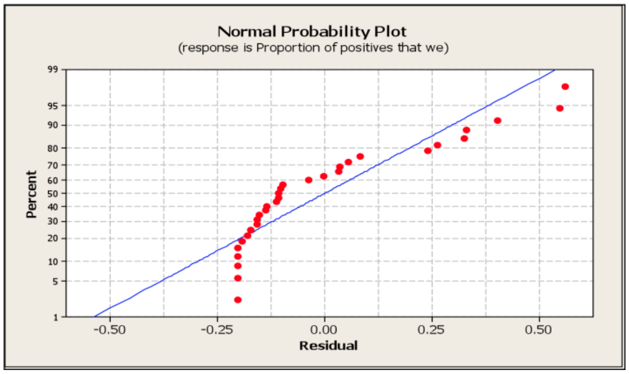
The proportion of positive blood samples from Anopheline mosquitoes that were from primates

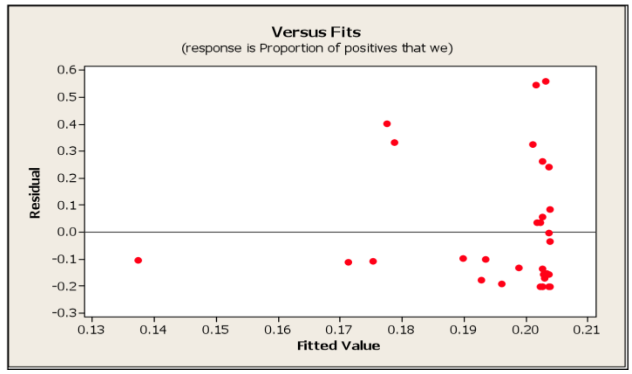
Total Blood Smear	Total positive tests	Positive Primate blood #	Proportion of positives that were from primates	arcsine	degrees
3338	2945	173	0.059	0.245	14.027
1070	930	85	0.091	0.307	17.597
112	109	5	0.046	0.216	12.367
1446	1433	133	0.093	0.310	17.737
1137	1122	15	0.013	0.116	6.640
164	148	35	0.236	0.508	29.098
285	276	145	0.525	0.811	46.454
19	19	0	0.000	0.000	0.000
510	508	33	0.065	0.258	14.766
802	629	2	0.003	0.056	3.233
2934	2741	182	0.066	0.261	14.932
2586	2511	1278	0.509	0.794	45.513
64	63	48	0.762	1.061	60.794
2711	2682	1552	0.579	0.864	49.526

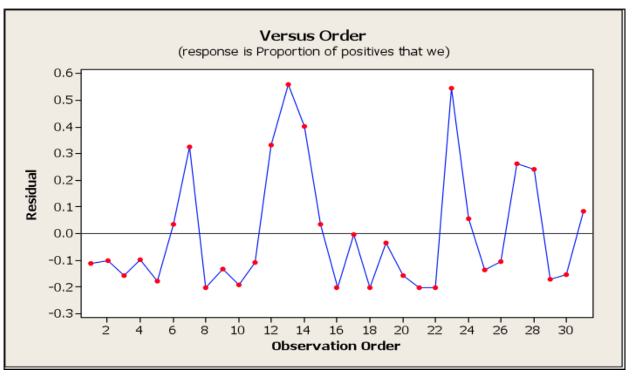
204	183	43	0.235	0.506	28.995
120	118	0	0.000	0.000	0.000
24	5	1	0.200	0.464	26.565
6	6	0	0.000	0.000	0.000
7	6	1	0.167	0.421	24.095
23	22	1	0.045	0.215	12.310
119	118	0	0.000	0.000	0.000
152	137	0	0.000	0.000	0.000
231	127	95	0.748	1.045	59.870
118	109	28	0.257	0.532	30.453
126	123	8	0.065	0.258	14.775
6830	6432	198	0.031	0.176	10.105
116	114	53	0.465	0.750	42.988
9	9	4	0.444	0.730	41.810
96	96	3	0.031	0.178	10.182
60	60	3	0.050	0.226	12.921
8	7	2	0.286	0.564	32.312

Non-Transformed Data:

LM: Linear Regression



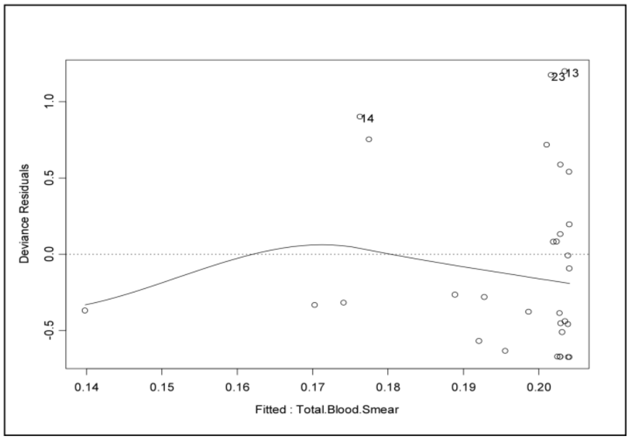


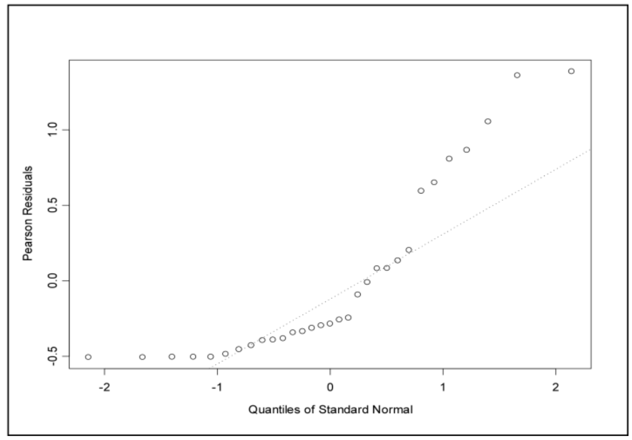


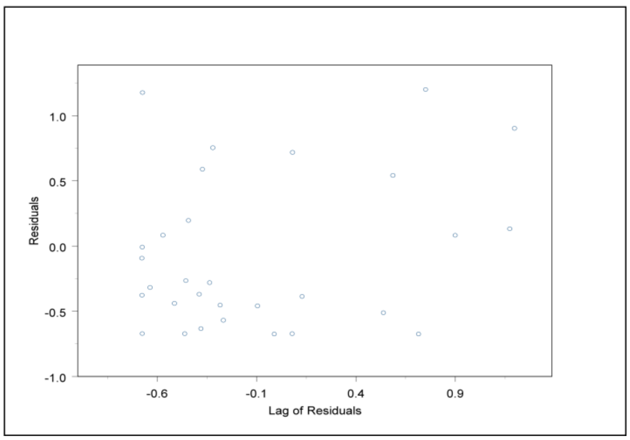
ANOVA Table:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Total Blood Smear	1	0.00621	0.00621	0.00621	0.11	0.739
Error	29	1.59659	1.59659	0.05505		
Total	30	1.60279				

GzLM: Logistic Regression





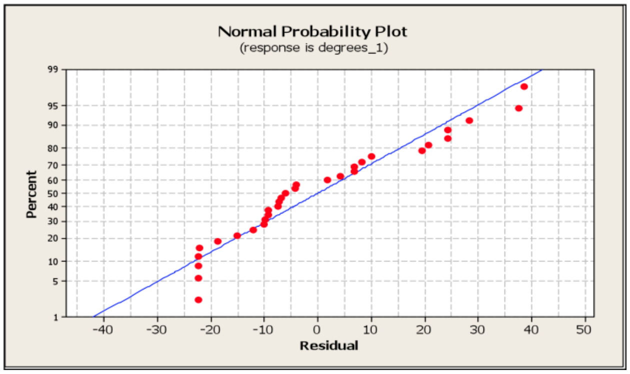


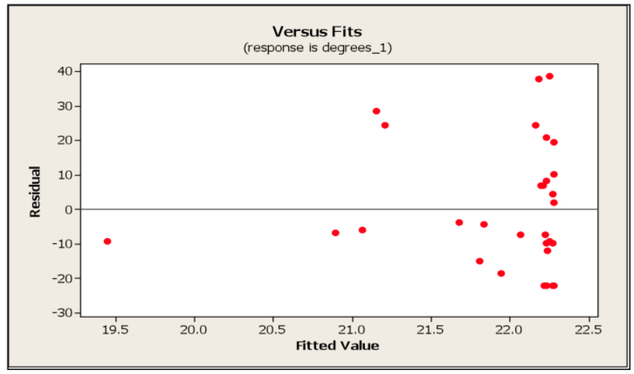
ANODEV Table:

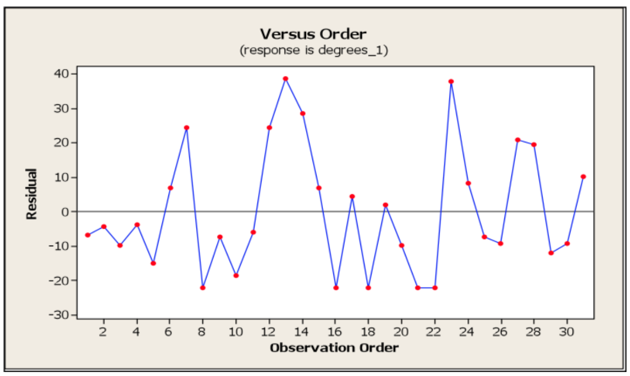
	Df	Dev	Res df	Res dev	P
Null			30	10.13312	
Total Blood Smear	1	0.04119883	29	10.09192	0.839154612

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Total Blood Smear	1	11.2	11.2	11.2	0.03	0.856
Error	29	9785.6	9785.6	337.4		
Total	30	9796.9				

Dataset 3

Reference:

Minakawa, Noboru, John I. Githure, John C. Beier, and Guiyun Yan. (2001). Anopheline Mosquito Survival Strategies During the Dry Period in Western Kenya. Journal of Medical Entomology, 38 (3): 388-392.

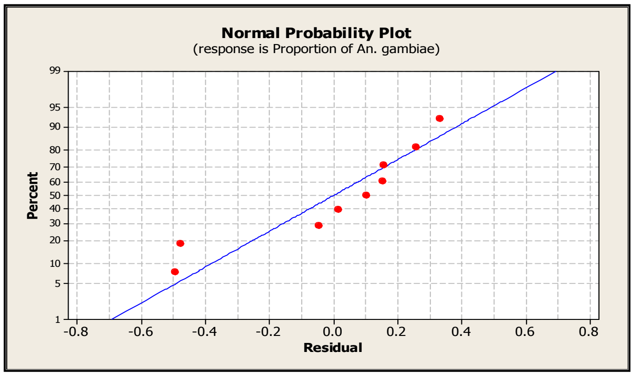
Raw Data:

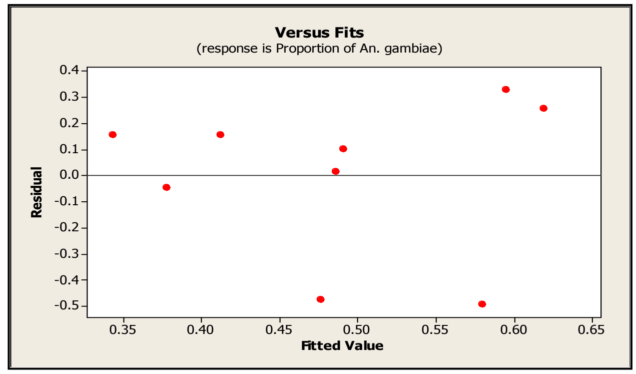
The proportion of surviving larvae that belong to *Anopheles gambiae* mosquitoes

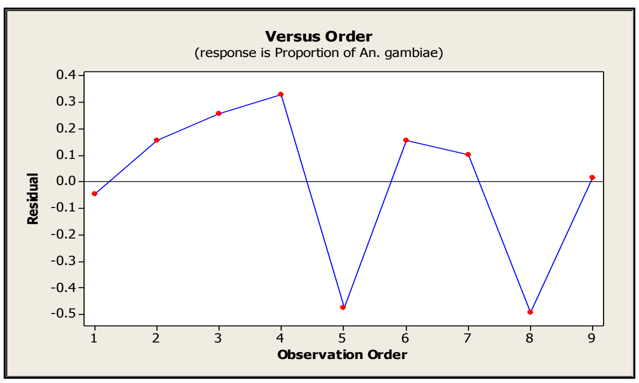
Number of larvae collected	Proportion of An. gambiae	arcsine	degrees
9	0.333	0.615	35.244
16	0.567	0.853	48.850
58	0.877	1.212	69.469
53	0.925	1.293	74.106
29	0	0.000	0.000
2	0.5	0.785	45.000
32	0.593	0.879	50.360
50	0.085	0.296	16.951
31	0.5	0.785	45.000

Non-Transformed Data:

LM: Linear Regression



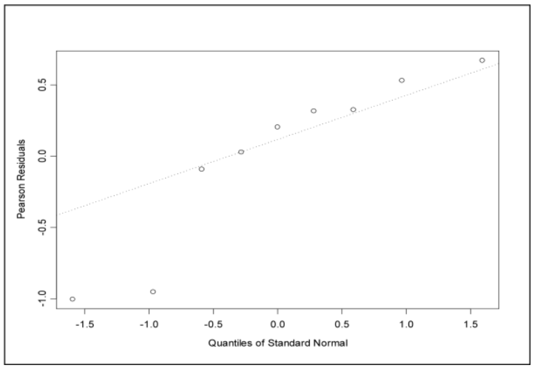


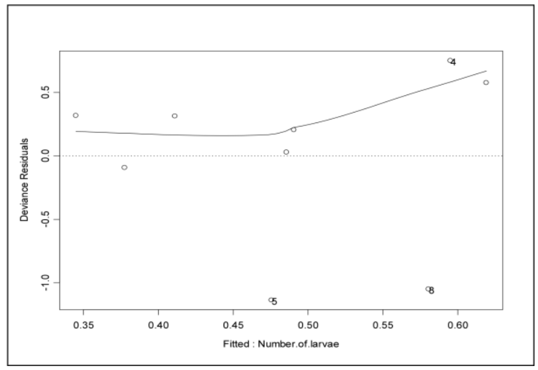


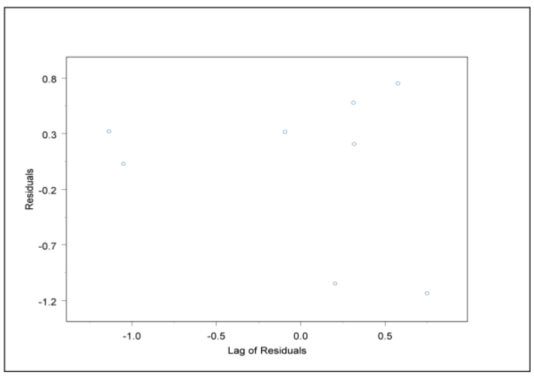
ANOVA Table:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Number of larvae collected	1	0.0760	0.0760	0.0760	0.75	0.415
Error	7	0.7084	0.7084	0.1012		
Total	8	0.7844				

GzLM: Logistic Regression:





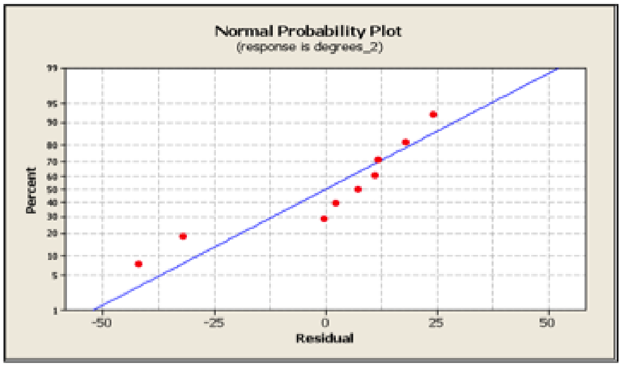


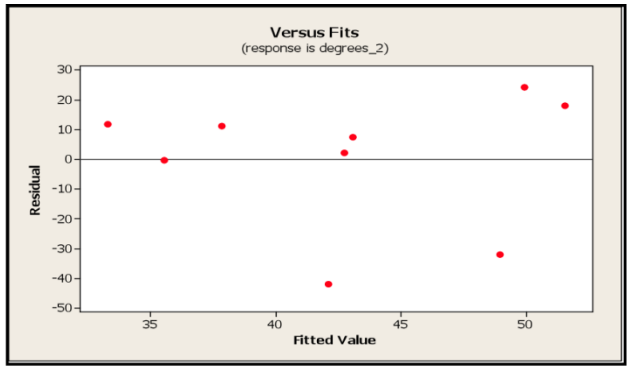
ANODEV Table:

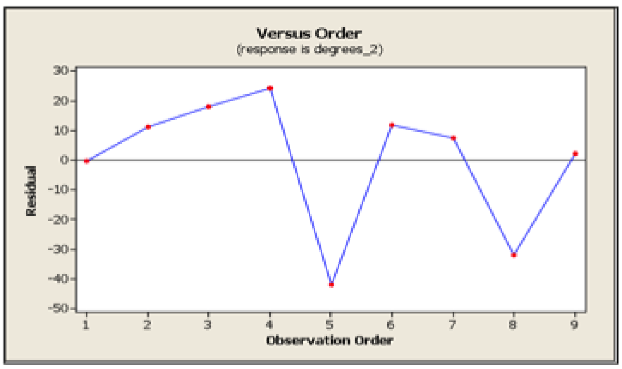
	Df	Dev	Res df	Res dev	P
Null			8	3.845196	
Number of larvae	1	0.307249	7	3.537947	0.579373238

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Number of larvae collected	1	334.2	334.2	334.2	0.58	0.470
Error	7	4017.7	4017.7	574.0		
Total	8	4351.8				

Dataset 4

Reference:

Marr, Allen G., and John L. Ingraham. (1962). Effect of Temperature on the Composition of Fatty Acids in *Escherichia Coli.* Department of Bacteriology, University of California, Davis, California, 84: 1260-1267.

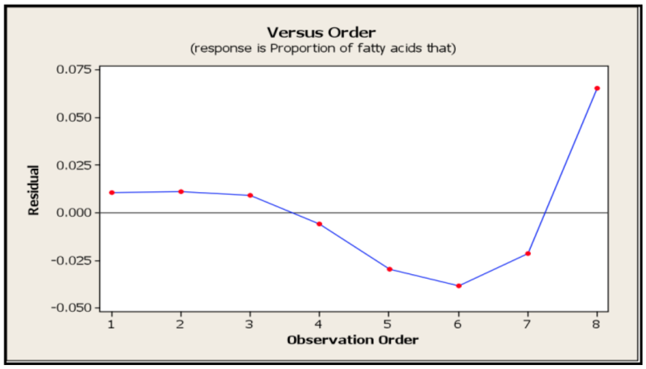
Raw Data:

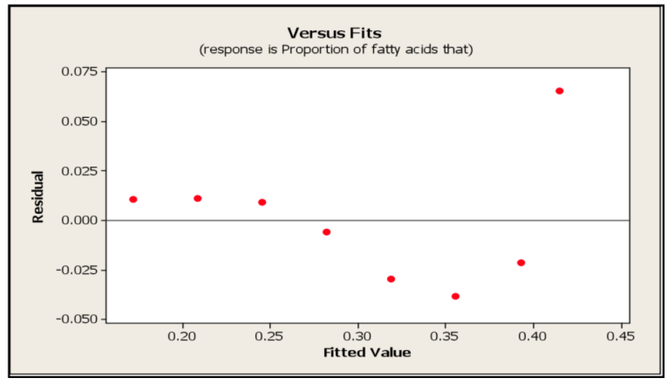
The proportion of fatty acid content specific to Palmitic Acid in *E. coli*, with change in temperature

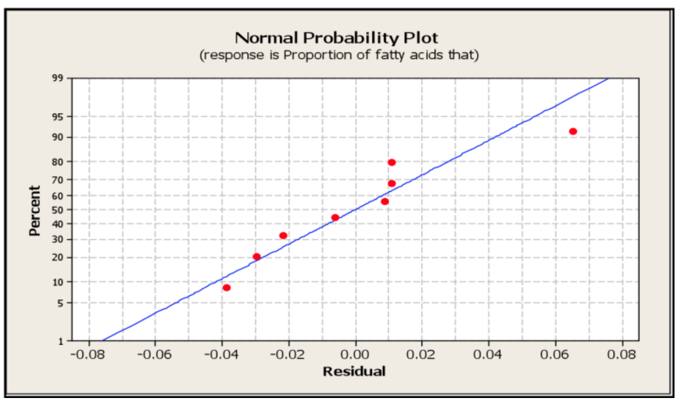
Temperature (^{O}C)	Proportion of fatty acids that is Palmitic acid	arcsine	degrees
10	0.182	0.441	25.253
15	0.219	0.487	27.903
20	0.254	0.528	30.264
25	0.276	0.553	31.692
30	0.289	0.568	32.520
35	0.317	0.598	34.265
40	0.371	0.655	37.524
43	0.48	0.765	43.854

Non-Transformed

LM: Linear Regression



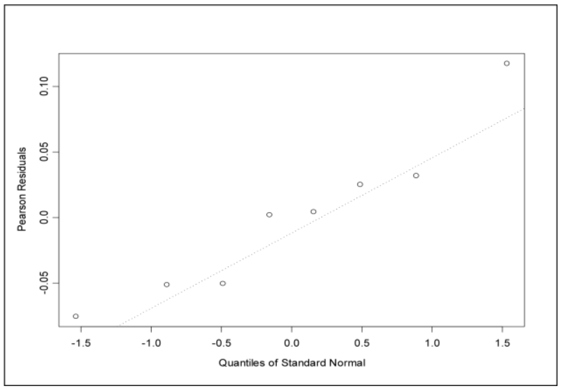


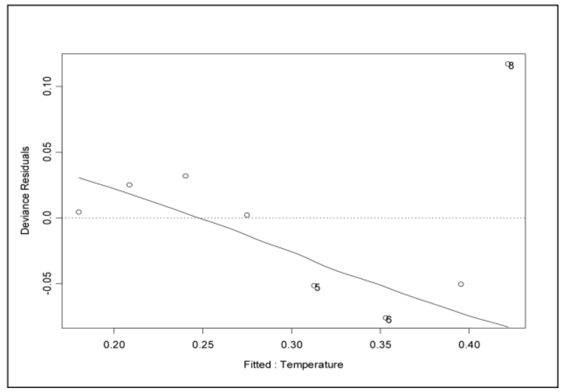


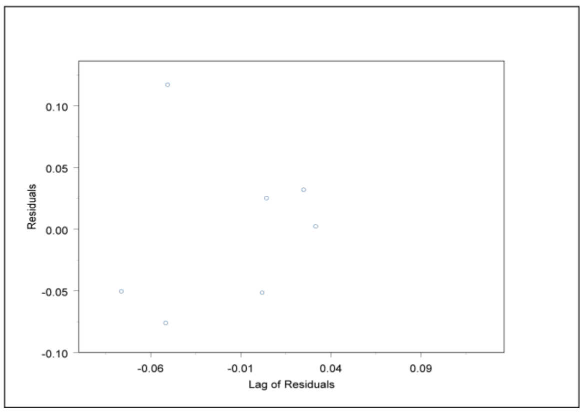
ANOVA

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Temp	1	0.053548	0.053548	0.053548	43.05	0.001
Error	6	0.007462	0.007462	0.001244		
Total	7	0.061010				

GzLM: Logistic Regression





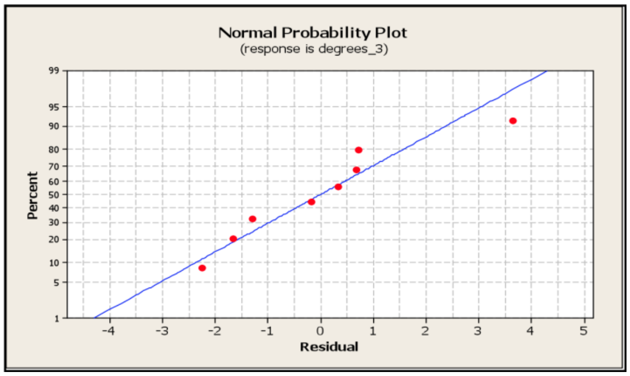


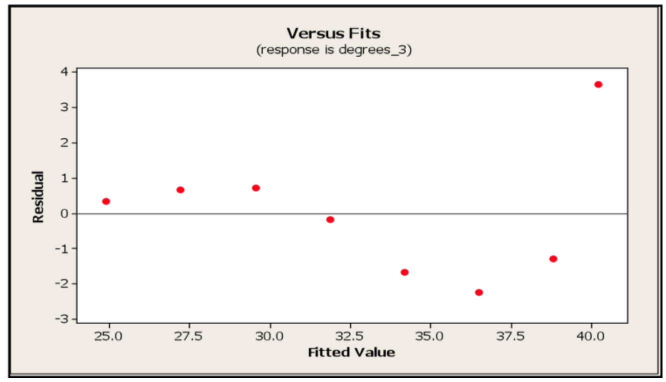
ANODEV

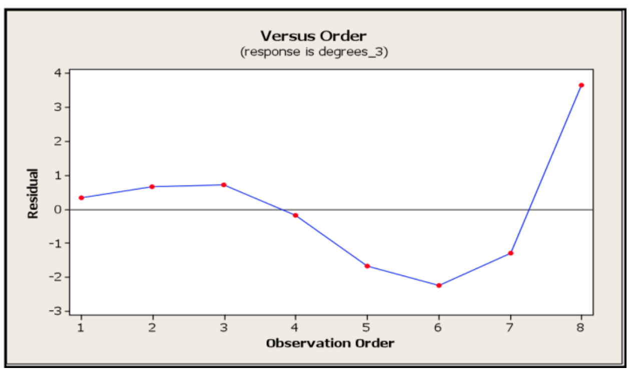
	Df	Dev	Res df	Res dev	P
Null			7	0.2863103	
Temperature	1	0.2599641	6	0.0263462	0.610144821

Arcsin Transformed Data

LM: Linear Regression







ANOVA

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Temp	1	211.40	211.40	211.40	53.22	0.000
Error	6	23.83	23.83	3.97		
Total	7	235.23				

Dataset 5

Reference:

Selong , Jason H., and Thomas E. McMahon. (2001). Effect of Temperature on Growth and Survival of Bull Trout, with Application of an Improved Method for Determining Thermal Tolerance in Fishes. Transactions of the American Fisheries Society, 130: 1026-1037.

Raw Data:

The proportion of nymphs occupying the 25 degree area at various hours after molting

Nymphs

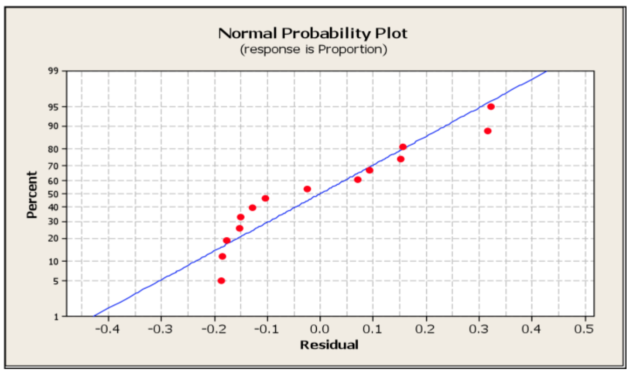
ambient temp 25 degrees

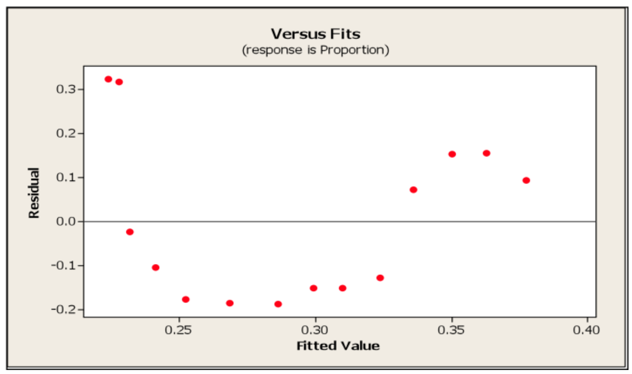
hours since molt	percent of nymphs at heat spot	Proportion	arc	degrees
3.6	54.6	0.546	0.831463302	47.63933805
7.3	54.4	0.544	0.829455152	47.52427948
11	20.8	0.208	0.473574358	27.13381201
19.9	13.7	0.137	0.379154456	21.72395013
30.1	7.5	0.075	0.277405516	15.89416531
45.4	8.3	0.083	0.2922392	16.74407275
62.1	9.9	0.099	0.320080164	18.33924252
74.4	14.7	0.147	0.393481083	22.54480537
84.5	15.9	0.159	0.410151253	23.49993576
97.4	19.6	0.196	0.458628632	26.27748499
108.9	40.7	0.407	0.691853401	39.64027992

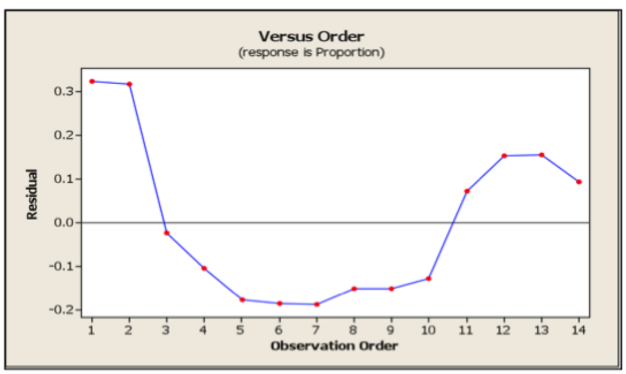
122.3	50.2	0.502	0.787398169	45.11459186
133.9	51.8	0.518	0.803402054	46.03154693
147.8	47	0.47	0.755380134	43.28009362

Non-Transformed

LM: Linear Regression



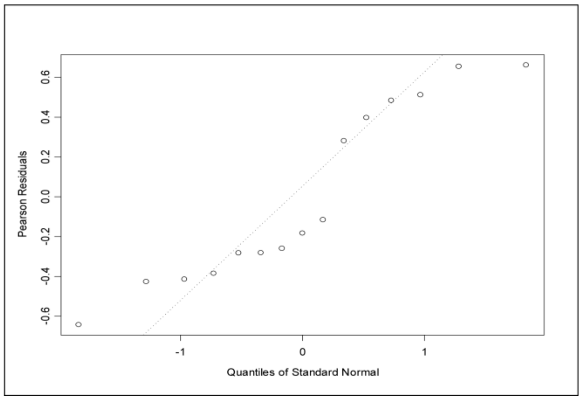


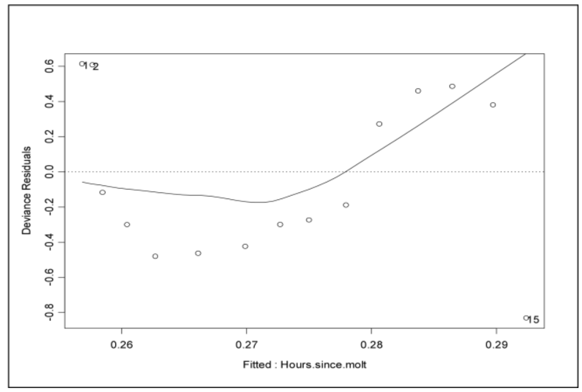


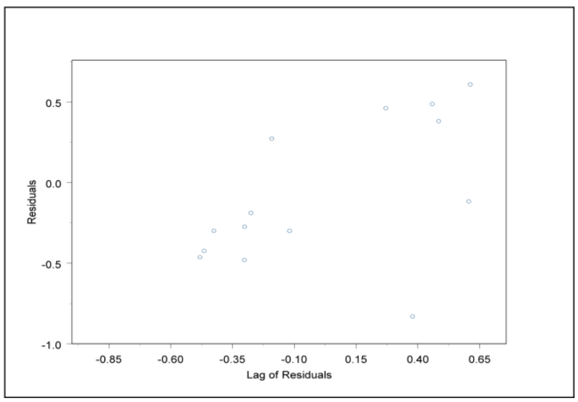
ANOVA

Source	DF	Seq	Adj SS	Adj MS	F	P
hours since molt	1	0.03612	0.03612	0.03612	0.99	0.340
Error	12	0.43902	0.43902	0.03658		
Total	13	0.47513				

GzLM: Logisitic Regression





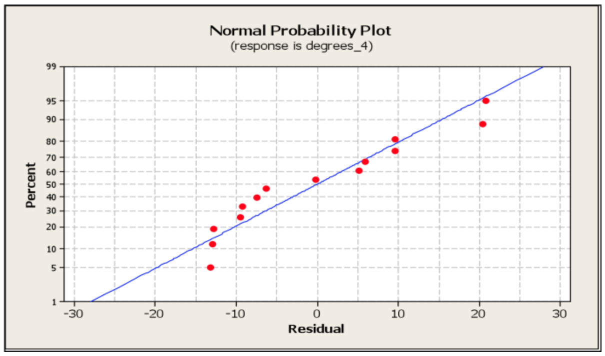


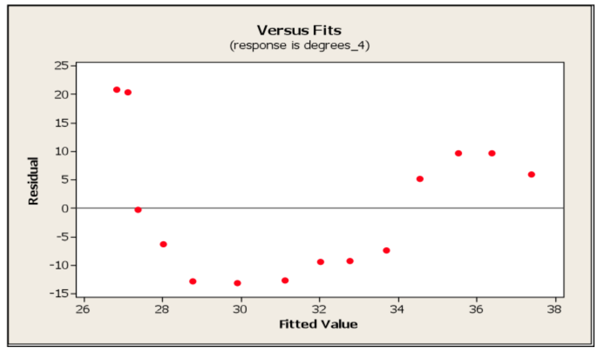
ANODEV

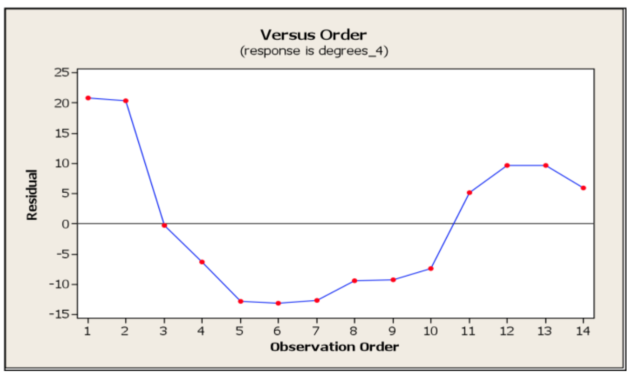
	Df	Dev	Res df	Res dev	P
Null			14	3.044213	
Hours since molt	1	0.01040625	13	3.033807	0.918747921

Arcsin Transformed

LM: Linear Regression







ANOVA

Source	DF	Seq SS	Adj SS	Adj MS	F	P
hours since molt	1	171.3	171.3	171.3	1.10	0.315
Error	12	1868.1	1868.1	155.7		
Total	13	2039.4				

Summary:

LM: prop vs. deg

	Plots			P-Value	
Data Set	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	no diff	no diff	no diff	no	down 0.362
2	no diff	no diff	no diff	no	up 0.117
3	worse	no diff	no diff	no	up 0.055
4	no diff	no diff	no diff	no	down 0.001
5	no diff	no diff	no diff	no	down 0.025

LM: Non-Transformed vs. GLM: prop

	Plots			P-Value	
Data Set	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	no diff	worse	better	no	down 0.362
2	no diff	no diff	no diff	no	up 0.117
3	worse	no diff	no diff	no	up 0.164
4	no diff	no diff	no diff	yes- sig to nonsign	up 0.609
5	no diff	no diff	no diff	no	up 0.578

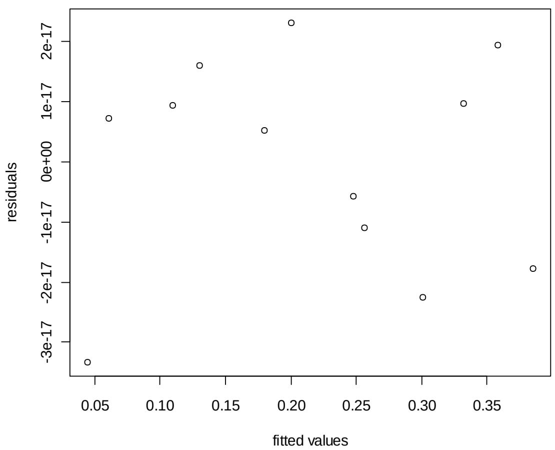
Appendix B – Stacey Camus

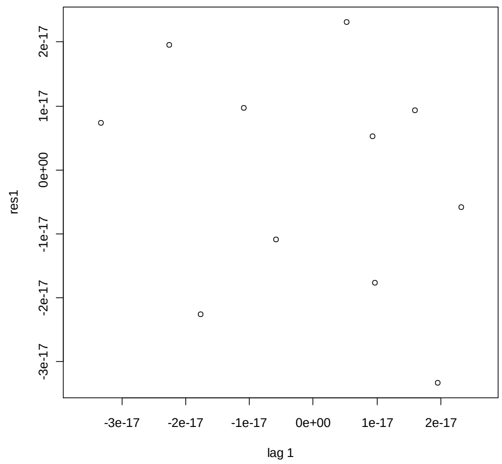
Dataset 1: Lipids in Animal Tissues.

Data Adapted from:
Folch, J., Lees, M., and Sloane-Stanley, G.H. 1957. A simple method for the isolation and purification of total lipids from animal tissues. The Journal of Biological Chemistry 226(1): 497-509.

Dataset 1			
White Matter Lipids (mg/175ml of tissue extract)	Protein Content (%)	Propotion of Protein Content	Arcsine-Sqrt Tranformation
6.1	2.32	0.02	8.76
4.5	2.82	0.03	9.67
35.8	2.42	0.02	8.95
30.1	2.69	0.03	9.44
38.5	1.67	0.02	7.43
33.2	1.98	0.02	8.09
25.6	2.01	0.02	8.15
24.8	2.01	0.02	8.15
20	2.06	0.02	8.25
18	2.14	0.02	8.41
11	2.1	0.02	8.33
13	1.98	0.02	8.09

General linear model with no arcsine transformation for dataset 1





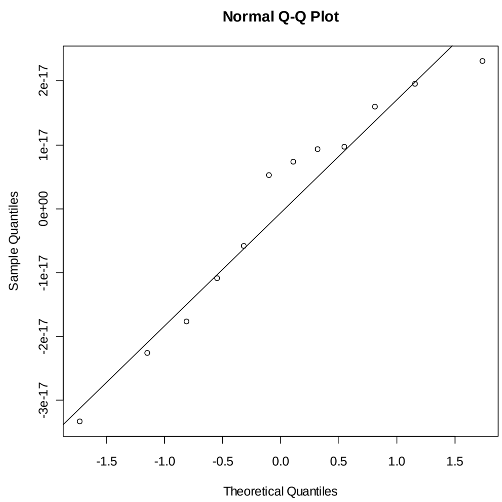


Table 1: ANOVA Table for GLM with no arcsine transformation for dataset 1.

Parameters	**Degrees of Freedom**	**Sum of Squares**	**Mean Squares**	**F-value**	**P-value**
White Matter Lipids	1	0.145	0.145	$4.12e^{32}$	<0.0001 ($2e^{-16}$)
Residuals	10	0	0		

**NOTE: F-test was a perfect fit so results may be unreliable.

General Linear Model with arcsine transformation for dataset 1

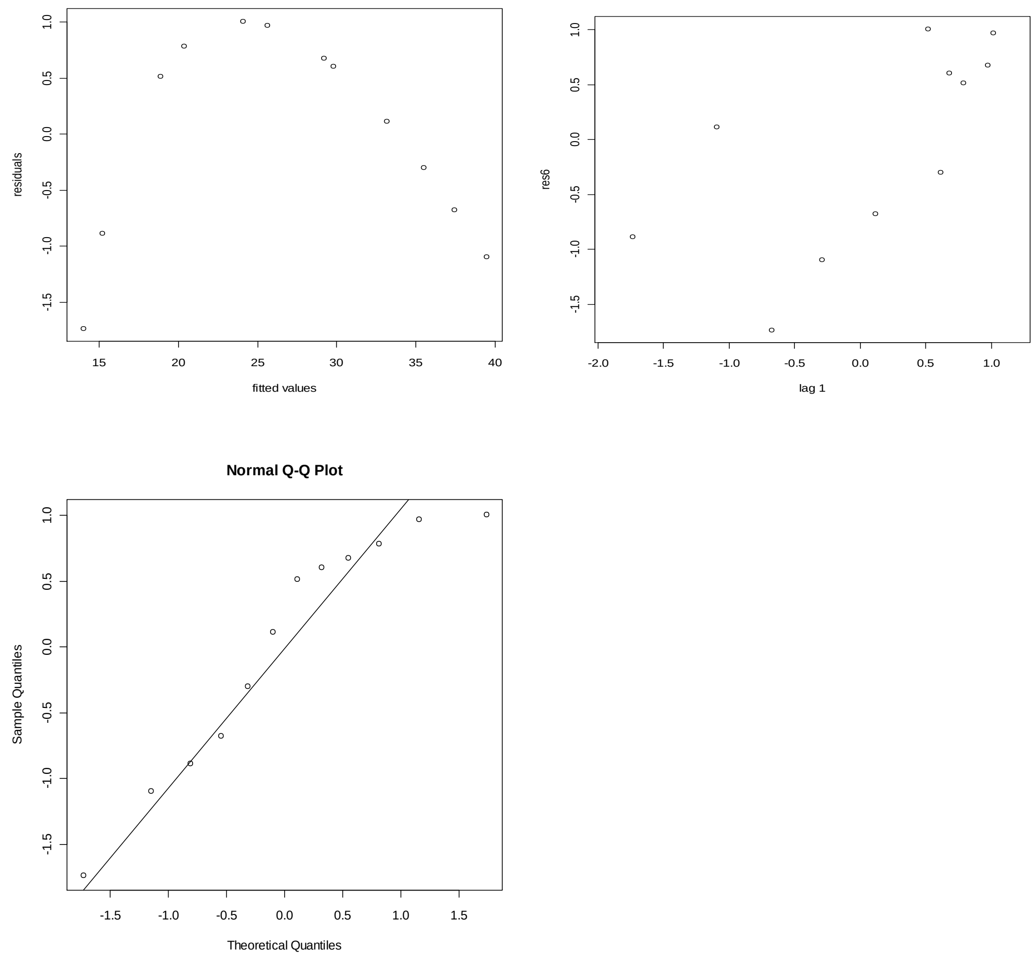
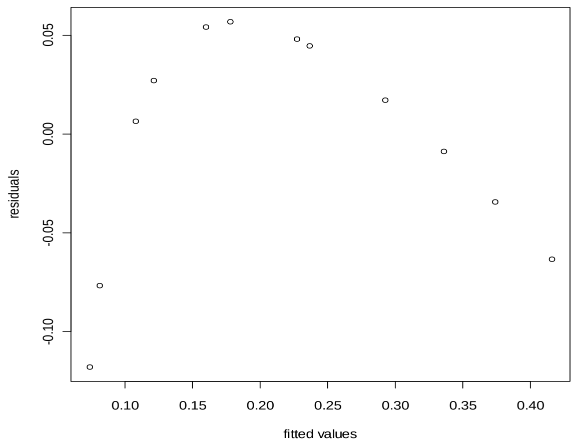
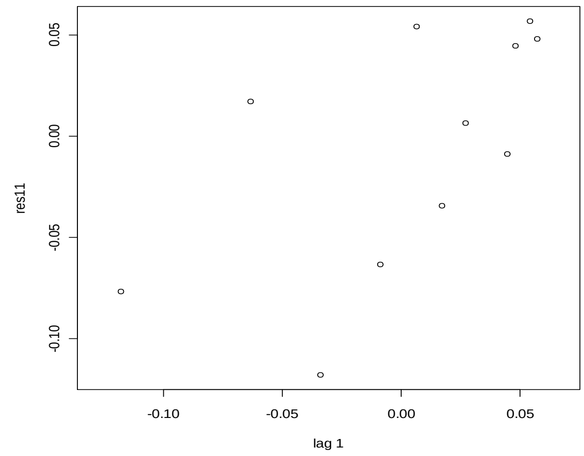


Table 2: ANOVA Table for the general Linear Model with arcsine transformation of dataset 1

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
White Matter Lipids	1	815.96	883.44	883.44	<0.0001 (4e^{-11})
Residuals	10	9.24	0.92		

Generalized Linear Model with binomial error structure for dataset 1





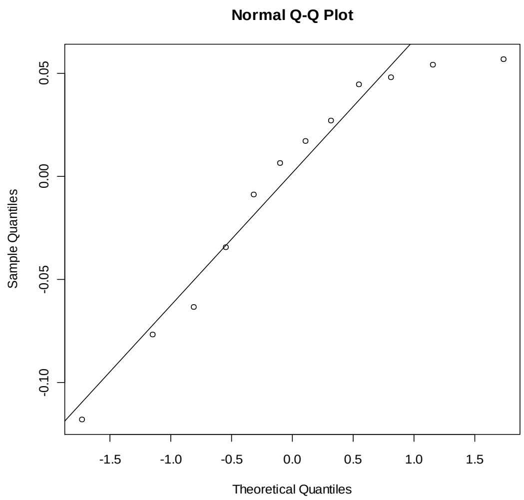


Table 3: ANODEV for Generalized Linear Model with binomial error structure for dataset 1

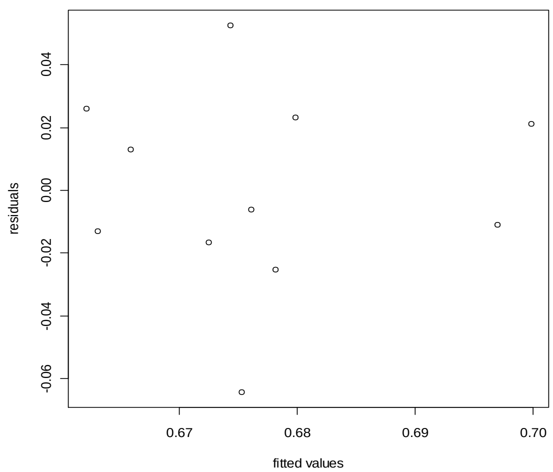
Parameters	Degrees of Freedom	Deviance	Residual Deviance	P-value
White Matter Lipids	1	815.96	883.44	0.3437
Null		9.24	0.92	

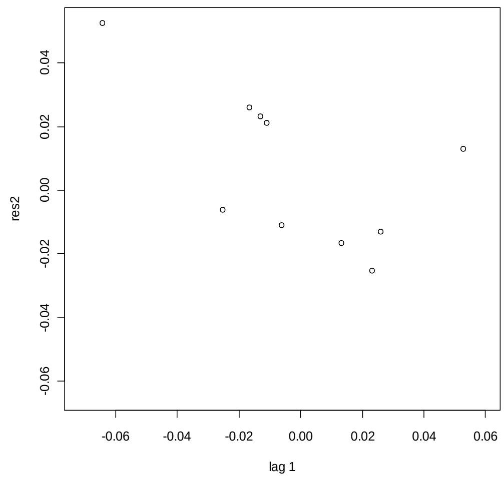
Dataset 2: Water Turnover in Mule Deer

Data Adapted from:
Knox, K.L., Nagy, J.G. Nagy, and Brown, R.D. Water turnover in Mule Deer. Journal of Wildlife Management. 33(2): 389-393.

Dataset 2: Water Turnover in Mule Deer			
age (days)	**Body Water (% of body weight)**	**Propotion of Body Water**	**Arcsine-Sqrt Tranformation**
105	72.1	0.721	58.11577939
120	68.6	0.686	55.91950772
231	67	0.67	54.93843704
220	65.3	0.653	53.90916821
211	70.3	0.703	56.97690374
300	65	0.65	53.72880156
305	68.8	0.688	56.04306622
250	65.6	0.656	54.08990059
285	67.9	0.679	55.48870994
240	72.7	0.727	58.50030546
235	61.1	0.611	51.41326491

General Linear Model with no arcsine transformation for dataset 2





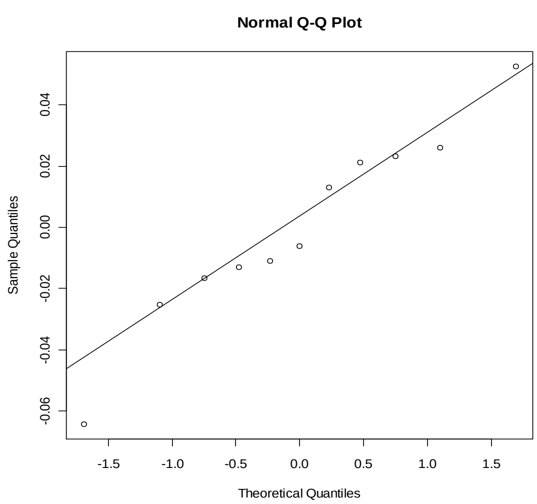
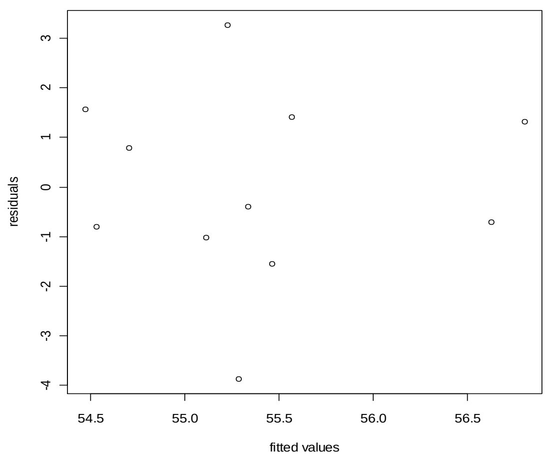
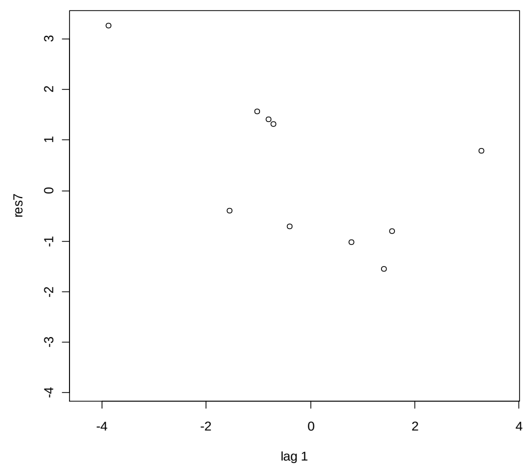


Table 4: ANOVA for General Linear Model with no arcsine transformation for dataset 2

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Age	1	0.0015	0.0015	1.357	0.2739
Residuals	9	0.0010	0.0011		

General Linear Model with an arcsine transformation for dataset 2





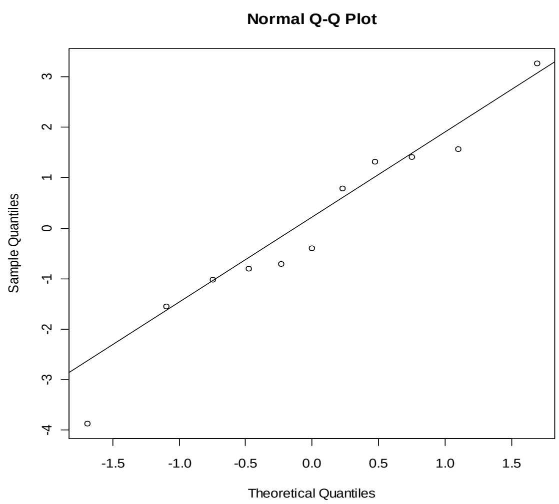
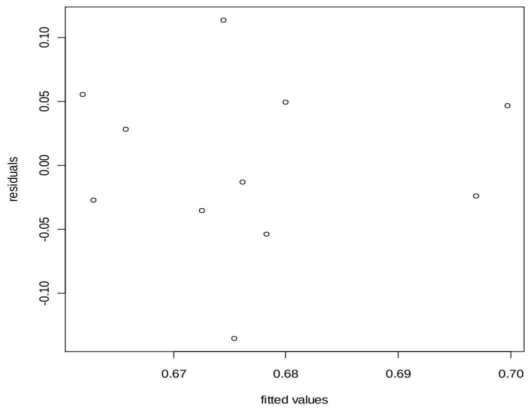
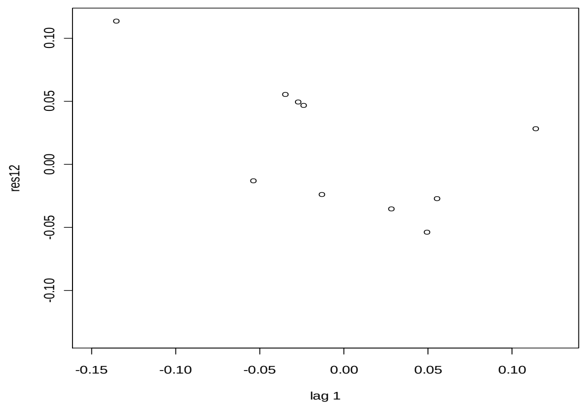


Table 5: ANOVA for General Linear Model with an arcsine transformation for dataset 2

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Age	1	5.73	5.73	1.38	0.2695
Residuals	9	37.26	4.14		

Generalized Linear Model: Logistic Regression with binomial error structure for dataset 2





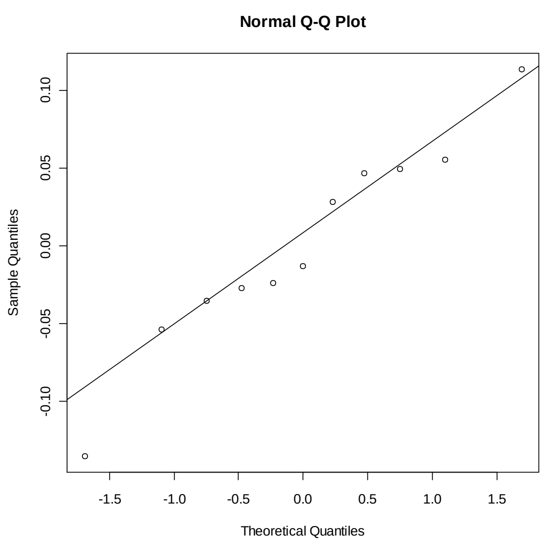


Table 6: ANODEV for Generalized Linear Model: Logistic Regression with binomial error structure for dataset 2

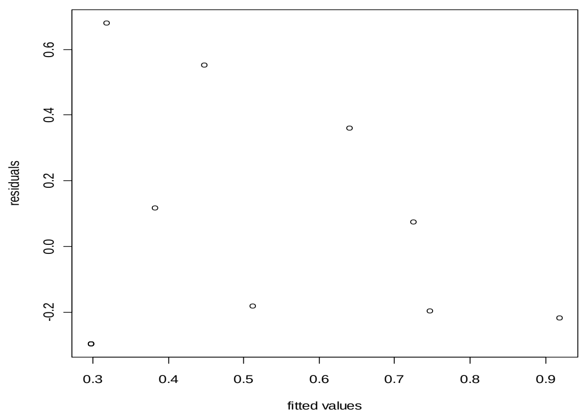
Parameters	Degrees of Freedom	Residual Deviance	P-value
Age	9	0.052	0.9338
Null	10	0.045	

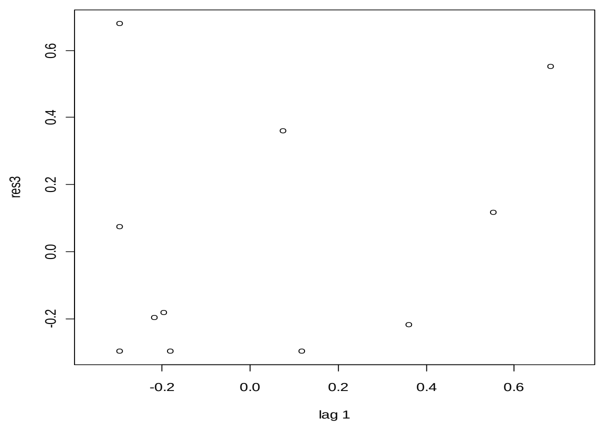
Dataset 3: Productivity of beavers

Data Adapted from:
Henry, D.B. and Bookhout, T.A. 1969. Productivity of beavers in Northeastern Ohio. Journal of Wildlife Management 33(4): 927-932.

Dataset 3: Productivity of beavers			
Number of Placental Scars	**Beaver Conception Rate for Previous Year (%)**	**Propotion of Conception**	**Arcsine-sqrt Tranformation**
0	0	0	0
0	0	0	0
10	33	0.33	35.06156296
21	55	0.55	47.86958524
29	70	0.7	56.78908924
16	100	1	90
20	80	0.8	63.43494882
0	0	0	0
4	50	0.5	45
7	100	1	90
1	100	1	90
0	0	0	0

General Linear Model with no arcsine transformation for dataset 3





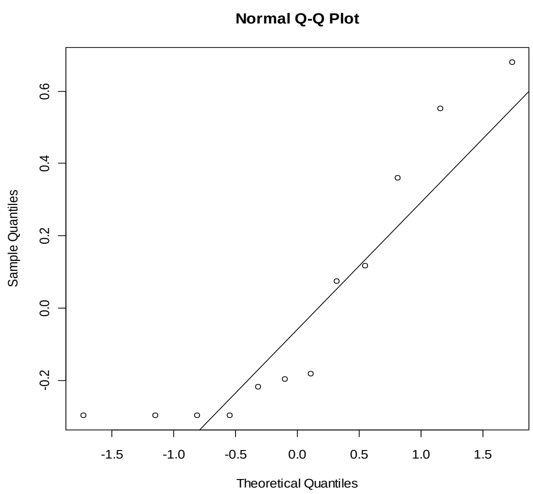
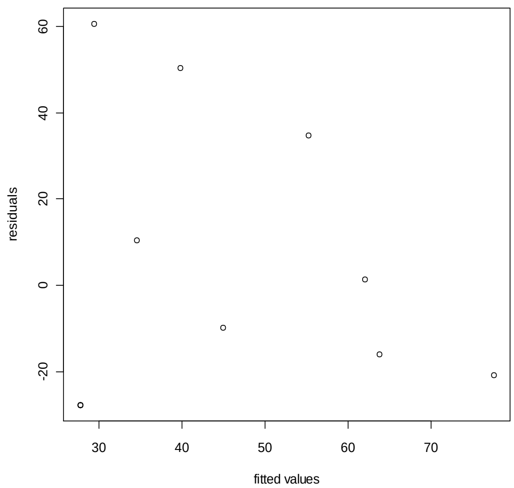
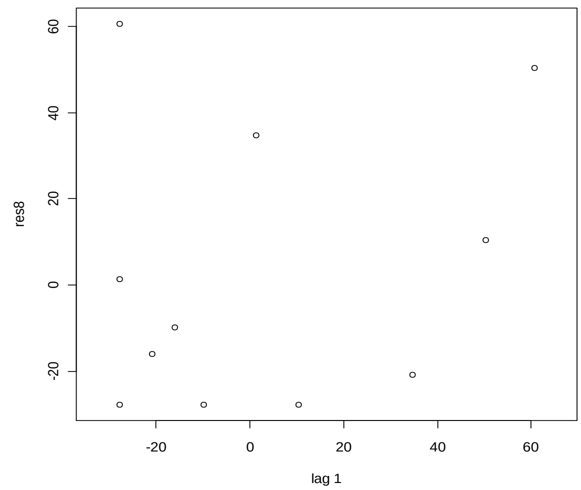


Table 7: ANOVA for General Linear Model with no arcsine transformation for dataset 3

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Placental Scars	1	0.52	0.52	3.73	0.0823
Residuals	10	1.39	0.13		

General Linear Model with an arcsine transformation for dataset 3





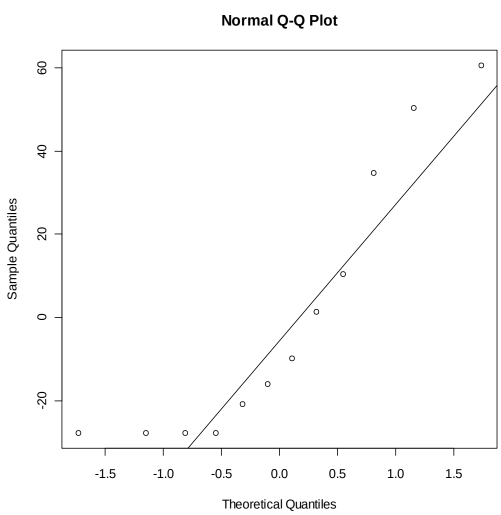
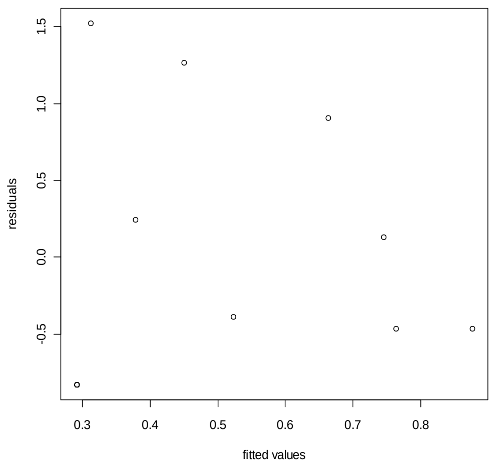
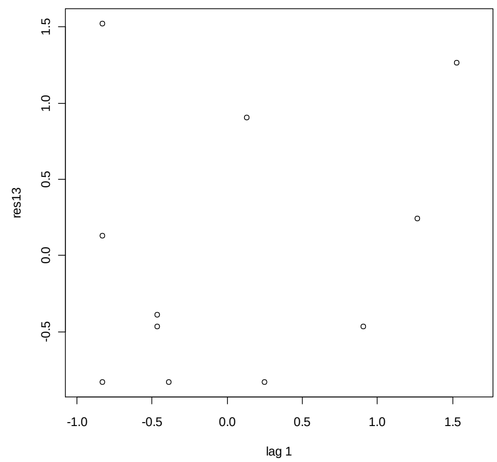


Table 8: ANOVA for General Linear Model with an arcsine transformation for dataset 3

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Placental Scars	1	3352.40	3352.40	2.94	0.1167
Residuals	10	1136.90	1136.90		

Generalized Linear Model: Logistic Regression with binomial error structure for dataset 3





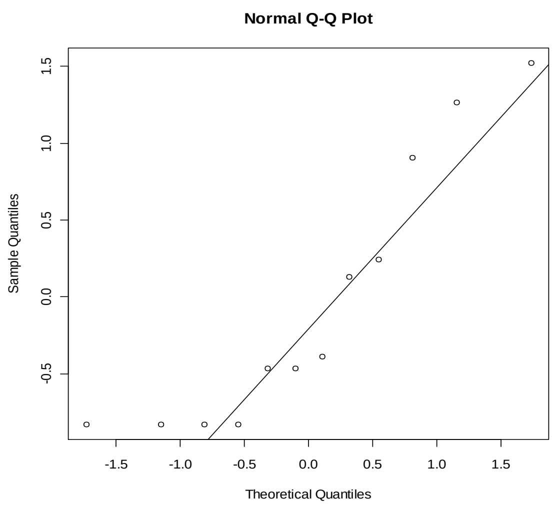


Table 9: ANODEV for Generalized Linear Model: Logistic Regression with binomial error structure for dataset 3

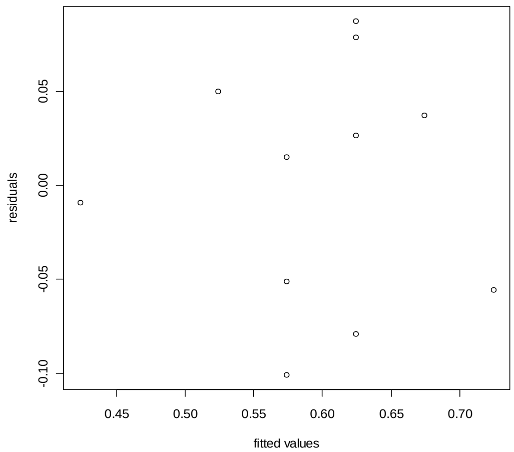
Parameters	Degrees of Freedom	Residual Deviance	P-value
Placental Scars	10	8.17	0.137
Null	11	10.38	

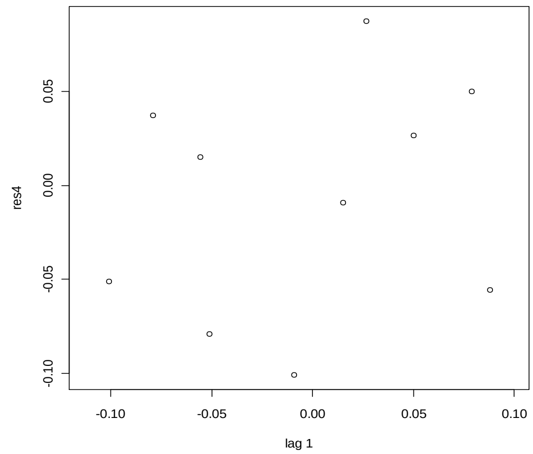
Dataset 4: Carolina Wren Incubation During the Month of July

Data Adapted from:
Laskey, A.R. 1948. Some nesting data on the Carolina Wren at Nashville, Tennessee. Bird-Banding 19(3): 101-121.

Dataset 4: Carolina Wren Incubation During the Month of July			
mean temperature (Degrees Fahrenheit)	**Time spent on nest (Percent)**	**propotion of time spent on nest**	**arcsine-sqrt tranformation**
77	71.2	0.712	57.54369154
78	54.5	0.545	47.58180355
79	52.3	0.523	46.31826812
79	47.3	0.473	43.45226113
82	41.4	0.414	40.04793816
79	58.9	0.589	50.12664354
76	66.9	0.669	54.87753501
78	71.2	0.712	57.54369154
78	65.1	0.651	53.78888372
80	57.4	0.574	49.25552069
78	70.3	0.703	56.97690374

General Linear Model with no arcsine transformation for dataset 4





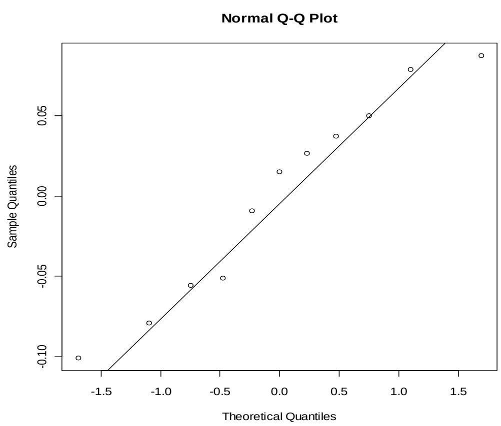
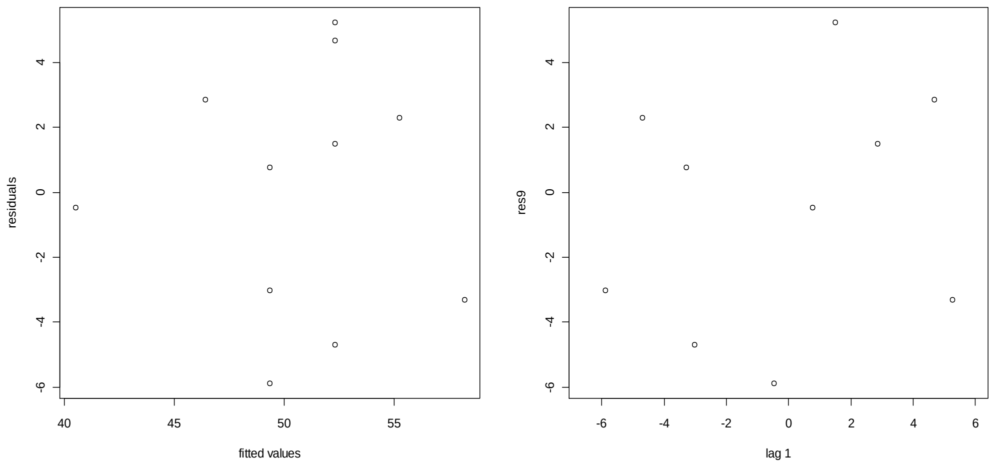


Table 10: ANOVA for General Linear Model with no arcsine transformation for dataset 4

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Temperature	1	0.062	0.062	13.67	0.00493
Residuals	9	0.041	0.0046		

General Linear Model with an arcsine transformation for dataset 4



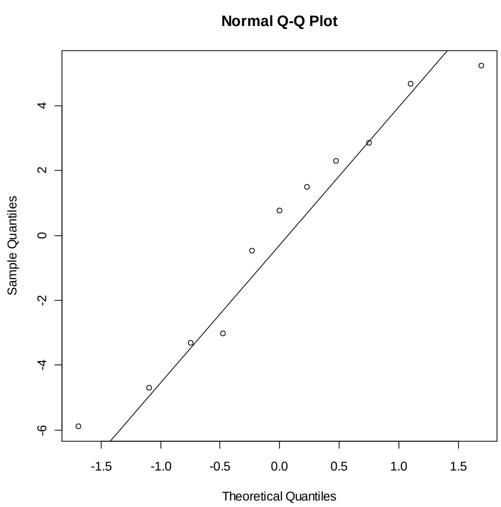
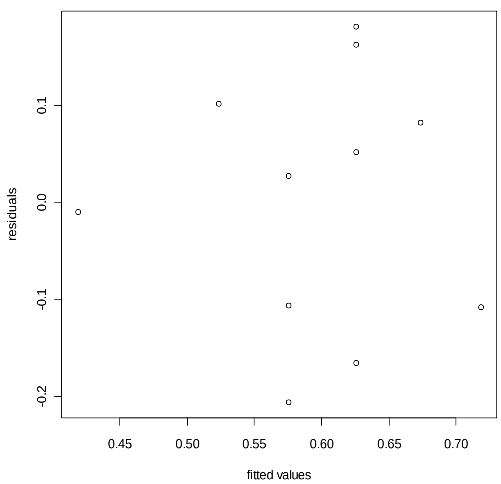
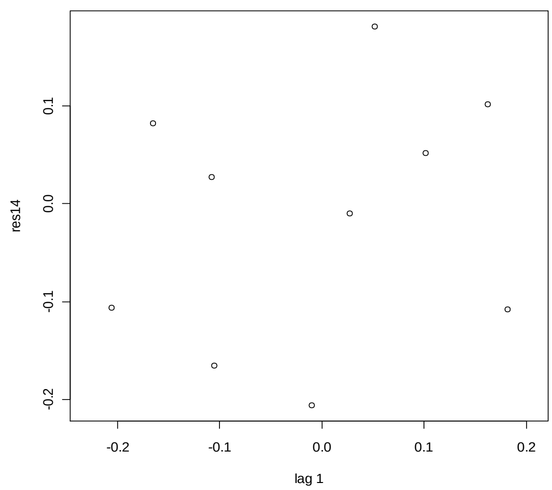


Table 11: ANOVA for General Linear Model with an arcsine transformation for dataset 4

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Temperature	1	214.31	214.31	13.48	0.005142
Residuals	10	143.08	15.89		

Generalized Linear Model: Logistic Regression with binomial error structure for dataset 4





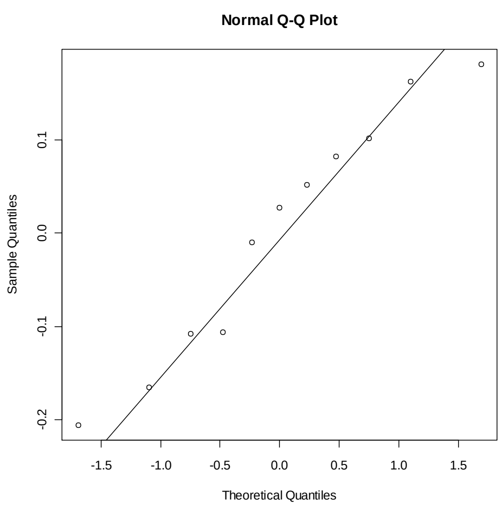


Table 12: ANODEV for Generalized Linear Model: Logistic Regression with binomial error structure for dataset 4

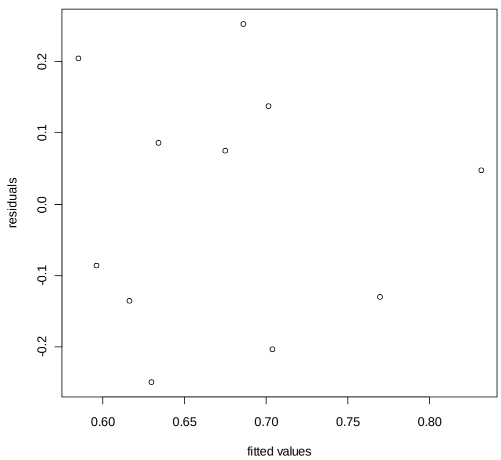
Parameters	Degrees of Freedom	Residual Deviance	P-value
Temperature	9	0.17	0.6105
Null	10	0.43	

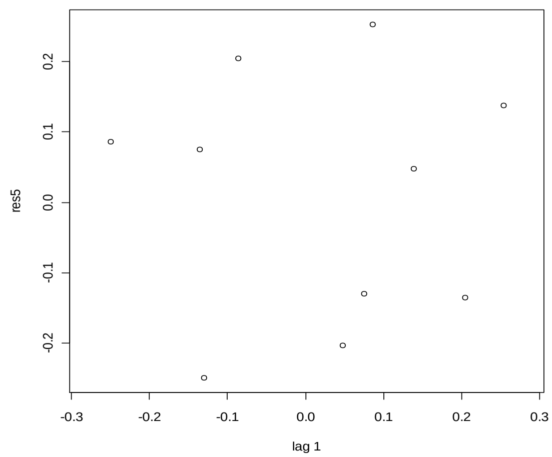
Dataset 5: Deer Browsing on mountain maple.

Data Adapted from:
Krefting, L.W., Stenlund, M.H., and Seemel, R.K. 1966. Effect of simulated and natural deer browsing on mountain maple. Journal of Wildlife Management. 30(3): 481-488.

Dataset 5: Deer Browsing on mountain maple			
Number of Twigs	**Total Twig Length Browsed**	**Propotion of Twig Length Browsed**	**Arcsine-Sqrt tranformation**
5524	50	0.5	45
7729	88	0.88	69.73209894
5488	84	0.84	66.42182152
5224	94	0.94	75.82118171
4323	72	0.72	58.05194057
4255	38	0.38	38.05672982
6663	64	0.64	53.13010235
5033	75	0.75	60
4015	48	0.48	43.85377861
3478	79	0.79	62.72527132
3672	51	0.51	45.572996

General Linear Model with no arcsine transformation for dataset 5





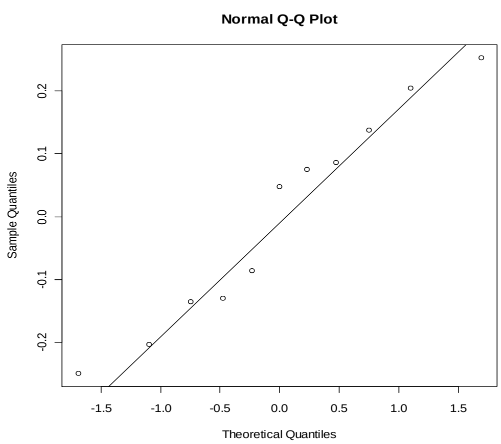
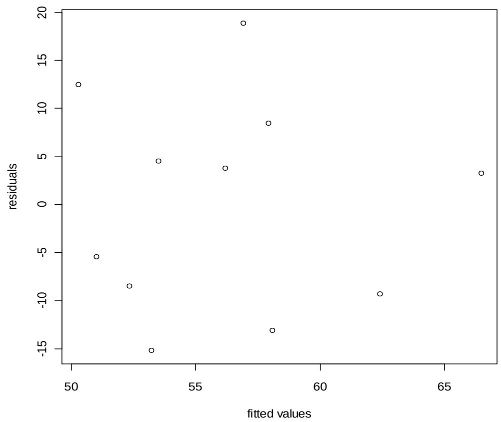
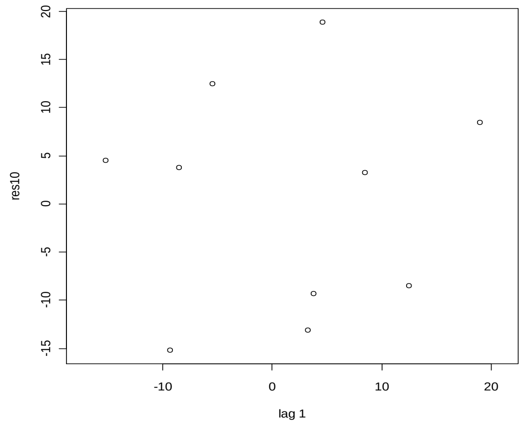


Table 13: ANOVA for General Linear Model with no arcsine transformation for dataset 5

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Number of Twigs	1	0.056	0.057	1.77	0.2155
Residuals	9	0.29	0.032		

General Linear Model with an arcsine transformation for dataset 5





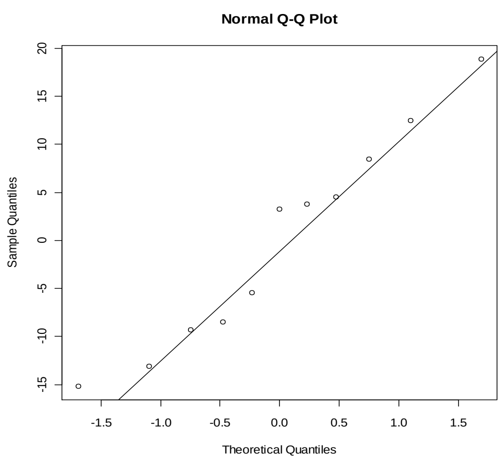
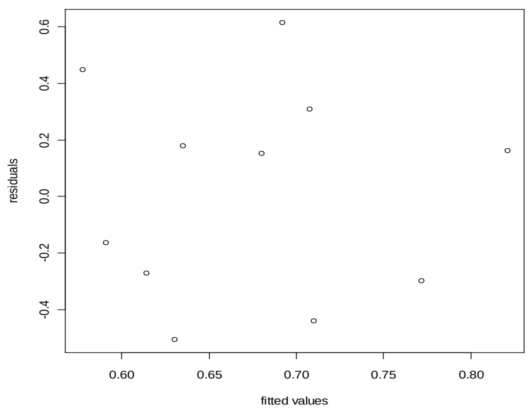
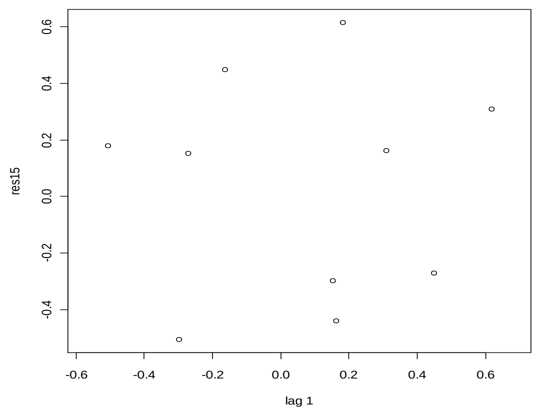


Table 14: ANOVA for General Linear Model with an arcsine transformation for dataset 5

Parameters	Degrees of Freedom	Sum of Squares	Mean Squares	F-value	P-value
Number of Twigs	1	243.91	243.91	1.801	0.2124
Residuals	9	1218.53	135.39		

Generalized Linear Model: Logistic Regression with binomial error structure for dataset 5





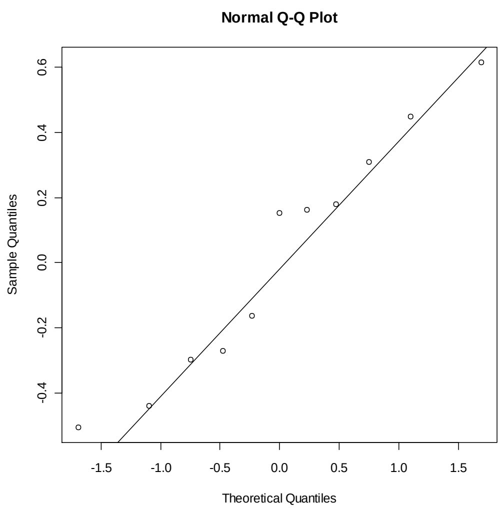


Table 15: ANODEV for Generalized Linear Model: Logistic Regression with binomial error structure for dataset 5

Parameters	Degrees of Freedom	Residual Deviance	P-value
Number of Twigs	9	1.40	0.6032
Null	10	1.67	

Table 16: Summary of influence on the p-value and final decision on the null hypothesis of arcsine transformed data and logistic regression compared to a general linear model.

DATASET 1	p-value		Change in Decision	
	Increase	Decrease	Yes	No
Arcsine	x			x
Logistic	x		x	
DATASET 2	p-value		Change in Decision	
	Increase	Decrease	Yes	No
Arcsine		x		x
Logistic	x			x
DATASET 3	p-value		Change in Decision	
	Increase	Decrease	Yes	No
Arcsine	x			x
Logistic	x			x
DATASET 4	p-value		Change in Decision	
	Increase	Decrease	Yes	No
Arcsine	x			x
Logistic	x		x	
DATASET 5	p-value		Change in Decision	
	Increase	Decrease	Yes	No
Arcsine		x		x
Logistic	x			x

Table 17: Summary of influence on the residual plots of arcsine transformed data and logistic regression compared to a general linear model.

DATASET 1	Normality			Homogeneity			Independence		
	better	worse	no change	better	worse	no change	better	worse	no change
Arcsine		x			x				x
Logistic		x			x		x		
DATASET 2	**Normality**			**Homogeneity**			**Independence**		
	better	worse	no change	better	worse	no change	better	worse	no change
Arcsine			x			x			x
Logistic			x	x			x		
DATASET 3	**Normality**			**Homogeneity**			**Independence**		
	better	worse	no change	better	worse	no change	better	worse	no change
Arcsine			x			x	x		
Logistic			x			x			x
DATASET 4	**Normality**			**Homogeneity**			**Independence**		
	better	worse	no change	better	worse	no change	better	worse	no change
Arcsine			x			x			x
Logistic			x			x			x
DATASET 5	**Normality**			**Homogeneity**			**Independence**		
	better	worse	no change	better	worse	no change	better	worse	no change
Arcsine			x			x			x
Logistic			x			x			x

R CODE EXAMPLE

```
#####LINEAR MODEL with no data transformation
##dataset 1
al<-read.delim("ds1.txt",header=T)
read.delim("ds1.txt",header=T)
GENERAL LINEAR MODEL: model1<-with(al,lm(pp~l))
GENERALIZED LINEAR MODEL: model<-with(al,glm(pp~l)
summary(model1)
anova(model1)
#RESIDUAL PLOTS
fit1<-fitted(model1)
res1<-resid(model1)
par(mfrow=c(1,3))
plot(fit1,res1,ylab="residuals",xlab="fitted values")
qqnorm(res1)
qqline(res1)
lag.plot(res1,diag=FALSE,do.lines=FALSE)
```

Appendix C: Holly Caravan

Data Set 1

Reference:

Srivastava, A. and Dunbar, R.I.M. (1996) The mating system of Hanuman langurs: a problem in optimal foraging. *Behav Ecol Sociobiol*. 39, 219-226

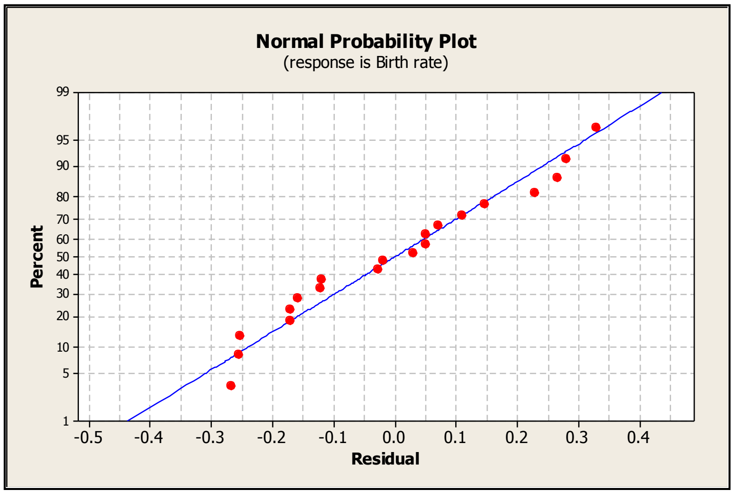
Raw Data:

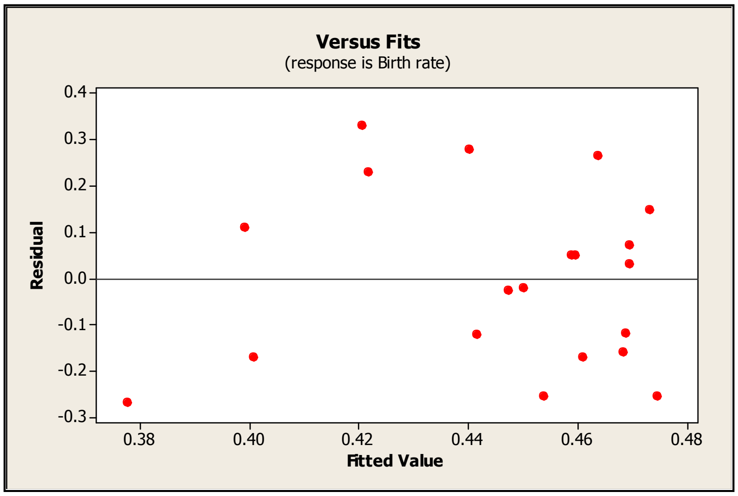
Temperature	**Birth Rate**	**Arcsine**
28	**0.22**	**27.9721**
27.7	**0.62**	**51.94327**
24	**0.2**	**26.56505**
21.4	**0.72**	**58.05194**
26.8	**0.31**	**33.83316**
26.9	**0.35**	**36.2712**
25	**0.51**	**45.573**
23.3	**0.43**	**40.97608**
27	**0.5**	**45**
17.9	**0.65**	**53.7288**
22.8	**0.42**	**40.39655**
21.7	**0.32**	**34.4499**
27	**0.54**	**47.29428**
25.1	**0.51**	**45.573**
25.9	**0.73**	**58.69355**
17.7	**0.75**	**60**
25.4	**0.29**	**32.58271**
9.5	**0.11**	**19.36971**
13.9	**28.65818**	**32.58271**

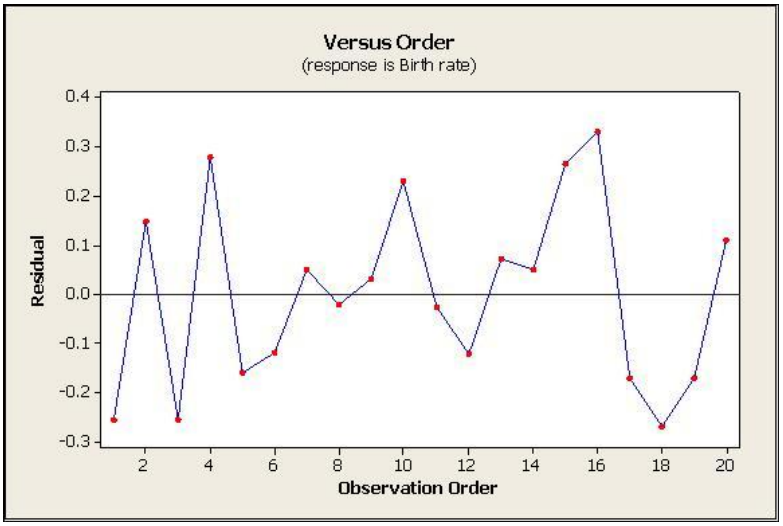
13.6	**0.51**	**45.573**

Non-Transformed Data:

LM: Linear Regression



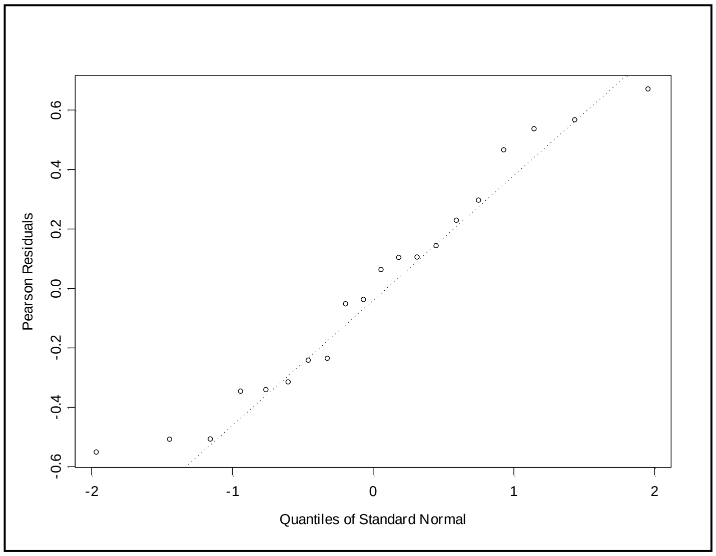


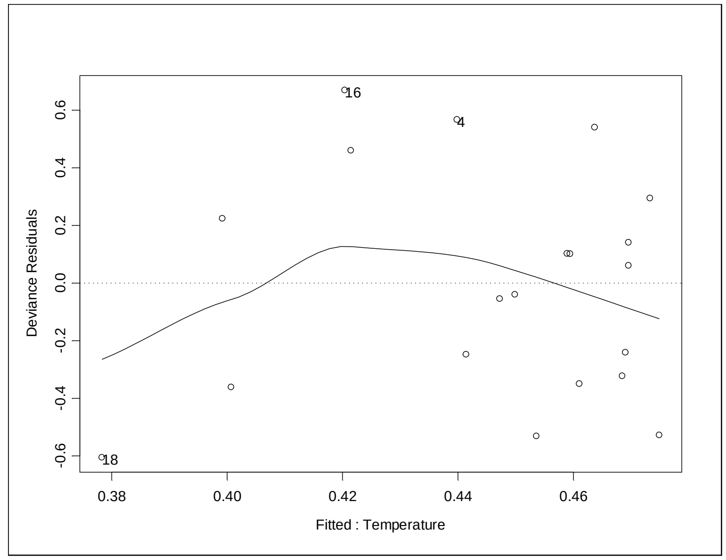


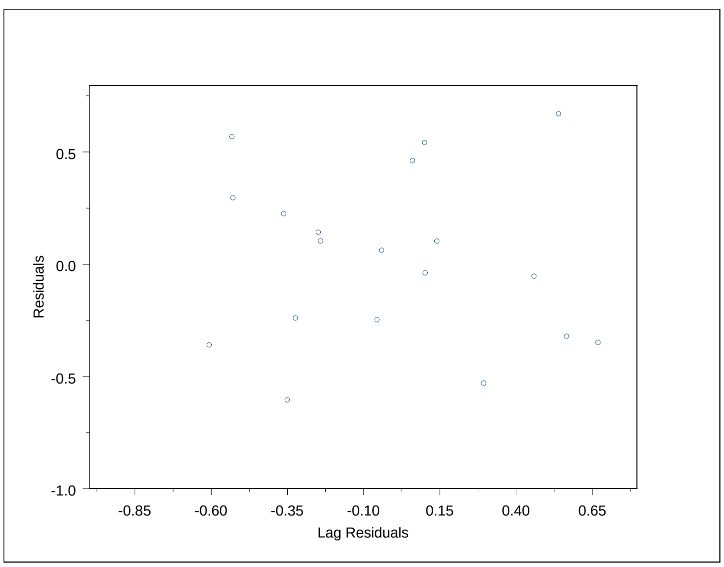
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	1	0.01482	0.01482	0.40	0.536
Residuals	18	0.66926	0.03718		

GLM: Logistic Regression





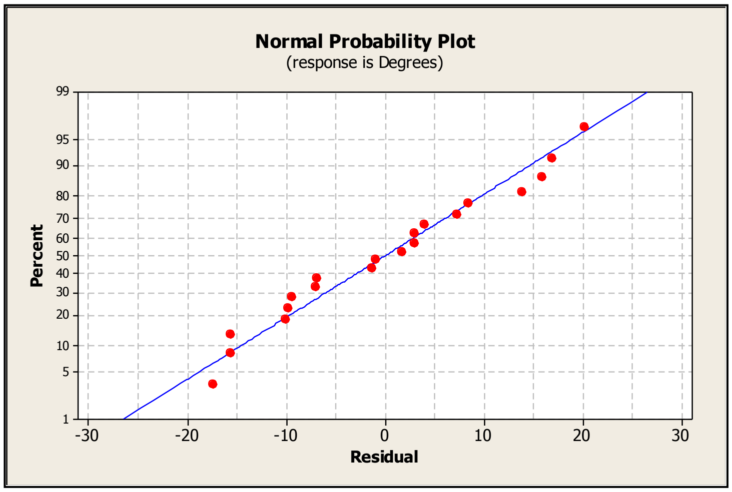


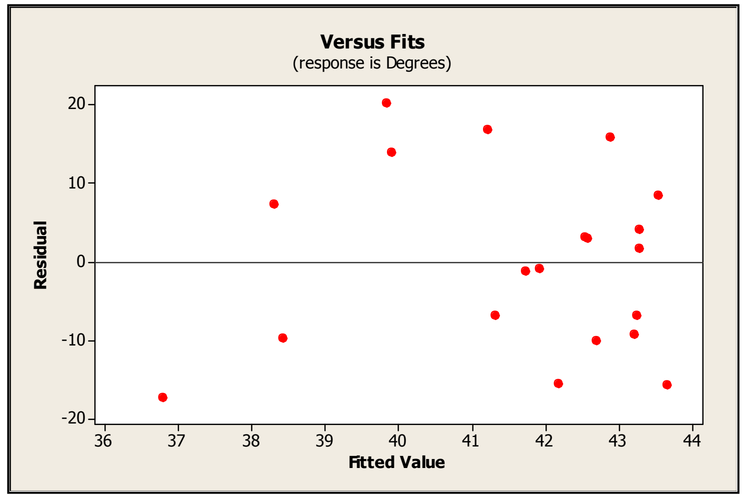
ANODEV Table

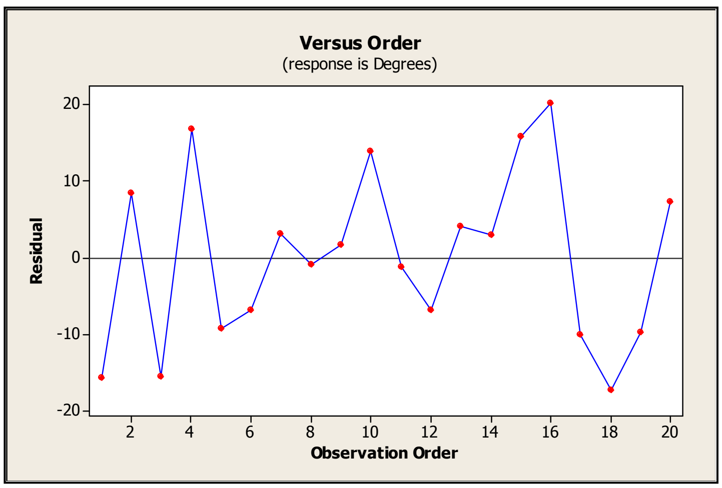
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			19	2.923616	
Temperature	1	0.0603345	18	2.863281	0.805968

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	1	74.2	74.2	0.54	0.471
Residuals	18	2469.1	137.2		

Data Set 2

Reference:

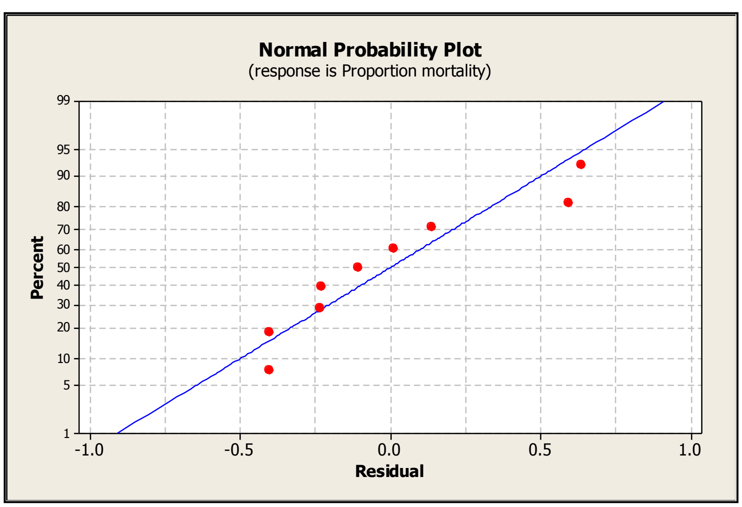
Kfir, R. and Luck, R.F. (1979) Effects of constant and variable temperature extremes on sex ratio and progeny production by *Aphytis melinus* and *A. lingnanensis* (Hymenoptera: Aphelinidae). *Ecol Entomol.* 4, 335-344

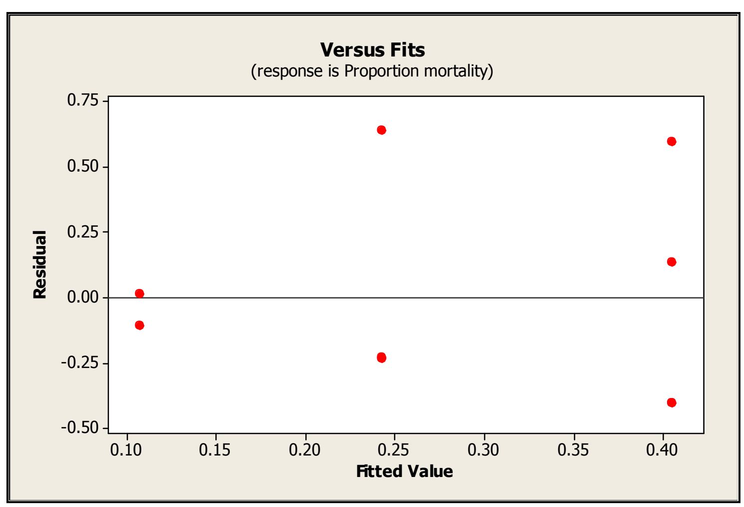
Raw Data:

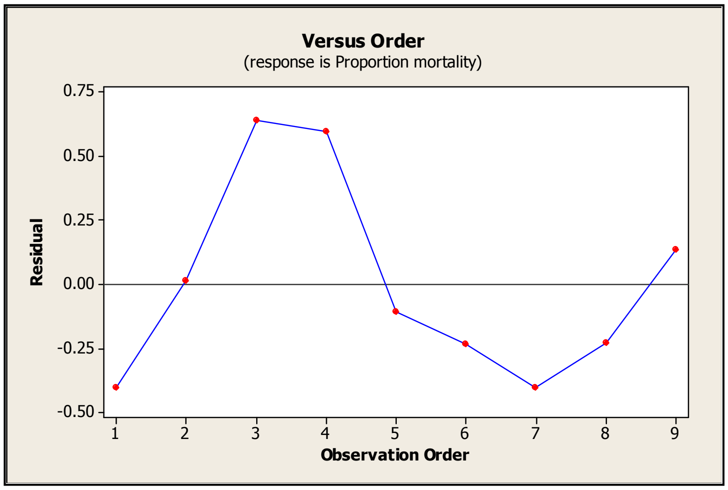
Time of Exposure (hours)	Percent mortality of *Aphytis melinus*	Arcsine
7	0	0
1.5	0.12	20.26790106
4	0.88	69.3209894
7	1	90
1.5	0	0
4	0.01	5.739170477
7	0	0
4	0.012	6.289059328
7	0.54	47.29428287

Non-Transformed Data:

LM: Linear Regression



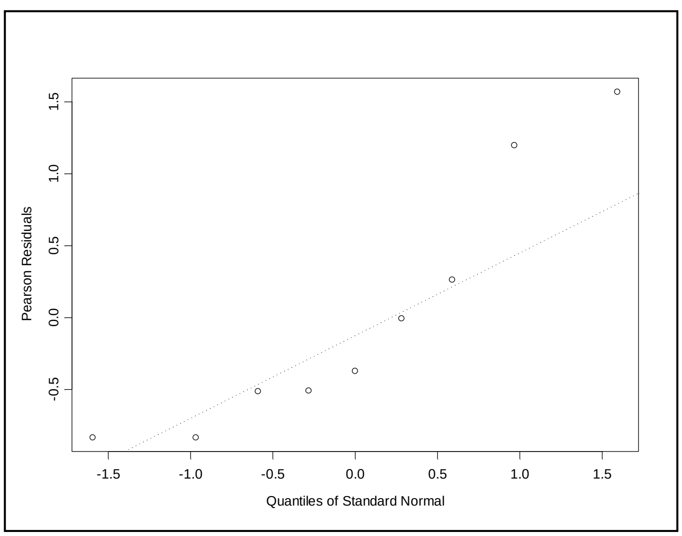


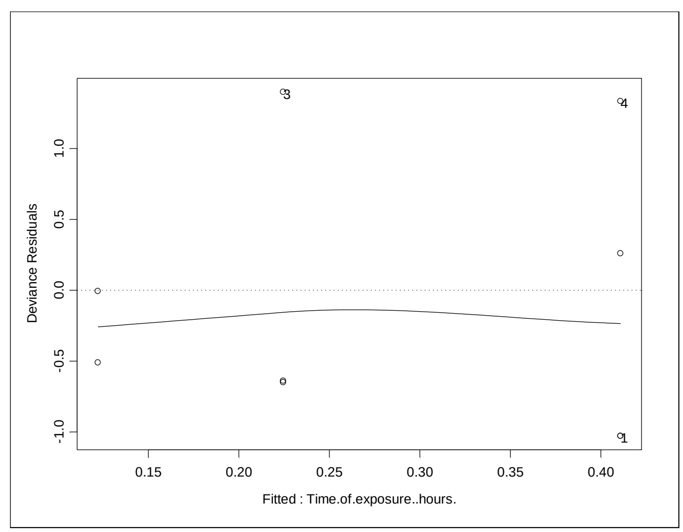


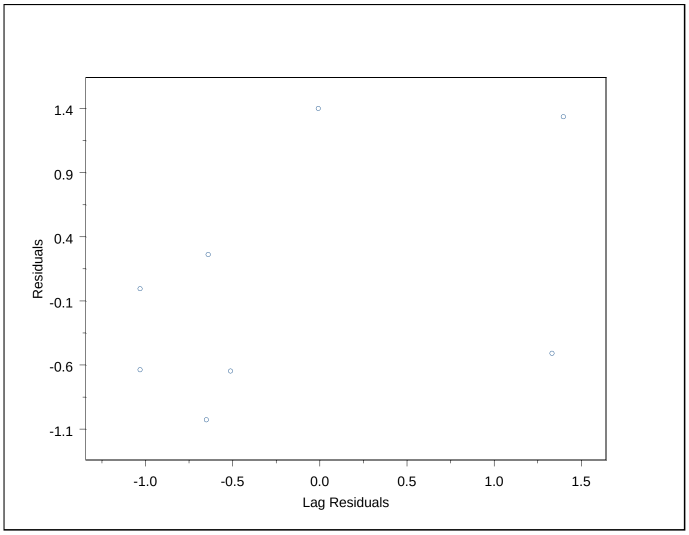
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Time Exposure	1	0.1258	0.1258	0.72	0.425
Residuals	7	1.2255	0.1751		

GLM: Logistic Regression





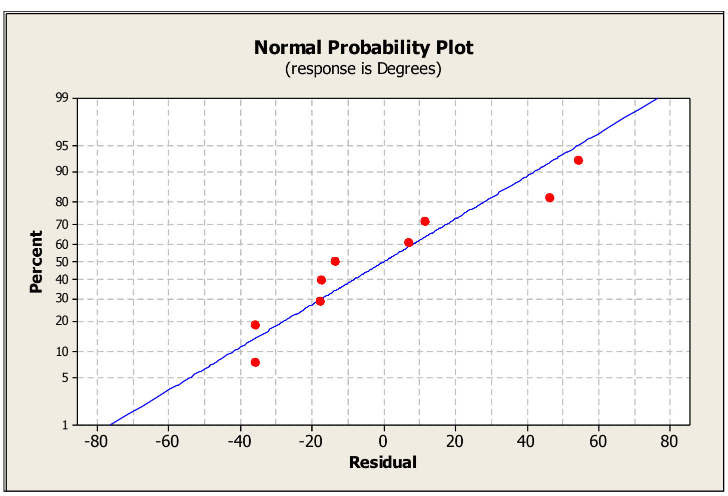


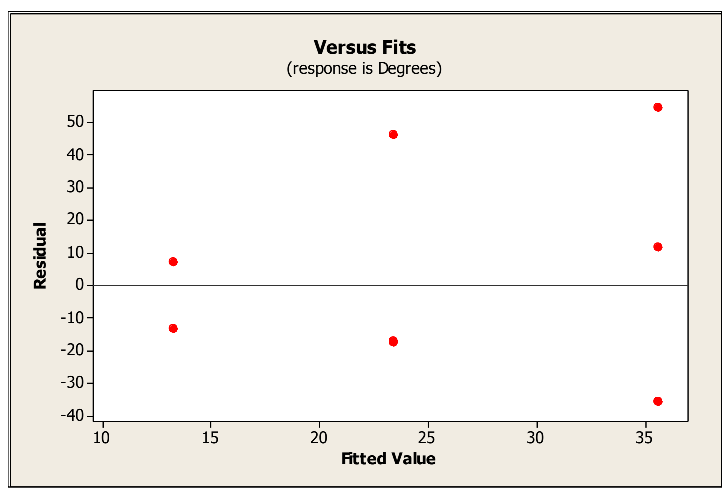
ANODEV Table

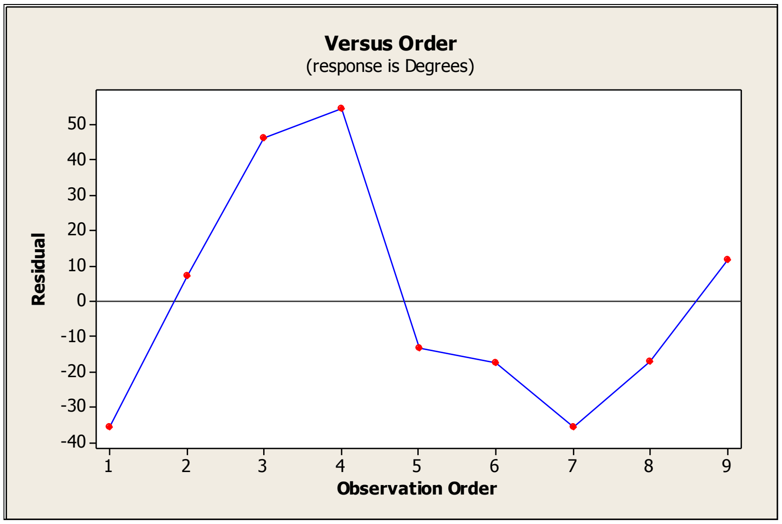
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			8	7.661931	
Time Exposure1		0.6535173	7	7.009414	0.419214115

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Time Exposure	1	712	712	0.58	0.472
Residuals	7	8607	1230		

Data Set 3

Reference:

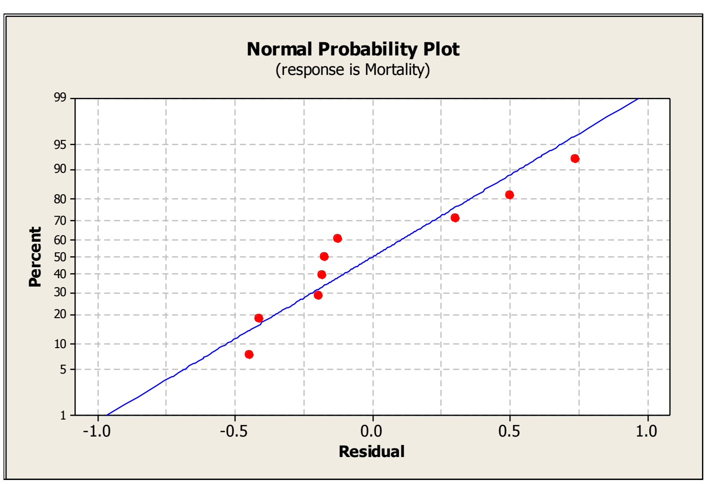
Kfir, R. and Luck, R.F. (1979) Effects of constant and variable temperature extremes on sex ratio and progeny production by *Aphytis melinus* and *A. lingnanensis* (Hymenoptera: Aphelinidae). *Ecol Entomol.* 4, 335-344

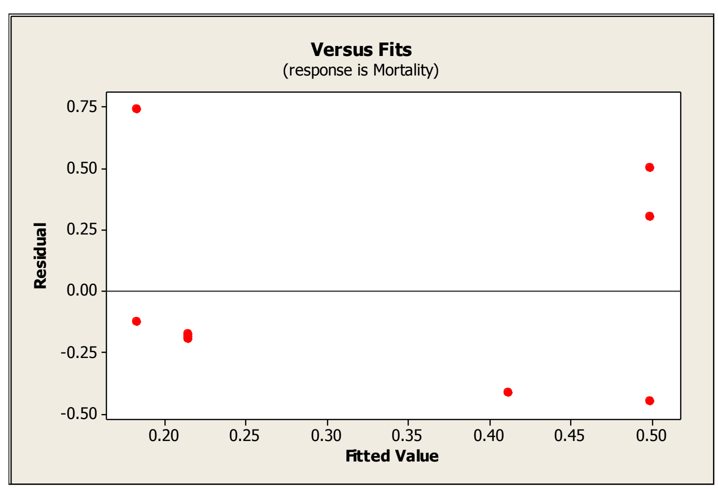
Raw Data:

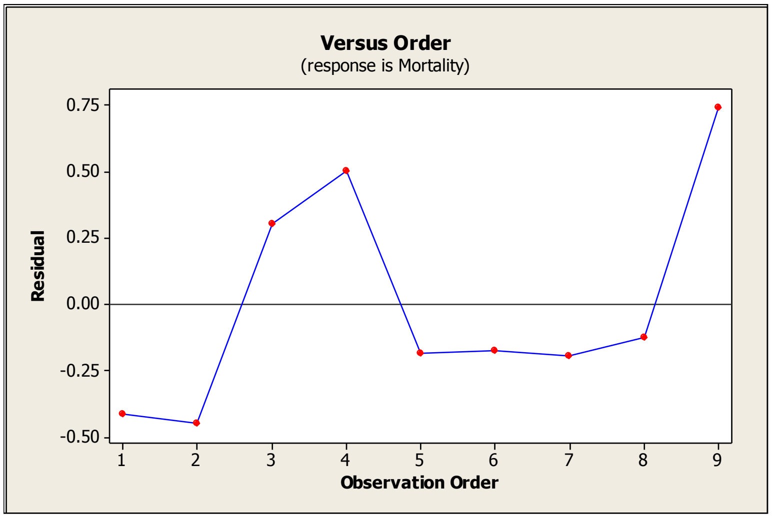
Temperature at Exposure	**Percent mortality of *Aphytis lingnanensis***	**Arcsine**
27	**0**	**0**
38	**0.05**	**12.92097**
38	**0.80**	**63.43495**
38	**1**	**90**
2	**0.03**	**9.974222**
2	**0.04**	**11.53696**
2	**0.02**	**8.130102**
-2	**0.056**	**13.6885**
-2	**0.922**	**73.78248**

Non-Transformed Data:

LM: Linear Regression



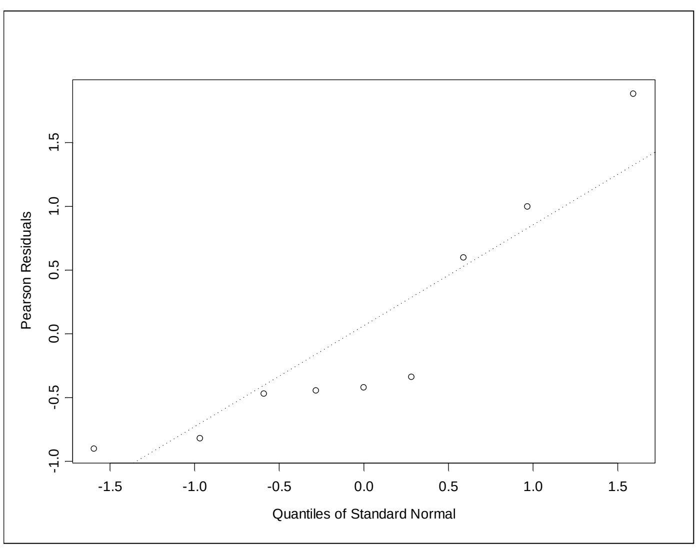


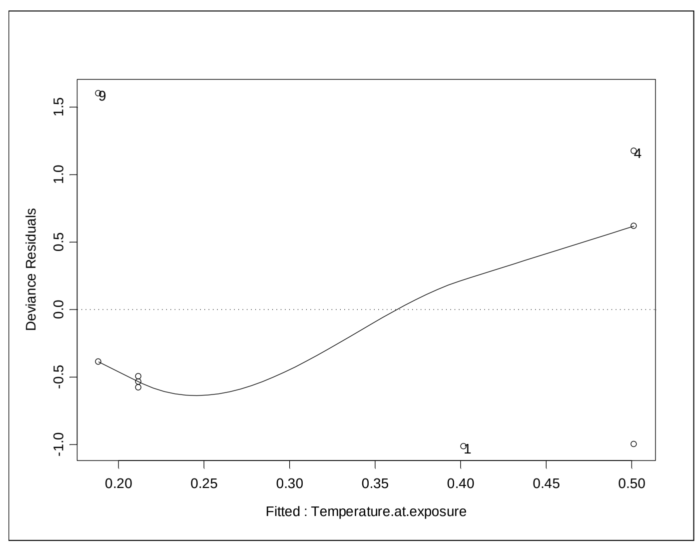


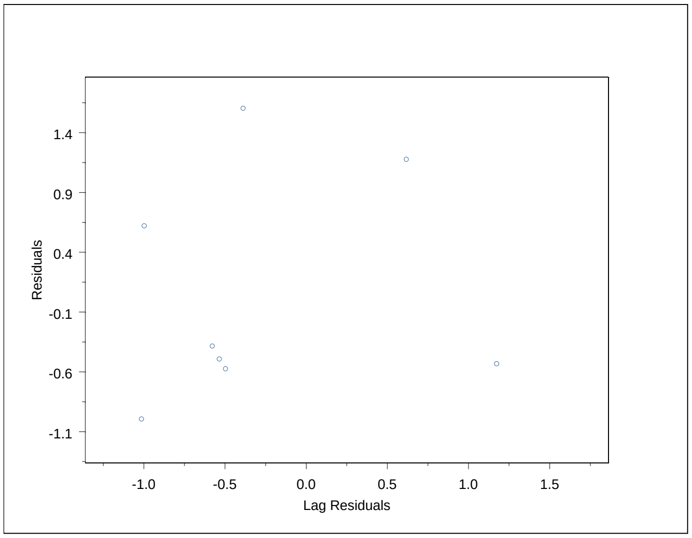
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	1	0.1748	0.1748	0.89	0.377
Residuals	7	1.3777	0.1968		

GLM: Logistic Regression





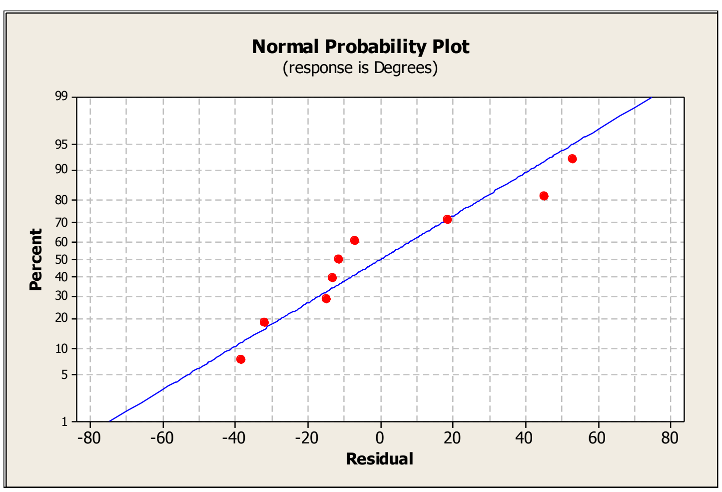


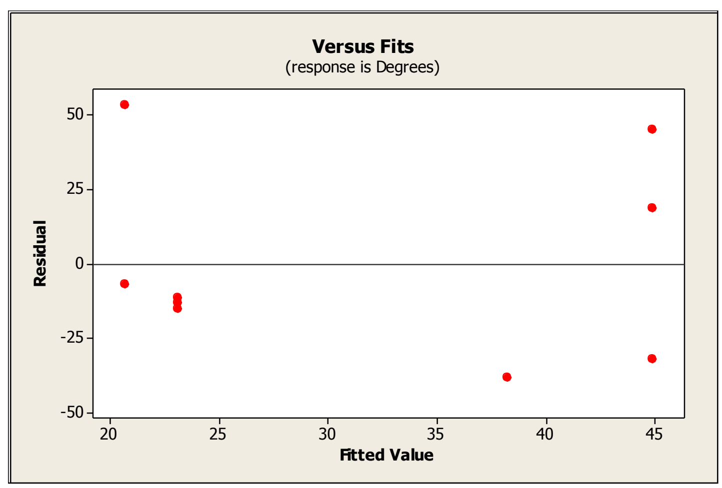
ANODEV Table

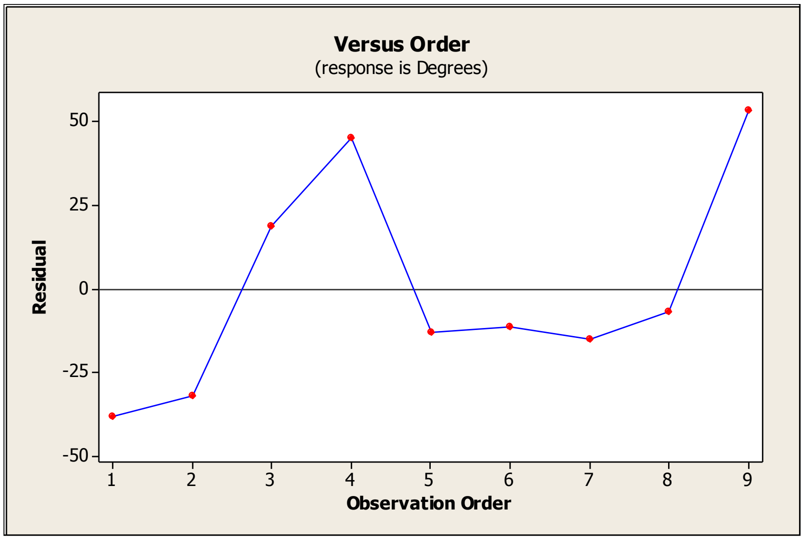
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			8	8.161568	
Temperature	1	0.8007553	7	7.360812	0.370868

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	1	1031	1031	0.87	0.381
Residuals	7	8261	1180		

Data Set 4

Reference:

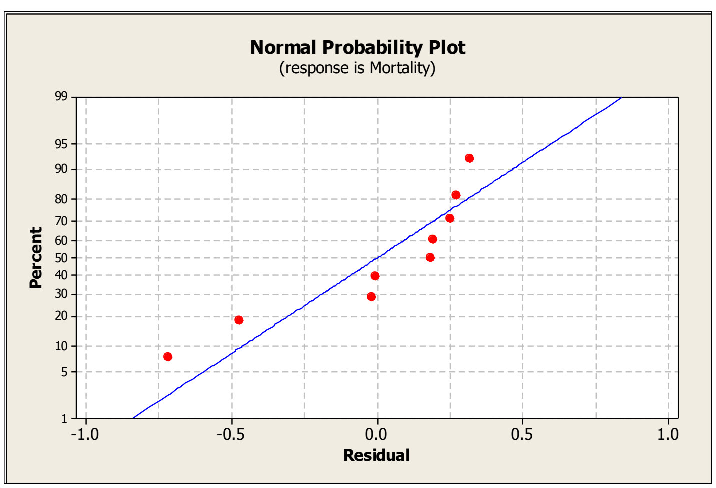
Smith, L.B. and Barker, P.S. (1991) Effect of temperature on the development, oviposition, and mortality of *Tribolium audax* Halstead and *Tribolium madens* (Charpentier) (Coleoptera: Tenebrionidae). *Can J Zool*. 69, 1189-1193

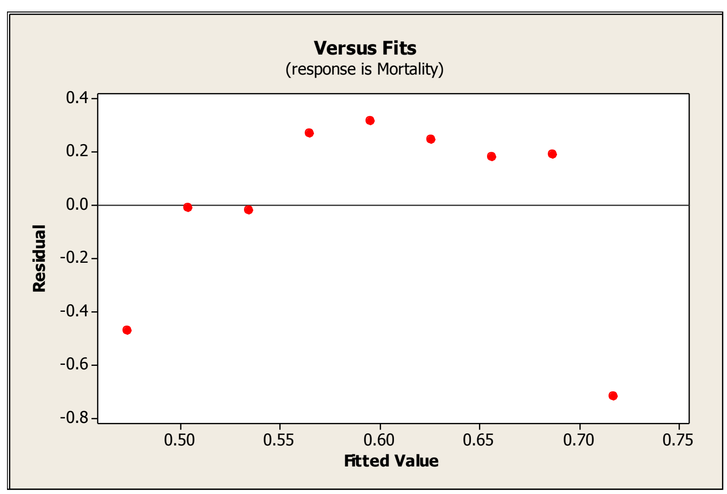
Raw Data:

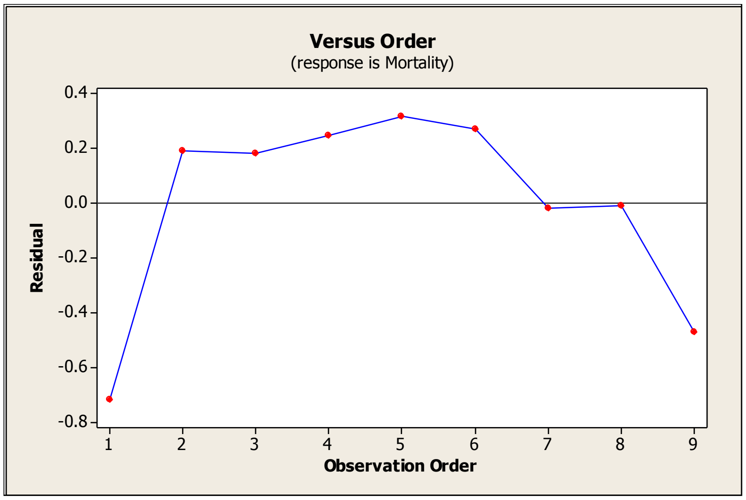
Temperature	Percent survival	Arcsine
37.5	0	0
35	0.878	69.55641
32.5	0.84	66.42182
30	0.874	69.20871
27.5	0.912	72.74362
25	0.838	66.26593
22.5	0.517	45.97422
20	0.495	44.71352
17.5	0	0

Non-Transformed Data:

LM: Linear Regression



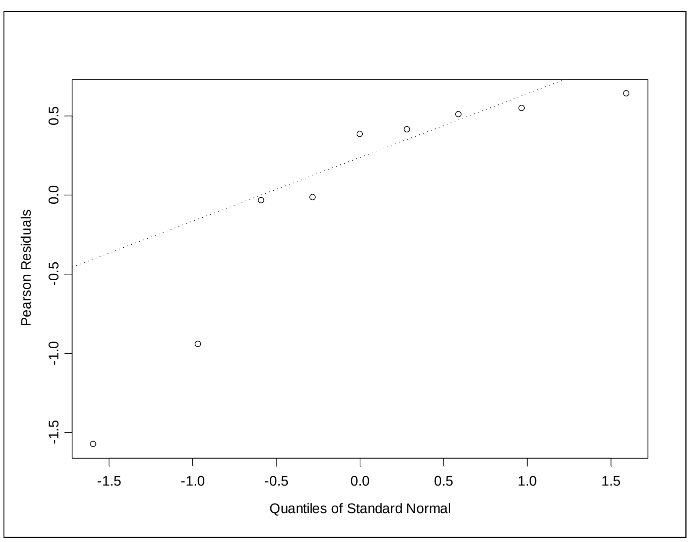


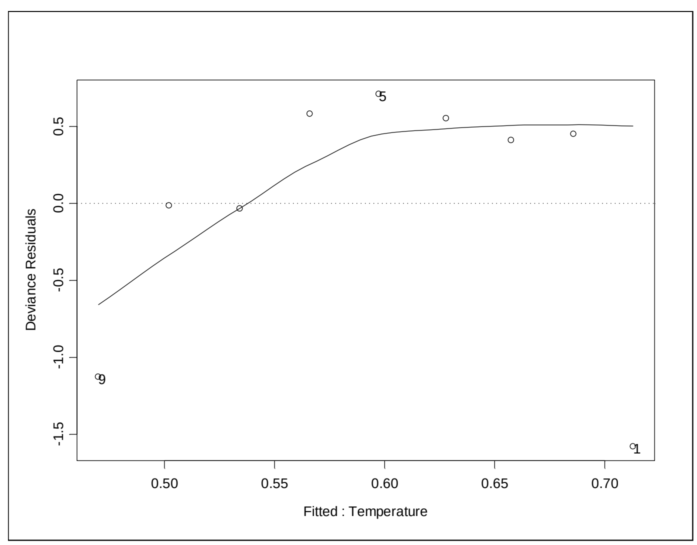


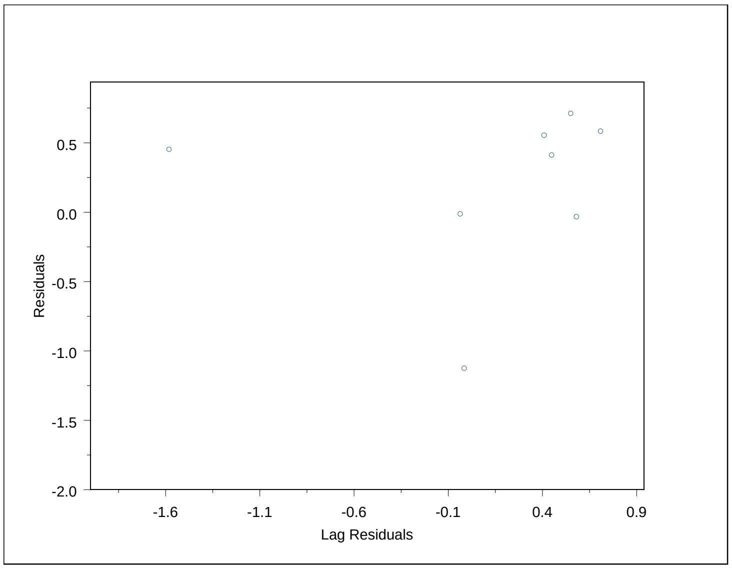
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	1	0.0559	0.0559	0.37	0.560
Residuals	7	1.0458	0.1494		

GLM: Logistic Regression





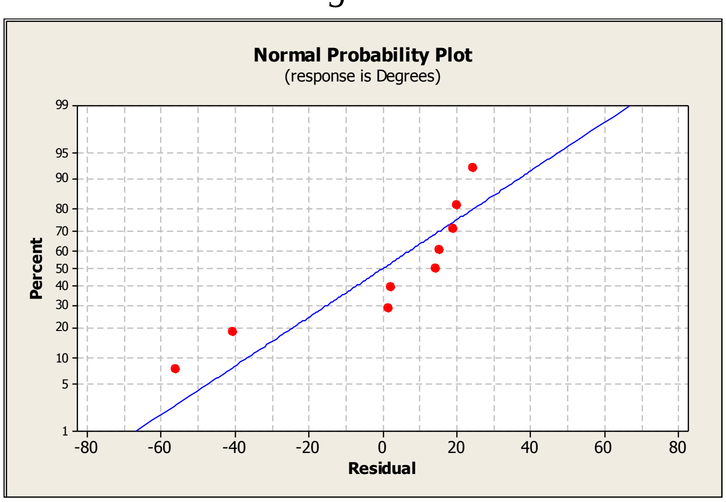


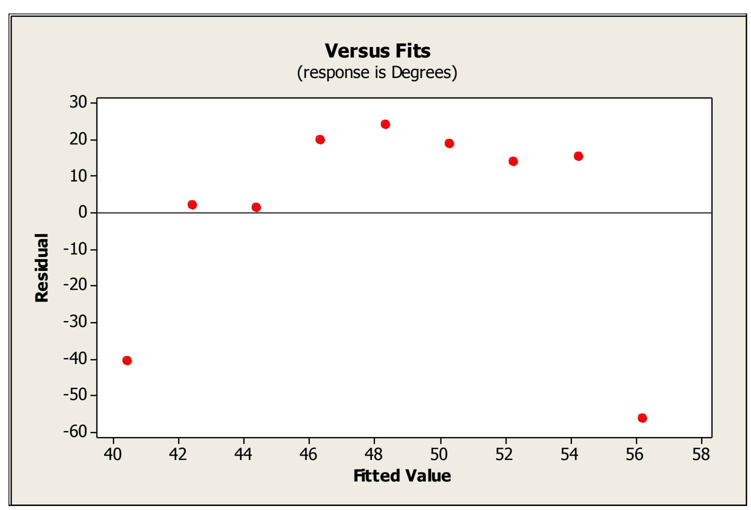
ANODEV Table

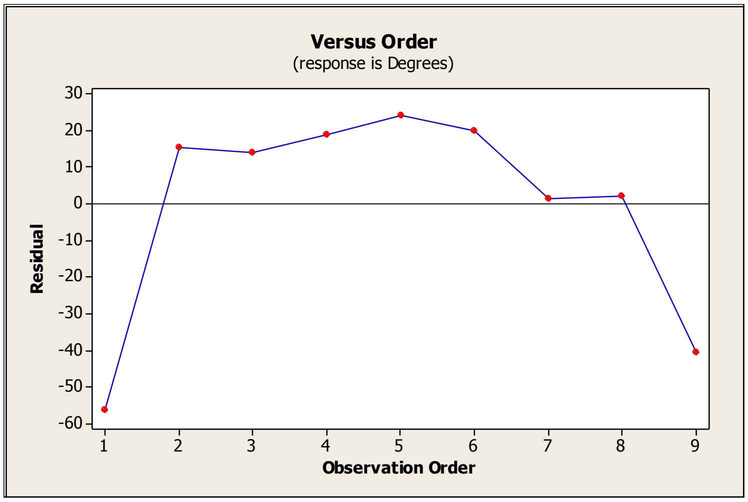
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			8	5.518942	
Temperature	1	0.2337401	7	5.285202	0.628764

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperature	1	233.5	233.5	0.25	0.634
Residuals	7	6588.2	941.2		

Data Set 5

Reference:

Barrett, R.W. and Chaing, H.C. (1966) The effects of thermal homogeneity in the environment on the activity and development of *Oncopeltus fasciatus* (Dallus). *Ecology*. 48, 590-598

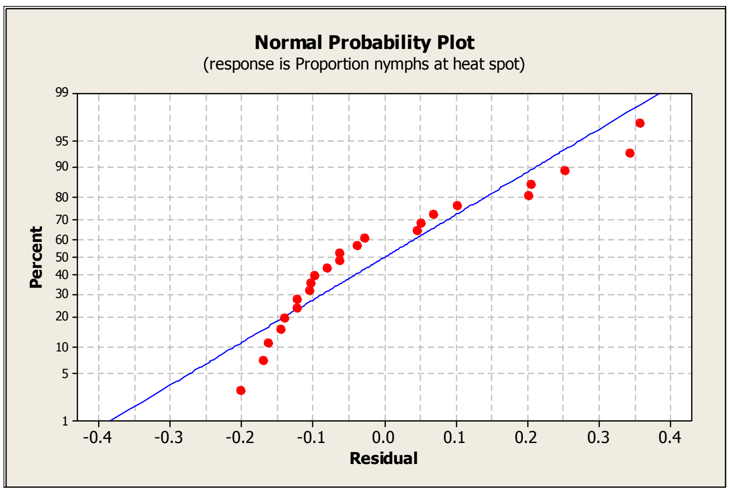
Raw Data:

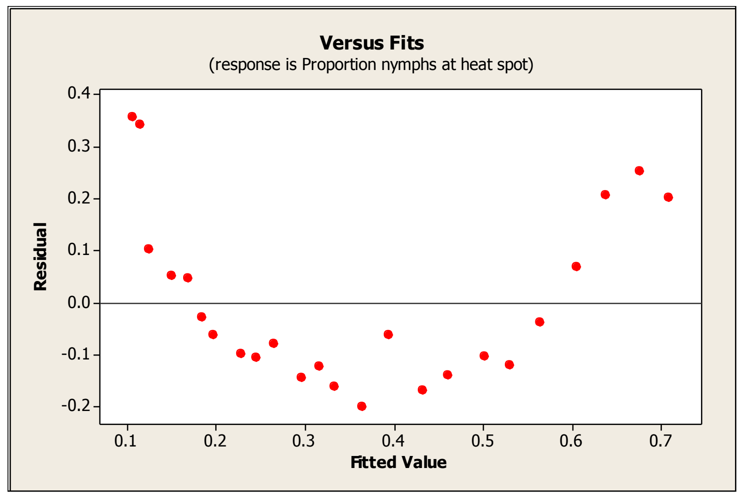
Hours since molting to 5th instar	**Proportion of nymphs at heat spot**	**Arcsine**
3	**0.463**	**42.8781**
6	**0.457**	**42.5332**
9.7	**0.226**	**28.3850**
18.7	**0.201**	**26.636**
25.3	**0.215**	**27.6248**
30.7	**0.157**	**23.3428**
35	**0.135**	**21.556**
46	**0.130**	**21.134**
52	**0.140**	**21.972**
58.7	**0.185**	**25.474**
70	**0.152**	**22.946**
76.7	**0.193**	**26.060**
82.7	**0.171**	**24.426**
93.7	**0.163**	**23.811**
104	**0.331**	**35.122**
117.3	**0.262**	**30.787**
127.3	**0.320**	**34.449**
142	**0.399**	**39.173**
151.7	**0.408**	**39.698**

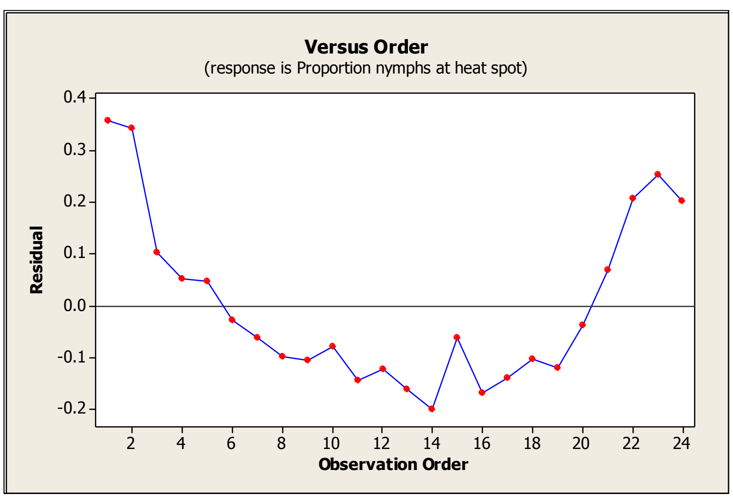
163.8	**0.526**	**46.490**
178.3	**0.675**	**55.243**
189.3	**0.843**	**66.657**
203	**0.928**	**74.435**
214.6	**0.911**	**72.642**

Non-Transformed Data:

LM: Linear Regression



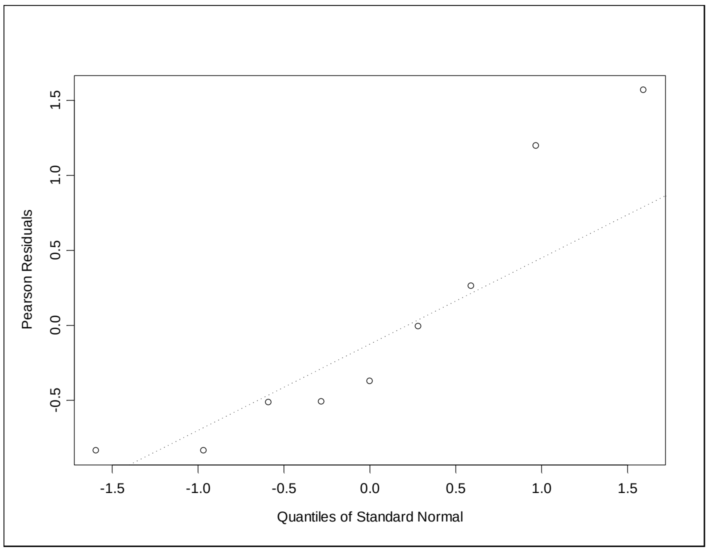


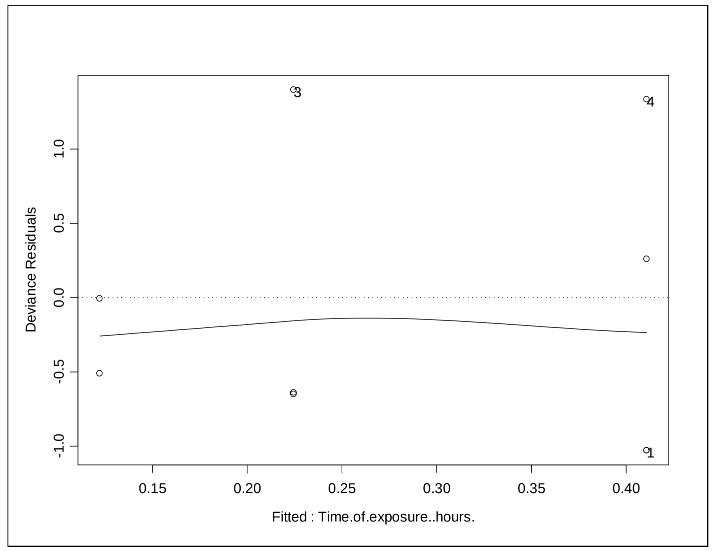


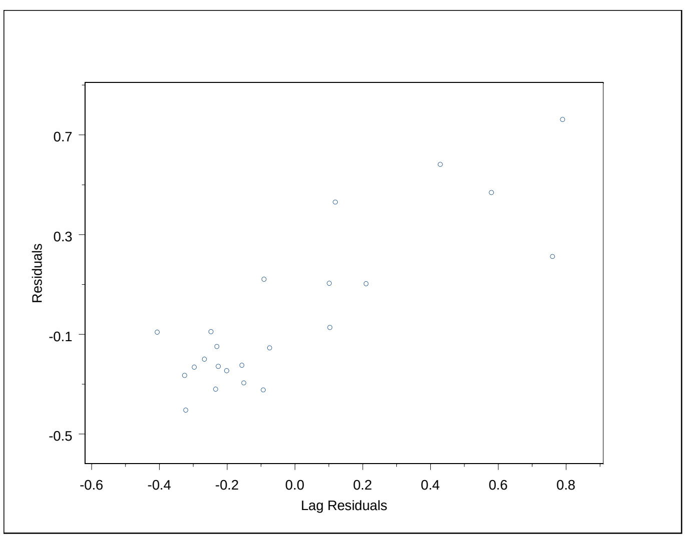
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hours	1	0.83045	0.83045	29.22	0.000
Residuals	22	0.62534	0.02842		

GLM: Logistic Regression





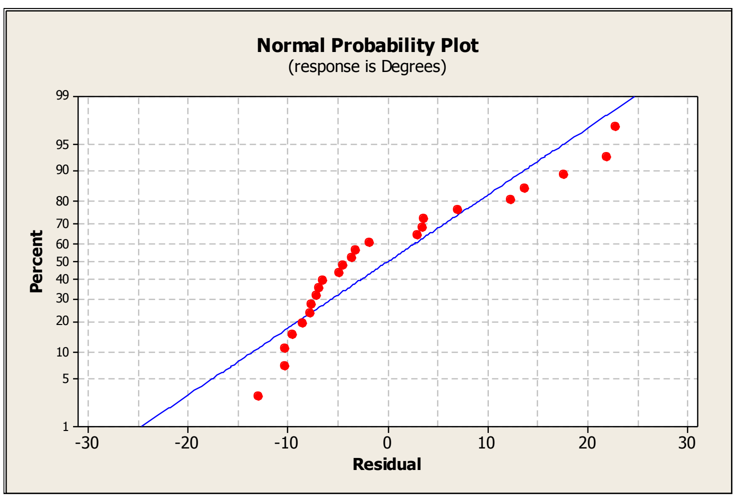


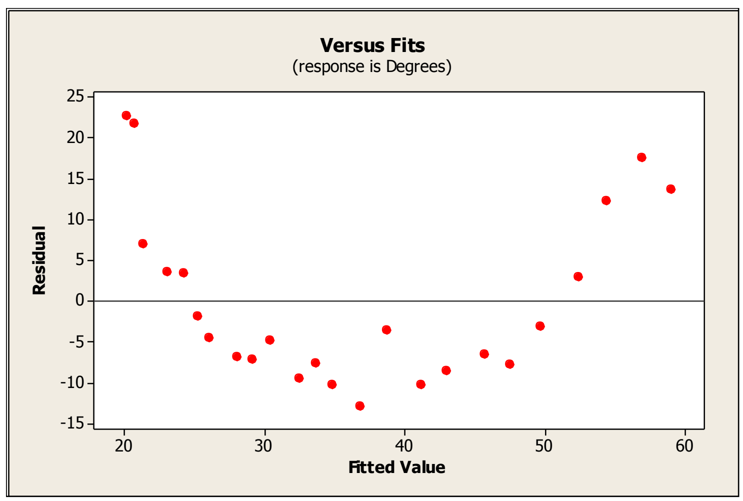
ANODEV Table

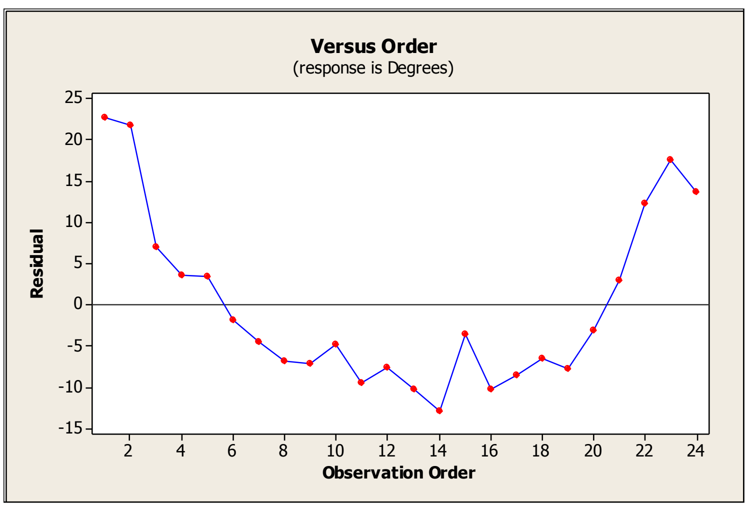
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			23	6.591917	
Hour	1	3.708801	22	2.883116	0.054126

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hour	1	3448.1	3448.1	29.21	0.000
Residuals	22	2597.3	118.1		

Summary Table:

	LM: prop vs. deg					LM: prop vs. GLM: prop				
	Plots			P-Value		Plots			P-Value	
Data Set	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	no diff	no diff	no diff	no	loss, 0.065	no diff	no diff	no diff	no	gain, 0.269968
2	no diff	worse	worse	no	gain, 0.047	no diff	no diff	no diff	no	loss (0.005785885
3	no diff	no diff	no diff	no	gain, .004	no diff	no diff	no diff	no	loss (0.006132)
4	no diff	no diff	no diff	no	gain, 0.074	no diff	worse	no diff	no	gain (0.068764)
5	no diff	no diff	no diff	no	no	better	no diff	no diff	yes - to not sign.	gain (0.054126)

Appendix D: Rebecca Doyle

Data set 1

Reference:

Flavell R.B., M.D. Bennett, J.B. Smith, D.B. Smith (1974) Genome size and proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet*. 12, 257–269

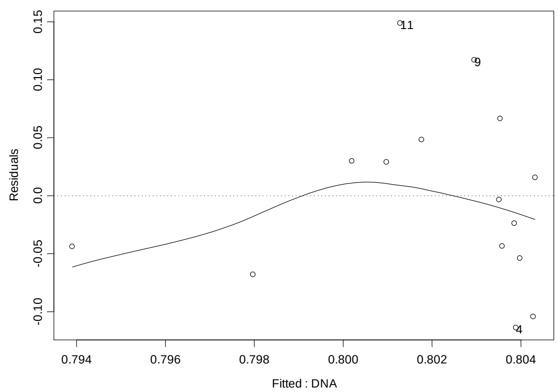
Raw Data:

Species	DNA content (pg)	%_rep_seq	Proportion	Arcsine	Degrees
Poa trivialis	6.9	82	0.82	0.961411	55.08479375
Tropaeolam majus	7.3	70	0.7	0.775397	44.427004
Pisum sativum	9.9	75	0.75	0.848062	48.59037789
Helianthus annuus	10.7	69	0.69	0.761489	43.63010887
Zea mays	11	78	0.78	0.894666	51.2605754
Hordeum vulgare	13.4	76	0.76	0.863313	49.46419789
Poa annua	13.8	87	0.87	1.055202	60.4586395
Triticum monococcum	14	80	0.8	0.927295	53.13010235
Secale cereale	18.9	92	0.92	1.16808	66.92608193
Vicia faba	29.3	85	0.85	1.015985	58.21166938
Allium cepa	33.5	95	0.95	1.253236	71.80512766
Triticum aestivum	36.2	83	0.83	0.979108	56.098738
Arena sativa	43	83	0.83	0.979108	56.098738
Tulipa kaufmanniana	62.5	73	0.73	0.818322	46.88639405
Hyacinth orientalis	98.1	75	0.75	0.848062	48.59037789

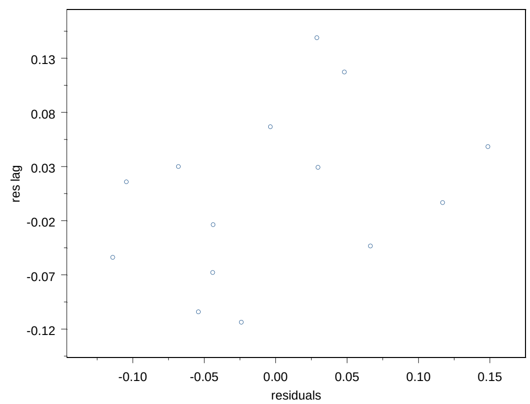
Non-Transformed Data:

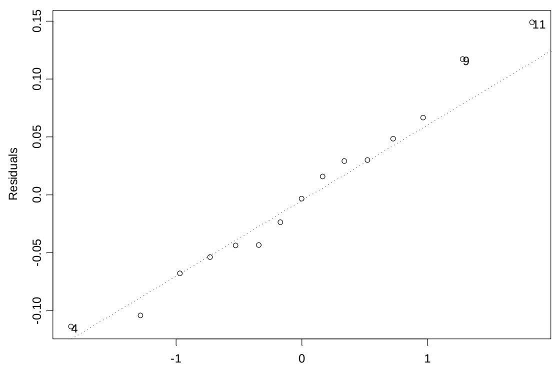
LM: Linear Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





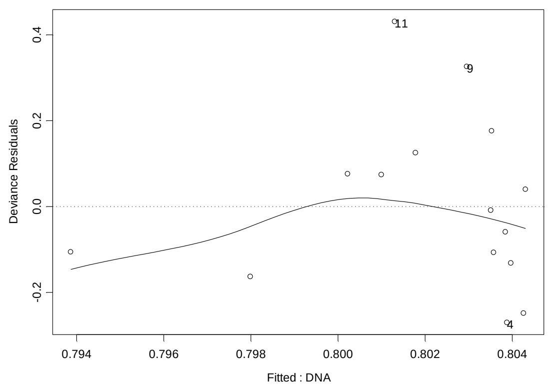
Normal QQ Plot

ANOVA Table:

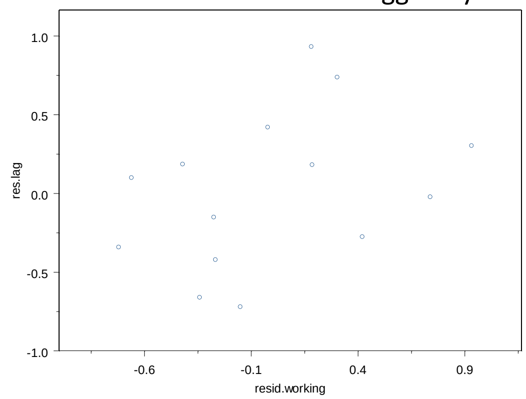
	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
DNA	1	0.000116	0.000116	0.018802	0.893036
Residuals	13	0.080324	0.006179		

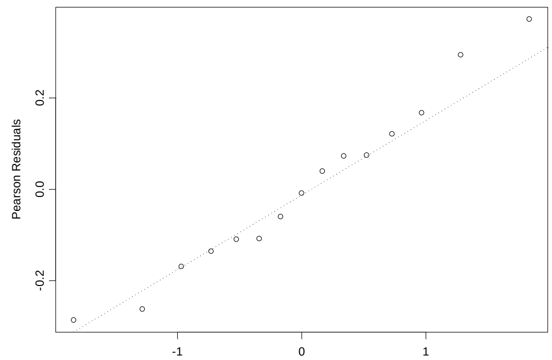
GLM: Logistic Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





Normal QQ Plot

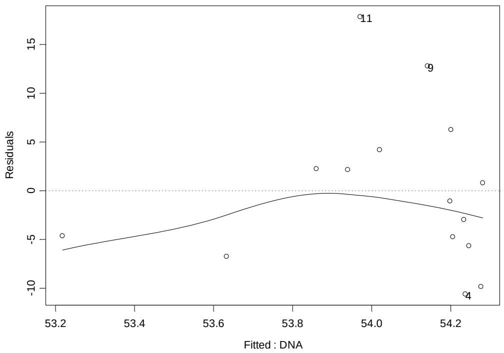
ANODEV Table

	Df	Deviance	Resid. Df	Resid. Dev	p-value
NULL			14	0.556858	
DNA	1	0.000727	13	0.55613	0.978485

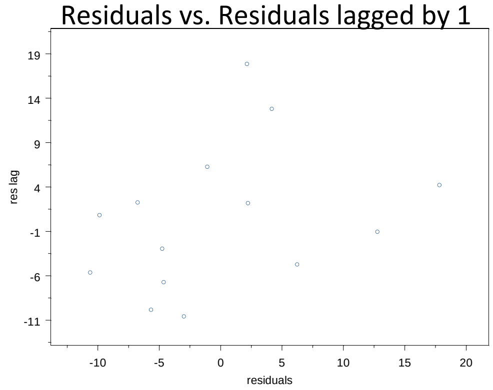
Arcsine Transformed Data:

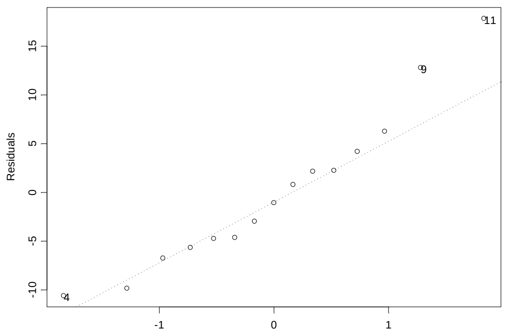
LM: Linear Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





Normal QQ Plot

ANOVA Table

	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
DNA	1	1.2106	1.21057	0.017694	0.896217
Residuals	13	889.4476	68.41905		

Data set 2

Reference:

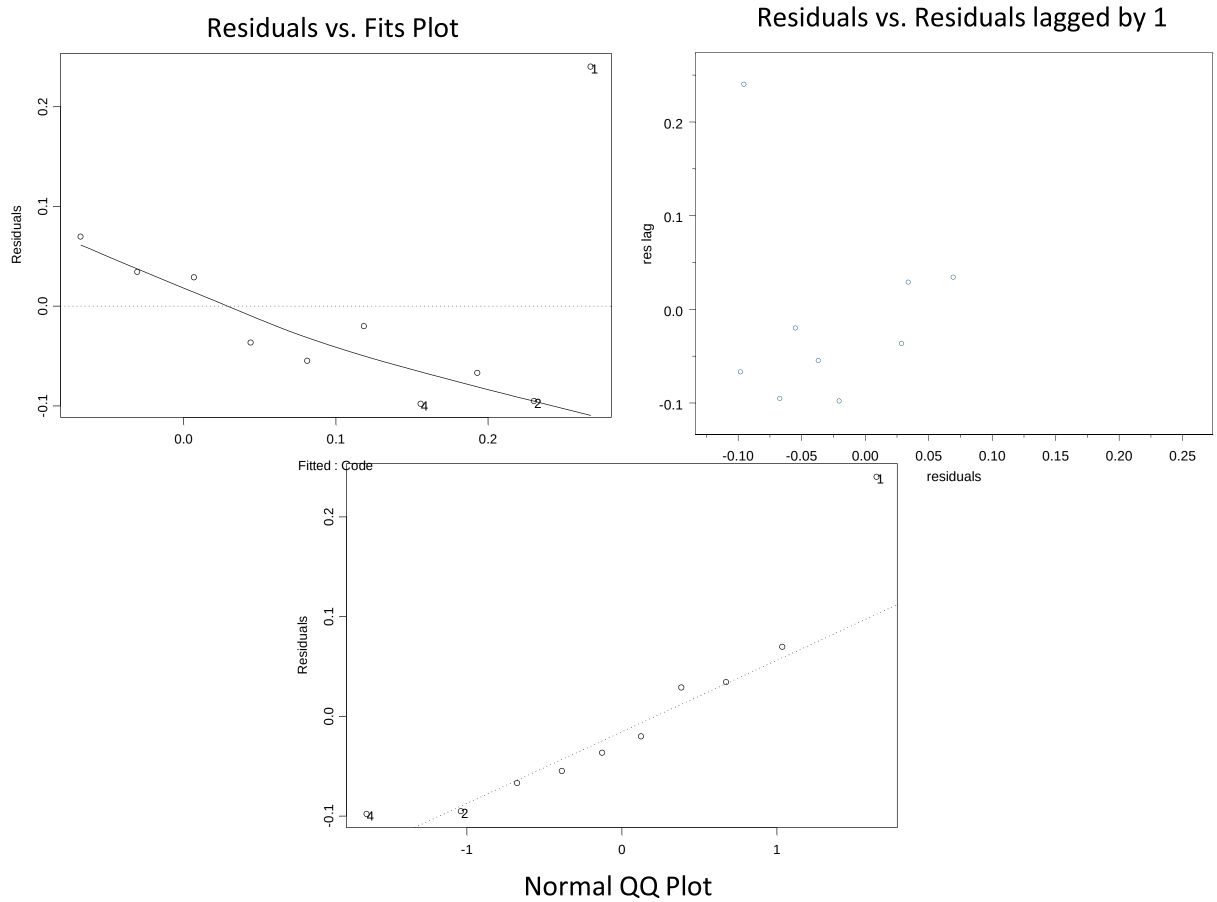
Bryan, W.B., L.W. Finger, F. Chayes (1969) Estimating proportions in petrographic mixing by least squares approximation. *Science* 163, 926 – 927

Raw Data:

Chemical	Code	%_Ox	P_Ox	Arcsine	Degrees
SiO2	1	50.75	0.5075	0.532280899	30.49744901
Al2O2 + Cr2O2	2	13.5	0.135	0.135413462	7.758619889
FeO	3	12.6	0.126	0.126335801	7.238508176
MgO	4	5.78	0.0578	0.057832232	3.313542808
CaO	5	9.84	0.0984	0.09855949	5.647042802
Na2O	6	2.65	0.0265	0.026503103	1.518515922
K2O	7	0.74	0.0074	0.007400068	0.423992638
TiO2	8	3.57	0.0357	0.035707588	2.045894064
P2O2	9	0.38	0.0038	0.003800009	0.217724486
MnO	10	0.19	0.0019	0.001900001	0.108862047

Non-Transformed Data:

LM: Linear Regression

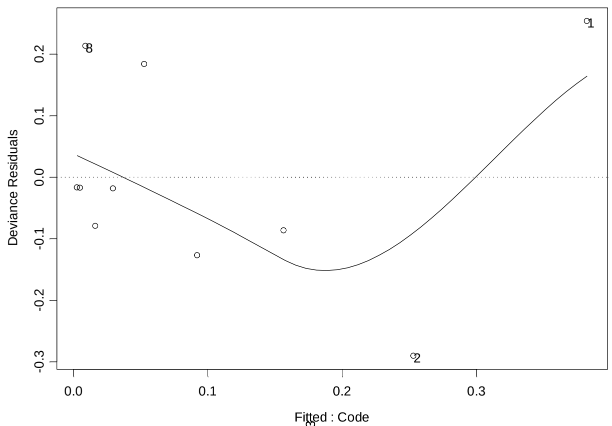


ANOVA Table:

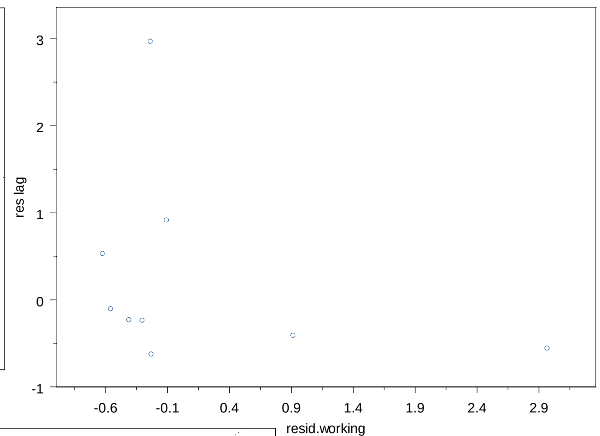
	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Code	1	0.114368	0.114368	9.905999	0.013652
Residuals	8	0.092362	0.011545		

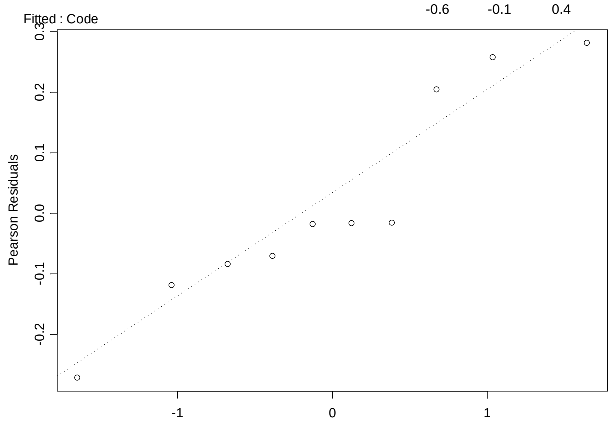
GLM: Logistic Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





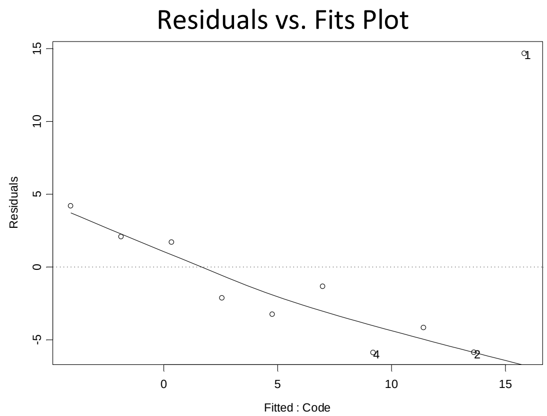
Normal QQ Plot

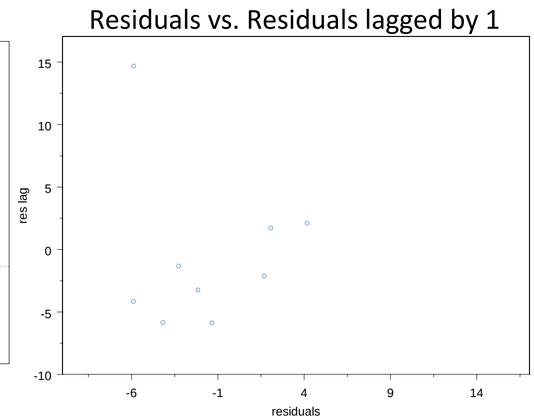
ANODEV Table

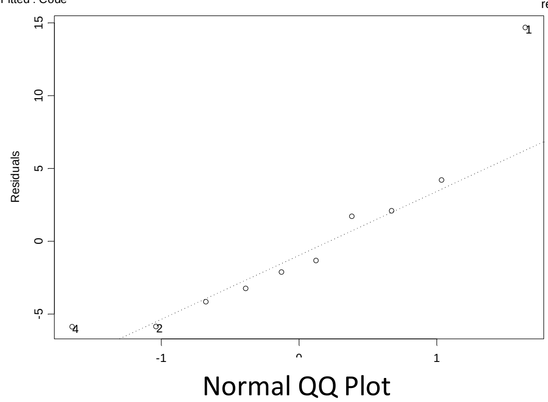
	Df	Deviance	Resid. Df	Resid. Dev	p-value
NULL			9	1.76408	
Code	1	1.5054	8	0.25868	0.219842

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Code	1	403.809	403.809	9.415987	0.015382
Residuals	8	343.0837	42.8855		

Data set 3

Reference:

Faraggi D., P. Izikson, B. Reiser (2003) Confidence intervals for the 50 per cent response dose. *Stat. Med.* 22, 1977– 88.

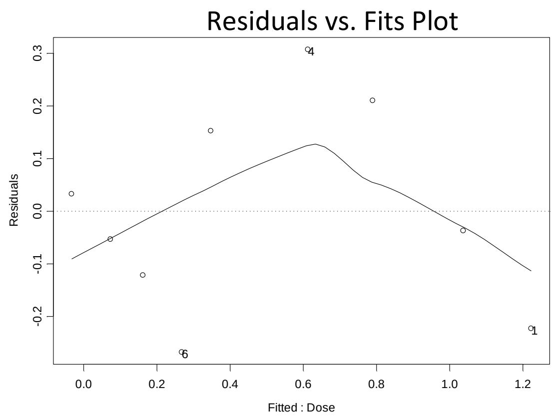
Data adapted from: Hewlett P.S., R.L. Plackett (1950) Statistical aspects of the independent joint action of poisons, particularly insecticides.II. Examination of data for agreement with the hypothesis. *Annals of Applied Biology* 37, 527–552

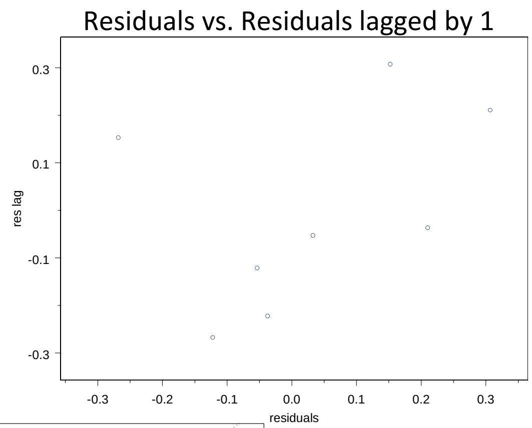
Raw Data:

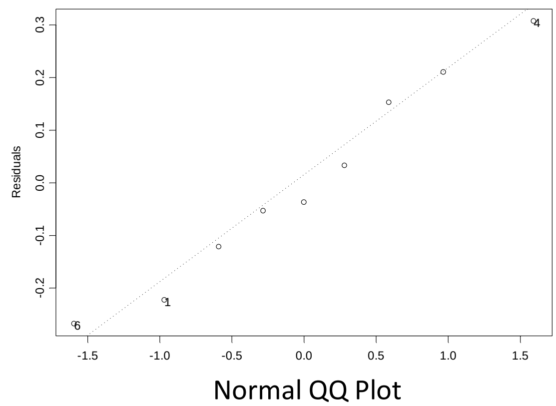
Dose	Nsub	Naff	Proportion	Arcsine	Degrees
1.909853	47	47	1	1.570796	90
1.699809	50	50	1	1.570796	90
1.420038	50	50	1	1.570796	90
1.220113	50	46	0.92	1.16808	66.92608
0.920026	50	25	0.5	0.523599	30
0.830042	50	0	0	0	0
0.710068	50	2	0.04	0.040011	2.292443
0.609958	50	1	0.02	0.020001	1.145992
0.490004	50	0	0	0	0

Non-Transformed Data:

LM: Linear Regression





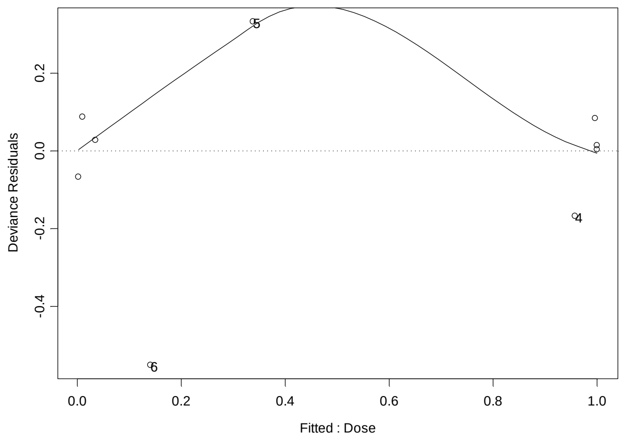


ANOVA Table:

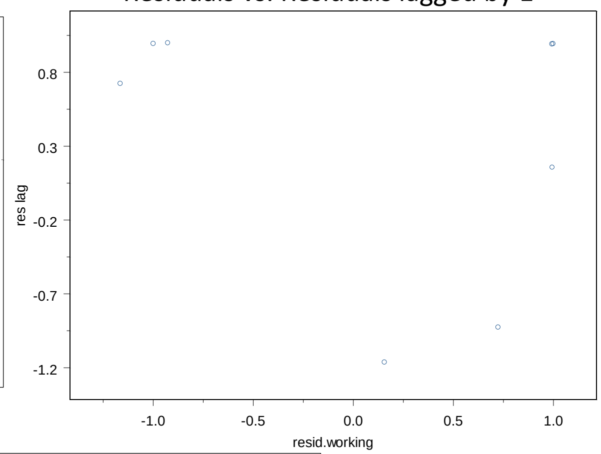
	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Dose	1	1.564967	1.564967	36.108	0.000537
Residuals	7	0.303389	0.043341		

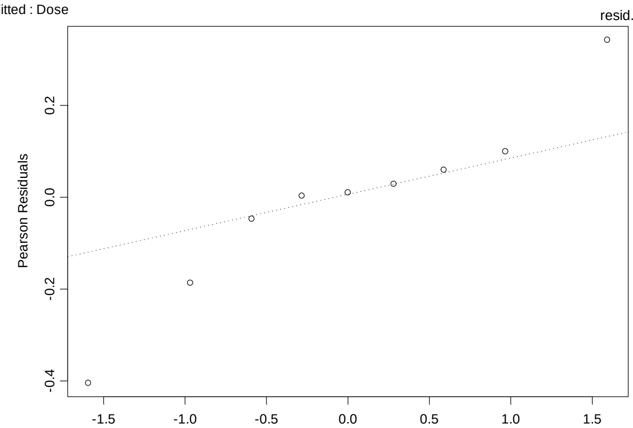
GLM: Logistic Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





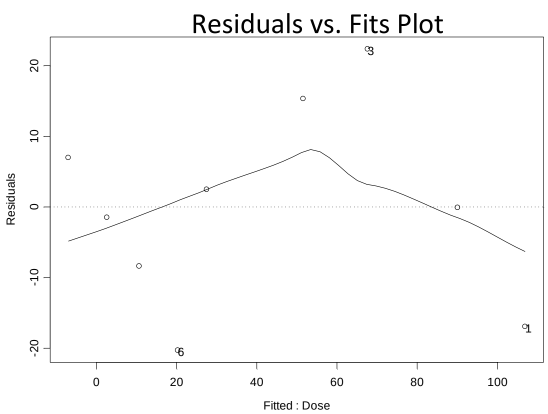
Normal QQ Plot

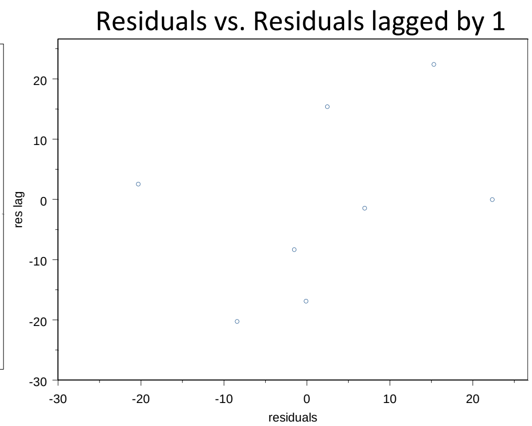
ANODEV Table

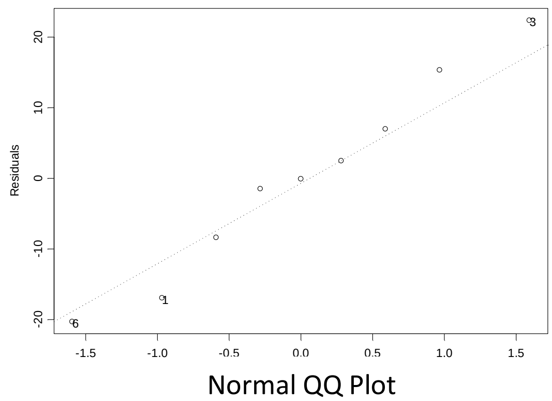
	Df	Deviance	Resid. Df	Resid. Dev	p-value
NULL			8	10.00067	
Dose	1	9.53812	7	0.46255	0.002012

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Dose	1	12883.14	12883.14	57.75616	0.000126
Residuals	7	1561.43	223.06		

Data set 4

Reference:

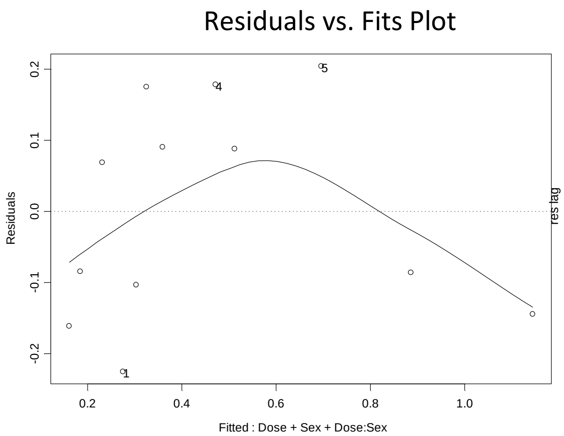
Venables, W.N., and B.D. Ripley (1997) Modern Applied Statistics with S-PLUS, second edition, Springer-Verlag: New York, 230 pgs

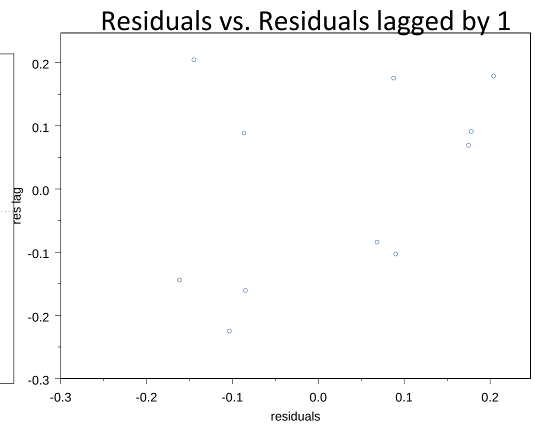
Raw Data:

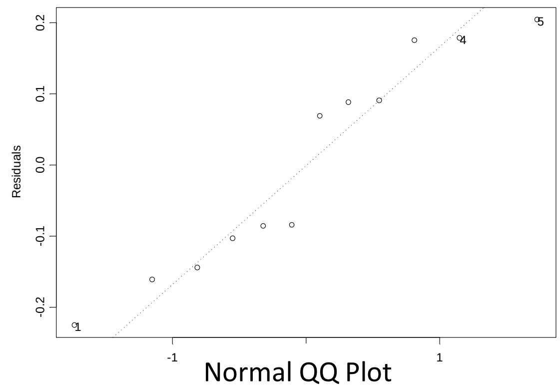
Killed	Number	Dose	Gender	Proportion	Arcsine	Degrees
1	20	1	M	0.05	0.050021	2.865983983
4	20	2	M	0.2	0.201358	11.53695903
9	20	4	M	0.45	0.466765	26.74368395
13	20	8	M	0.65	0.707584	40.54160187
18	20	16	M	0.9	1.11977	64.15806724
20	20	32	M	1	1.570796	90
0	20	1	F	0	0	0
2	20	2	F	0.1	0.100167	5.739170477
6	20	4	F	0.3	0.304693	17.45760312
10	20	8	F	0.5	0.523599	30
12	20	16	F	0.6	0.643501	36.86989765
16	20	32	F	0.8	0.927295	53.13010235

Non-Transformed Data:

LM: Linear Regression





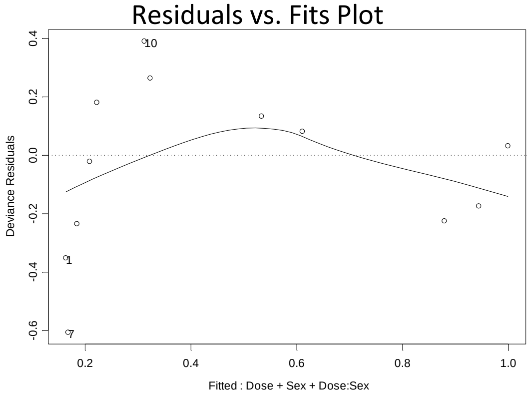


ANOVA Table:

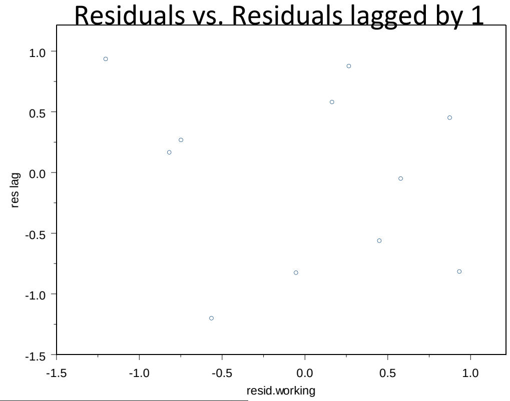
	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Dose	1	0.930086	0.930086	30.03815	0.000587
Sex	1	0.075208	0.075208	2.42894	0.157731
Dose:Sex	1	0.007623	0.007623	0.24619	0.633111
Residuals	8	0.247708	0.030964		

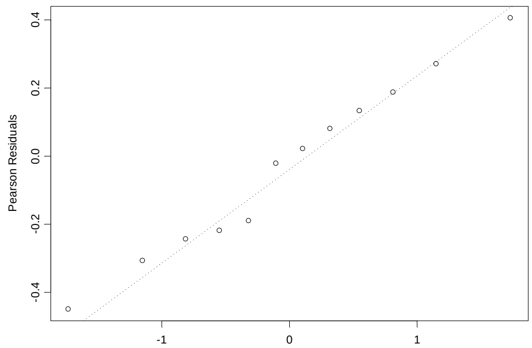
GLM: Logistic Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





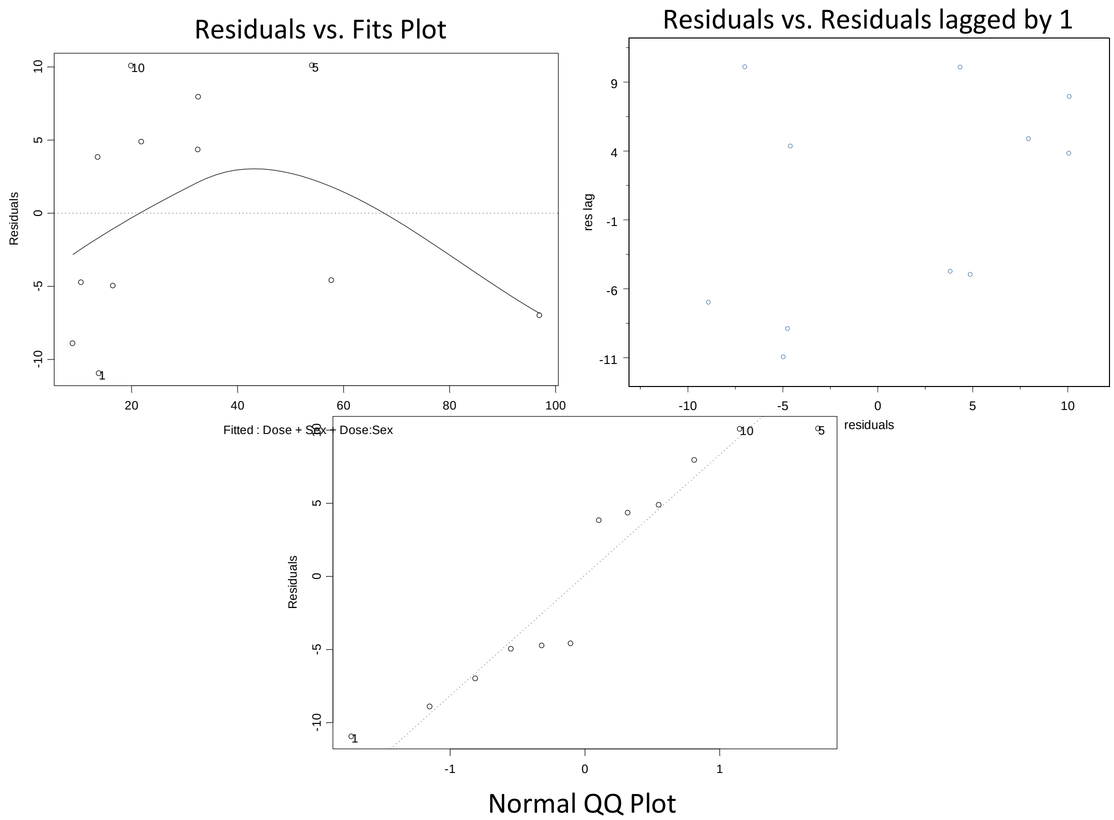
Normal QQ Plot

ANODEV Table

	Df	Deviance	Resid. Df	Resid. Dev	p-value
NULL			11	6.24378	
Dose	1	4.392522	10	1.851258	0.036097
Sex	1	0.452859	9	1.398399	0.50098
Dose:Sex	1	0.490189	8	0.90821	0.483843

Arcsine Transformed Data:

LM: Linear Regression



ANOVA Table

	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Dose	1	6376.198	6376.198	79.69458	1.97E-05
Sex	1	715.328	715.328	8.94071	0.017331
Dose:Sex	1	431.928	431.928	5.39857	0.048654
Residuals	8	640.063	80.008		

Data set 5

Reference:

Martin, P.A., T.J. Reimers, J.R. Lodge, and P.J. Dzuik (1974) The effect of ratios and numbers of spermatozoa mixed from two males on proportions of offspring. *J. Reprod. Fert*. 39, 251-258

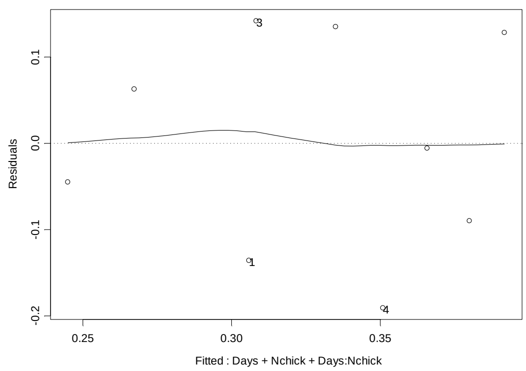
Raw Data:

Days egg collection after insemination	Nchick	%chick sired	Proportion	Arcsine	Degrees
2 to 6	92	17	0.17	0.17083	9.787819
2 to 6	190	36	0.36	0.368268	21.1002
2 to 6	96	45	0.45	0.466765	26.74368
7 to 11	77	16	0.16	0.160691	9.206896
7 to 11	32	33	0.33	0.336304	19.26878
7 to 11	99	52	0.52	0.546851	31.33225
12 to 16	20	20	0.2	0.201358	11.53696
12 to 16	35	29	0.29	0.294227	16.85796
12 to 16	30	47	0.47	0.489291	28.0343

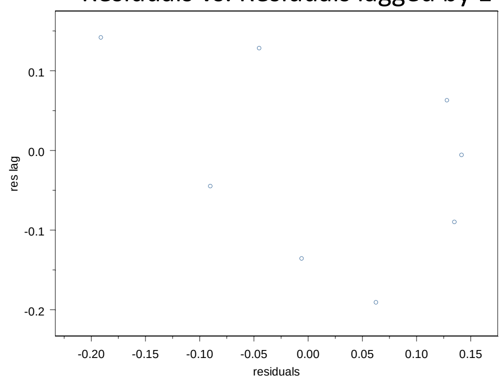
Non-Transformed Data:

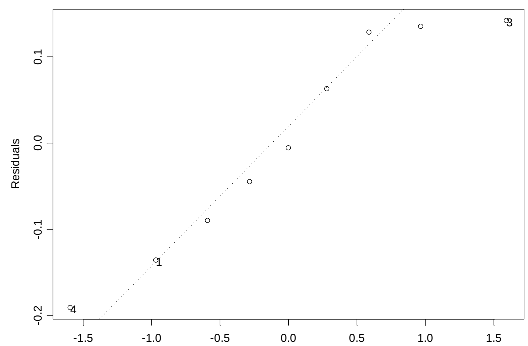
LM: Linear Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





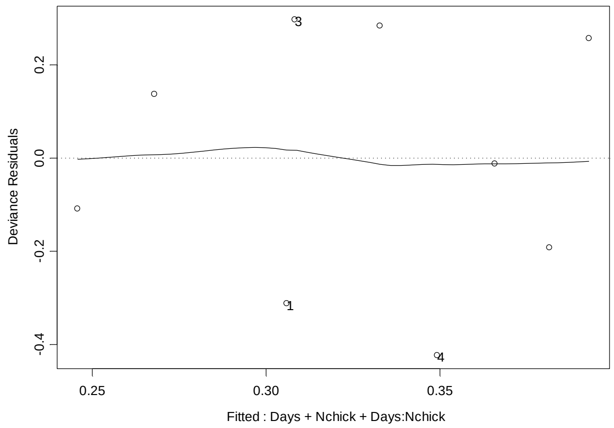
Normal QQ Plot

ANOVA Table:

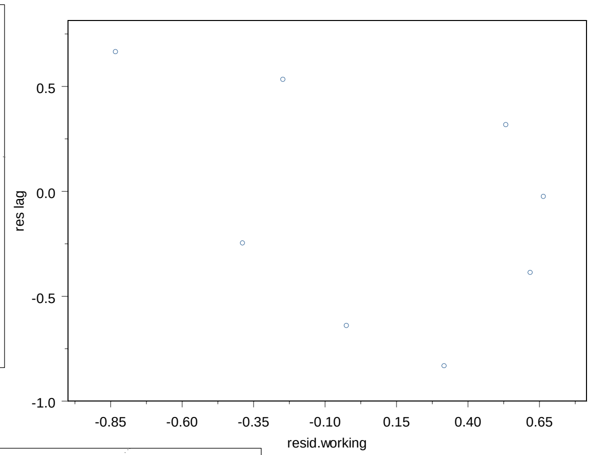
	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Days	2	0.000422	0.000211	0.005119	0.994903
Nchick	1	0.00972	0.009719	0.235652	0.660609
Days:Nchick	2	0.010079	0.005039	0.122178	0.889179
Residuals	3	0.123735	0.041245		

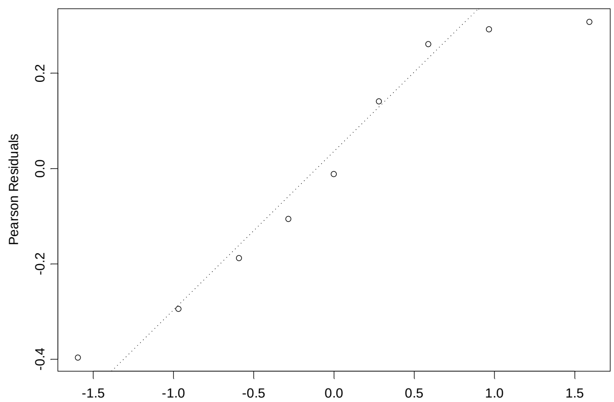
GLM: Logistic Regression

Residuals vs. Fits Plot



Residuals vs. Residuals lagged by 1





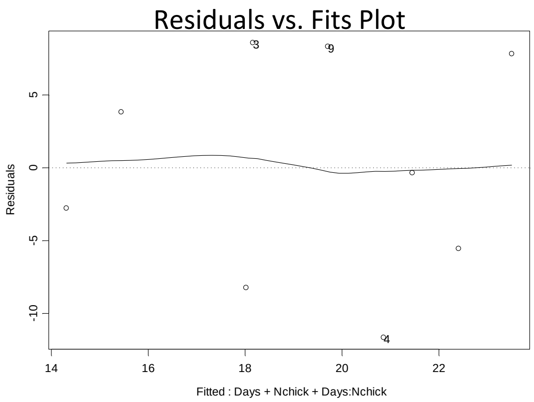
Normal QQ Plot

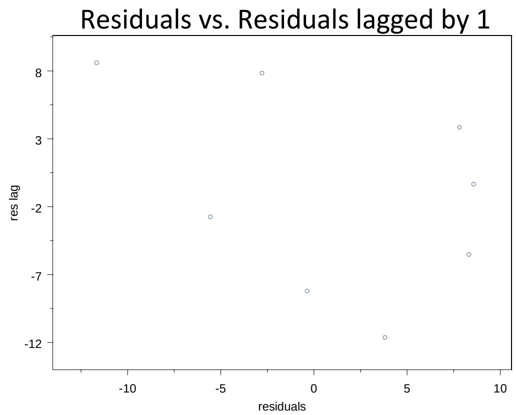
ANODEV Table

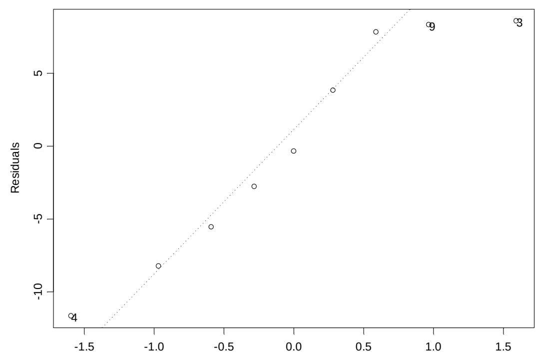
	Df	Deviance	Resid. Df	Resid. Dev	p-value
NULL			8	0.671609	
Days	2	0.001915	6	0.669694	0.999043
Nchick	1	0.043652	5	0.626042	0.834502
Days:Nchick	2	0.047432	3	0.57861	0.976563

Arcsine Transformed Data:

LM: Linear Regression







Normal QQ Plot

ANOVA Table

	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
Days	2	1.9553	0.9776	0.006362	0.993672
Nchick	1	36.3329	36.3329	0.236427	0.660105
Days:Nchick	2	38.9774	19.4887	0.126818	0.885378
Residuals	3	461.0251	153.675		

Summary Table:

Data Set	Untransformed vs. Transformed - Linear Model				
	Plots			P-Value	
	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	NOD	NOD	NOD	No	Gain (0.0032)
2	NOD	NOD	NOD	No	Gain (0.00173)
3	NOD	NOD	NOD	No	Loss (0.000401)
4	NOD	NOD	NOD	Yes	Loss (0.58445)
5	NOD	NOD	NOD	No	Loss (0.0038)

Data set	Untransformed LM vs. Logistic Regression GzLM				
	Plots			P-Value	
	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	NOD	NOD	NOD	No	Gain (0.08545)
2	Better	Better	Worse	Yes	Gain (0.20619)
3	Worse	Better	Worse	No	Gain (0.001475)
4	Better	Better	Better	No	Loss (0.1493)
5	NOD	NOD	NOD	No	Gain (0.08738)

Appendix E: Karla Letto

Data set 1:

Reference:

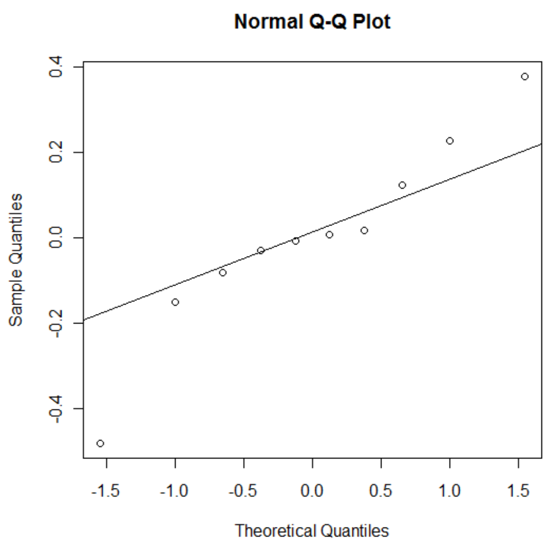
Formal, S. B. *et al.* (1958) Experimental shigella infections: characteristics of a fatal infection produce in guinea pigs. *J. Bacteriol.* 75, 604–610

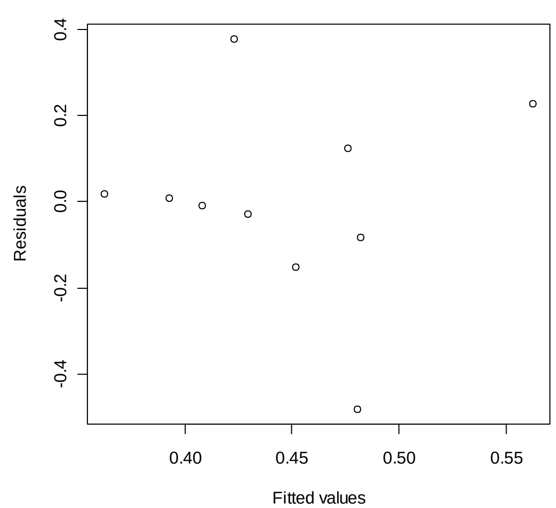
Raw Data:

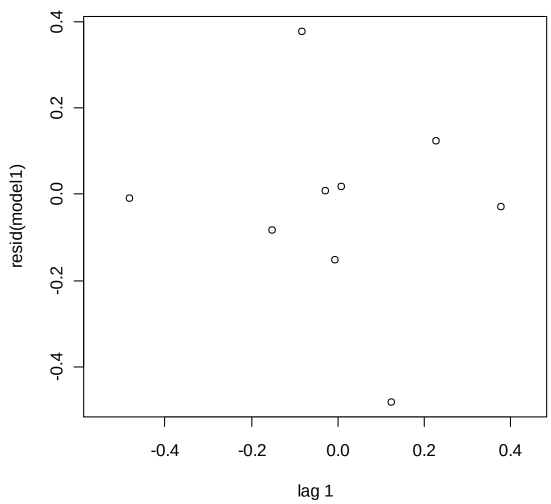
Dose	Mortality	Arcsine
1.40E+07	0.38	38.05673
1.20E+07	0.4	39.23152
9.60E+06	0.4	39.23152
1.00E+07	0.8	63.43495
6.10E+06	0.4	39.23152
8.10E+06	0.3	33.21091
1.10E+07	0.4	39.23152
6.20E+06	0	0
6.50E+06	0.6	50.76848
8.00E+05	0.79	62.72527

Non-Transformed Data:

LM: Linear Regression



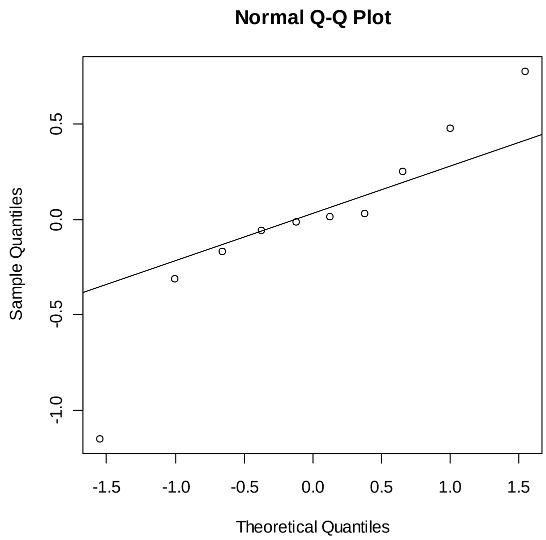


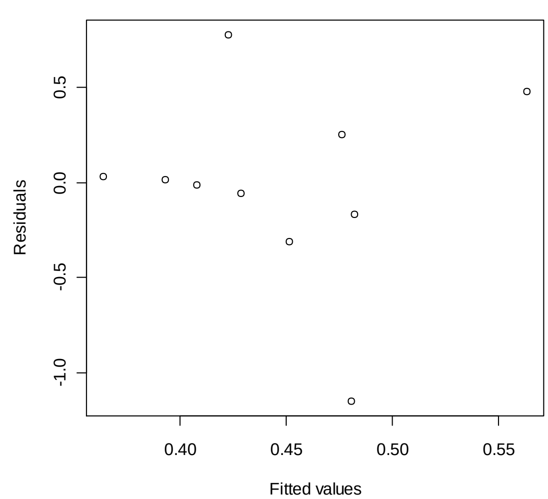


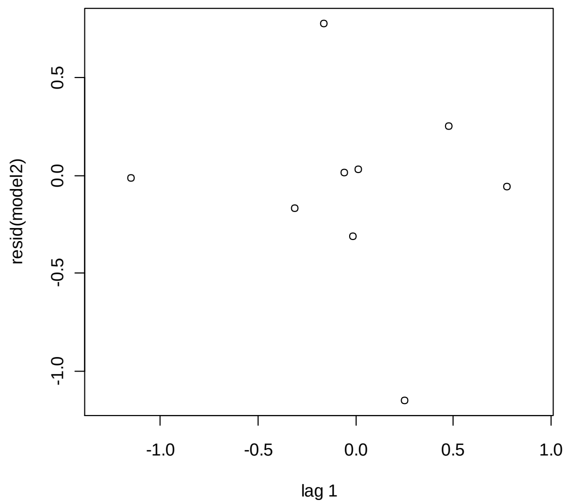
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Dose	1	0.02911	0.029107	0.4941	0.5021
Residuals	8	0.47130	0.058913		

GLM: Logistic Regression





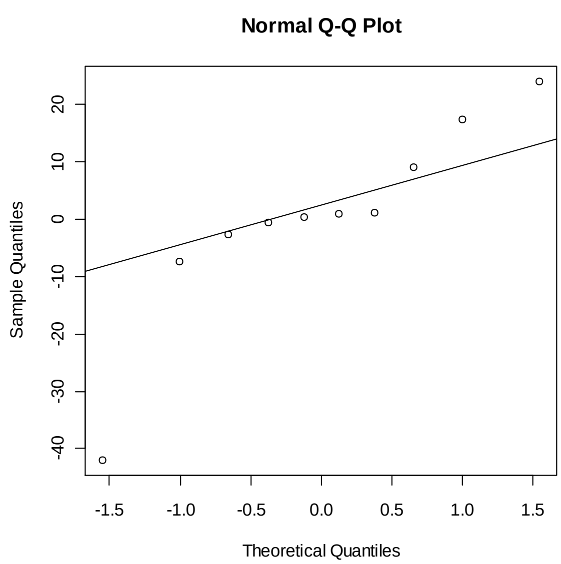


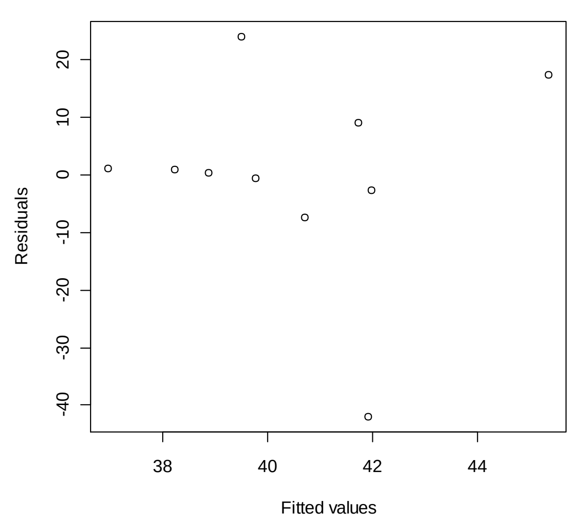
ANODEV Table

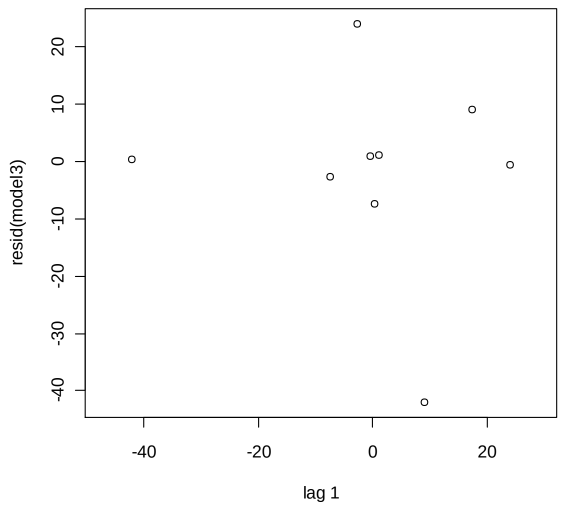
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			9	2.4417	
Dose	1	0.11795	8	2.3237	0.7313

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Dose	1	51.28	51.28	0.1476	0.7109
Residuals	8	2779.92	347.49		

Data Set 2:

Reference:

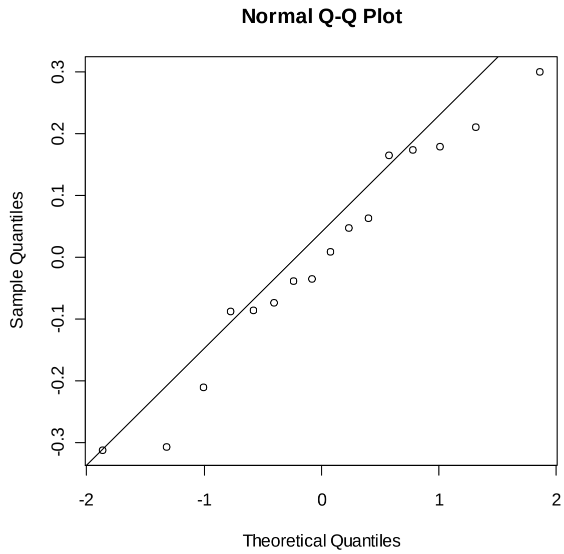
Robbins, J. D. et al. (1965) Relationship of dietary mineral intake to urinary mineral excretion and the incidence of urinary calcium in lambs. *J. Anim. Sci.* 24, 76-82

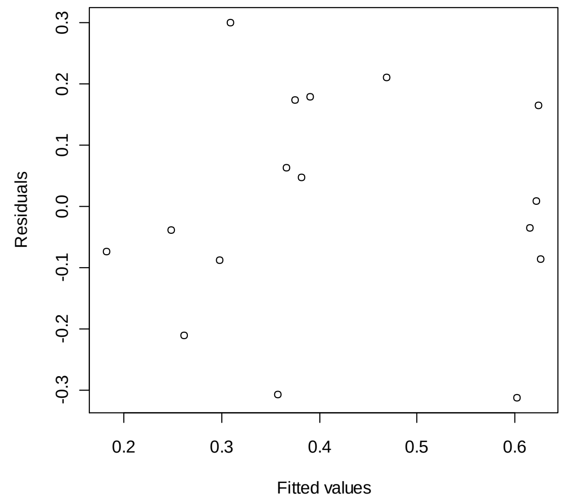
Raw Data:

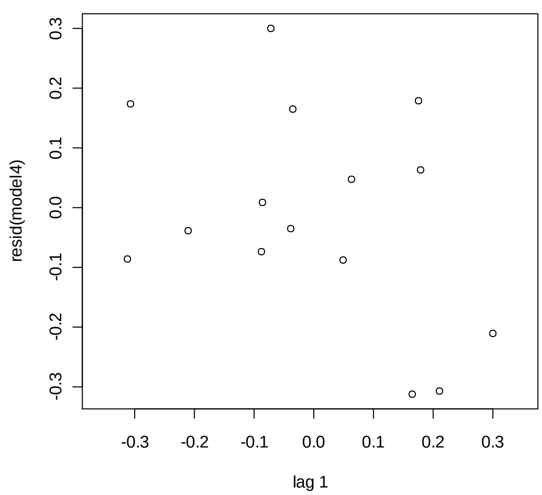
Potassium Intake (mg)	Occurrence of urolithiasis	Arcsine
3.5	0.63	52.53503
3.3	0.54	47.29428
4.4	0.29	32.58271
3.4	0.79	62.72527
3.8	0.58	49.60345
20.6	0.21	27.27473
20	0.05	12.92097
17.8	0.61	51.35452
23.6	0.11	19.36971
18.3	0.21	27.27473
14.5	0.43	40.97608
15.2	0.43	40.97608
14.1	0.57	49.02392
14.8	0.55	47.86959
15.6	0.05	12.92097
10.5	0.68	55.5501

Non-Transformed Data:

LM: Linear Regression



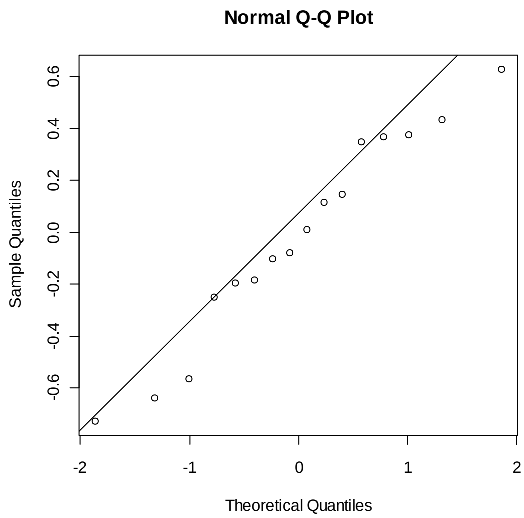


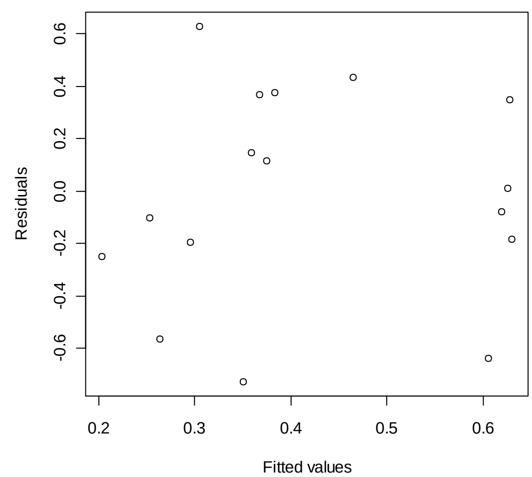


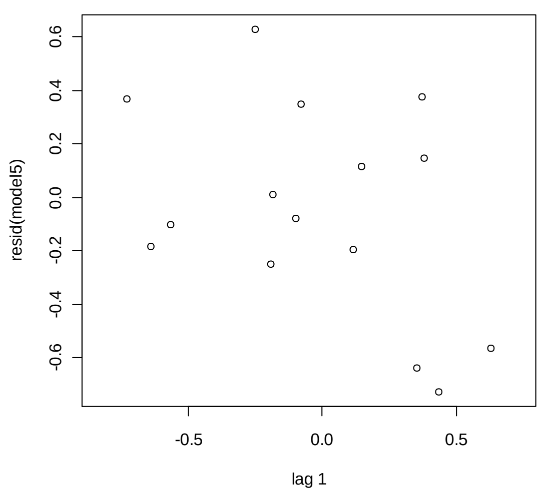
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
K	1	0.34748	0.34748	9.8912	0.007162 **
Residuals	14	0.49182	0.03513		

GLM: Logistic Regression





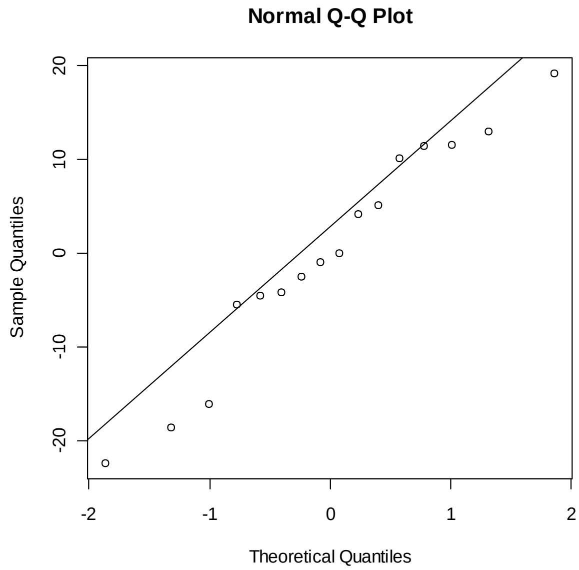


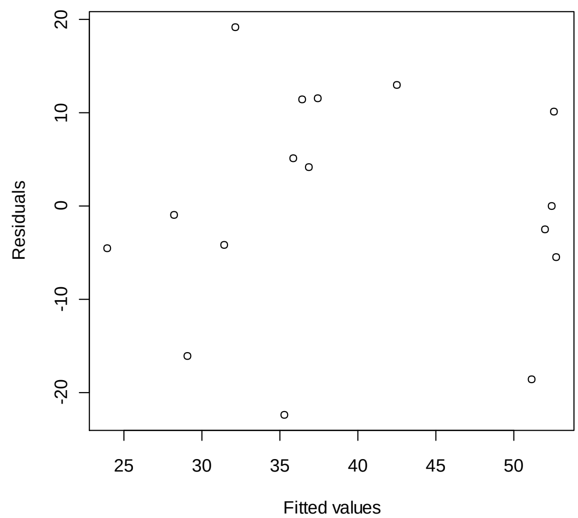
ANODEV Table

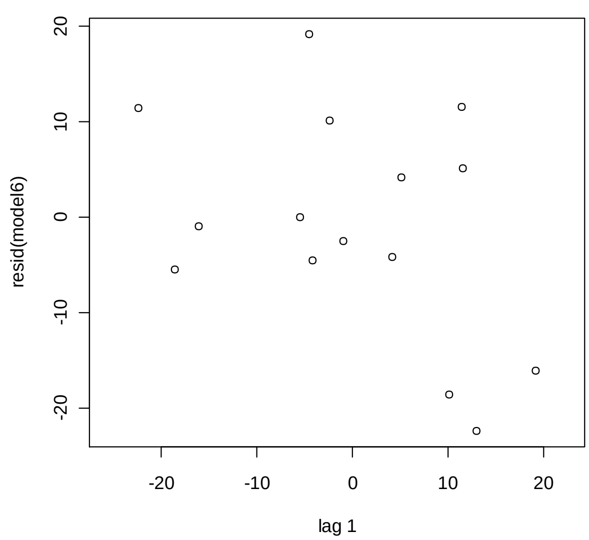
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			15	3.8749	
K	1	1.4526	14	2.4223	0.2281

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
K	1	1469.8	1469.82	9.6595	0.00771 **
Residuals	14	2130.3	152.16		

Data Set 3:

Reference:

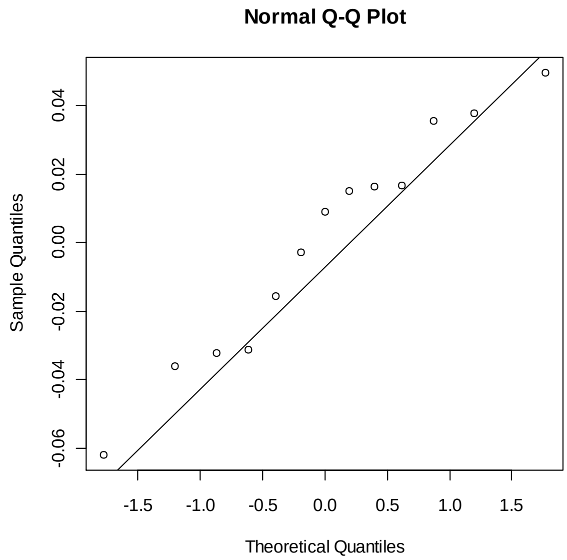
Rosner, B., eds (1995) *Fundamentals of Biostatistics*, Wadsworth Publishing Company

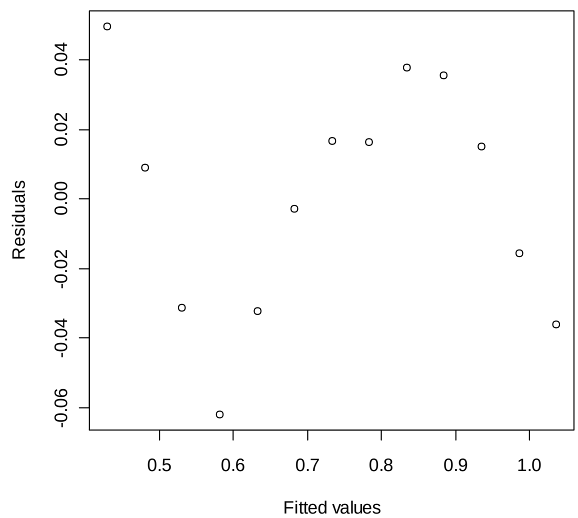
Raw Data:

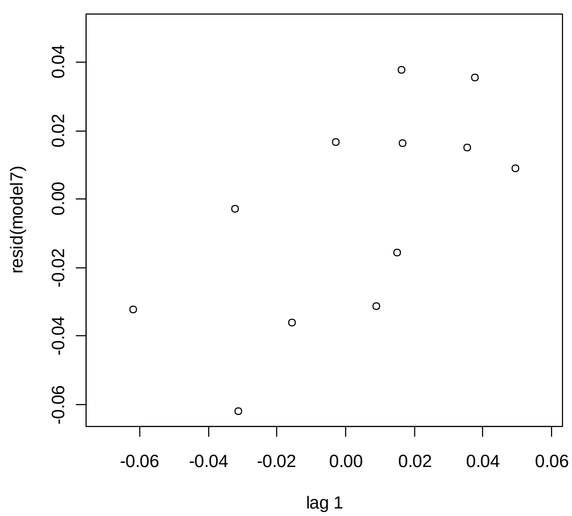
Month	Proportion of disease free infants	Arcsine
0	1	90
1	0.97	80.02578
2	0.95	77.07903
3	0.92	73.57006
4	0.872	69.03664
5	0.8	63.43495
6	0.75	60
7	0.68	55.5501
8	0.6	50.76848
9	0.52	46.14622
10	0.5	45
11	0.49	44.427
12	0.48	43.85378

Non-Transformed Data:

LM: Linear Regression



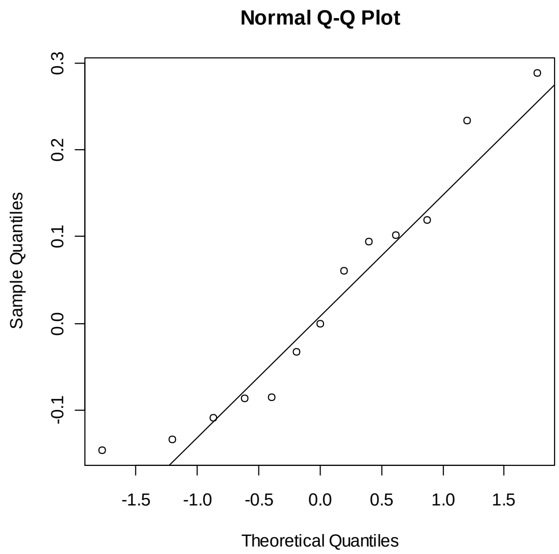


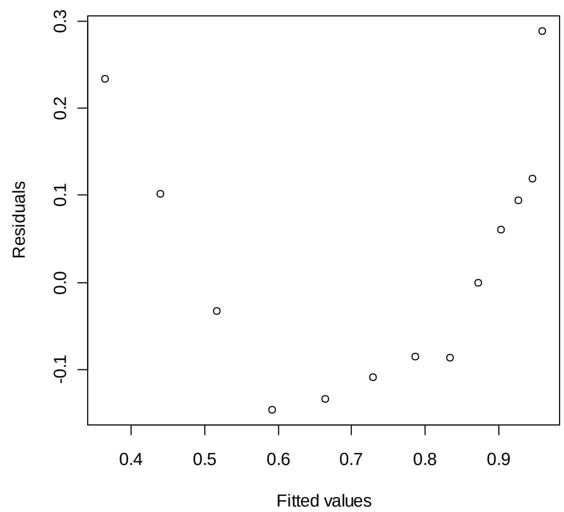


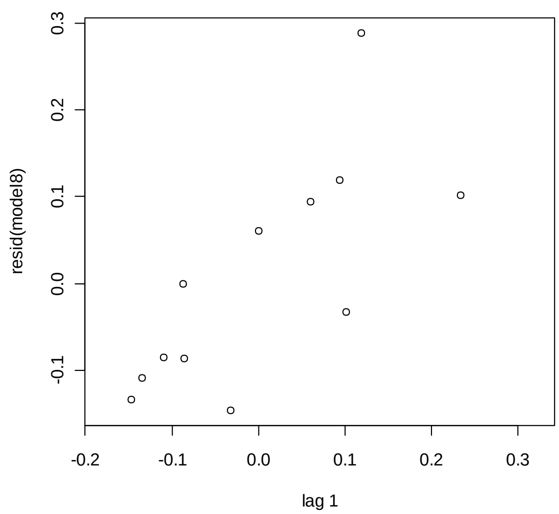
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
month	1	0.46344	0.46344	380.73	6.964e-10 ***
Residuals	11	0.01339	0.00122		

GLM: Logistic Regression





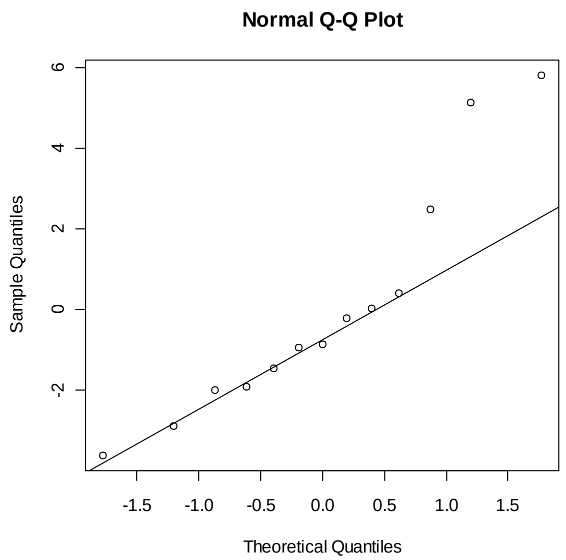


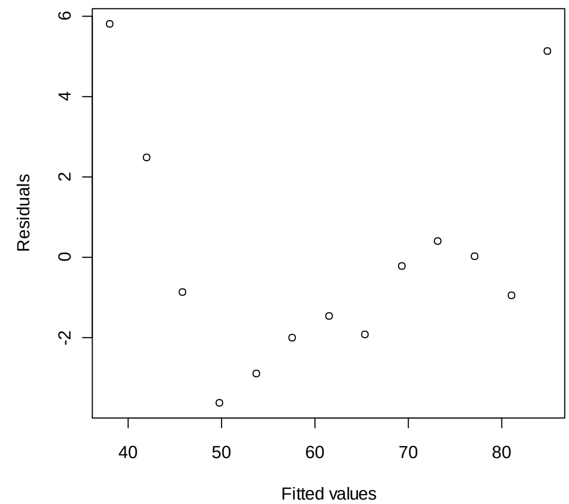
ANODEV Table

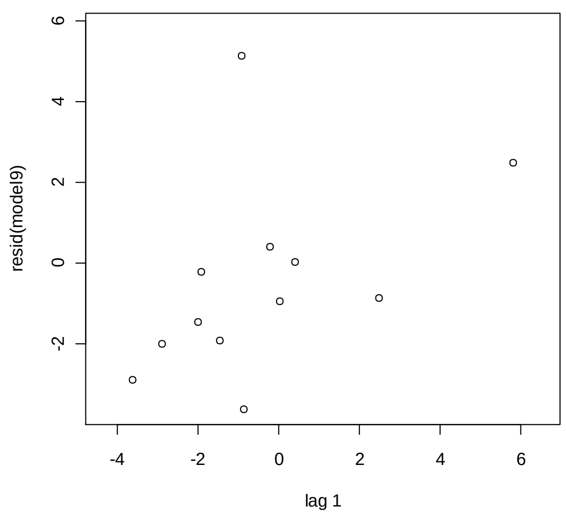
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			12	2.82449	
month	1	2.5813	11	0.24316	0.1081

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
month	1	2768.8	2768.84	307.03	2.198e-09 ***
Residuals	11	99.2	9.02		

Data Set 4:

Reference:

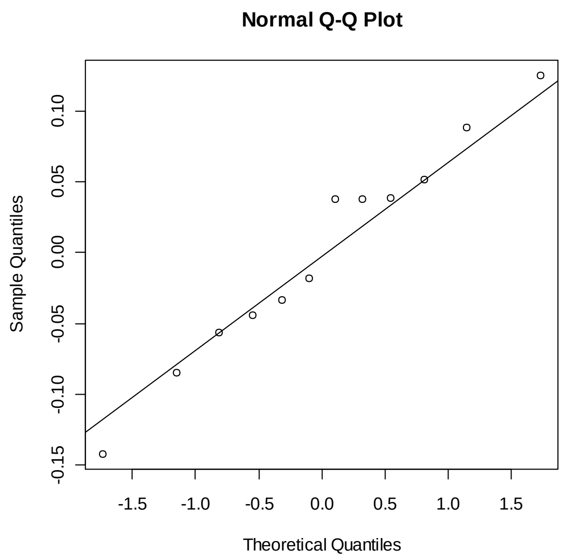
Fox, C. and Sullivan, E. C. (1925) A comparative study of rat-flea data for several seaports of the United States. *Public Health Reports* 40, 1909-1934

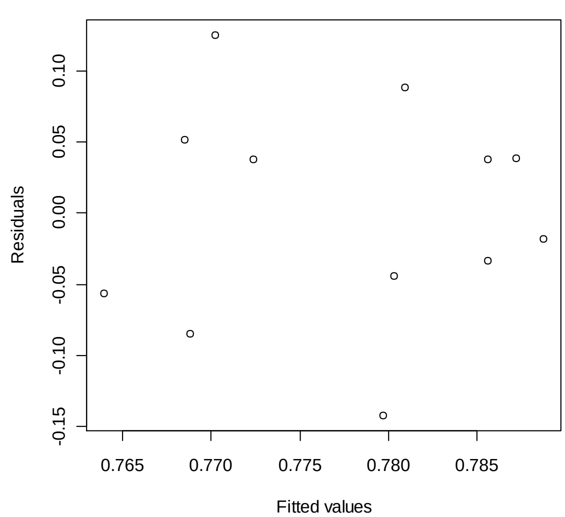
Raw Data:

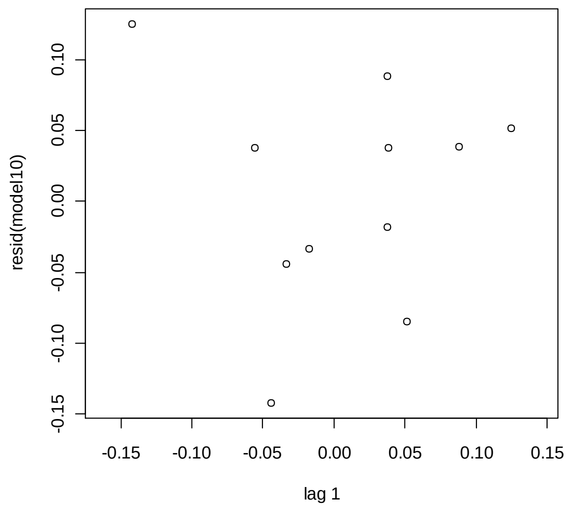
Monthly mean temp (°F)	percentage of rats with fleas	Arcsine
51	0.684	55.79616
50.8	0.82	64.89591
51.9	0.895	71.09276
58	0.638	53.01081
58.4	0.736	59.0821
61.8	0.752	60.1325
63.8	0.771	61.40995
61.8	0.823	65.12035
62.8	0.826	65.34628
58.8	0.869	68.78066
53.3	0.81	64.15807
47.9	0.708	57.29115

Non-Transformed Data:

LM: Linear Regression



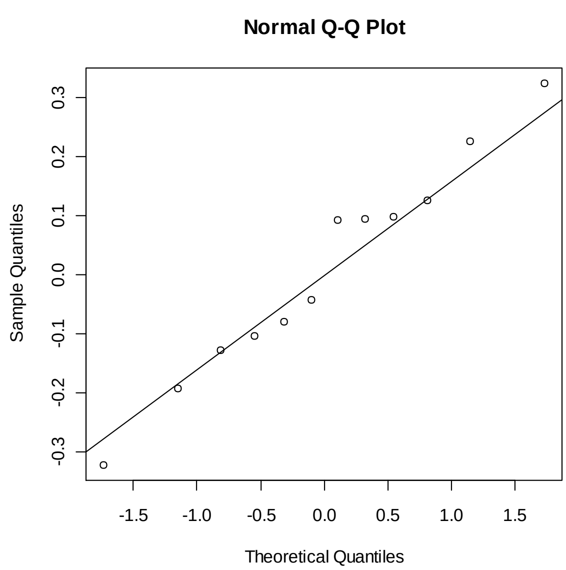


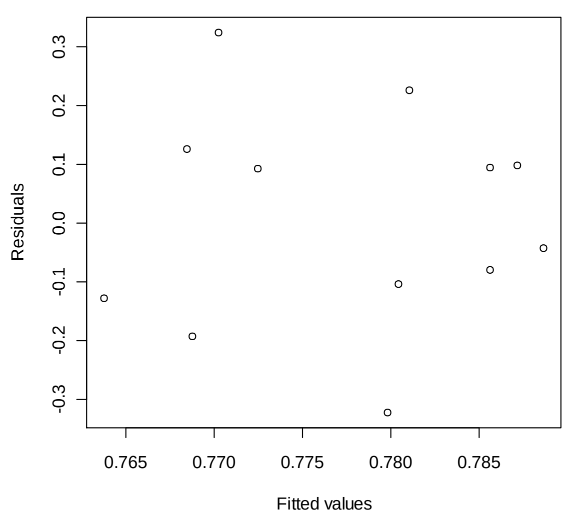


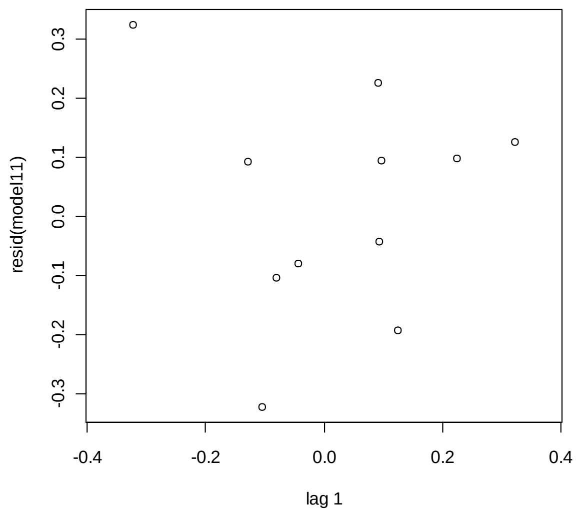
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
temp	1	0.000797	0.0007970	0.1243	0.7317
Residuals	10	0.064114	0.0064114		

GLM: Logistic Regression





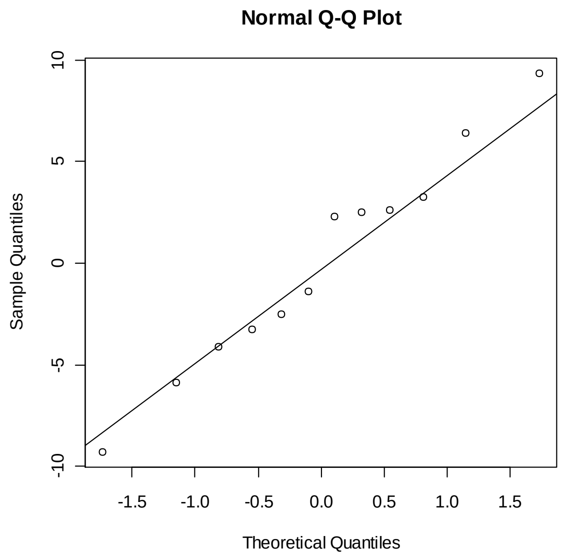


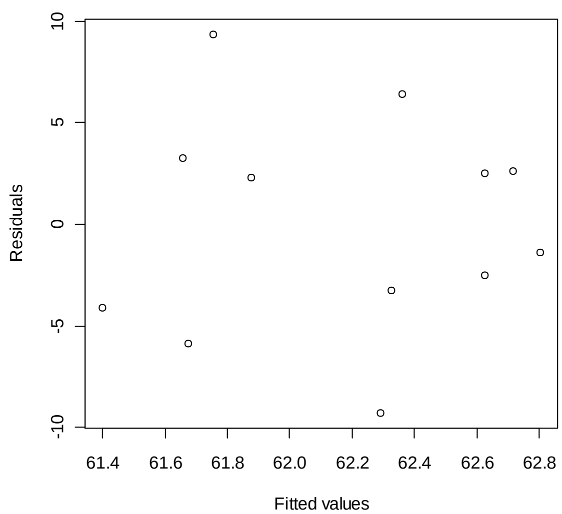
ANODEV Table

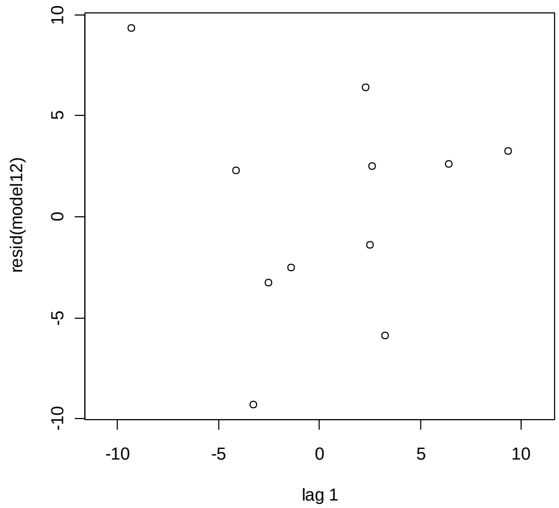
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			11	0.37925	
temp	1	0.0046023	10	0.37465	0.946

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
temp	1	2.556	2.5562	0.0815	0.781
Residuals	10	313.520	31.3520		

Data Set 5:

Reference:

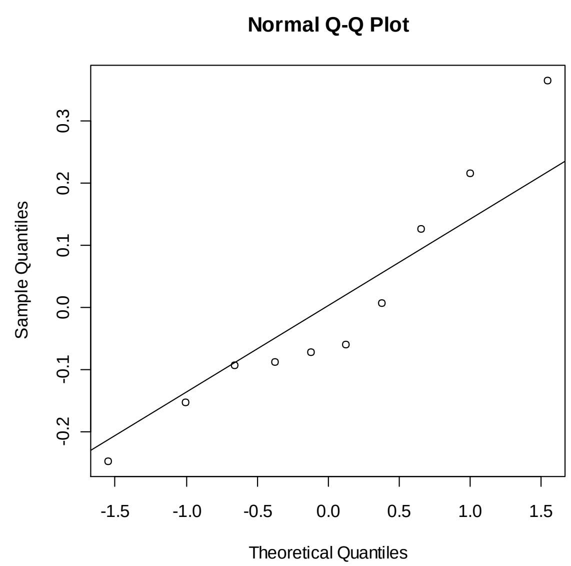
Sealander, J. A. (1951) Survival of Peromyscus in relation to environmental temperature and acclimation at highland low temperatures. *Am. Midi. Nat*. 46**,** 257-311

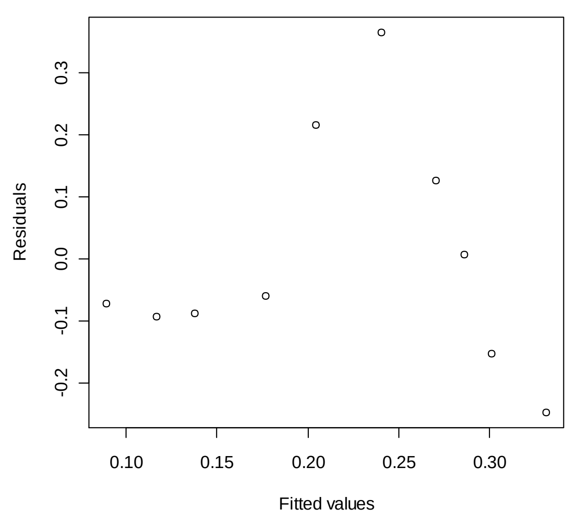
Raw Data:

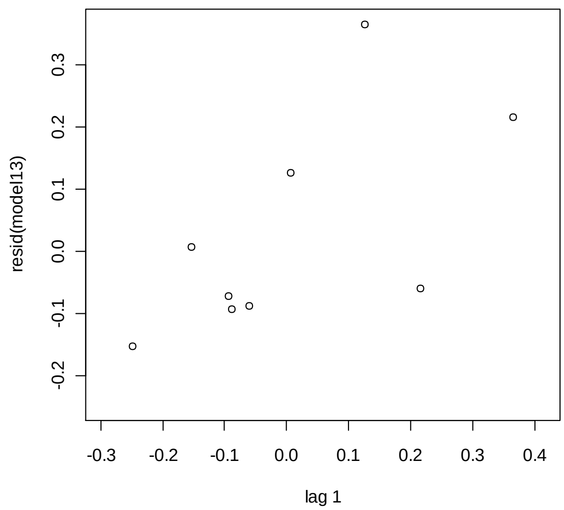
Mean air temp. (°C)	**Mean proportion initial weight loss**	**Arcsine**
-30	0.0171	7.513915
-21	0.0234	8.799116
-14	0.0491	12.80216
-1	0.1167	19.97523
8	0.4205	40.42557
20	0.6059	51.11393
30	0.3972	39.06769
35	0.2929	32.76553
40	0.1475	22.58523
50	0.0835	16.79592

Non-Transformed Data:

LM: Linear Regression



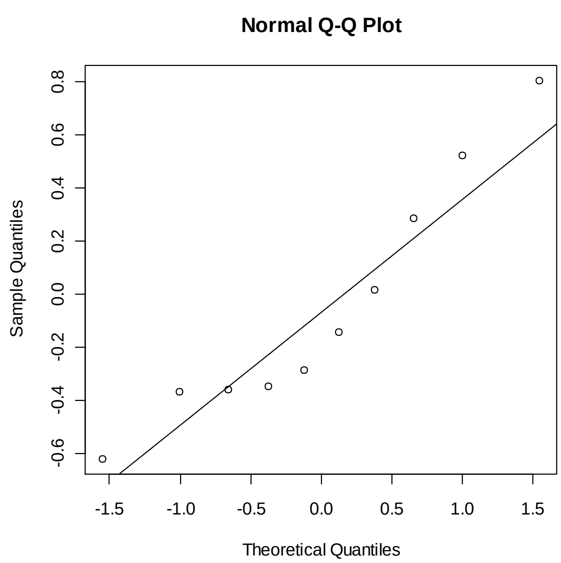


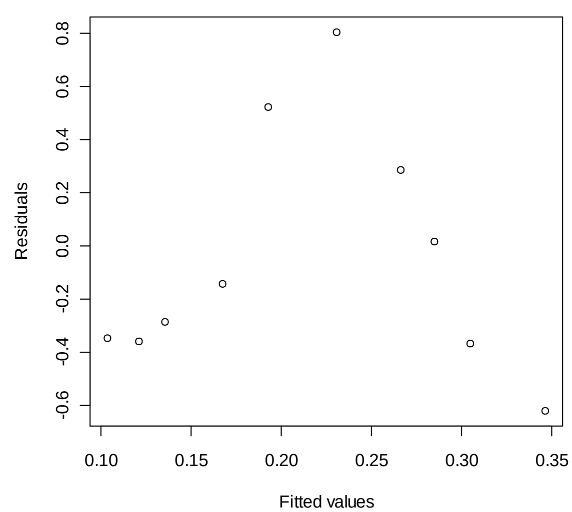


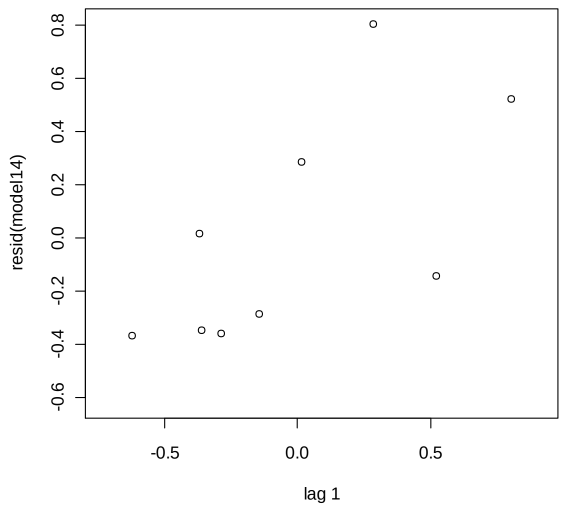
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
temp	1	0.062579	0.062579	1.6327	0.2372
Residuals	8	0.306627	0.038328		

GLM: Logistic Regression





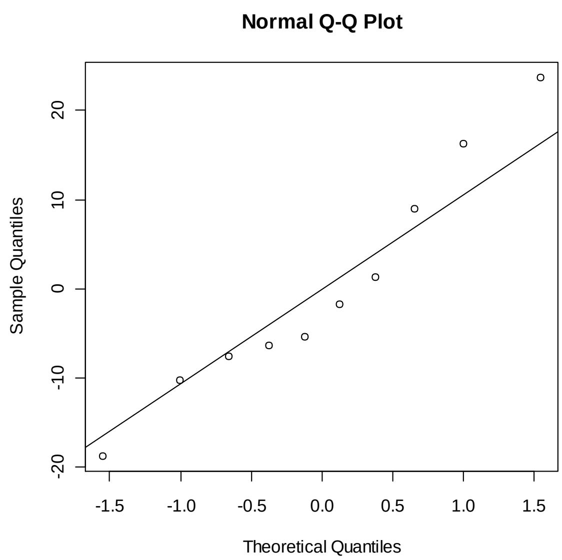


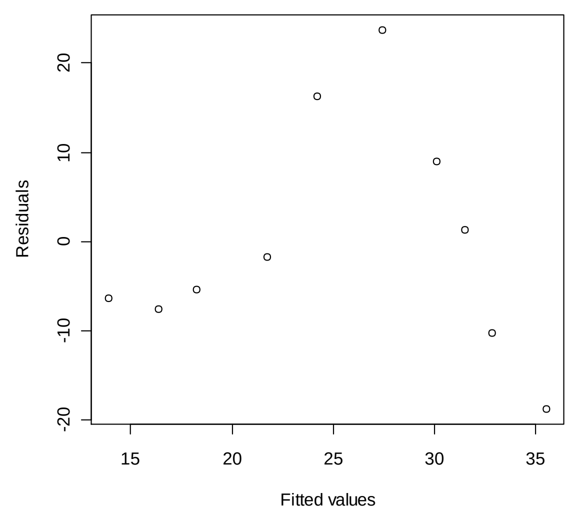
ANODEV Table

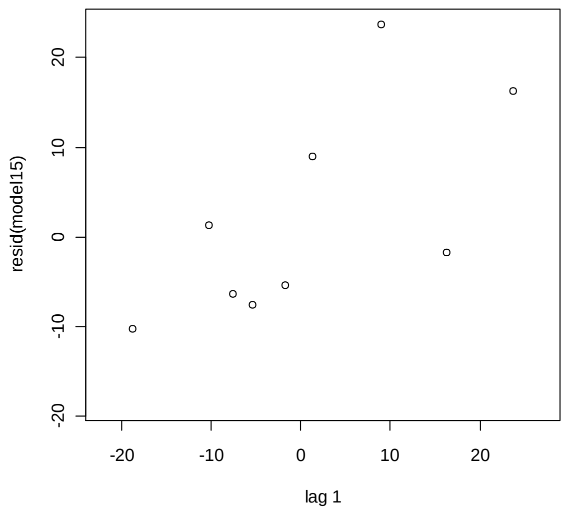
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			9	2.2463	
temp	1	0.38312	8	1.8632	0.5359

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
temp	1	499.98	499.98	2.6787	0.1403
Residuals	8	1493.18	186.65		

Summary of Results

Table 1: Change in residual plots and p-value for the arcsine transformation and logistic regression compared to the linear regression of the non-transformed data for data sets 1 through 5.

	Arcsine					**GLM: Logistic Regression**				
	Plots			**P-Value**		**Plots**			**P-Value**	
Data Set	**Normality**	**Homogeneity**	**Independence**	**Change in decision?**	**increase or decrease?**	**Normality**	**Homogeneity**	**Independence**	**Change in decision?**	**Increase or decrease?**
1	no diff	no diff	no diff	no	increase by 0.2088	no diff	no diff	no diff	no	increase by 0.2292
2	no diff	no diff	no diff	no	increase by 5.48E-4	no diff	no diff	no diff	yes	increase by 0.2209
3	better	worse	worse	no	increase by 1.50E-9	better	worse	no diff	yes	increase by 0.1081
4	no diff	no diff	no diff	no	increase by 0.0493	no diff	no diff	no diff	no	increase by 0.2143
5	no diff	no diff	better	no	decrease by 0.0969	no diff	no diff	better	no	increase by 0.2987

Appendix F: Olivia Puckrin

Data Set #1

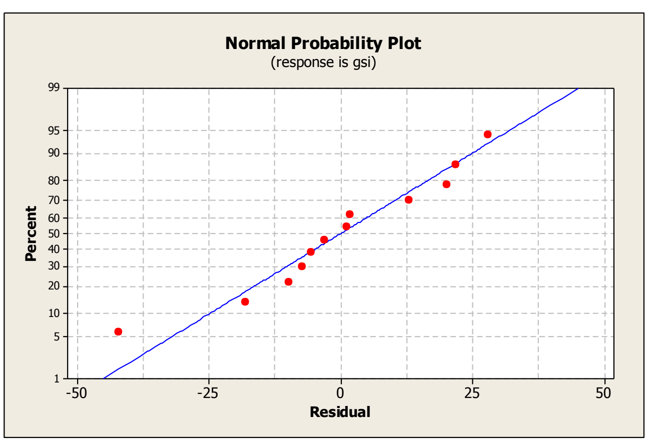
Reference:

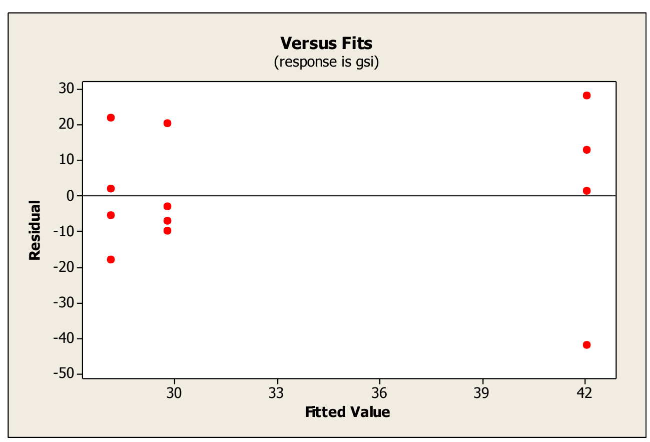
Arslan, T., Phelps, R.P., and Osborne, J.A. (2009) Effects of oestradiol-17β or 17α-methyltestosterone administration on gonadal differentiation of largemouth bass *Micropterus salmoides* (Lacepede). *Aquaculture Research.* 40, 1813-1822.

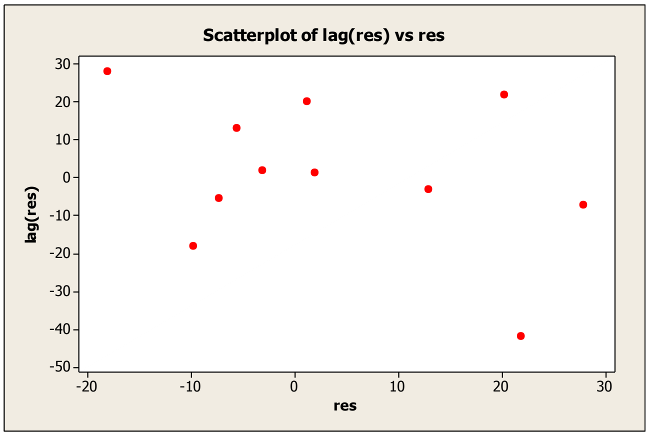
Raw Data:

Treatment	GSI groups	Proportion of GSI Group	Arcsine
Control	<1.0	0	0
Control	1-2	0.5	28.64789
Control	>2	0.5	28.64789
200mg 30 days	<1.0	0.433	24.80907
200mg 30 days	1-2	0.3	17.18873
200mg 30 days	>2	0.267	15.29797
45 days	<1.0	0.55	31.51268
45 days	1-2	0.225	12.89155
45 days	>2	0.225	12.89155
60 days	<1.0	0.7	40.10705
60 days	1-2	0.1	5.729578
60 days	>2	0.2	11.45916

Non-Transformed Data:



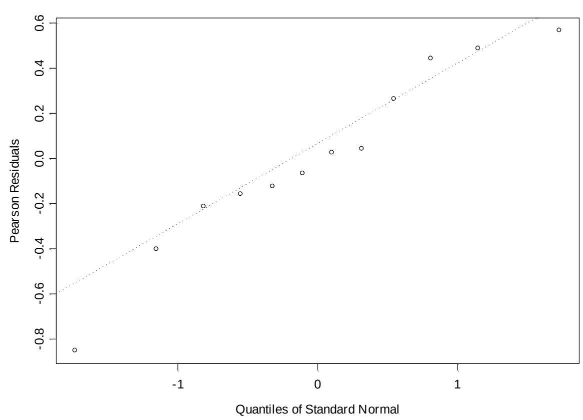


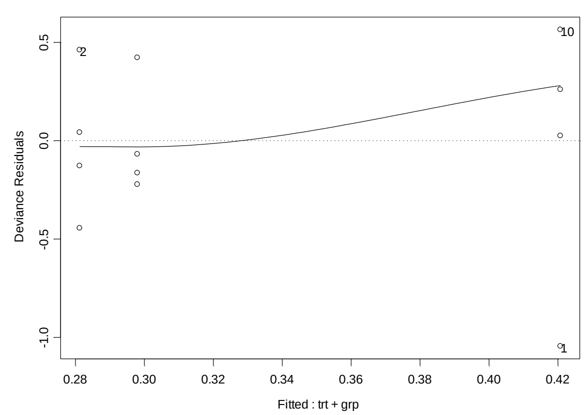


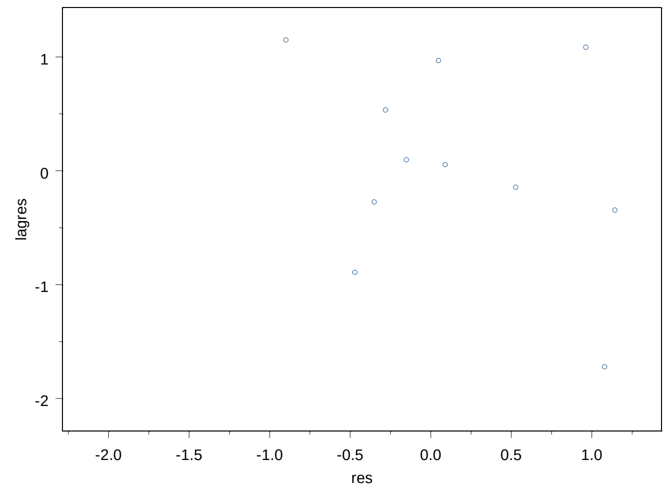
ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
trt	3	0.0	0.0	0.0	0.00	1.000
grp	2	464.1	464.1	232.1	0.34	0.726
Error	6	4127.8	4127.8	688.0		
Total	11	4591.9				

GLM: Logistic Regression





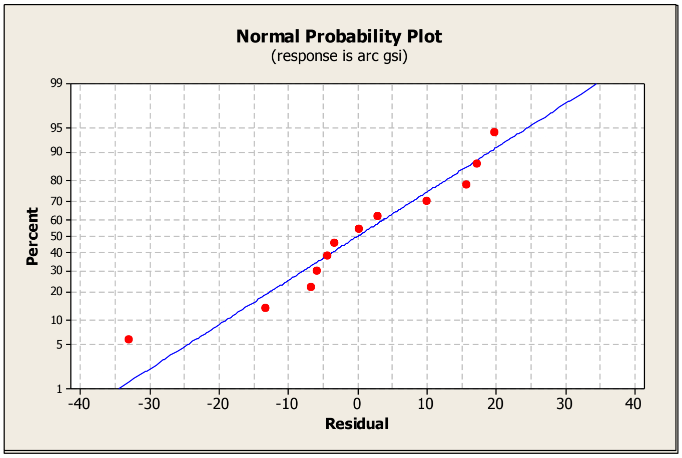


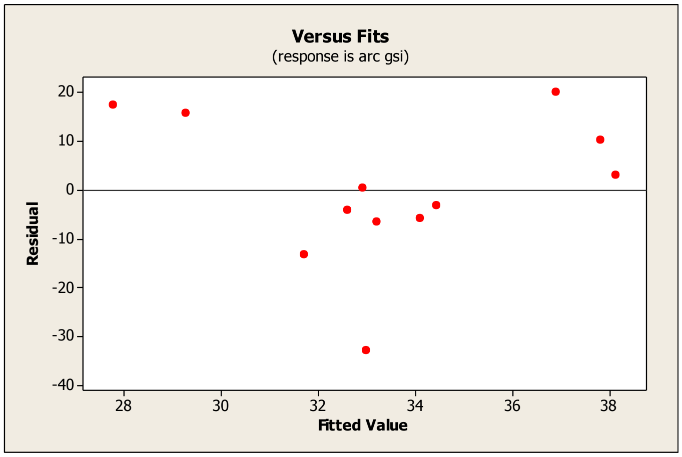
ANODEV Table

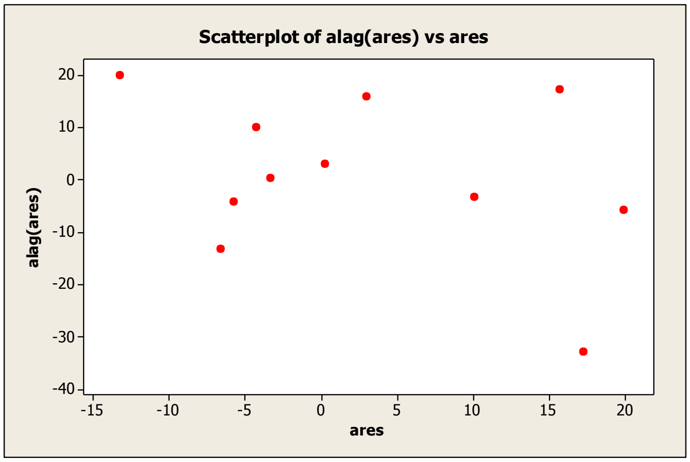
```
     Df  Deviance Resid. Df Resid. Dev  P value
NULL                    11   2.371602
 trt  3 0.0000000        8   2.371602  1.0000
 grp  2 0.2057551        6   2.165847  0.9022374
```

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
trt	3	50.9	50.9	17.0	0.04	0.987
grp	2	57.5	57.5	28.8	0.07	0.932
Error	6	2420.3	2420.3	403.4		
Total	11	2528.7				

Data Set #2

Reference:

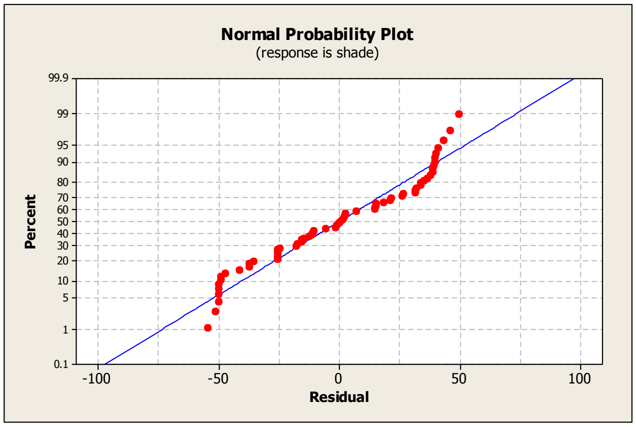
Johnson, L.T., Tank, J.L., and Dodds, W.K. (2009) The influence of land-use on stream biofilm nutrient limitation across eight North American ecoregions. *Can. J. Fish. Aquat. Sci.* 66, 1081-1094.

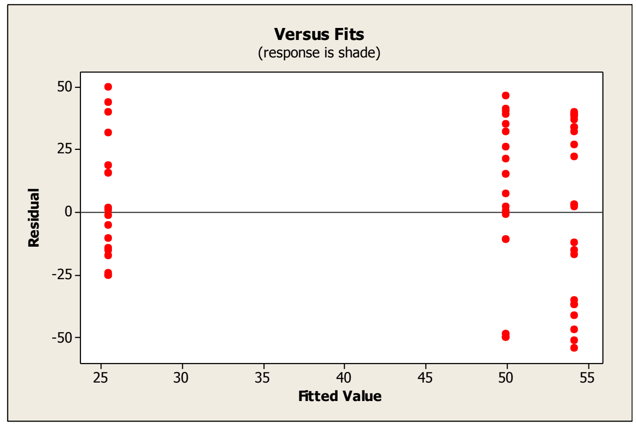
Raw Data:

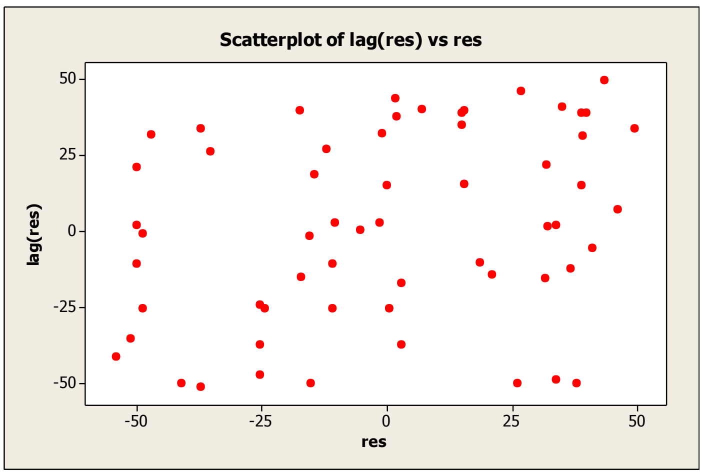
Land-use Classification	Proportion Shade	Arcsine
REF	0.76	60.66612575
REF	0.86	68.02724022
REF	0.07	15.34170855
AGR	0	0
AGR	0.26	30.65729899
AGR	0.2	26.56505118
URB	0.91	72.54239688
URB	0.85	67.213502
URB	0.65	53.72880156
REF	0.93	74.65829145
REF	0.93	74.65829145
REF	0.94	75.82118171
AGR	0.41	39.8151201
AGR	0.41	39.8151201
AGR	ND	ND
URB	0.9	71.56505118
URB	0.57	49.02392312
URB	0.96	78.46304097
REF	0.81	64.15806724
REF	0.42	40.39655189
REF	0.91	72.54239688
AGR	ND	ND
AGR	0.65	53.72880156
AGR	0.08	16.42994019
URB	ND	ND
URB	0.52	46.14622139
URB	0	0
REF	0.39	38.6454835
REF	0.37	37.46496893
REF	0.57	49.02392312
AGR	0.15	22.786498

AGR	0.44	41.55394871
AGR	0.11	19.3697123
URB	0.71	57.41729374
URB	0	0
URB	0.76	60.66612575
REF	0.19	25.84193276
REF	0.03	9.974221794
REF	0.17	24.3500636
AGR	0	0
AGR	0.01	5.739170477
AGR	0	0
URB	0.39	38.6454835
URB	0.39	38.6454835
URB	0	0
REF	0.13	21.13429221
REF	0	0
REF	ND	ND
AGR	0	0
AGR	ND	ND
AGR	0	0
URB	0.01	5.739170477
URB	ND	ND
URB	0	0
REF	0.92	73.57005981
REF	0.56	48.44605129
REF	0.88	69.73209894
AGR	0.75	60
AGR	0.69	56.16684133
AGR	0.27	31.30644625
URB	0.82	64.89590975
URB	0.49	44.427004
URB	0.01	5.739170477
REF	0.88	69.73209894
REF	0.17	24.3500636
REF	0.57	49.02392312
AGR	0.24	29.33387425
AGR	0.1	18.43494882
AGR	0.57	49.02392312
URB	0.89	70.6302877
URB	0.65	53.72880156
URB	0.5	45

Non-Transformed Data:



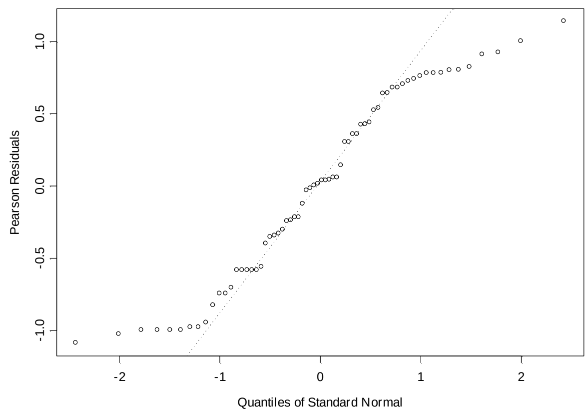


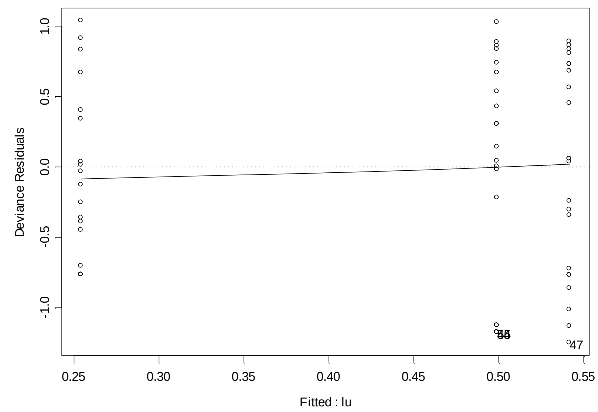


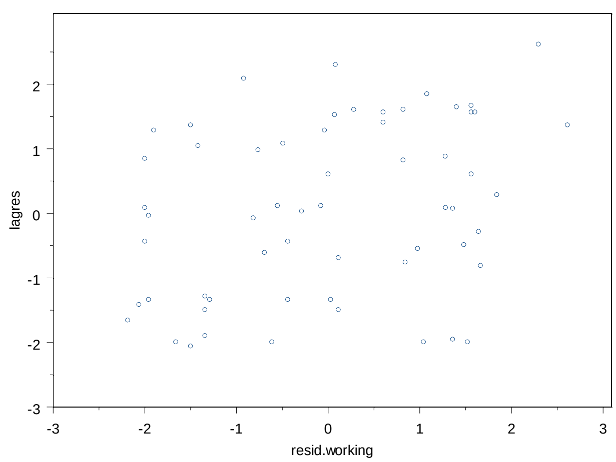
ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
lu	2	10381	10381	5191	5.08	0.009
Error	63	64356	64356	1022		
Total	65	74738				

GLM: Logistic Regression





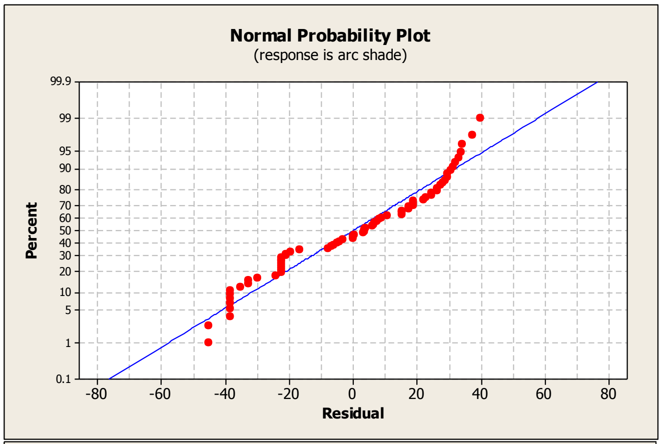


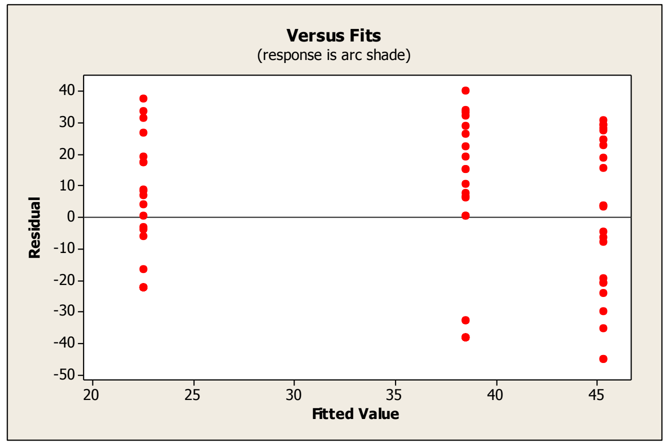
ANODEV Table

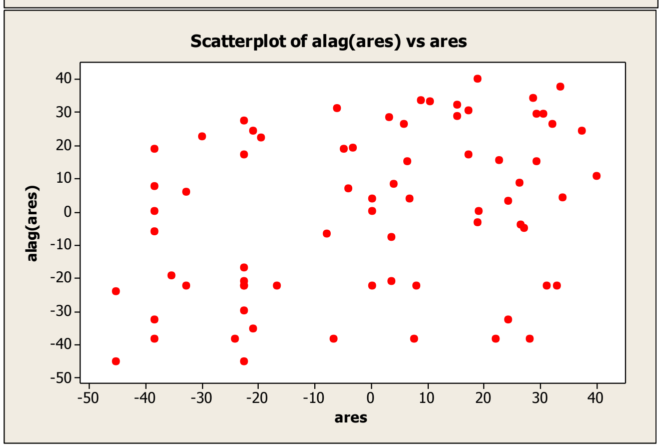
	Df	Deviance	Resid. Df	Resid. Dev	P Value
NULL			65	37.19463	0.112102
landuse	2	4.376684	63	32.81794	

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
lu	2	6566.3	6566.3	3283.1	5.20	0.008
Error	69	43541.3	43541.3	631.0		
Total	71	50107.6				

Data Set #3

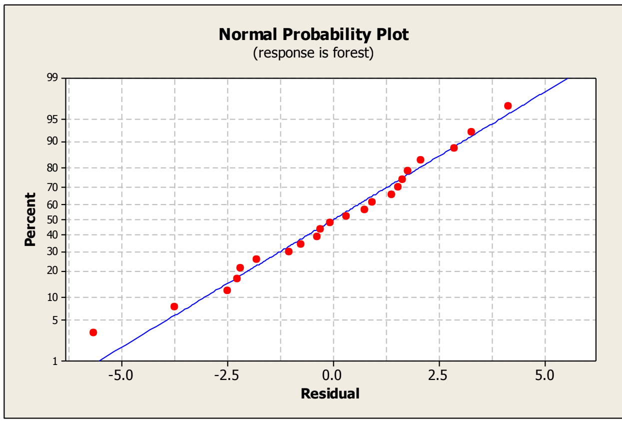
Reference:

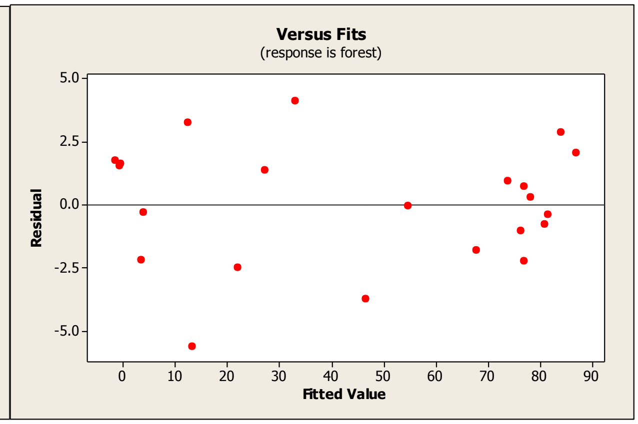
Kohzu, A., Tayasu, I., Yoshimizu, C., *et al.*. (2009) Nitrogen-stable isotopic signatures of basal food items, primary consumers and omnivores in rivers with different levels of human impact. *Ecol. Res.* 24, 127-136.

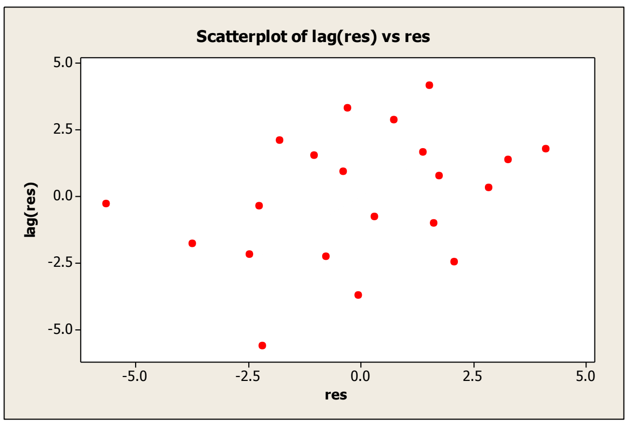
Raw Data:

Human Impact Area	Proportion Forest	Arcsine
19.9	0.748	59.86786
12.5	0.811	64.23117
17	0.746	59.73606
13.1	0.801	63.50664
15.7	0.785	62.37511
10	0.869	68.78066
17	0.776	61.75216
93.4	0.001	1.812154
59.7	0.371	37.52429
92.5	0.008	5.131548
17.6	0.752	60.1325
92.3	0.011	6.020304
65.3	0.286	32.32965
79.7	0.157	23.34285
88	0.036	10.93742
79	0.075	15.89417
88.5	0.012	6.289059
70.4	0.195	26.20525
7.1	0.891	70.72203
25.9	0.659	54.27101
46.5	0.428	40.86031
38.5	0.547	47.69688

Non-Transformed Data:



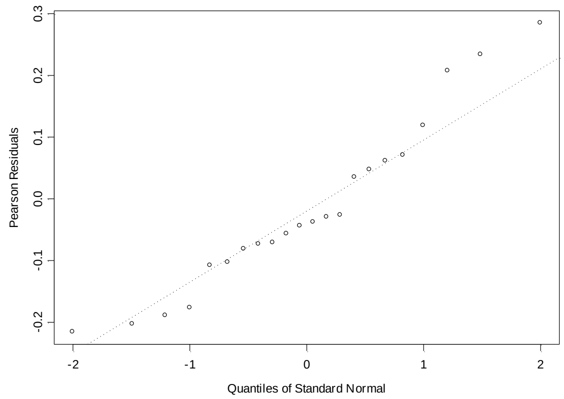


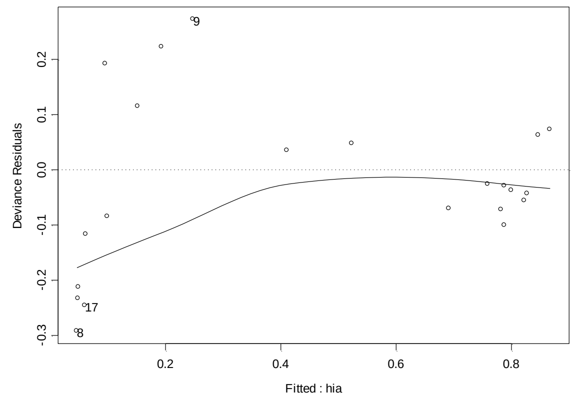


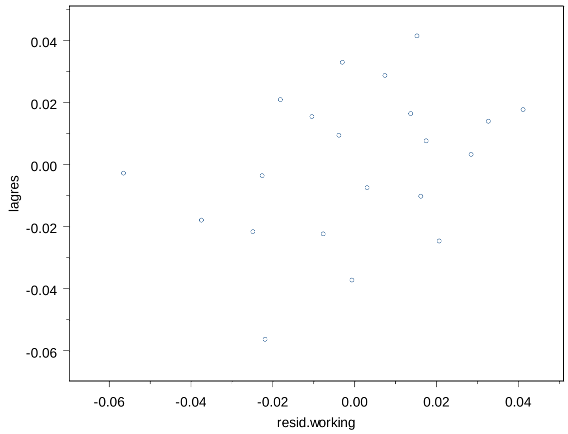
ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
hia	1	24288	24288	24288	4094.86	0.000
Error	20	119	119	6		
Total	21	24406				

GLM: Logistic Regression





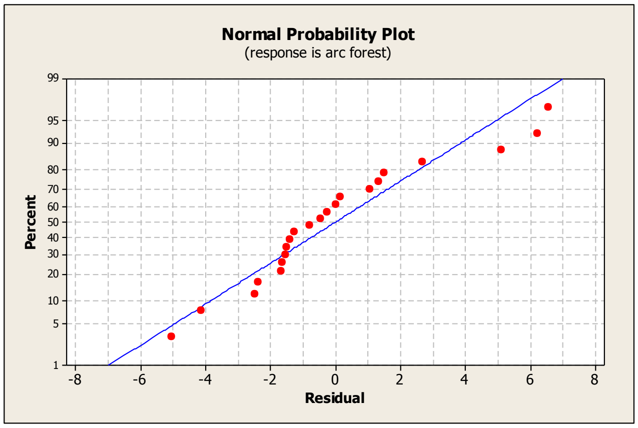


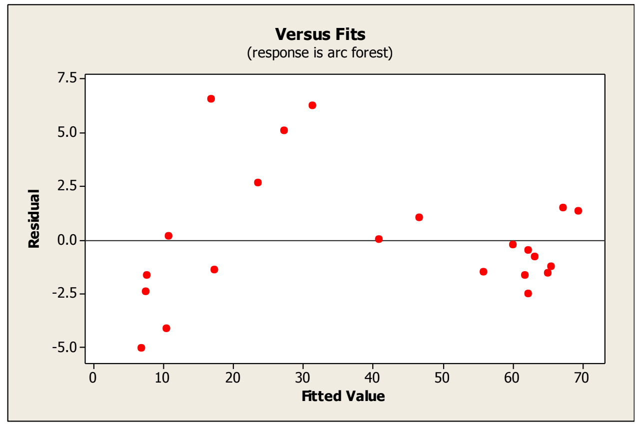
ANODEV Table

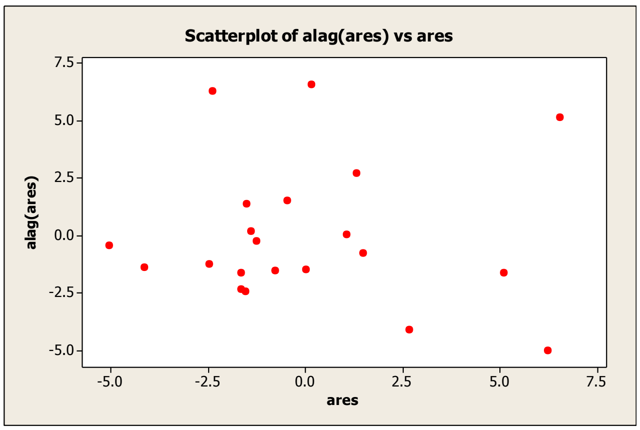
	Df	Deviance	Resid. Df	Resid. Dev	P value
NULL			21	11.78154	
hia	1	11.30065	20	0.48090	0.000775

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
hia	1	12073	12073	12073	1276.73	0.000
Error	20	189	189	9		
Total	21	12262				

Data Set #4

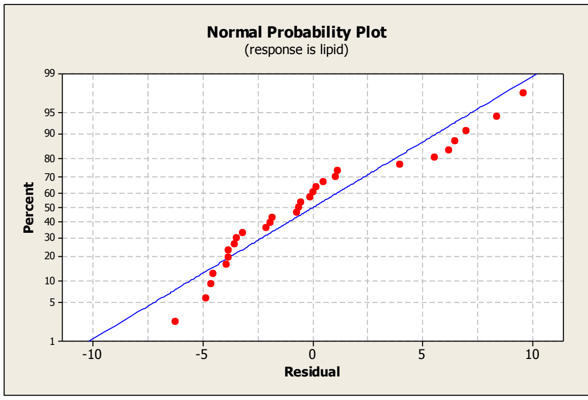
Reference:

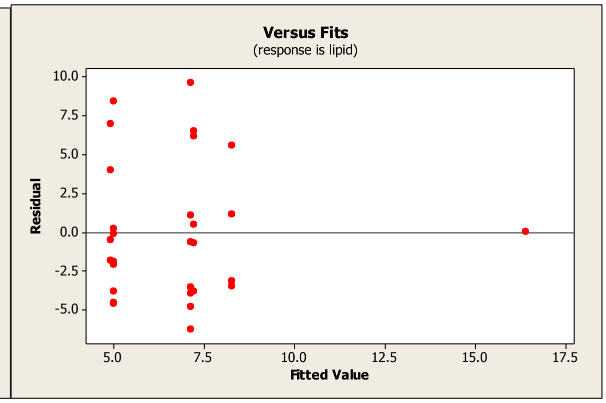
Wiemeyer,, S.N., Belisle, A.A., and Gramlich, F.J. (1978) Organochlorine residues in potential food items of Maine Bald eagles (*Haliaeetus leucocephalus*), 1966 and 1974. *Bull. Environ. Contam. Toxicol.* 78, 64-72.

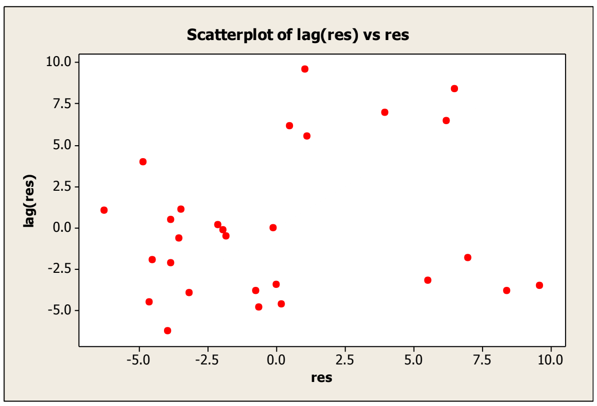
Raw Data:

County	Year	Proportion Lipid Wt.	Arcsine
Washington	1966	0.044	12.10839
Washington	1966	0.031	10.14082
Washington	1966	0.119	20.17958
Washington	1966	0.089	17.35725
Washington	1974	0.023	8.722994
Washington	1974	0.065	14.77068
Washington	1974	0.036	10.93742
Washington	1974	0.167	24.12046
Washington	1974	0.082	16.63995
Washington	1974	0.009	5.443741
Washington	1974	0.032	10.30485
Hancock	1966	0.051	13.0518
Hancock	1966	0.138	21.80714
Hancock	1966	0.094	17.8541
Hancock	1966	0.048	12.65553
Penobscot	1974	0.164	23.88919
Lincoln	1966	0.049	12.78889
Lincoln	1966	0.031	10.14082
Lincoln	1966	0.005	4.054807
Lincoln	1966	0.004	3.626123
Lincoln	1966	0.052	13.18142
Lincoln	1966	0.029	9.804905
Lincoln	1966	0.012	6.289059
Lincoln	1966	0.134	21.47284
Lincoln	1974	0.137	21.72395
Lincoln	1974	0.134	21.47284
Lincoln	1974	0.077	16.11038
Lincoln	1974	0.034	10.62562
Lincoln	1974	0.065	14.77068

Non-Transformed Data:



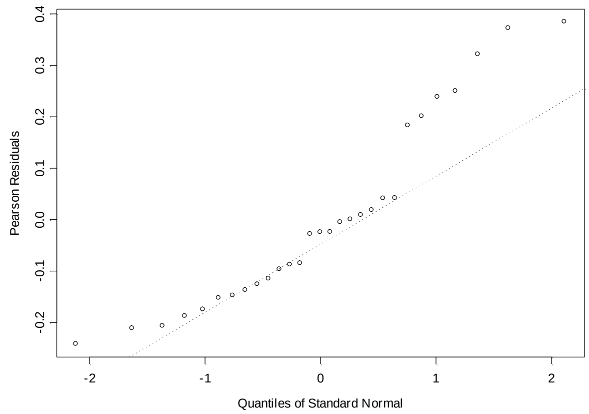


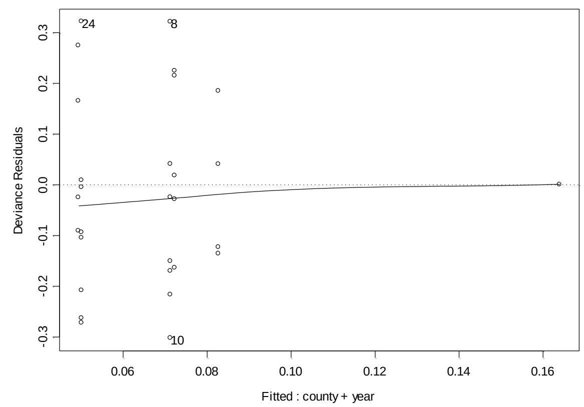


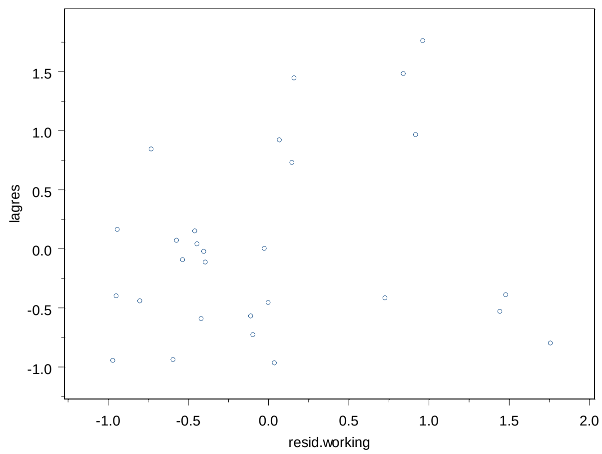
ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
county	3	114.39	110.94	36.98	1.66	0.203
yr	1	27.34	27.34	27.34	1.23	0.279
Error	24	535.44	535.44	22.31		
Total	28	677.17				

GLM: Logistic Regression





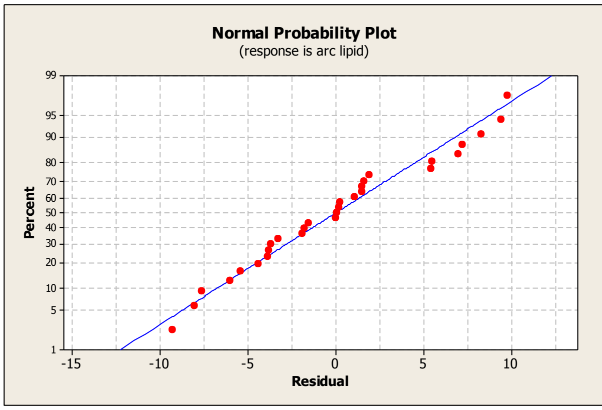


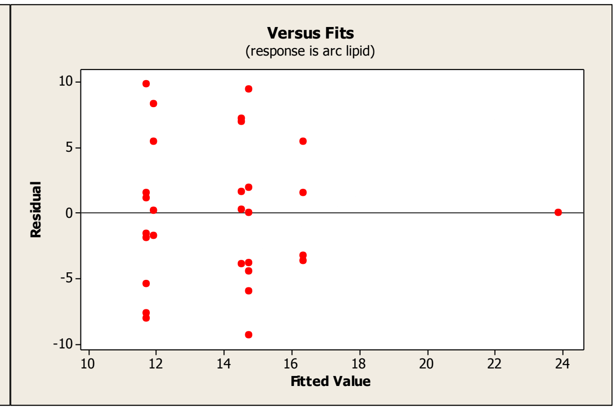
ANODEV Table

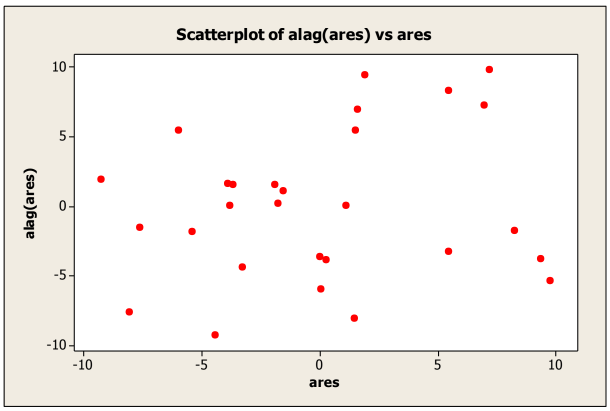
	Df	Deviance	Resid. Df	Resid. Dev	P value
NULL			28	1.101358	
county	3	0.1421378	25	0.959220	0.98634
year	1	0.0481212	24	0.911099	0.826365

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
county	3	140.25	142.22	47.41	1.46	0.252
yr	1	44.72	44.72	44.72	1.37	0.253
Error	24	781.73	781.73	32.57		
Total	28	966.71				

Data Set #5

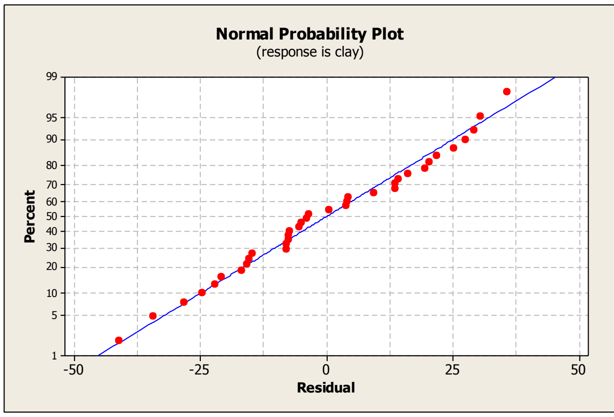
Reference:

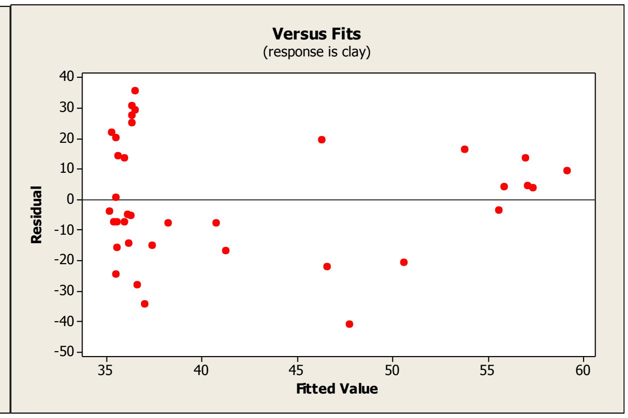
Yeager, K.M., Brinkmeyer, R., Rakocinski, C.F., Schindler, K.J., and Santschi, P.H. (2010) Impacts of dredging activities on the accumulation of dioxins in surface sediments of the Houston Ship Channel, Texas. *J. Coastal Res.* 26, 743-752.

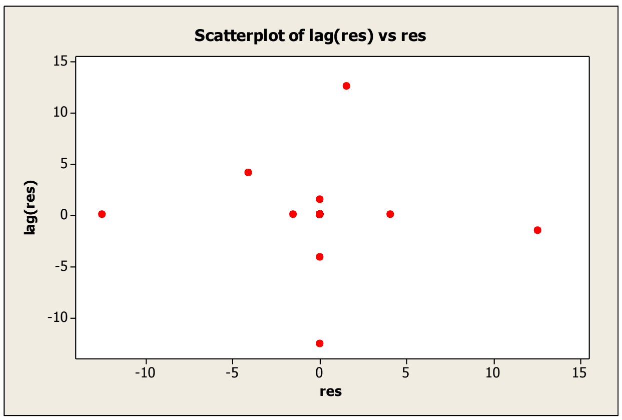
Raw Data:

Water Depth	Proportion Clay	Arcsine
0.49	0.28	31.94806
0.61	0.28	31.94806
0.61	0.198	26.42154
0.87	0.286	32.32965
0.9	0.496	44.77081
0.66	0.498	44.88541
0.57	0.558	48.33065
0.41	0.571	49.0818
0.31	0.312	33.95693
1.05	0.216	27.69456
1.36	0.084	16.84763
2.57	0.304	33.4605
4.38	0.329	35.00061
8.61	0.244	29.60143
11.57	0.299	33.14837
9.48	0.067	15.00146
4.74	0.244	29.60143
1.95	0.221	28.0412
0.579	0.109	19.27797
1.13	0.308	33.70916
1.19	0.615	51.64854
1.28	0.722	58.17969
1.16	0.638	53.01081
0.579	0.36	36.8699
1.16	0.669	54.87754
1.31	0.657	54.15023
1	0.312	33.95693
15.4	0.599	50.71001
16.2	0.706	57.16527
13.9	0.7	56.78909
8.41	0.658	54.2106
16.3	0.615	51.64854
16.5	0.613	51.53084
17.8	0.685	55.85781
15.2	0.521	46.20357
1.65	0.026	9.279177

Non-Transformed Data:



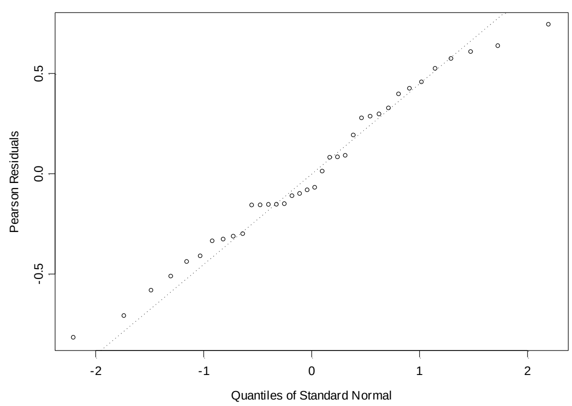


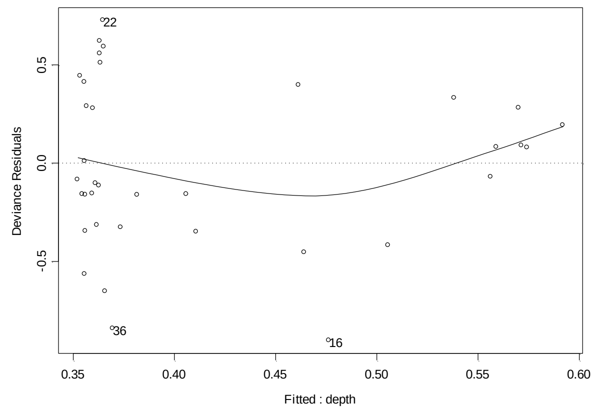


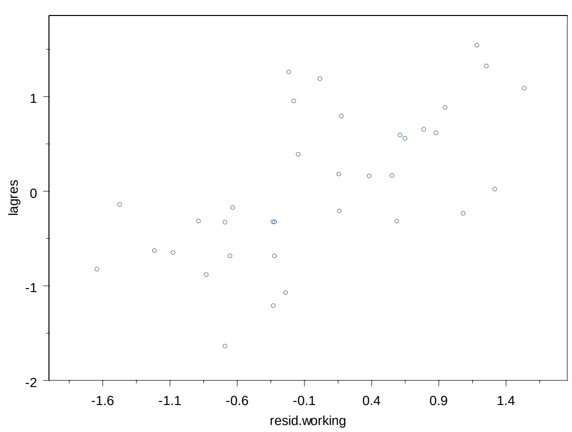
ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
depth	1	2449.7	2449.7	2449.7	6.30	0.017
Error	34	13227.6	13227.6	389.0		
Total	35	15677.3				

GLM: Logistic Regression





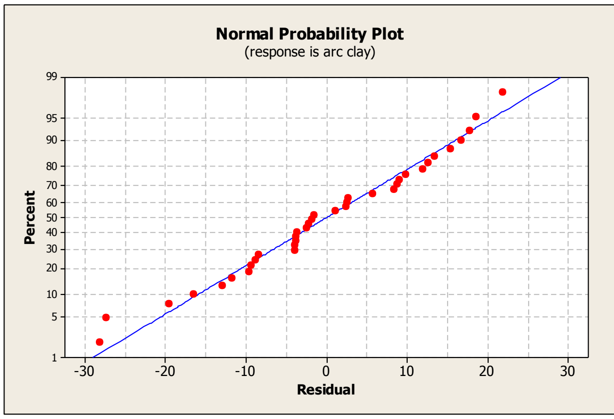


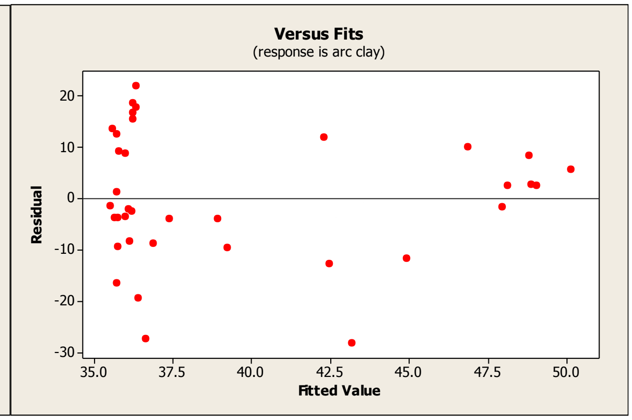
ANODEV Table

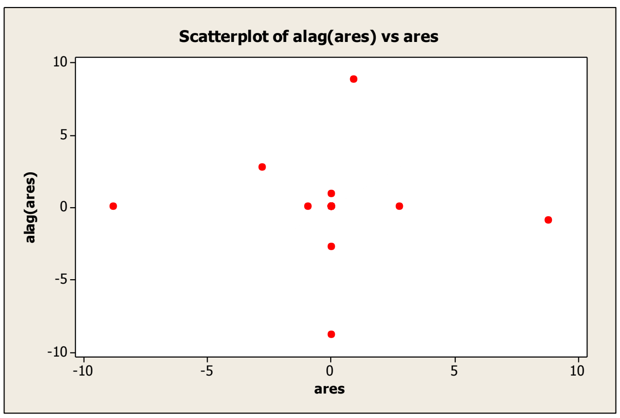
	Df	Deviance	Resid. Df	Resid. Dev	P value
NULL			35	7.022609	
depth	1	0.9997986	34	6.022811	0.317359

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

Source	DF	Seq SS	Adj SS	Adj MS	F	P
depth	1	908.9	908.9	908.9	5.68	0.023
Error	34	5443.3	5443.3	160.1		
Total	35	6352.3				

Summary Tables:

	LM: prop vs. deg				
	Plots			P-Value	
Data Set	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	no diff	no diff	no diff	no	gain (0.013), loss (0.206)
2	no diff	no diff	no diff	yes- sig. to not sig.	gain (0.071)
3	worse	no diff	worse	no	no change
4	no diff	no diff	better	no	gain (0.049), loss (0.026)
5	no diff	no diff	worse	no	gain (0.006)

	LM: prop vs. GLM: prop				
	Plots			P-Value	
Data Set	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	no diff	no diff	worse	no	no change, gain (0.176)
2	no diff	no diff	better	yes- sig. to not sig.	gain (0.103)
3	worse	no diff	better	no	gain (0.000775)
4	no diff	no diff	no diff	no	gain (0.78334, 0.547365)
5	no diff	better	worse	yes- sig. to not sig.	gain (0.300359)

Appendix G: Melanie Underwood

Data Set 1

Reference:

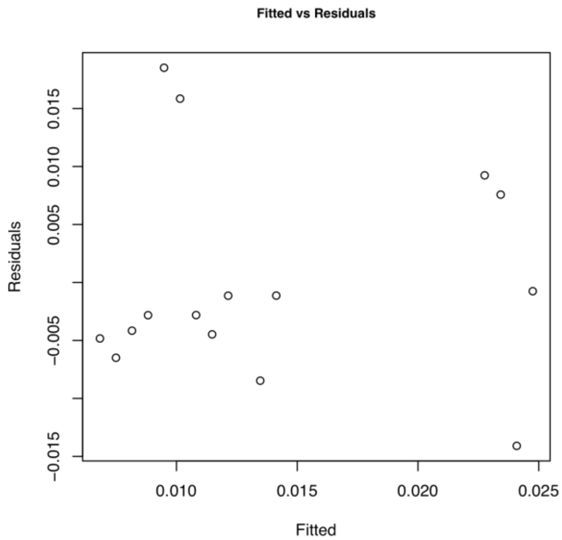
Cramer, S.P. and Vigg, S.C. (1999) Estimation of probable harvest rates on Cowlitz river Chinook salmon during the 1940's and 1950's. Technical Report. S.P. Cramer & Associates Inc.

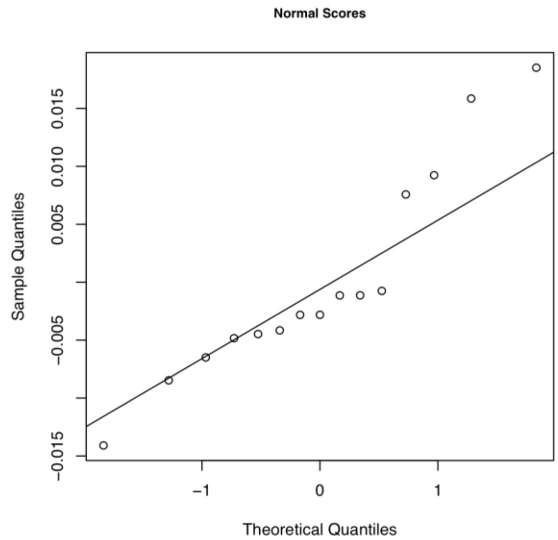
Raw Data:

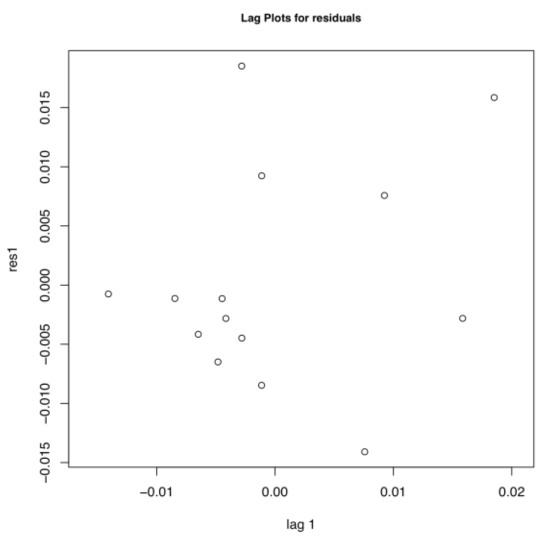
Year	Survival Rate at year 2	Arcsine
1965	0.024	8.912
1966	0.010	5.739
1967	0.031	10.141
1968	0.032	10.305
1981	0.013	6.547
1982	0.005	4.055
1984	0.011	6.020
1985	0.007	4.799
1986	0.008	5.132
1987	0.026	9.279
1988	0.028	9.633
1989	0.006	4.443
1990	0.004	3.626
1991	0.001	1.812
1992	0.002	2.563

Non-Transformed Data:

LM: Linear Regression



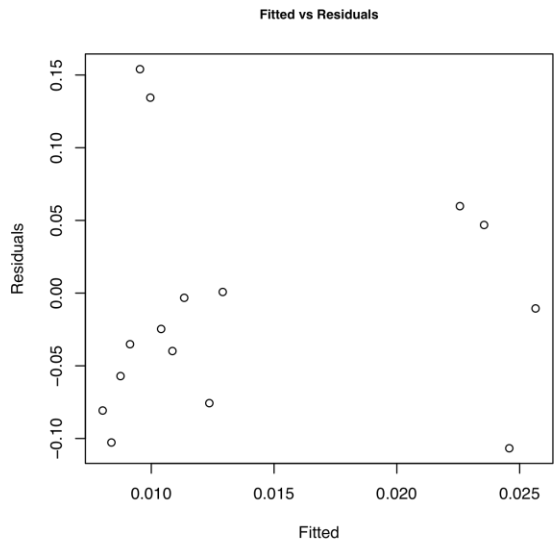


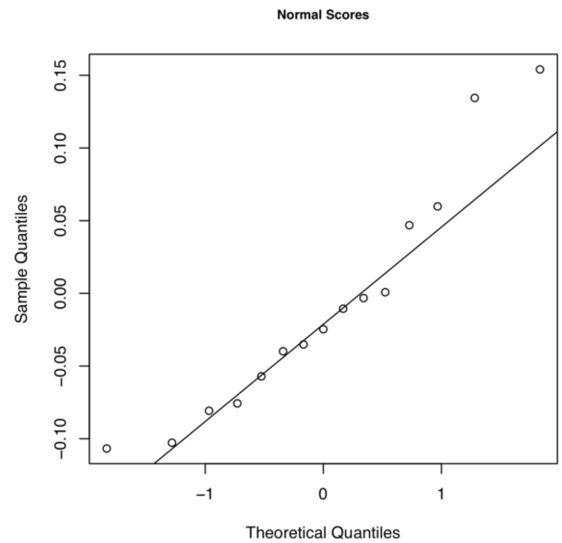


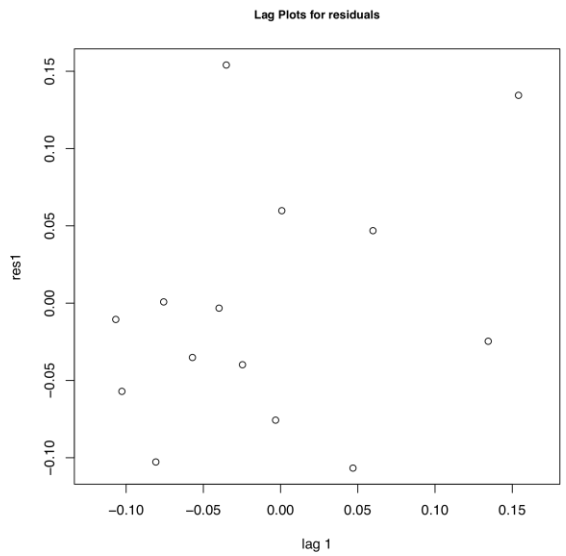
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	0.00060545	0.00060545	6.9726	0.02037
Year	1	0.0005929	0.0005929	6.828	0.02147
Residuals	13	0.00112883	8.68331E-05		

GLM: Logistic Regression





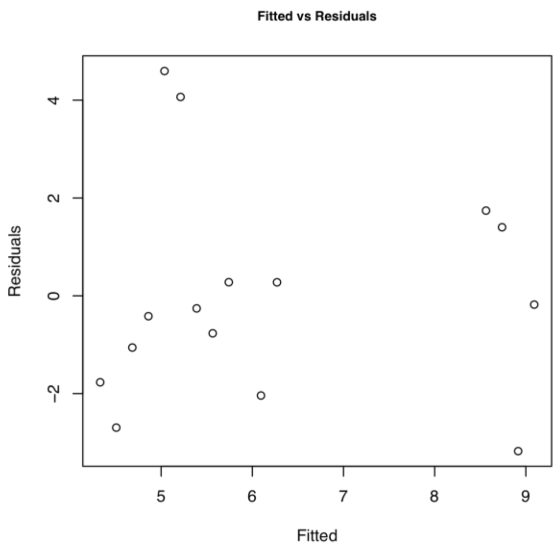


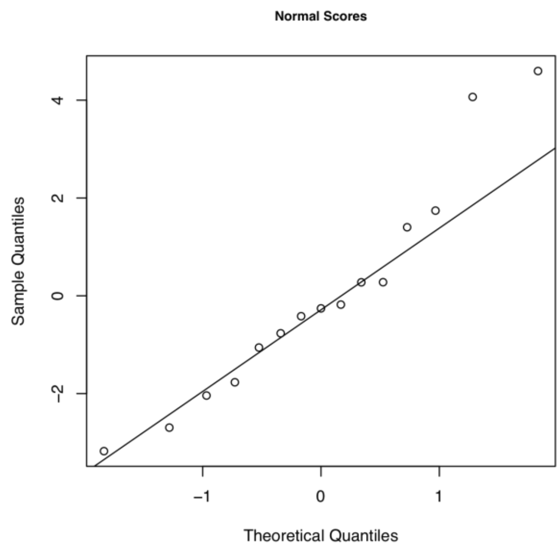
ANODEV Table

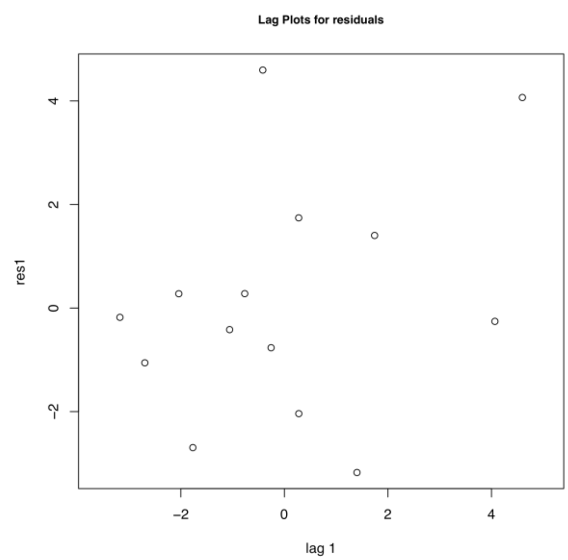
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			14	0.128634	
Year	1	0.040052	13	0.088582	0.8414

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
(Intercept)	1	43.328	43.328	8.112	0.01371
Year	1	41.831	41.831	7.8317	0.01507

Residuals	13	69.436	5.341230769

Data Set 2

Reference:

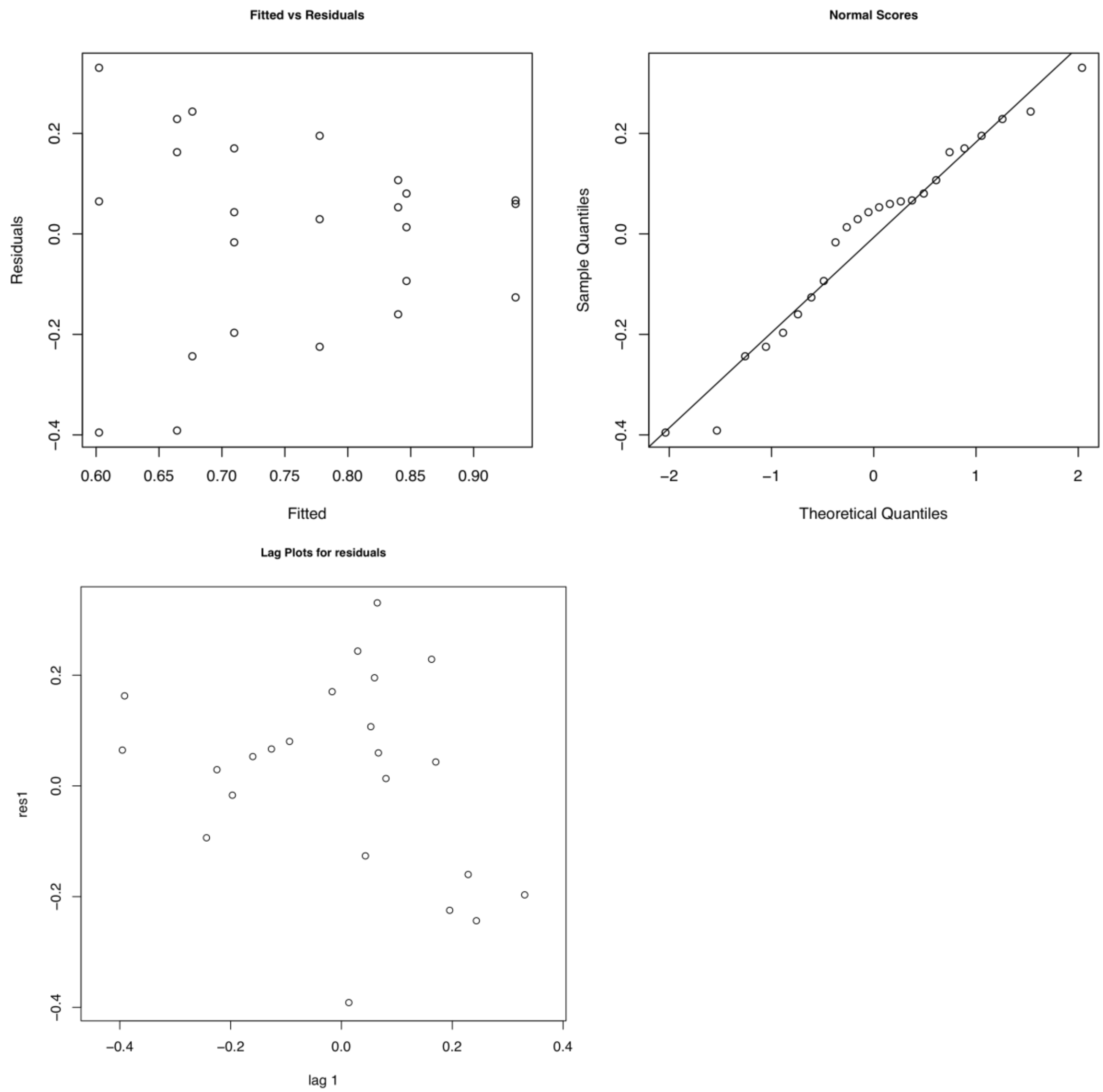
Boutin, C. *et al.* (2010). Measuring variability in phytotoxicity testing using crop and wild plant species. *Environmental Toxicology and Chemistry*, 29: 327–337.

Raw Data:

Species	Max Germination	Arcsine
B. perennis	0.947	1.338495364
B. perennis	0.893	1.237554162
B. perennis	0.68	0.96953211
C. cyanus	0.893	1.237554162
C. cyanus	0.827	1.141828152
C. cyanus	0.273	0.549773383
D. purpurea	0.86	1.187299323
D. purpurea	0.927	1.297211028
D. purpurea	0.753	1.050668637
l. helenium	0.433	0.718196017
l. helenium	0.92	1.284039775
P. vulgaris	0.807	1.115957366
P. vulgaris	0.553	0.83849792
P. vulgaris	0.973	1.405731003
R. crispus	0.993	1.487032405
R. crispus	1	1.570796327
R. crispus	0.807	1.115957366
R. hirta	0.753	1.050668637
R. hirta	0.88	1.217054721
R. hirta	0.693	0.983543958
R. hirta	0.513	0.798399629
S. canadensis	0.933	1.308971536
S. canadensis	0.667	0.955670216
S. canadensis	0.207	0.472341362

Non-Transformed Data:

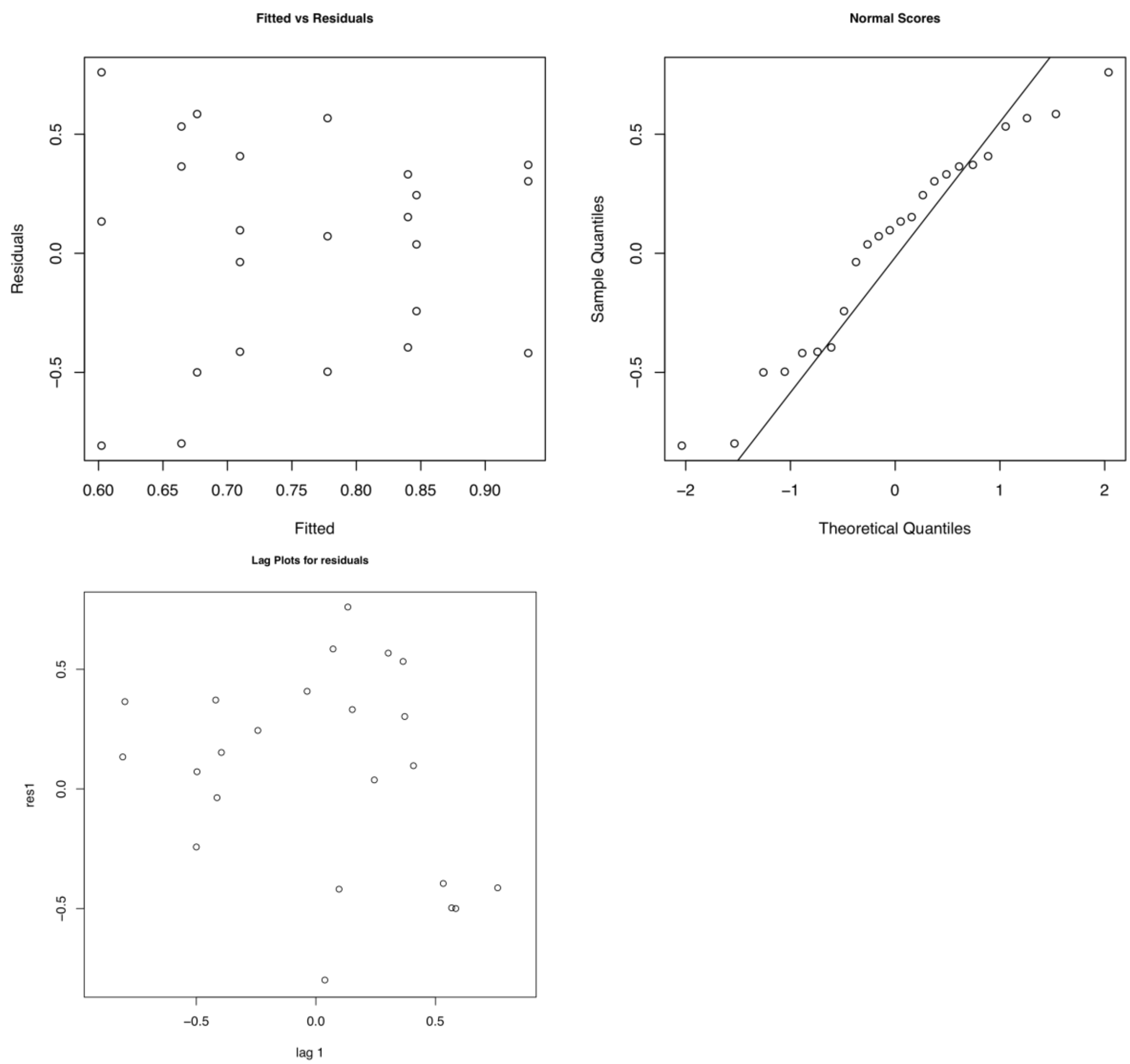
LM: Linear Regression



ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	2.15053	2.15053	40.0635	1.00E-05
Species	7	0.25875	0.036964286	0.6886	0.6806
Residuals	16	0.85885	0.053678125		

GLM: Logistic Regression

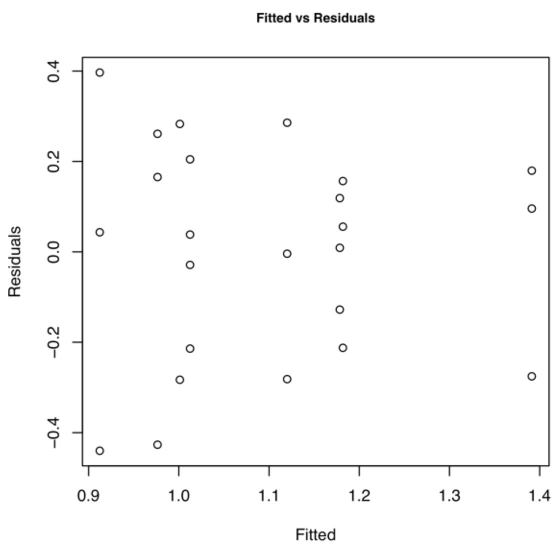


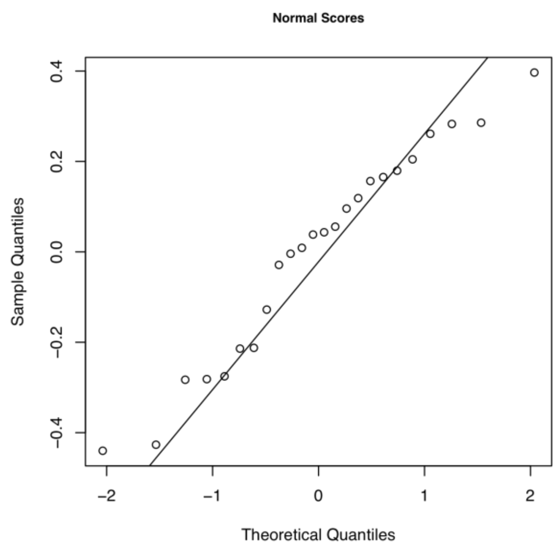
ANODEV Table

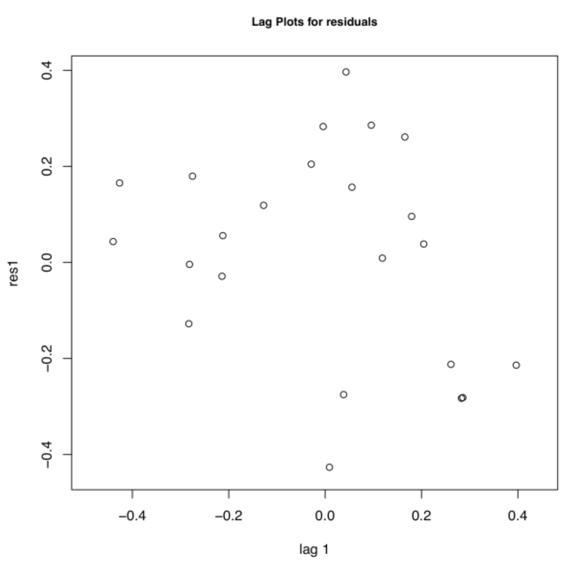
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			23	6.1549	
Species	7	1.5226	16	4.6323	0.9815

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	4.1658	4.1658	52.8254	1.88E-06
Species	7	0.4958	0.070828571	0.8982	0.5311
Residuals	16	1.2618	0.0788625		

Data Set 3

Reference:

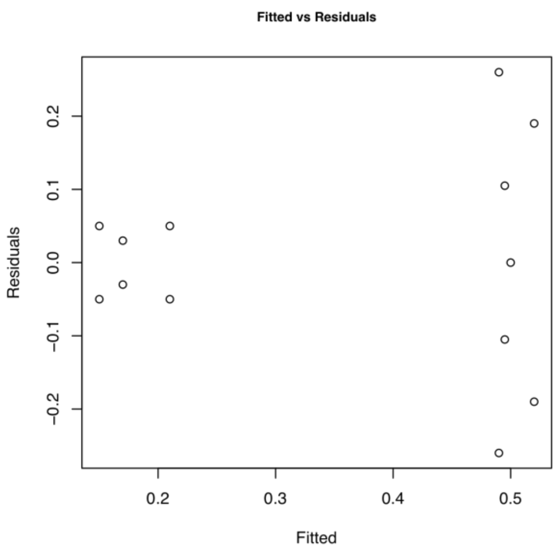
Tang, Y.A. (1970). Evaluation of balance between fishes and available fish foods in multispecies fish culture ponds in Taiwan. *Transactions of the American Fisheries Society*. 99(4):708

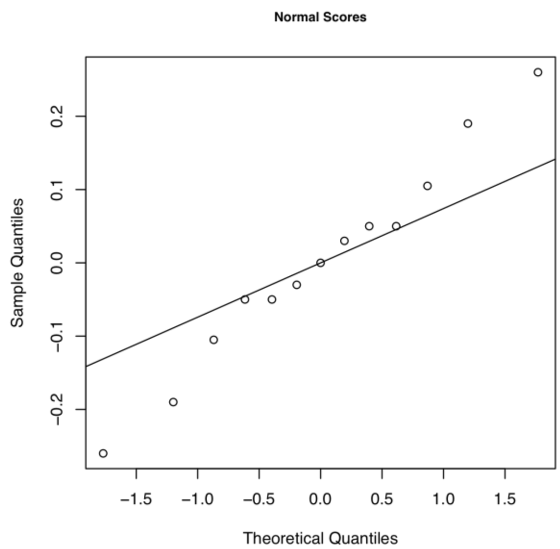
Raw Data:

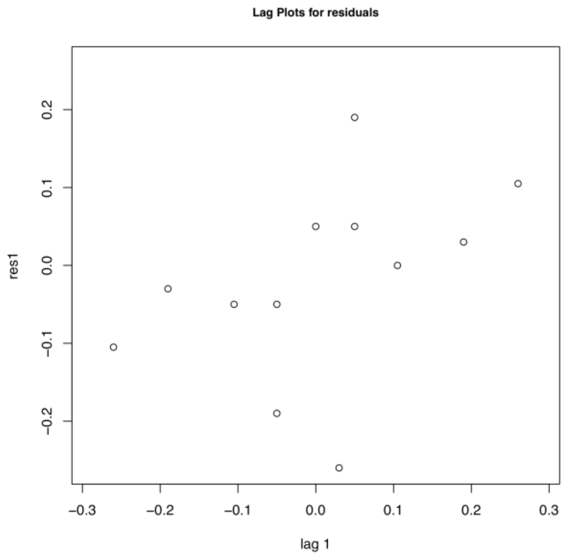
Species	Ave Mortality	Arcsine
Silver Carp	0.14	21.97275978
Grey mullet	0.33	35.06156296
Bighead carp	0.16	23.57817848
Grass carp	0.1	18.43494882
Common carp	0.39	38.6454835
Sea perch	0.23	28.65818058
Silver Carp	0.2	26.56505118
Grey mullet	0.71	57.41729374
Bighead carp	0.26	30.65729899
Grass carp	0.2	26.56505118
Black carp	0.5	45
Common carp	0.6	50.76847952
Sea perch	0.75	60

Non-Transformed Data:

LM: Linear Regression



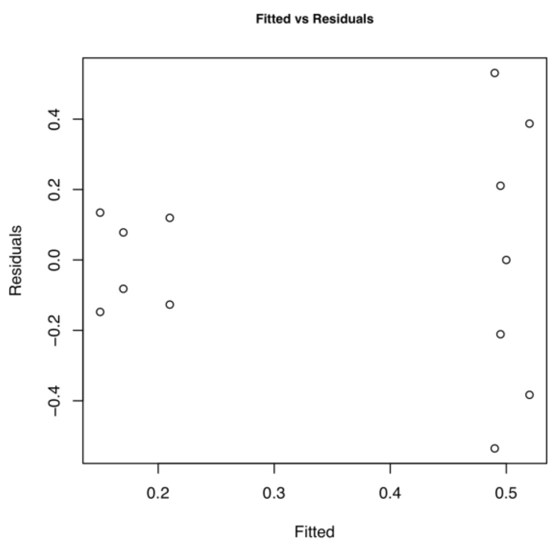


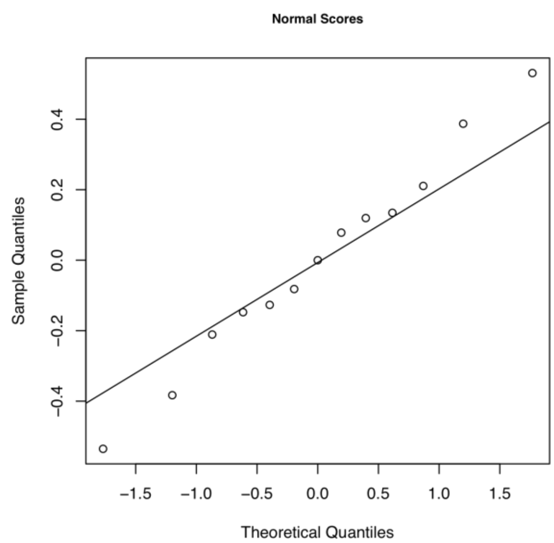


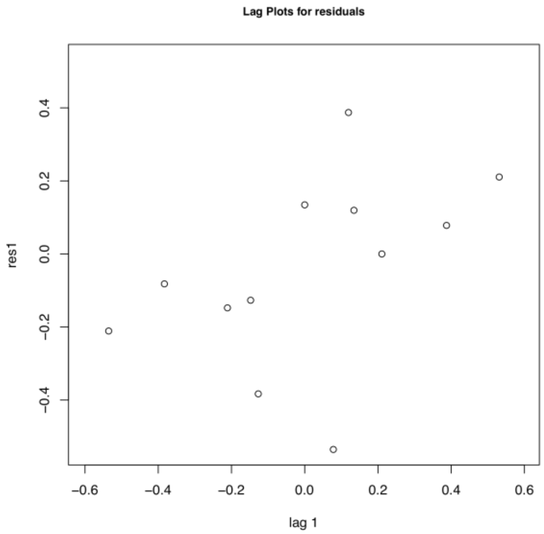
ANOVA Table:

	Df	Sum of Sq.	Mean Sq.	F-Value	p-value
NULL	1	0.0882	0.0882	2.1936	0.1891
Species	6	0.34552	0.057586667	1.4322	0.3369
Residuals	6	0.24125	0.040208333		

GLM: Logistic Regression





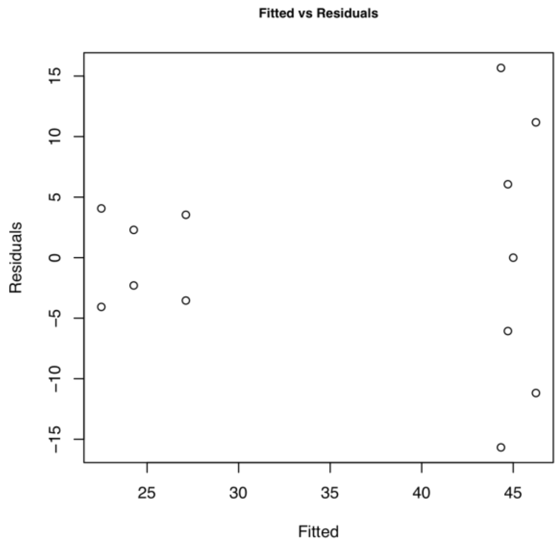


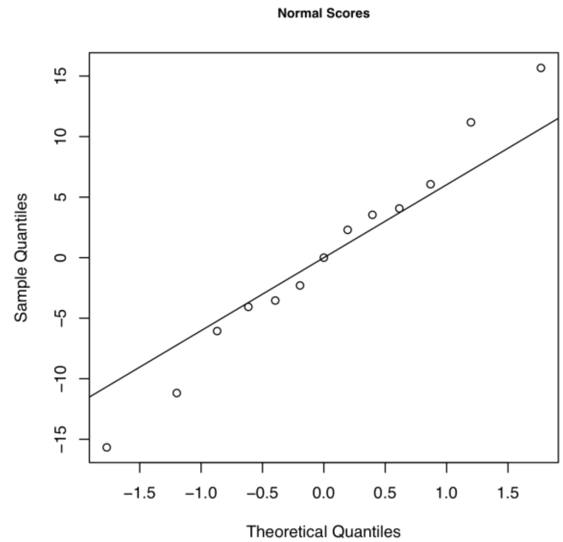
ANODEV Table

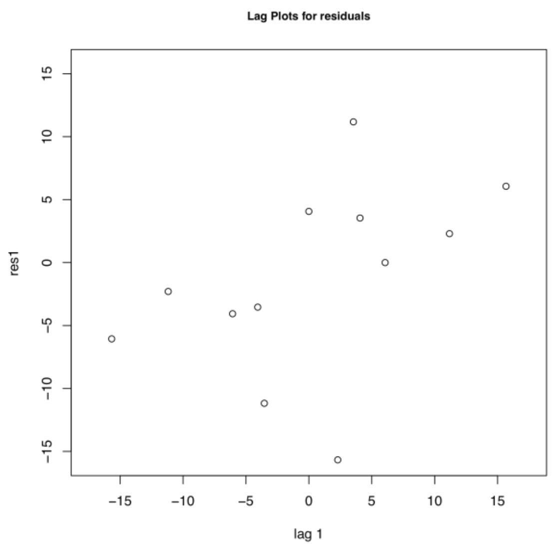
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			12	2.6254	
Species	6	1.5882	6	1.0372	0.9534

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Null	1	1470.74	1470.74	9.9917	0.01954
Species	6	1376.91	229.485	1.559	0.30162
Residuals	6	883.18	147.1966667		

Data Set 4

Reference:

Heman, M.L. (1969). Manipulation of fish populations through reservoir drawdown. *Transactions of the American Fisheries Society*. 98(2):293

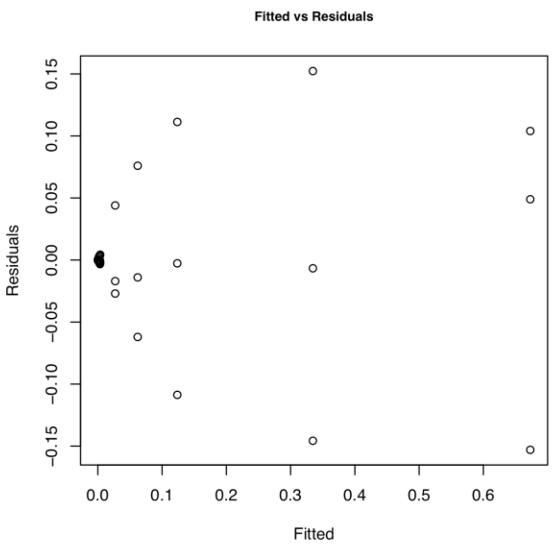
Raw Data:

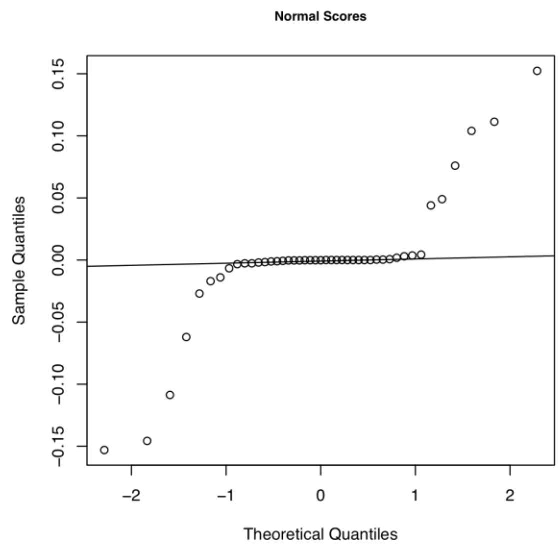
Food Item	Percent by weight	Arcsine
Bluegill	0.189	25.76883369
Redear sunfish	0	0
Largemouth bass	0.01	5.739170477
Unidentified fish	0.015	7.034933874
Crayfish	0.777	61.82093243
Ephemeroptera	0.007	4.799319192
Diptera	0.0005	1.281279367
Hemiptera	0	0
Odonata	0.001	1.812153747
Trichoptera	0	0
Unidentified insect	0.0005	1.281279367
Annelida	0	0
Tadpole	0	0
Plant material	0.001	1.812153747
Unidentified animal remains	0	0
Bluegill	0.328	34.939618
Redear sunfish	0	0
Largemouth bass	0.071	15.45362038
Unidentified fish	0.121	20.35590093
Crayfish	0.52	46.14622139
Ephemeroptera	0	0
Diptera	0	0
Hemiptera	0	0
Odonata	0	0
Trichoptera	0	0
Unidentified insect	0	0
Annelida	0	0
Tadpole	0.048	12.65553258
Plant material	0.008	5.131547949
Unidentified animal remains	0	0
Bluegill	0.487	44.25507092
Redear sunfish	0	0
Largemouth bass	0	0

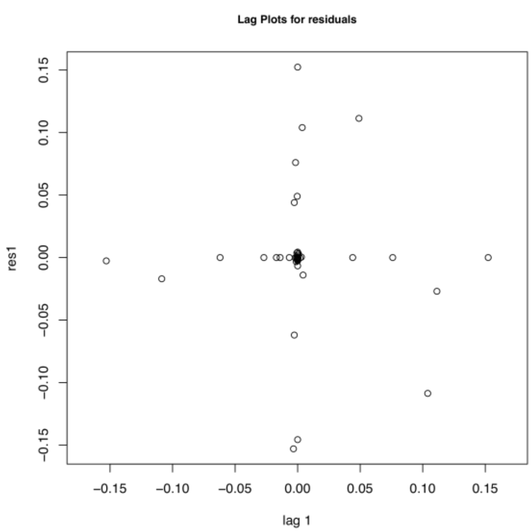
Unidentified fish	0.235	28.99727259
Crayfish	0.722	58.1796884
Ephemeroptera	0.003	3.139790321
Diptera	0.003	3.139790321
Hemiptera	0.0005	1.281279367
Odonata	0.005	4.054807228
Trichoptera	0.001	1.812153747
Unidentified insect	0	0
Annelida	0	0
Tadpole	0.138	21.80713846
Plant material	0.002	2.563200041
Unidentified animal remains	0	0

Non-Transformed Data:

LM: Linear Regression



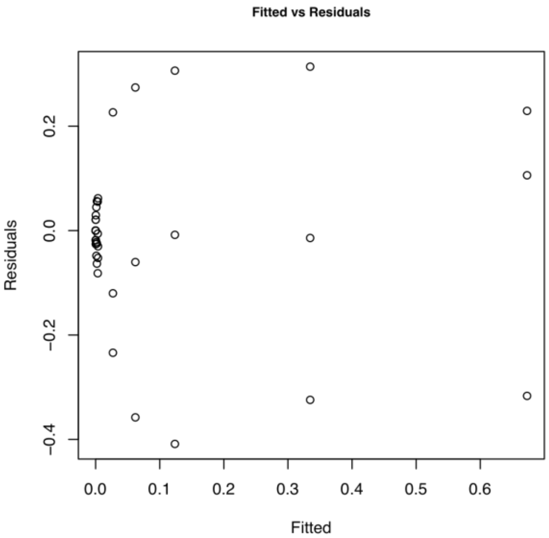


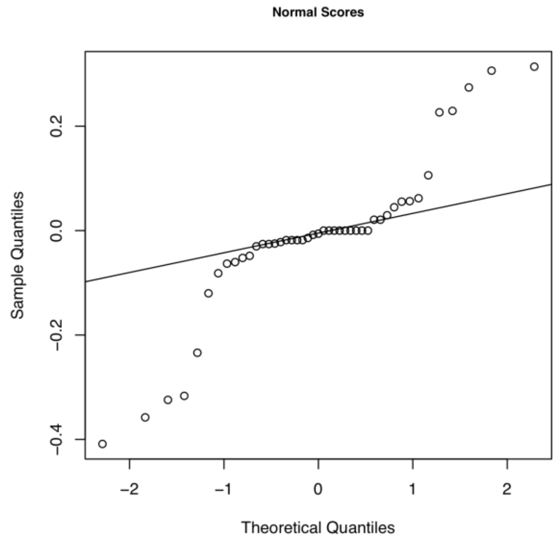


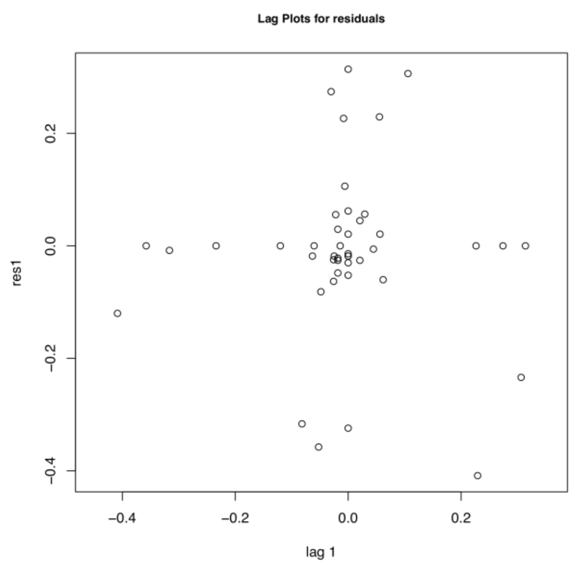
ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	0	0	0	1
Food.Item	14	1.45133	0.103666429	26.323	4.94E-13
Residuals	30	0.11815	0.003938333		

GLM: Logistic Regression





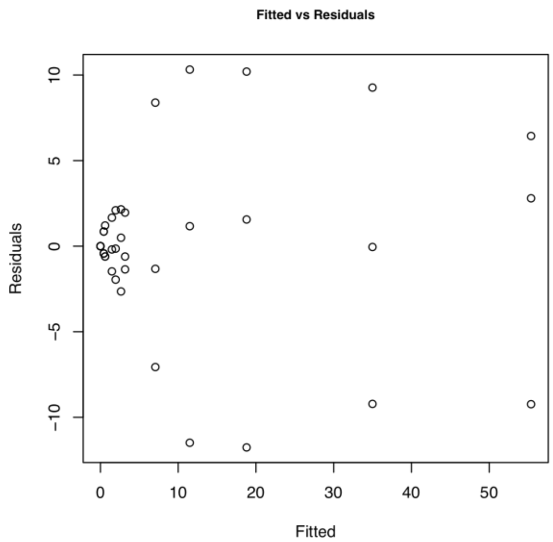


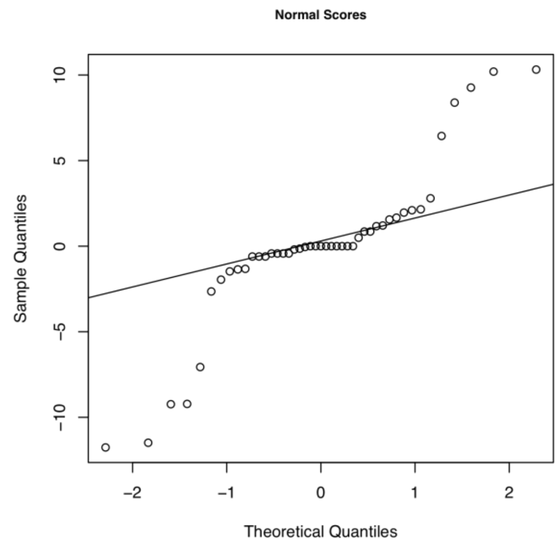
ANODEV Table

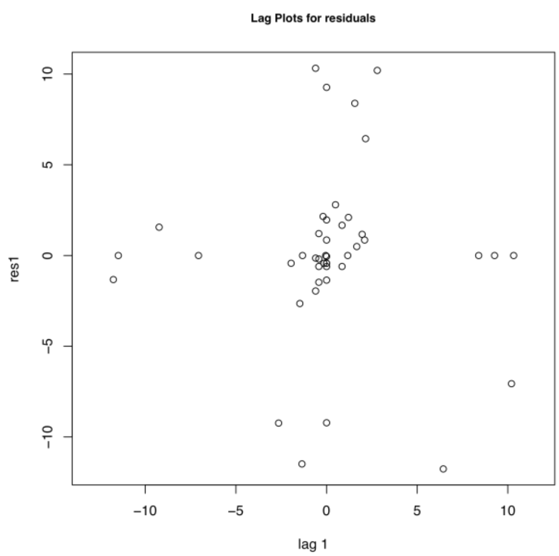
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			44	14.07484	
Food.Item	14	13.08436	30	0.99048	0.5199

Arcsine Transformed Data:

LM: Linear Regression







\

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	0	0	0	1
Food.Item	14	10718.8	765.6285714	24.236	1.49E-12

Residuals	30	947.7	31.59

Data Set 5

Reference:

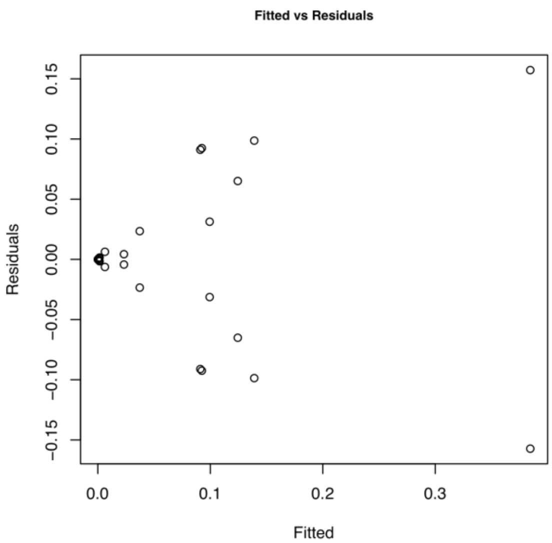
Gaskin, D. E. and Cawthorn, M. W. (1967) Diet and feeding habits of the sperm whale (Physeter Catodon L.) in the cook strait region of New Zealand. *New Zealand Journal of Marine and Freshwater Research* 1(2):156 — 179

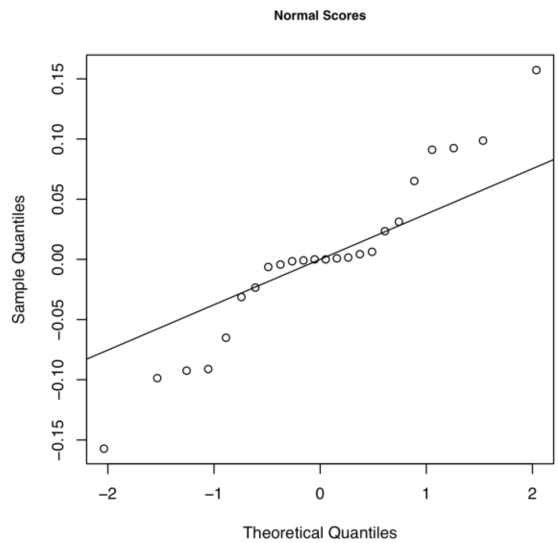
Raw Data:

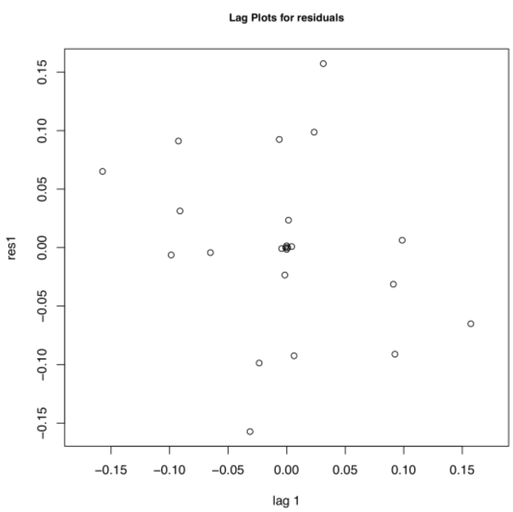
Beak Type	Percentage	Acrsine
Ai	0.1895	25.8054017
Aii	0.2272	28.46716246
Aiii	0.0681	15.12701894
Bi	0.1822	25.26775022
Bii	0	0
Biii	0.0126	6.445018472
C	0.2377	29.17934032
D	0.0607	14.26302966
E	0.0031	3.191744512
F	0	0
G	0	0
Unidenitfied	0.0189	7.901892963
Ai	0.0593	14.09414577
Aii	0.5417	47.39201245
Aiii	0.1307	21.19385339
Bi	0	0
Bii	0.1849	25.46756014
Biii	0	0
C	0.0404	11.5952969
D	0.0138	6.746307708
E	0	0
F	0	0
G	0.0017	2.363035358
Unidenitfied	0.0275	9.545526256

Non-Transformed Data:

LM: Linear Regression





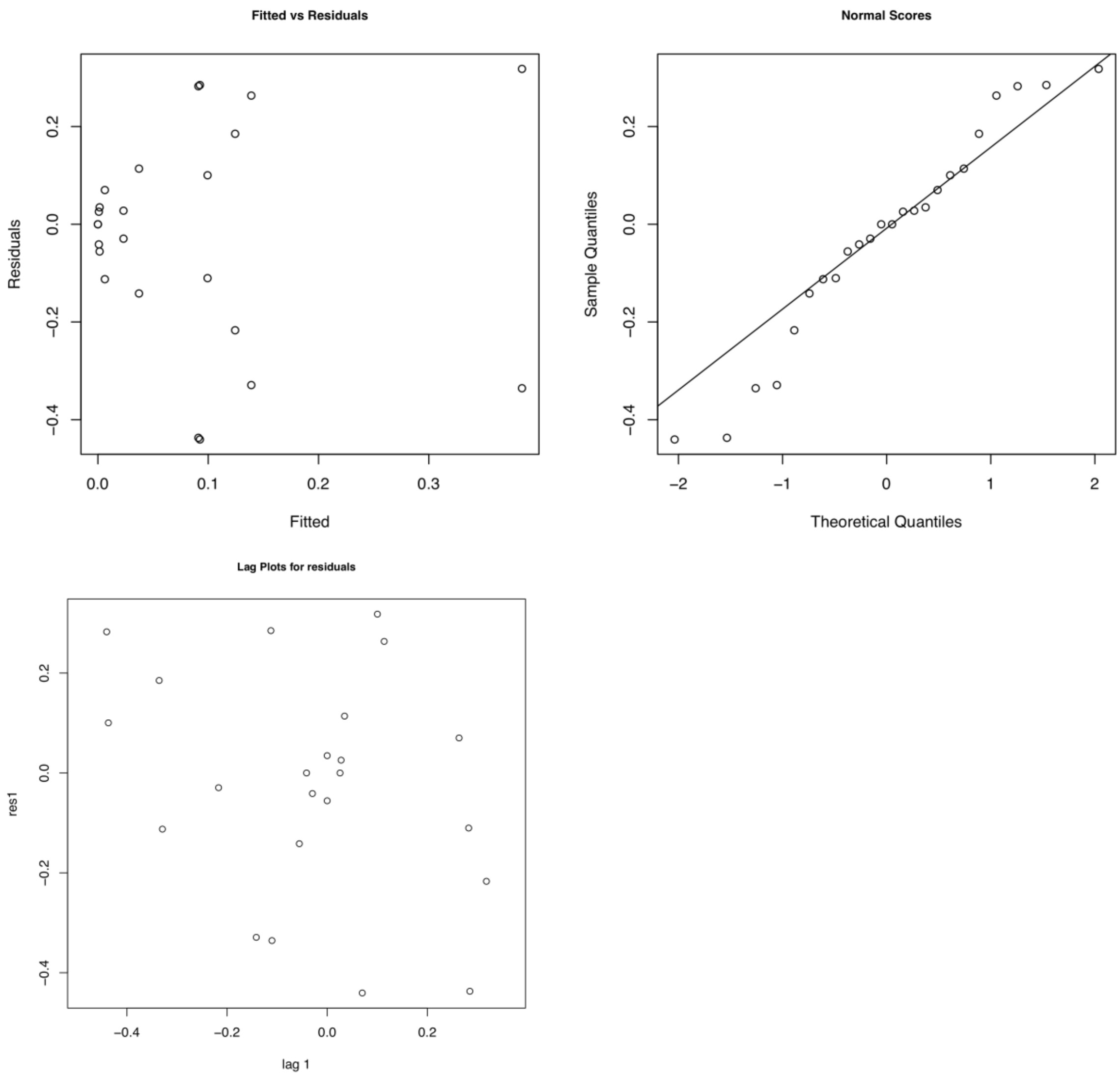


ANOVA Table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	0.030951	0.030951	3.2503	0.09656

Beak.Type	11	0.255948	0.023268	2.4435	0.07003
Residuals	12	0.114269	0.009522417		

GLM: Logistic Regression



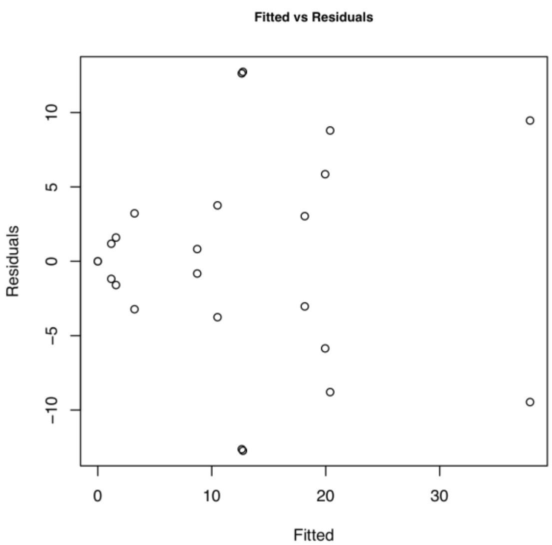
ANODEV Table

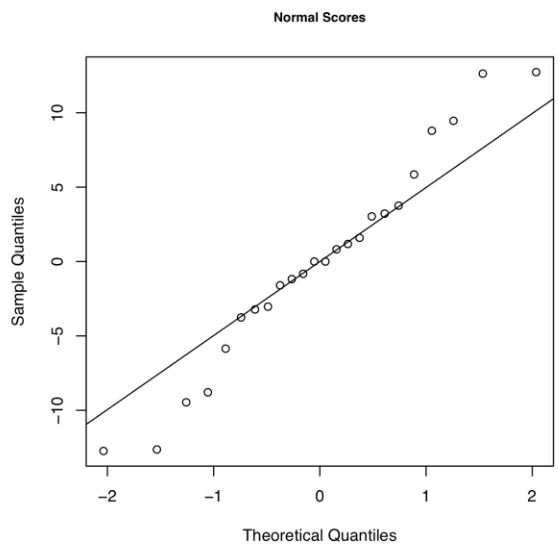
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			23	4.0359	

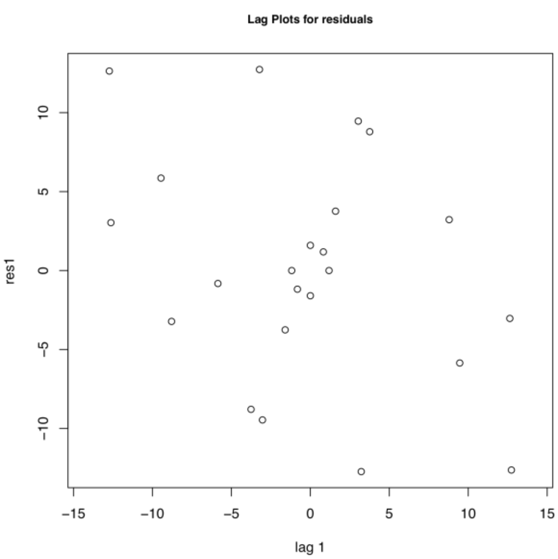
Food.Item	11	2.936	12	1.0999	9.92E-01

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NULL	1	795.99	795.99	8.5099	0.01291
Beak.Type	11	2606.64	236.9672727	2.5334	0.0626
Residuals	12	1122.44	93.53666667		

Data Summary:

Data Set	**LM: Non-transformed vs. LM: Transformed**				
	Plots			P-Value	
	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	NOD	NOD	Better	No	Loss (0.0064)
2	Better	NOD	Worse	No	Loss (0.1495)
3	NOD	NOD	Better	No	Loss (0.02538)
4	Better	Better	NOD	No	Gain (1.00E-12)
5	Better	Better	Better	No	Loss (0.00743)

Data Set	**LM: Non-transformed vs. GLM: Non-transformed**				
	Plots			P-Value	
	Res. Vs. Fits	Res. Vs. Res lag	Normality	Change in decision?	Loss or Gain?
1	NOD	NOD	Better	Yes	Gain (0.82633)
2	NOD	NOD	Worse	No	Gain (0.4504)
3	NOD	NOD	NOD	No	Gain (0.65178)
4	NOD	NOD	NOD	Yes	Gain (0.52)
5	Better	NOD	NOD	No	Gain (0.929)

Appendix H

Data set 1

Reference:

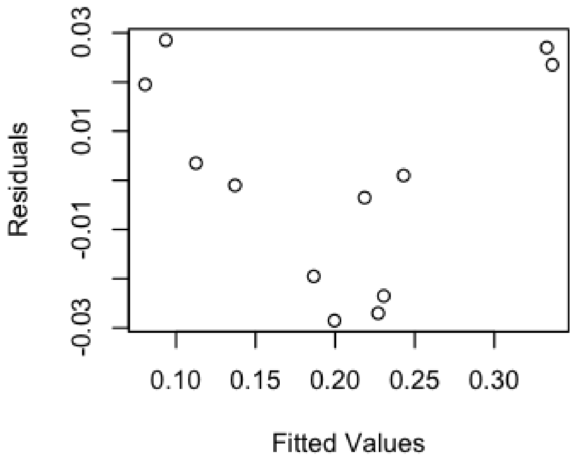
Mayr, E. (1939) The Sex Ratio in Wild Birds. *The American Naturalists* 73 (745) 156-179

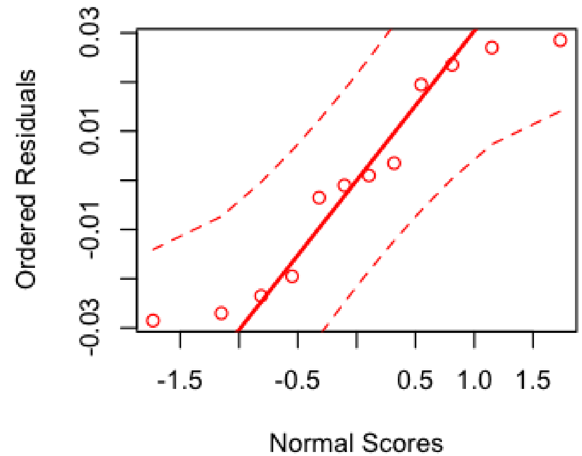
Raw Data:

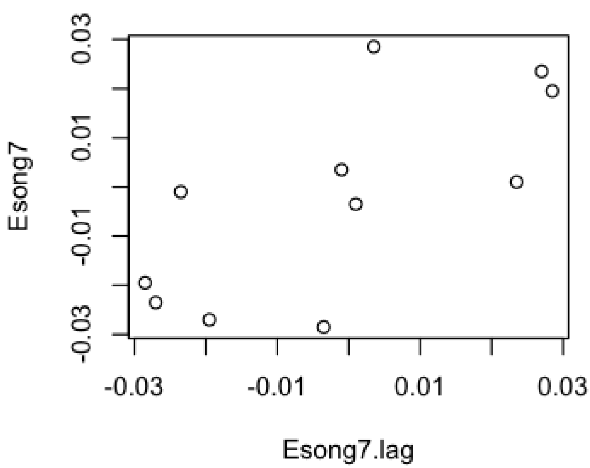
Year	Male Population Decrease from April to June	Male Arcsine	Female Population Decrease from April to June	Female Arcsine
1930	10.0	18.43	16.7	24.12
1931	12.2	20.44	17.1	24.43
1932	11.6	19.91	21.5	27.62
1933	13.6	21.64	24.4	29.60
1934	20.7	27.06	36.0	36.87
1935	20.0	26.56	36.0	36.87

Non-Transformed Data:

LM: Linear Regression



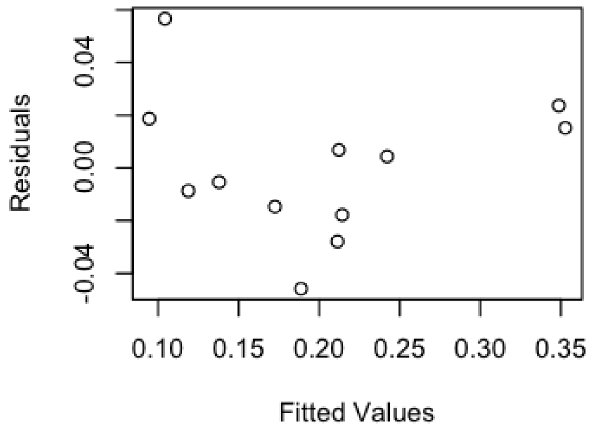


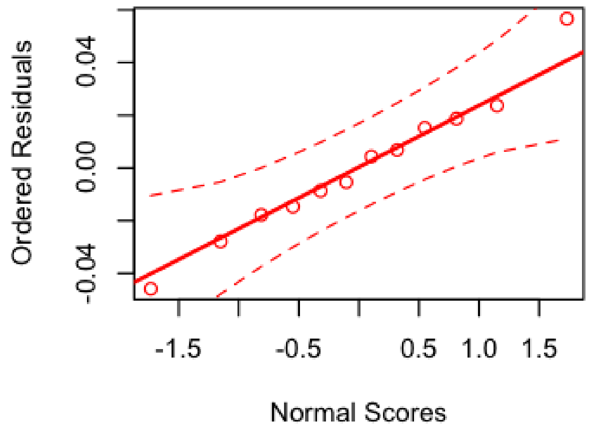


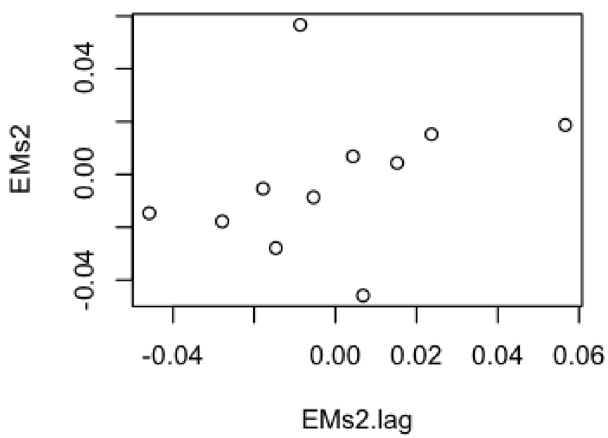
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SEX	1	0.033708	0.033708	33.8842	0.002113 **
fyear	5	0.043894	0.008779	8.8246	0.016037 *
Residuals	5	0.004974	0.000995		

GLM: Logistic regression





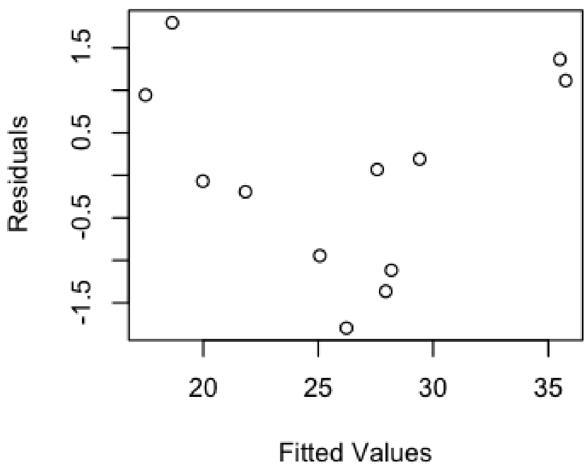


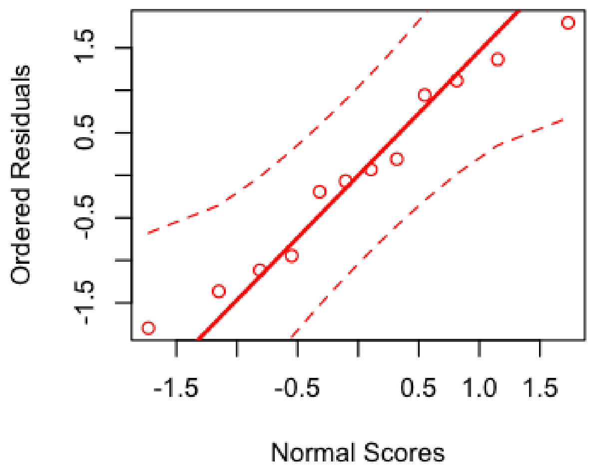
ANODEV Table

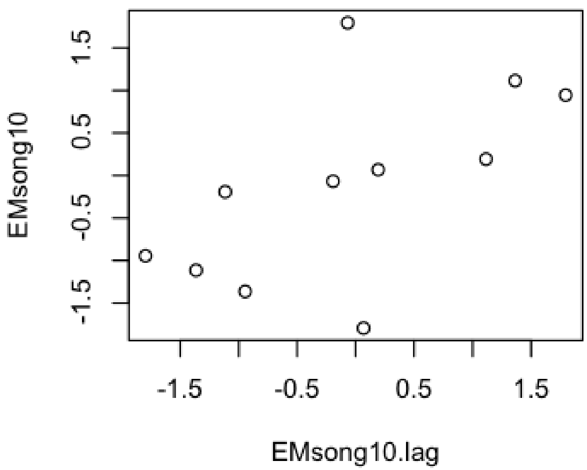
	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			11	0.49531	
SEX	1	0.21287	10	0.28244	0.6445
fyear	5	0.27450	5	0.00793	0.9981

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SEX	1	172.164	172.164	59.212	0.0005912 ***
fyear	5	216.636	43.327	14.902	0.0050291 **
Residuals	5	14.538	2.908		

Data set 2

Reference:

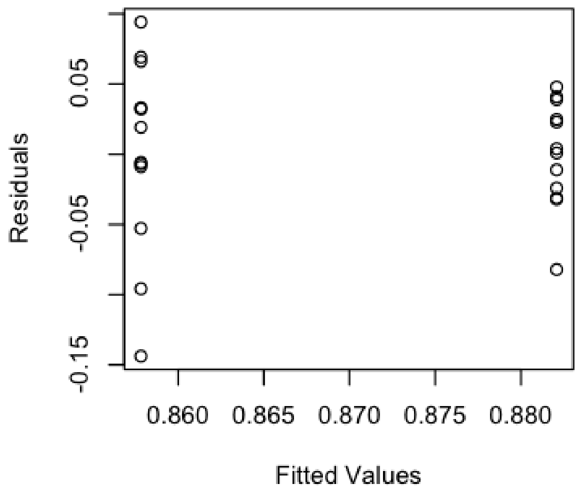
Kosin, I. L. (1958) Effect of simulated airplane sounds on the reproductive functions of the male domestic chicken. *J. Appl. Physiol.* 12(2) : 217-220.

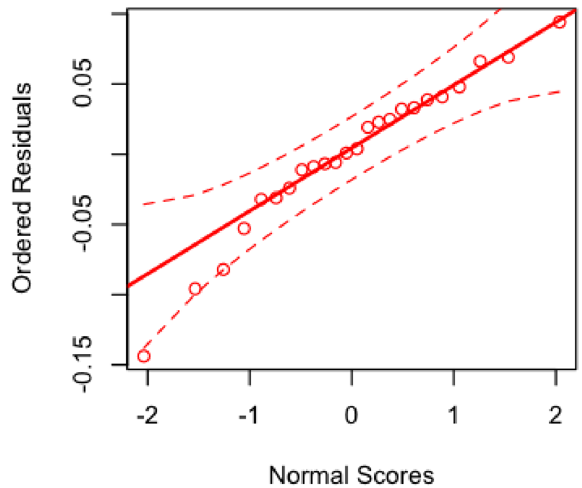
Raw Data:

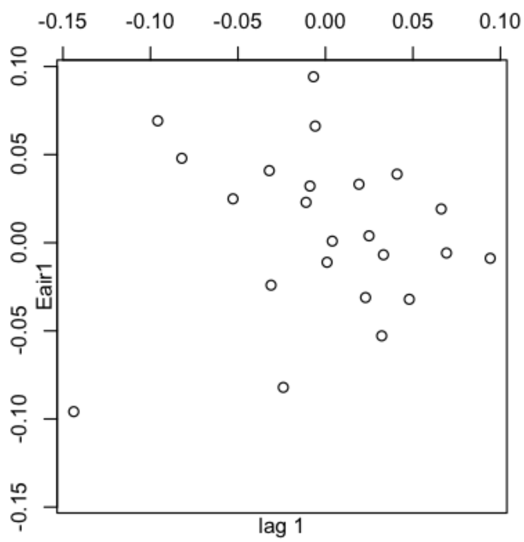
Condition	Percent Normal Sperm	Arcsine
Control	92.1	73.676
Control	92.3	73.890
Control	85.0	67.214
Control	93.0	74.658
Control	80.0	63.43
Control	85.8	67.863
Control	85.1	67.294
Control	90.5	72.048
Control	87.1	68.951
Control	88.3	69.998
Control	88.6	70.267
Control	90.7	72.244
Experiment	80.5	63.795
Experiment	89.0	70.630
Experiment	84.9	67.133
Experiment	95.2	77.344
Experiment	85.1	67.294
Experiment	89.1	70.722
Experiment	87.7	69.47
Experiment	92.4	73.997
Experiment	85.2	67.374
Experiment	92.7	74.325
Experiment	76.2	60.800
Experiment	71.4	57.670

Non-Transformed Data:

LM: Linear Regression



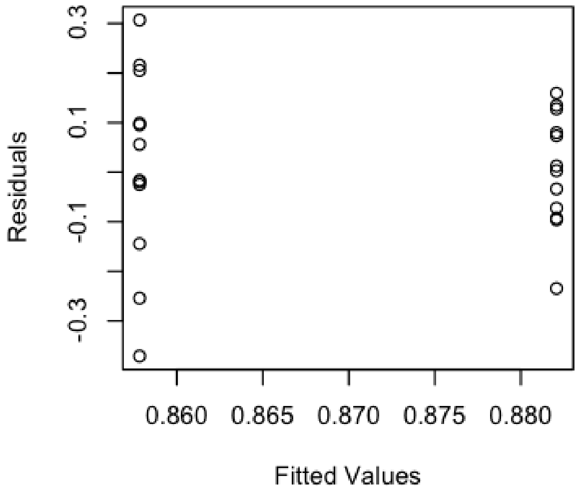


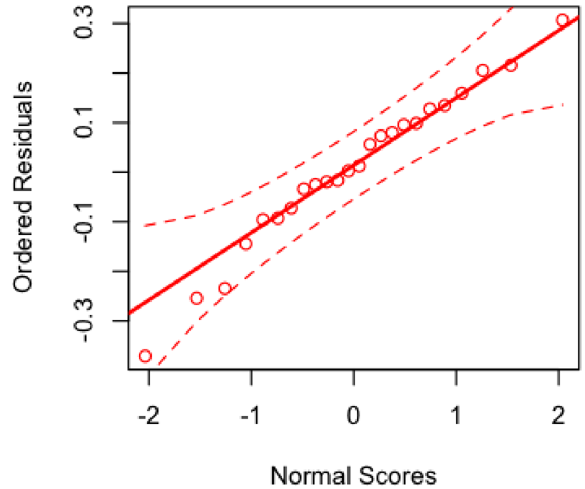


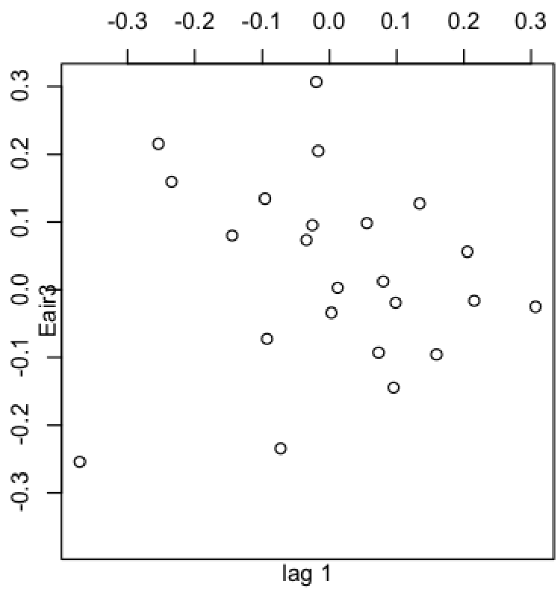
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
CTRLEXPT	1	0.003528	0.0035284	1.1179	0.3018
Residuals	22	0.069437	0.0031562		

GLM: Logistic Regression





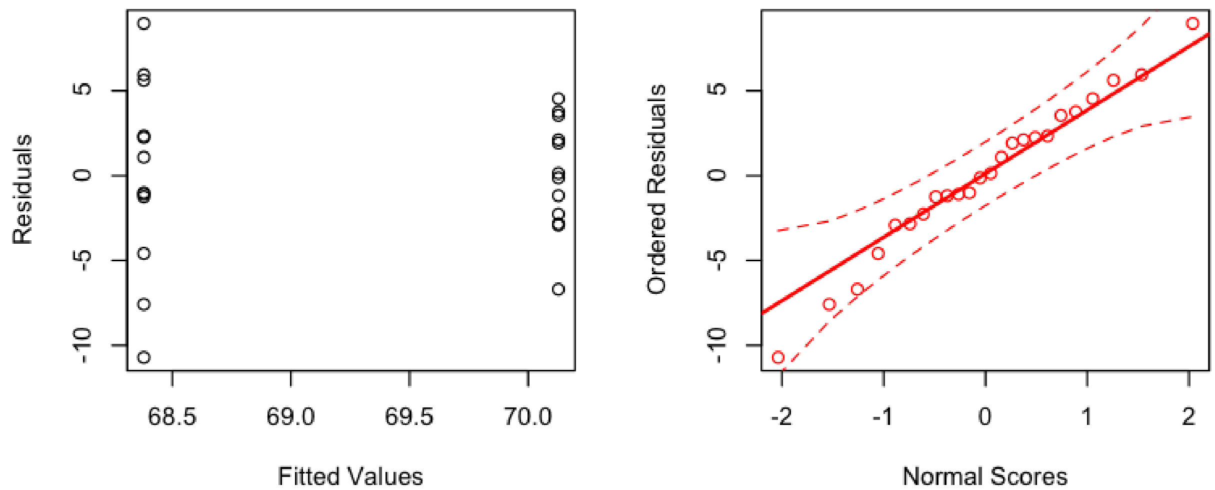


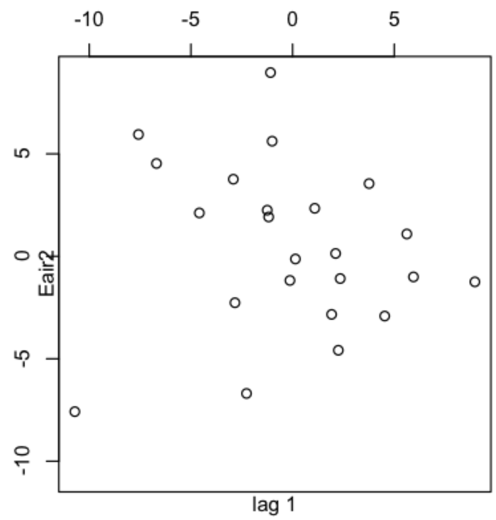
ANODEV Table

	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			23	0.61140	
CTRLEXPT	1	0.031228	22	0.58017	0.8597

Arcsine Transformed Data

LM: Linear Regression





ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
CTRLEXPT	1	18.34	18.343	0.8421	0.3687
Residuals	22	479.21	21.782		

Data set 3

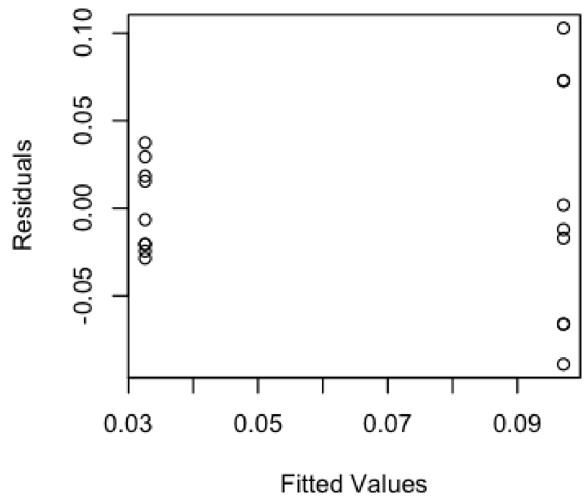
Reference:

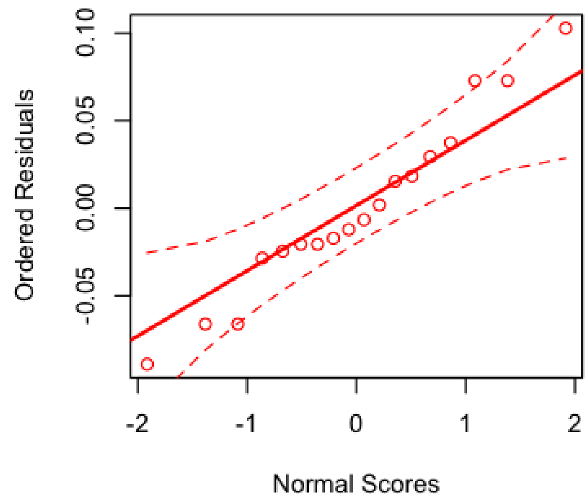
Lombard L.S. and Witte E.J. (1959) Frequency and types of tumors in mammals and birds of the Philadelphia Zoological Garden. *Cancer Research* 19(2), 127-141

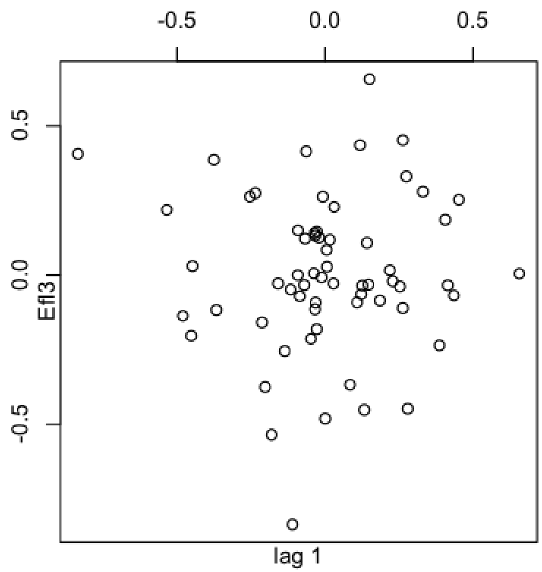
Raw Data:

Years	Percent Tumors	Arcsine
1901-1934	1.2	6.289
1901-1934	2.6	9.279
1901-1934	4.8	12.66
1901-1934	5.1	13.05
1901-1934	0.8	5.132
1901-1934	7.0	15.34
1901-1934	6.2	14.42
1901-1934	1.2	6.289
1901-1934	0.4	3.626
1935-1955	3.1	10.14
1935-1955	9.9	18.34
1935-1955	8.0	16.43
1935-1955	8.5	16.95
1935-1955	17.0	24.35
1935-1955	20.0	26.57
1935-1955	17.0	24.35
1935-1955	3.1	10.14
1935-1955	0.8	5.132

Non-Transformed Data:
LM: Linear Regression



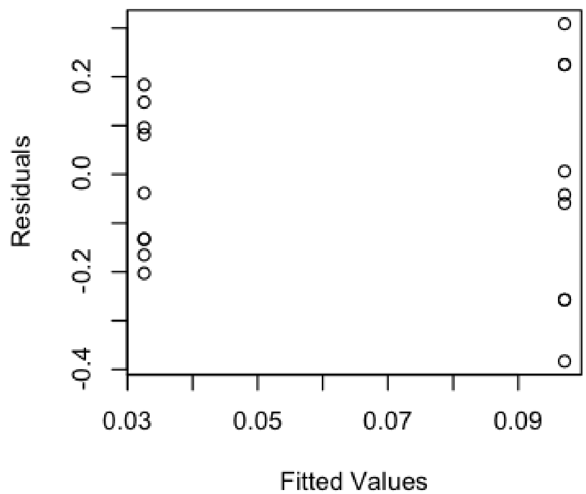


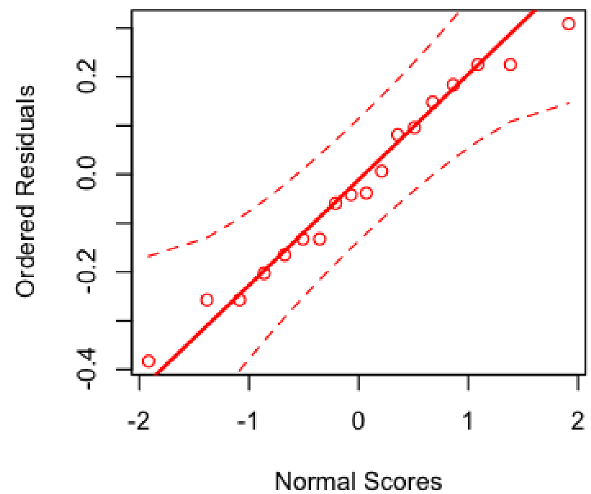


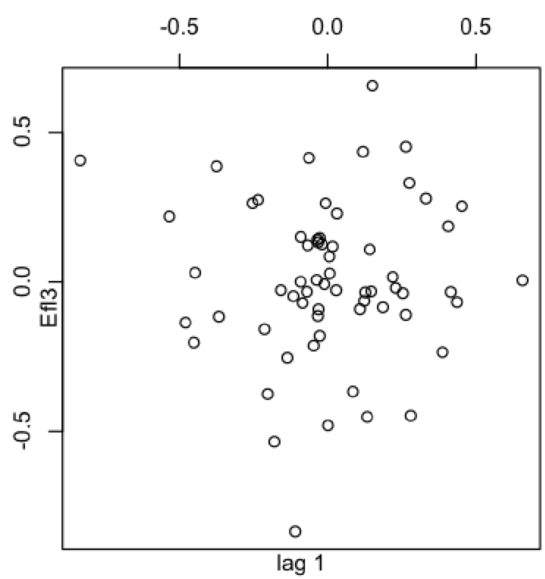
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Years	1	0.018753	0.0187534	6.8992	0.01832 *
Residuals	16	0.043491	0.0027182		

GLM: Logistic Regression





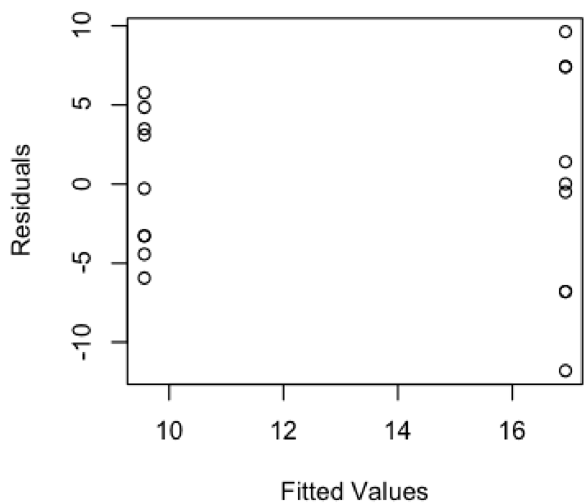


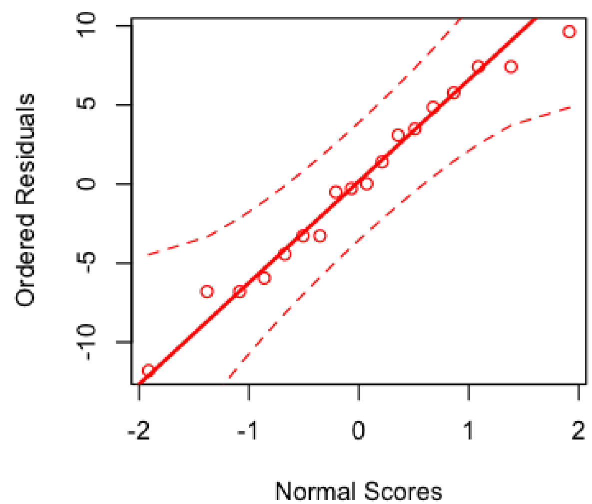
ANODEV Table

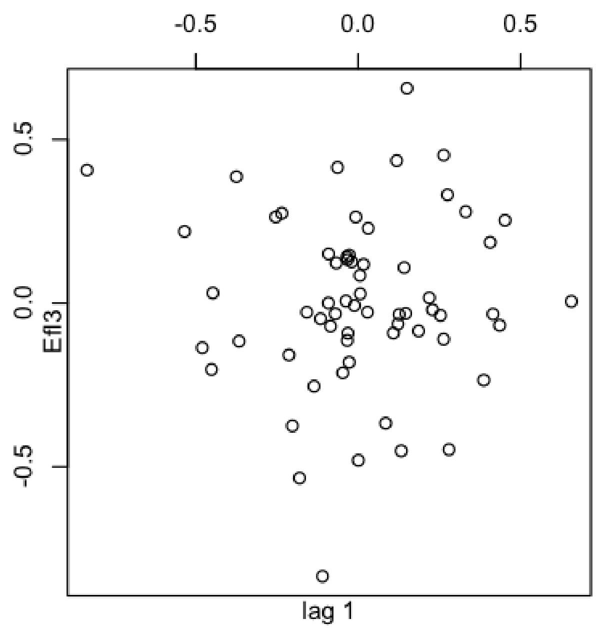
	Df	Deviance Resid.	Df	Resid. Dev	P(>|Chi|)
NULL			17	0.98038	
Years	1	0.32263	16	0.65775	0.57

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Years	1	244.32	244.32	6.6069	0.02054 *
Residuals	16	591.68	36.98		

Data set 4

Reference:

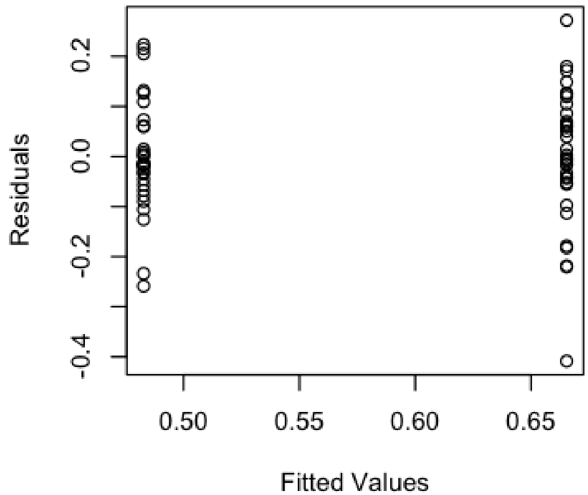
Nice, M.M. (1957) Nesting success in altricial birds. *The Auk* 74(3), 305-321

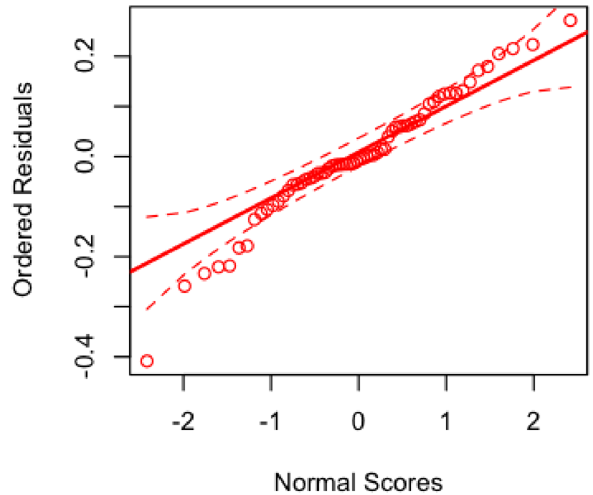
Raw Data:

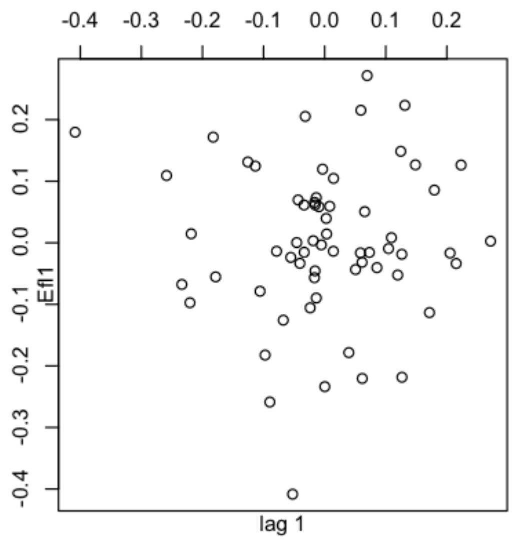
Nest Type	Percent Fledge Success	Arcsine
Open	42.6	40.74
Open	46.6	43.05
Open	68.8	56.04
Open	45.1	42.2
Open	54.4	47.52
Open	44.9	42.07
Open	69.8	56.66
Open	54.2	47.41
Open	49.1	44.48
Open	59.2	50.30
Open	22.4	28.2
Open	39.3	38.82
Open	46.9	43.2
Open	49.7	45
Open	48.6	44.20
Open	46.4	42.94
Open	60.9	51.30
Open	70.6	57.17
Open	61.4	51.59
Open	35.7	36.69
Open	41.5	40.1
Open	24.9	29.93
Open	48.3	44.03
Open	43.7	41.38
Open	46.7	43.11
Open	55.6	48.22
Open	46.9	43.2
Open	40.4	39.47
Open	37.7	37.88
Open	45.9	42.65
Cavity	61	51.35
Cavity	48.7	44.3
Cavity	70.5	57.10
Cavity	66.8	54.82
Cavity	93.7	75.46
Cavity	73.5	59.02
Cavity	62.2	52.06
Cavity	71.6	57.80
Cavity	73.1	58.76
Cavity	64.9	53.67

Nest Type	Percent Fledge Success	Arcsine
Cavity	72.4	58.31
Cavity	65.6	54.1
Cavity	77	61.3
Cavity	68	55.55
Cavity	44.7	41.96
Cavity	79.2	62.87
Cavity	81.4	64.45
Cavity	79	62.73
Cavity	55.2	47.98
Cavity	83.7	66.19
Cavity	48.3	44.03
Cavity	56.8	48.91
Cavity	44.5	41.84
Cavity	72.7	58.5
Cavity	65	53.7
Cavity	63.2	53
Cavity	62.5	52.2
Cavity	75.1	60.07
Cavity	84.5	66.82
Hole	25.7	30.46
Hole	61.3	51.53
Hole	78.5	62.38
Hole	66.2	54.45
Hole	66	54.33

Non-Transformed Data:
LM: Linear Regression



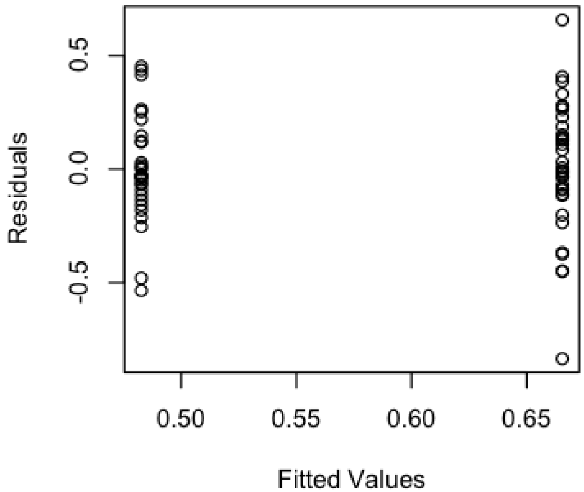


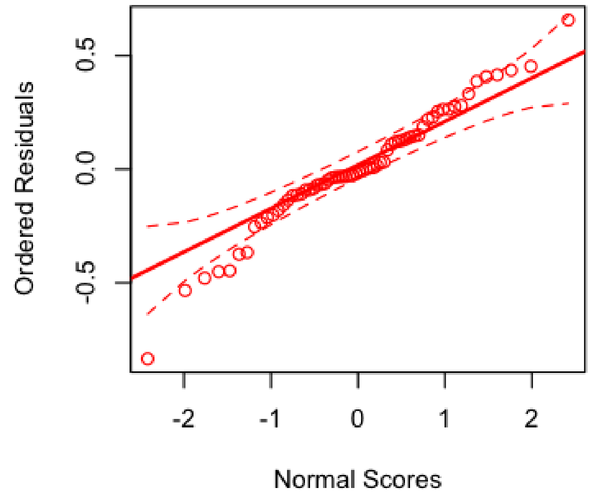


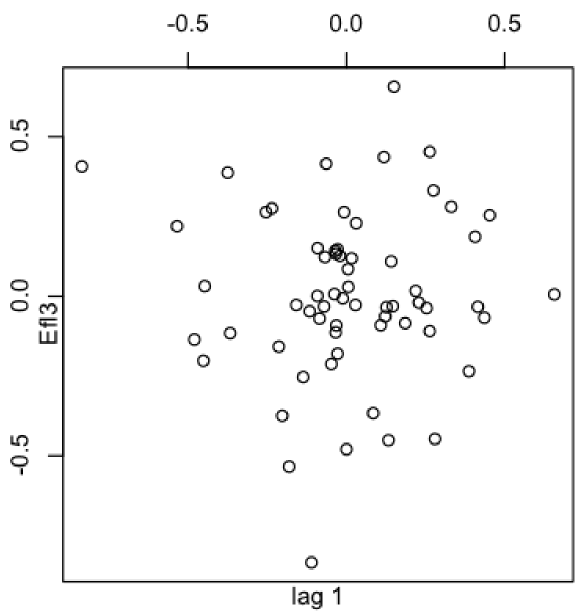
ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NEST	1	0.53203	0.53203	33.762	2.322e-07 ***
Residuals	62	0.97702	0.01576		

GLM: Logistic Regression





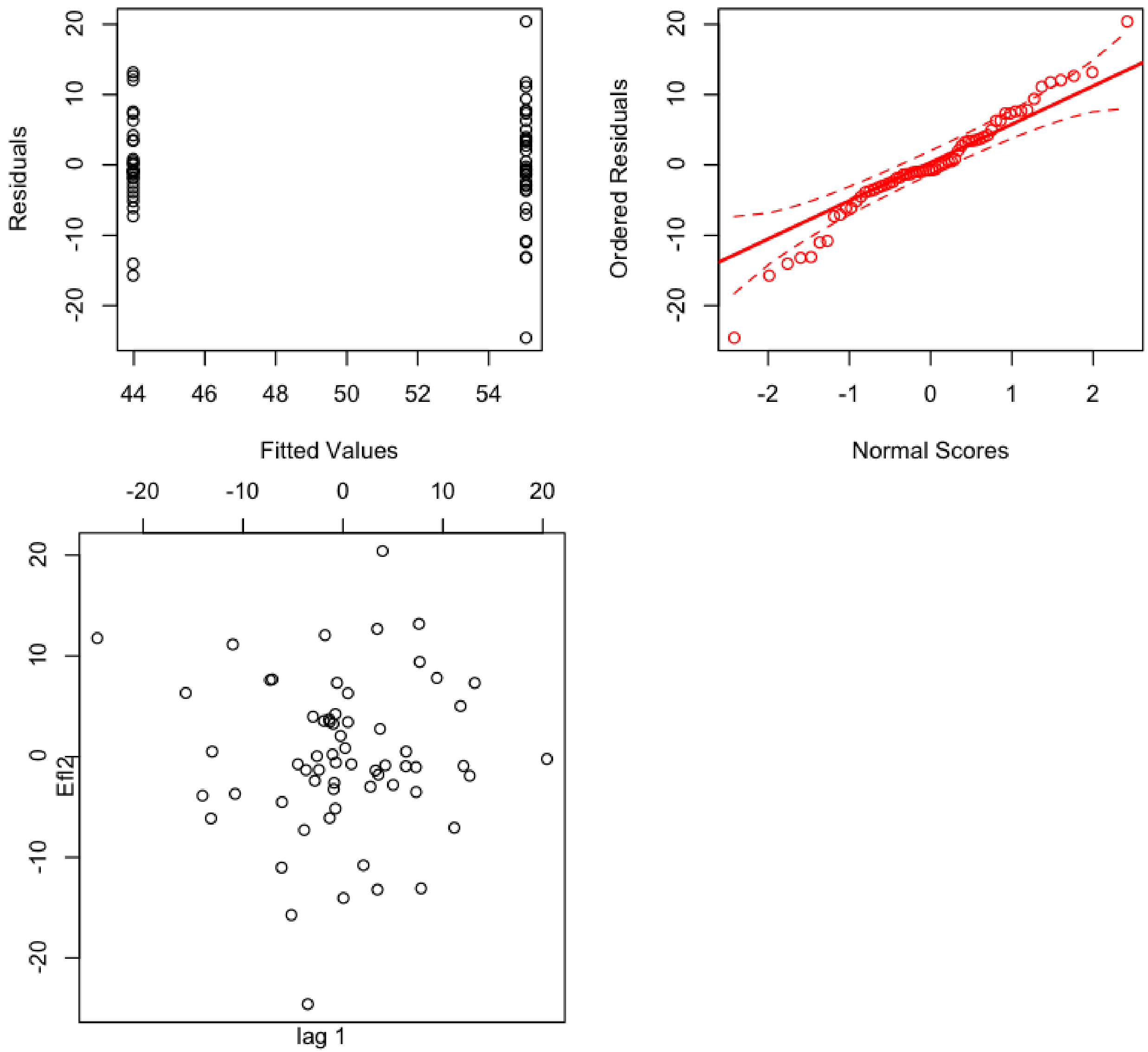


ANODEV Table

	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			63	6.5127	
NEST	1	2.1922	62	4.3205	0.1387

Arcsine Transformed Data:

LM: Linear Regression



ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NEST	1	1952.2	1952.20	32.694	3.319e-07 ***
Residuals	62	3702.1	59.71		

Data set 5

Reference:

Huston T.M. and Carmon J.L. (1958) Influence of High Environmental Temperature on Fertility and Hatchability of Eggs of Domestic Fowl. *Physiological Zoology*, 31(3) 232-235

Raw Data:

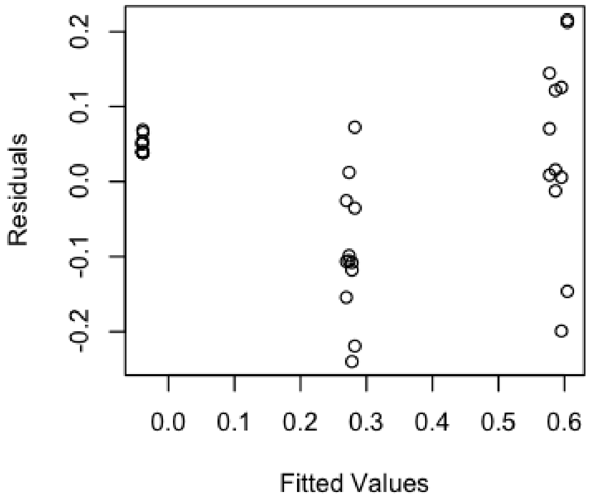
Parent Ambient Temperature Treatments:
1 = Male High x Female High
2 = Male High x Female Variable
3 = Male Variable x Female High
4 = Male Variable x Female Variable

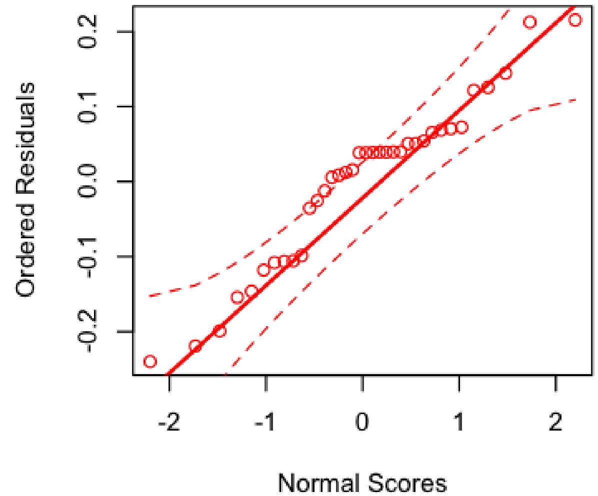
	Week	Percent Fertility	Arcsine
1	1	58.6	49.95
2	1	60.2	50.89
3	1	39.6	39.00
4	1	45.8	42.59
1	1	64.8	53.61
2	1	57.4	49.26
3	1	72.1	58.12
4	1	81.7	64.67
1	1	72.2	58.18
2	1	70.8	57.29
3	1	60.1	50.8
4	1	82	64.90
1	2	16.3	23.81
2	2	16.8	24.20
3	2	16	23.58
4	2	6.3	14.54
1	2	11.5	19.82
2	2	17.5	24.73
3	2	3.8	11.24
4	2	24.7	29.80
1	2	24.4	29.60
2	2	28.6	32.33
3	2	17	24.35
4	2	35.5	36.57
1	3	0	0
2	3	0	0
3	3	0	0
4	3	0	0
1	3	0	0
2	3	3	9.974
3	3	0	0
4	3	1.1	6.020
1	3	2.7	9.458

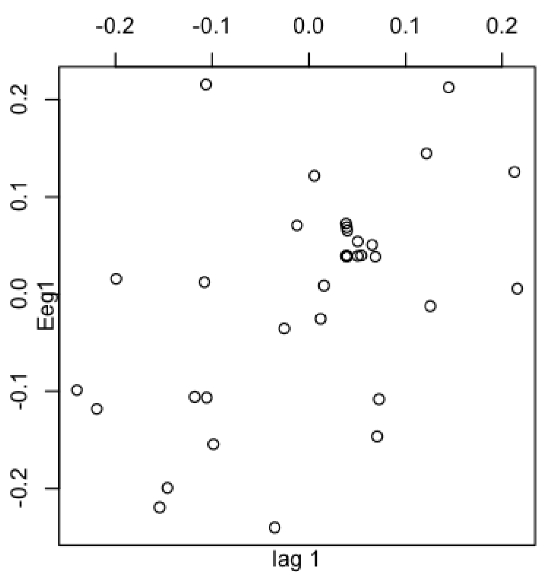
	Week	Percent Fertility	Arcsine
2	3	0.1	1.812
3	3	1.5	7.035
4	3	1.1	6.020

Non-Transformed Data:

LM: Linear Regression



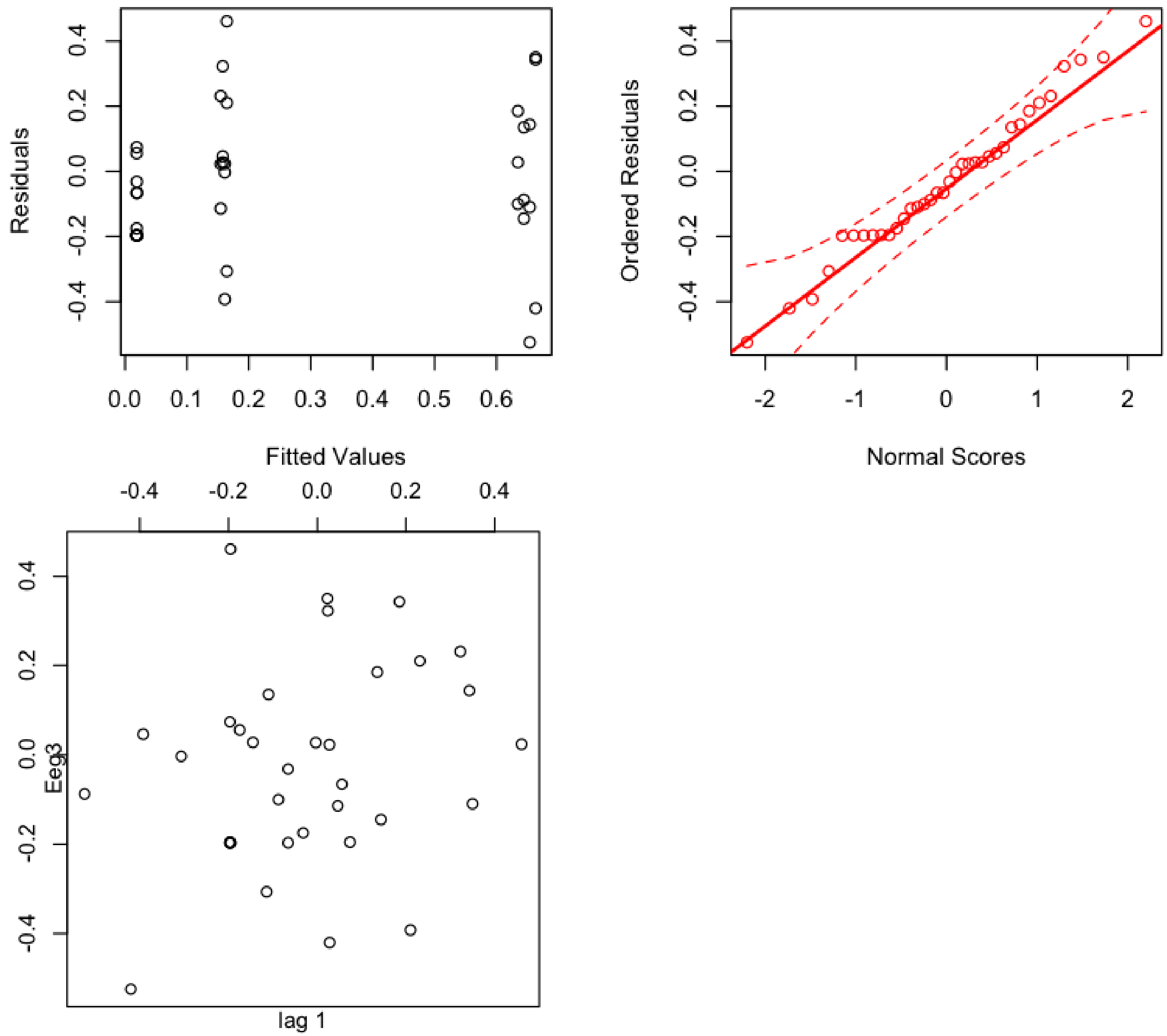




ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TREAT	1	0.00084	0.00084	0.0619	0.8051
WEEK	1	2.38014	2.38014	176.1068	1.478e-14 ***
TREAT:WEEK	1	0.00066	0.00066	0.0490	0.8262
Residuals	32	0.43249	0.01352		

GLM: Logistic Regression

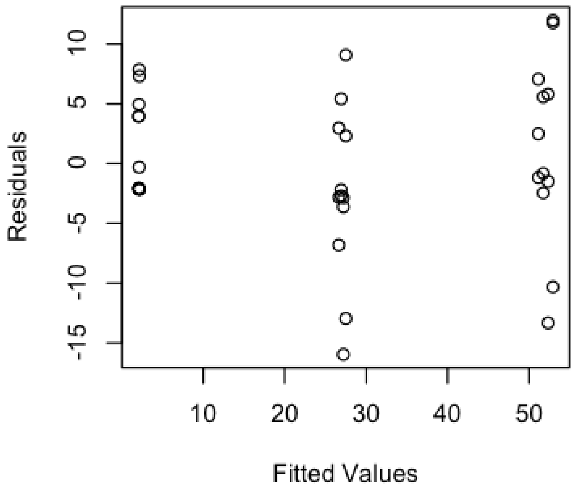


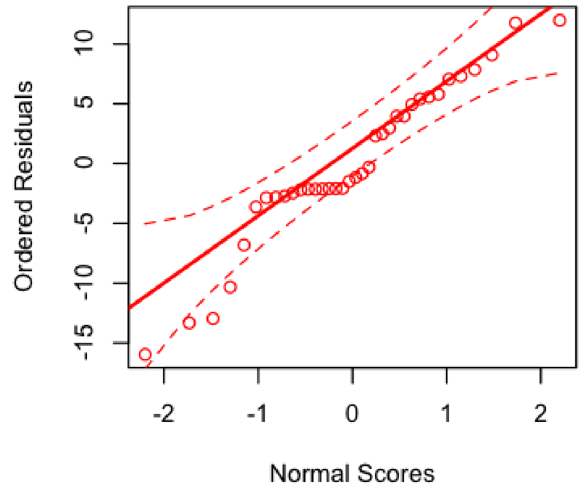
ANODEV Table

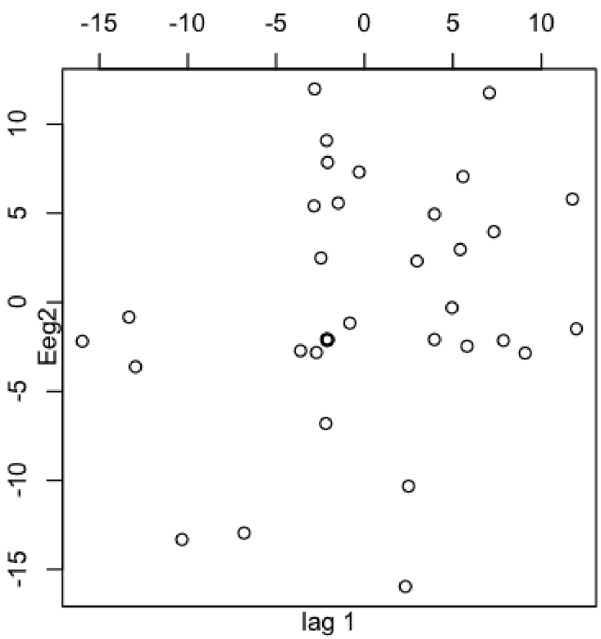
	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)
NULL			35	15.8325	
TREAT	1	0.0042	34	15.8283	0.9484092
WEEK	1	14.0521	33	1.7762	0.0001778 ***
TREAT:WEEK	1	0.0006	32	1.7756	0.9804378

Arcsine Transformed Data:

LM: Linear Regression







ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TREAT	1	3.7	3.7	0.0767	0.7837
WEEK	1	14947.7	14947.7	312.4498	<2e-16 ***
TREAT:WEEK	1	2.9	2.9	0.0614	0.8060
Residuals	32	1530.9	47.8		

Summary plots of data analyzed in Appendix H

Data Set	LM: Percent vs Arcsine				
	plots			p-value	
	Res vs Fit	Res vs Lag	Normality	Change decision?	Loss or Gain?
1	no change	no change	no change	no	-0.0015, -0.011
2	no change	better	no change	no	+0.0669
3	better	no change	better	no	+0.0022
4	no change	better	no change	no	+9.97e-6
5	no change	worse	no change	no	-0.021, -1.27e-14, -0.02

Data Set	LM: Percent vs Logistic				
	plots			p-value	
	Res vs Fit	Res vs Lag	Normality	Change decision?	Loss or Gain?
1	no change	no change	better	Yes	+0.643, +0.982
2	no change	better	better	No	+0.552
3	Better	no change	better	Yes	+0.3868
4	no change	no change	better	Yes	+0.1387
5	better	better	better	Yes	+0.143, +0.00018, +0.154