APPENDIX G

AVIAN QUALITY ASSURANCE/CONTROL AND AVIAN OBSERVER EFFICIENCY ANALYSIS

Avian Quality Assurance/Control

Introduction

A Quality Assurance Work Plan (QAWP) was developed to comply with standards set forth in the contract between the New Jersey Department of Environment Protection (NJDEP) Division of Science, Research, & Technology (DSRT) and Geo-Marine, Inc. (GMI) to conduct Ocean/Wind Power Ecological Baseline Studies (EBS) in the coastal and nearshore waters of New Jersey. A copy of the QAWP and Quality Assurance/Quality Control (QA/QC) data for the project can be obtained from NJDEP. A brief summary of avian QA/QC protocols and results is provided along with an analysis of avian observer efficiency.

Quality Assurance/Quality Control Summary

Avian Boat Surveys

Offshore shipboard and small boat coastal avian survey data was checked at the end of the end of each survey day by the senior seabird biologist. If an error(s) were found in the raw electronic data files corrections were made. The file name was changed (QAQC added to file name) and saved (see project database). The avian principal investigator (PI) then conducted a final QA/QC check of the data by randomly selecting 10% of the files for QA/QC.

Monthly error rates were generally low throughout the study for the avian shipboard surveys (**Table 1**). If an error was found, an additional 10% of the survey files were checked until no errors were detected. All files were re-checked in December 2009 when both initially checked files failed QA/QC.

Table 1
QA/QC Results for Avian Shipboard Surveys

Survey Month	No. Data Files	No. of QA/QC Checked Files	No. Fail	No. Pass	Error %
Jan 2008	129	15	1	14	6.67
Feb 2008	58	5	0	5	0.00
Mar 2008	275	28	1	18	3.57
Apr 2008	181	19	4	15	21.05
May 2008	136	14	0	14	0.00
June 2008	81	8	0	8	0.00
July 2008	62	8	1	7	12.50
Aug 2008	65	8	1	7	12.50
Sept 2008	119	13	1	12	7.69
Oct 2008	95	10	1	9	10.00
Nov 2008	50	6	0	6	0.00
Dec 2008	35	4	0	4	0.00
Jan 2009	47	5	0	5	0.00
Feb 2009	74	7	0	7	0.00
Mar 2009	40	4	0	4	0.00
Apr 2009	31	3	0	3	0.00
May 2009	68	8	0	4	0.00
June 2009	52	10	1	10	10.00
Aug 2009	81	8	0	8	0.00
Sept 2009	85	7	0	7	0.00
Oct 2009	39	4	0	4	0.00
Nov 2009	49	5	1	4	20.00
Dec 2009	20	2	2	0	100.00
TOTAL	1,872	201	14	187	7.49

Overall, 10.74% of the data files were checked and the overall failure rate was 7.49%. If a failure was identified the error was corrected in the data file and identified on the QA/QC log. The avian principal investigator then randomly selected and checked another 10% of the data files. This procedure continued until no errors were found. The QA/QC protocol for the small boat coastal survey data was identical to that conducted for the offshore avian ship surveys were similar to the shipboard surveys.

Avian shipboard electronic survey raw data files and QA/QC files were filed with the NJDEP. These files may be requested from NJDEP.

Aerial Surveys

The general QA/QC protocol followed that developed for the avian boat surveys. Each observer played-back voice recorded data and recorded that data. The data was then exchanged and each observer checked a 10% portion of the others. Error rates are not reported for this task since only one avian aerial survey was completed during the study.

Avian Radar

Avian Radar QA/QC was accomplished through detailed data analysis (see **Volume II, Section 5.4**).

Thermal Imaging

QA/QC protocols identified in the QAWP were followed for this task. Upon completion of the thermal imaging data analysis for all seasons a second biologist sampled a subset of the total dataset (15 minutes of each 150 minutes analyzed) for QA/QC purposes. The second biologist randomized the 15-minute samples within each 150-minute time block and conducted checks of recorded tracks and details (e.g. identification, direction, altitude, and size of flock). Upon completion of the QA/QC protocols there were no errors found and final analyses were then completed. If errors would have been recorded the second biologist would determine the cause of the errors and corrected them immediately.

Avian Observer Efficiency

Objective

The objective of this observer efficiency analysis was to calculate observer efficiencies and determine whether bird count observations collected by two independent observers were significantly different from each other over several months of data collection. Observer efficiency error (defined as the percent difference in total bird counts between the two observers) was used as the statistical metric, and was calculated for all months for which adequate observational data were available.

Data Format

The avian observational data consisted of species bird counts and total bird counts recorded by two observers: a primary observer ("Observer") and a QA/QC observer ("Tester"). Species-specific bird counts were summed over the number of species to obtain total bird counts. Bird counts were recorded by species over several survey time periods distributed over one or several days a month, for the given month. Bird count data (collected by the "Observer" and "Tester") were collected on the following dates (the number of survey periods for the given date given in parentheses):

September 2008: 9-13 (1), 9-14 (2), 9-15 (1), and 9-16 (1): n=5
October 2008: 10-14 (2), 10-17 (1): n=3
November 2008: 11-12 (3): n=3
January 2009: 1-12 (2), 1-13 (1): n=3
February 2009: 2-14 (1): n=1
June 2009: 6-2 (2), 6-3 (2), 6-4 (4): n=8
September 2009: 9-20 (1): n=1
October 2009: 10-1 (1): n=1
November 2009: 11-19 (1), 11-22 (1): n=2

For the current analysis (discussed in the Methodology section below), sample size (n) for each month was the cumulative number of survey periods conducted in that month (i.e., summation of the number in parentheses). In the following analysis, only those months with a sample size n>2 were examined. Those months for which n=2 or less were discarded from the analysis due to "insufficient data availability". Thus, the following 5 months were analyzed: September 2008 (n=5), October 2008 (n=3), November 2008 (n=3), January 2009 (n=3), and June 2009 (n=8).

Methodology

A two-sample one-way Analysis of Variance (ANOVA)/Kruskal-Wallis (K-W) analysis was conducted to assess sources of variance and test for significant differences in total bird counts between the two observers ("Observer" and "Tester"). Additional two-sample analyses included the two-sample t-test/Mann-Whitney test and the paired-sample t-test/Wilcoxon test. Tests for significant difference were conducted at the 95% confidence level (CL). These analyses were conducted for each of the above-listed 5 months of adequate avian observer data (n>2): September 2008 (n=5), October 2008 (n=3), November 2008 (n=3), January 2009 (n=3), and June 2009 (n=8). For each month, there were two samples (1=Observer, 2=Tester), with the number of monthly survey periods (= n) being 5, 3, 3, 3, and 8 for these 5 months, respectively. The names of the observer and tester were generally different for observation day (i.e., data point comprising the sample size). The measured attribute was TOTAL bird counts (i.e., summed over the individual species).

In addition, for each of these 5 months, a one-sample t-test/Wilcoxon test was conducted (at the 95% CL) to determine if the difference in total bird counts between the Observer and Tester was significant (i.e., significantly different from 0). Two types of differences were calculated and tested: 1) Difference = Observer bird count – Tester bird count and 2) % Observer Error = 100%*(Observer bird count – Tester bird count)/(Observer bird count)

In addition to the one-way analysis (which was conducted for each of the 5 individual months), a two-way ANOVA/K-W analysis was conducted (at the 95% CL) to test for significant differences in total bird counts among the 5 months (Factor 1), between the two observers (Factor 2), and two-way interactions (Month x Observer). For 5 months of data, there were (5 months)*(two observers) = 10 samples, with sample size n = number of survey periods in the given month.

In addition, separate one-way ANOVA/K-W analyses were conducted (at the 95% CL) to test for significant differences in "Difference" and in "% Observer Error" among the 5 months. The number of samples was equal to the number of months (=5), and sample size n = number of survey periods in the given month.

In initial testing for normality (Kolmogorov-Smirnov test) and homoscedacity (Bartlett test), if it was determined that the given dataset was normal (i.e., conforms to a Gaussian bell-shaped distribution) and homoscedastic (i.e., exhibits homogeneity of variances), then the parametric tests were applied to the samples. These tests include: one-sample t-test, two-sample t-test, paired-sample t-test, and multiple-sample ANOVA. In contrast, if it was determined that the dataset was non-normal or non-homoscedastic after two data transformations (e.g., using logarithmic, arcsine, or square-root transformations), then the non-parametric tests were applied. These tests included: one-sample Wilcoxon test, Mann-Whitney two-sample test, paired-sample Wilcoxon test, and multiple-sample Kruskal-Wallis (K-W) test.

Generally, nonparametric tests were conducted on the RANKS of the attribute data (e.g., sorted from lowest to highest values) rather than on the values themselves. For example, in the K-W test, the total bird counts were ranked from lowest to highest, and an ANOVA-type test was conducted on the RANKS (rather than on the values) of the total bird counts to assess differences among the months of available data (Factor 1), between the 2 observers (Factor 2), and any significance in the 2-way interaction (Month x Observer). In the event that the ANOVA or K-W test detected a significant difference (indicated by rejection of the Null hypothesis H_o of no difference), a subsequent Tukey test examined each pair-wise comparison (i.e., N*(N-1)/2 pairs for N samples) to determine the sources of variance (i.e., which specific sample pairs are significantly different from each other).

In addition to the above analyses, several types of Randomization Tests (which assumes datasets were collected serially/sequentially in space or time) were conducted on "Difference" and "% Observer Error" time-series data, in which the sample size (n) in the time series for each month is equal to the number of survey periods in that month. Two types of randomization tests (with associated null hypotheses H_o) were conducted:

1) **Mean Square Successive Difference Test (MSSDT) (parametric):** assumes normality in the underlying distribution of the data.
    a) Null hypothesis (H_o): Consecutive samples exhibit random variability (i.e., are not serially/sequentially correlated).

2) **Runs-up-and-down Test (nonparametric):** Runs, identified by successive directions of change in the same (either positive or negative) directions are counted in the serial/sequential data and are tested to determine whether the occurrence of these runs is random.
    a) H_o (2-tailed): Successive directions of change in the serial/sequential data occur randomly.
    b) H_o (1-tailed): Successive positive and negative changes in the serial/sequential data are not clustered and do not alternate regularly.

3) **(Optional): Runs-above-and-below-median Test (nonparametric):** Runs, identified by directions of deviation of the data values above or below the sample median, are counted in the serial/sequential data and are tested to determine whether the occurrence of these runs is random.
    a) H_o: The population from which the sample was drawn has a random distribution, with no underlying deterministic trend component.

The null hypothesis (H_o) of these Randomization Tests states that the serial/sequential variability and the occurrence of runs in the given dataset (assumed to be collected serially or sequentially in space or time) are random. Rejection of H_o (at the 95% confidence level CL) suggests that the data exhibit an underlying deterministic trend component. In these cases (i.e., rejection of H_o), linear or quadratic regression analysis is used to model the serial/sequential data and subtract out the modeled trend component from the data, leaving the residuals. The Randomization Tests are then conducted on the residual component only, to determine if the serial/sequential variability in the residuals is random (stationary, with a mean of zero).

Results

Results of the two-sample analyses (one-way ANOVA/K-W test, two-sample t-test/Mann-Whitney test, paired-sample t-test/Wilcoxon test), conducted on the "Observer" and "Tester" total bird counts, and of the one-sample t-test/Wilcoxon test (conducted on "Difference" and "% Observer Error"), for each of the five months, are given in the five month-specific Excel files: 1) ANOVA_obseff_9+2008.xls; (2) ANOVA_obseff_10+2008.xls; 3) ANOVA_obseff_11+2008.xls; 4) ANOVA_obseff_1+2009.xls; and 5) ANOVA_obseff_6+2009.xlsResults of the 1-way ANOVA/K-W tests conducted on "Difference" and "% Observer Error" across the 5 months, and of the 2-way ANOVA/K-W test conducted on total bird counts across the 5 months and between the 2 observers, are given in the Excel file: ANOVA_obseff_total.xls (see Statistics file in Volume II Appendix CD). The results of the observer efficiency analysis are summarized below.

<u>September 2008</u>

For the five samples, the two-sample dataset was determined to be non-normal therefore the non-parametric Mann-Whitney two-sample test, Wilcoxon paired-sample test, Kruskal-Wallis test, Wilcoxon one-sample test, and Runs-Up-and-Down Test were applied. In addition, the Mean Square Successive Difference Test (MSSDT) was conducted to provide a normal approximation.

Mann-Whitney: Results of the two-tailed U test show no significant difference in total bird counts between the two observers (acceptance of H_o: n_1=5, n_2=5, U=13, U_{crit}=23, P=0.9000).

Wilcoxon: Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o: T_m=6.5, T_p=8.5, Crit=-999, P=0.8750.

Kruskal-Wallis: Results of the two-tailed chi-squared (X^2) test show no significant difference in total bird counts between the two observers (acceptance of H_o): DF=1, X^2=0.0109, X^2_{crit}=3.841, P=0.9212 . For N=2 levels, there was 2*(2-1)/2 = 1 pair. A subsequent Tukey test for Rank Means on this one pair confirms this result: H_o is accepted.

All three of these tests confirm each other in accepting H_o, with P values ranging from 0.8750 to 0.9212.

1-sample Wilcoxon (Difference): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=6.0, T_p=12.0, Crit=-999, P=0.8000 (Line 255).

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.0663, C_{crit} = 0.509, P=0.3764 (Line 317).

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=3, u_{1crit}=1, u_{2crit}=999, P=0.8000 (Line 343).

Runs-Above-and-Below-Median Test:

One-sample Wilcoxon (% Observer Error): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=8.0, T_p=8.0, Crit=-999, P=1.0000 (Line 395).

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.0246, C_{crit} = 0.509, P=0.4100 (Line 457).

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=3, u_{1crit}=1, u_{2crit}=999, P=0.8000 (Line 483).

These 1-sample tests confirm each other in arriving at the same conclusions (acceptance of H_o, serial/sequential randomness) in the analyses of “Difference” and “% Observer Error”.

October 2008

For the three samples, the two-sample dataset was determined to be non-normal, therefore the non-parametric Mann-Whitney two-sample test, Wilcoxon paired-sample test, Kruskal-Wallis test, Wilcoxon one-sample test, and Runs-Up-and-Down Test were applied. In addition, the Mean Square Successive Difference Test (MSSDT) was conducted to provide a normal approximation.

Mann-Whitney: Results of the two-tailed U test show no significant difference in total bird counts between the two observers (acceptance of H_o): n_1=3, n_2=3, U=6, U_{crit}=999, P=0.4000.

Wilcoxon: Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=0.0, T_p=6.0, Crit=-999, P=0.2000.

Kruskal-Wallis: Results of the two-tailed chi-squared (X^2) test show no significant difference in total bird counts between the two observers (acceptance of H_o): DF=1, X^2=0.4286, X^2_{crit}=3.841, P=0.5187. For N=2 levels, there are 2*(2-1)/2 = 1 pair. A subsequent Tukey test for Rank Means on this one pair confirms this result: H_o is accepted.

All 3 of these tests confirm each other in accepting H_o, with P values ranging from 0.2000 to 0.5187; however, it should be cautioned that the sample size (n=3) for this month is very small, and the results of these analyses may not be as accurate compared to larger sample sizes.

One-sample Wilcoxon (Difference): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=6.0, Crit=-999, P=0.6500.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = -0.5, C_{crit} = 0.509, P=0.0545 (Line 307).

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=2, u_{1crit}=1, u_{2crit}=999, P=0.5051 (Line 331).

Runs-Above-and-Below-Median Test:

One-sample Wilcoxon (% Observer Error): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=4.0, Crit=-999, P=0.6500.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.4868, C_{crit} = 0.509, P=0.0611.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=1, u_{1crit}=1, u_{2crit}=999, P=0.3500.

These one-sample tests confirm each other in arriving at the same conclusions (acceptance of H_o, serial/sequential randomness) in the analyses of “Difference” and “% Observer Error”.

November 2008

For the three samples, the two-sample dataset was determined to be non-normal, hence the non-parametric Mann-Whitney two-sample test, Wilcoxon paired-sample test, Kruskal-Wallis test, Wilcoxon one-sample test, and Runs-Up-and-Down Test were applied. In addition, the Mean Square Successive Difference Test (MSSDT) was conducted to provide a normal approximation.

Mann-Whitney: Results of the two-tailed U test show no significant difference in total bird counts between the two observers (acceptance of H_o): n_1=3, n_2=3, U=5, U_{crit}=999, P=0.5000.

Wilcoxon: Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=3.0, Crit=-999, P=0.6500.

Kruskal-Wallis: Results of the two-tailed chi-squared (X^2) test show no significant difference in total bird counts between the two observers (acceptance of H_o): DF=1, X^2=0.0476, X^2_{crit}=3.841, P=0.8449 (Line 154). For N=2 levels, there are 2*(2-1)/2 = 1 pair. A subsequent Tukey test for Rank Means on this one pair confirms this result: H_o is accepted.

All 3 of these tests confirm each other in accepting H_o, with P values ranging from 0.5000 to 0.8449; however, it should be cautioned that the sample size (n=3) for this month is very small, and therefore the results of these analyses may not be as accurate compared to larger sample sizes.

1-sample Wilcoxon (Difference): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=6.0, Crit=-999, P=0.6500 (Line 247).

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.25, C_{crit} = 0.509, P=0.2282.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=1.5, u_{1crit}=1, u_{2crit}=999, P=0.6541.

Runs-Above-and-Below-Median Test:

1-sample Wilcoxon (% Observer Error): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=4.0, Crit=-999, P=0.6500.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.4861, C_{crit} = 0.509, P=0.0615.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=1, u_{1crit}=1, u_{2crit}=999, P=0.3500.

These one-sample tests confirm each other in arriving at the same conclusions (acceptance of H_o, serial/sequential randomness) in the analyses of “Difference” and “% Observer Error”.

January 2009

For the three samples, the two-sample dataset was determined to be non-normal, therefore the non-parametric Mann-Whitney 2-sample test, Wilcoxon paired-sample test, Kruskal-Wallis test, Wilcoxon one-sample test, and Runs-Up-and-Down Test were applied. In addition, the Mean Square Successive Difference Test (MSSDT) was conducted to provide a normal approximation.

Mann-Whitney: Results of the two-tailed U test show no significant difference in total bird counts between the two observers (acceptance of H_o): n_1=3, n_2=3, U=4.5, U_{crit}=999, P=0.5500.

Wilcoxon: Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=4.0, Crit=-999, P=0.6500.

Kruskal-Wallis: Results of the two-tailed chi-squared (X^2) test show no significant difference in total bird counts between the two observers (acceptance of H_o): DF=1, X^2=0.0000, X^2_{crit}=3.841, P=0.9900 (Line 158). For N=2 levels, there are 2*(2-1)/2 = 1 pair. A subsequent Tukey test for Rank Means on this 1 pair confirms this result: H_o is accepted.

All three of these tests confirm each other in accepting H_o, with P values ranging from 0.5500 to 0.9900; however, it should be cautioned that the sample size (n=3) for this month is very small, and therefore, the results of these analyses may not be as accurate compared to larger sample sizes.

One-sample Wilcoxon (Difference): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=4.0, Crit=-999, P=0.6500.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.1676, C_{crit} = 0.509, P=0.2947.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=2, u_{1crit}=1, u_{2crit}=999, P=0.6541.

Runs-Above-and-Below-Median Test:

One-sample Wilcoxon (% Observer Error): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=3.0, T_p=4.0, Crit=-999, P=0.6500.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = -0.1125, C_{crit} = 0.509, P=0.3391.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=2, u_{1crit}=1, u_{2crit}=999, P=0.6541 (Line 468).

These one-sample tests confirm each other in arriving at the same conclusions (acceptance of H_o, serial/sequential randomness) in the analyses of “Difference” and “% Observer Error”.

June 2009

For the eight samples, the two-sample dataset was determined to be non-normal therefore the non-parametric Mann-Whitney two-sample test, Wilcoxon paired-sample test, Kruskal-Wallis test, Wilcoxon one-sample test, and Runs-Up-and-Down Test were applied. In addition, the Mean Square Successive Difference Test (MSSDT) was conducted to provide a normal approximation.

Mann-Whitney: Results of the two-tailed U test show no significant difference in total bird counts between the two observers (acceptance of H_o): n_1=8, n_2=8, U=35.5, U_{crit}=51, P=0.4375.

Wilcoxon: Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=15.0, T_p=31.0, Crit=3.0, P=0.7250.

Kruskal-Wallis: Results of the two-tailed chi-squared (X^2) test show no significant difference in total bird counts between the two observers (acceptance of H_o): DF=1, X^2=0.1375, X^2_{crit}=3.841, P=0.7248. For N=2 levels, there are 2*(2-1)/2 = 1 pair. A subsequent Tukey test for Rank Means on this one pair confirms this result: H_o is accepted.

All three of these tests confirm each other in accepting H_o, with P values ranging from 0.4375 to 0.7248.

One-sample Wilcoxon (Difference): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=15.0, T_p=31.0, Crit=3.0, P=0.7250.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = 0.0518, C_{crit} = 0.509, P=0.3881.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=4.6667, u_{1crit}=2, u_{2crit}=999, P=1.0000.

Runs-Above-and-Below-Median Test:

One-sample Wilcoxon (% Observer Error): Results of the two-tailed test show no significant difference in total bird counts between the two observers (acceptance of H_o): T_m=15.0, T_p=31.0, Crit=3.0, P=0.7250.

Mean Square Successive Difference Test (MSSDT): Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): C = -0.2610, C_{crit} = 0.509, P=0.2193.

Runs-Up-and-Down Test: Variability in total bird counts exhibits serial/sequential randomness (acceptance of H_o): u=4.6667, u_{1crit}=2, u_{2crit}=999, P=1.0000.

These one-sample tests confirm each other in arriving at the same conclusions (acceptance of H_o, serial/sequential randomness) in the analyses of "Difference" and "% Observer Error".

Summary

Total Bird Counts

From the two-sample one-way analysis for each of the five months, there's no significant difference in total bird counts between the two observers ("Observer" and "Tester").

Monthly Difference

One-way ANOVA was run and the five-sample dataset was determined to be non-normal, therefore hence the non-parametric Kruskal-Wallis test was applied. Results of the two-tailed chi-squared (X^2) test show no significant difference in "Difference" (i.e., difference between the 2 observers' total bird counts) among the five months (acceptance of H_o): DF=4, X^2=2.7512, X^2_{crit}=9.488, P=0.6056. For N=5 levels, there are 5*(5-1)/2 = 10 pairs. A subsequent Tukey test for Rank Means on these 10 pairs confirms this result: H_o is accepted for all 10 pair-wise combinations.

Percent Observer Error

The five-sample dataset was determined to be non-normal therefore the non-parametric Kruskal-Wallis test was applied. Results of the two-tailed chi-squared (X^2) test show no significant difference in "% Observer Error" among the five months (acceptance of H_o): DF=4, X^2=2.6789, X^2_{crit}=9.488, P=0.6182 (Line 157). For N=5 levels, there are 5*(5-1)/2 = 10 pairs. A subsequent Tukey test for Rank Means on these 10 pairs confirms this result: H_o is accepted for all 10 pair-wise combinations.

Between Observer Difference

There was no significant difference in total bird counts (acceptance of H_o): DF1=1, DF2=70, F=1.3814, F_{crit}=3.841, P=0.2437. For N=2 levels, there are 2*(2-1)/2 = 1 pair. A subsequent Tukey test for Rank Means on this pair confirms this result: H_o is accepted.

Month x Observer Interaction

There was no significant difference (acceptance of H_o): DF1=4, DF2=70, F=0.9138, F_{crit}=9.488, P=0.9213 (Line 303). For N=5*2=10 levels, there are 10*(10-1)/2 = 45 pairs. A subsequent Tukey test for Rank Sums on these 45 pairs confirms this result: H_o is accepted for all 45 pair-wise combinations.

Conclusions

The following conclusions can be made. No significant difference exists in:

1) Total bird counts among the five months between the two observers.
2) Difference (i.e., the difference between the two observers' total bird counts) among the five months.
3) Percent Observer Error among the five months.

In addition, no significant interaction between month and observer (i.e., the differences in total bird counts with respect to one factor are independent of the other factor). For example, the sensitivity of total bird counts to changes in month is independent of observer, and likewise, the sensitivity of total bird counts to changes in observer is independent of month. Where there is no significant difference in total bird counts with respect to changes in a given factor or interaction, the Tukey test confirms this result for all pair-wise combinations.