TESTS FOR LONGITUDINAL CHANGE IN SKEWED SAMPLES WITH APPLICATION TO HIBERNIA SEDIMENT CHEMISTRY DATA

GWENDA MARY DROVER

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-83996-6
Our file Notre référence
ISBN: 0-612-83996-6

Tests For Longitudinal Change in Skewed Samples With Application to Hibernia Sediment Chemistry Data

by

Gwenda Mary Drover

A practicum report submitted to the
School of Graduate Studies in partial fulfilment
of the requirements for the degree of
Master of Applied Statistics

Department of Mathematics and Statistics
Faculty of Science
Memorial University of Newfoundland
May, 2002

St.John's Newfoundland Canada

Abstract

The Hibernia platform began drilling for oil on November 17th, 1997 in the Atlantic Ocean off the coast of Newfoundland. Measurements of organic and inorganic chemicals occurred in both 1995 and 1998 to determine if this drilling process resulted in a significant sediment change in the ocean floor surrounding the drill site. Statistical analysis and summary of the change in sediment chemistry data collected from the Hibernia Gravity Based Structure (GBS) were released in 1999 by Jacques Whitford Environment Ltd. St. John's, Newfoundland. The objective of this practicum is to more thoroughly examine these data collected before and after the Hibernia drilling process began and to determine whether the introduction of the GBS platform resulted in changes in the concentration of various chemicals in the surrounding sediment. This procedure involved the modification of existing statistical tests to accommodate for the asymmetrical nature of the data as well as existing longitudinal and spatial correlations. Combined data test results of this practicum do not in fact differ significantly from those reported in 1999 by Jacques Whitford Environment Ltd.

Acknowledgements

First, I would like to thank God for countless prayers answered.

I wish to thank my supervisor, Dr. Brajendra Sutradhar (Faculty of Mathematics and Statistics, Memorial University of Newfoundland) for all his time and help during the completion of my practicum.

I also wish to thank my employer Mr. Alton Hollett and Linda Bartholomew, my friend and co-worker at Newfoundland Statistics Agency, for their support and encouragement as I completed the Masters Program.

I would also like to thank my loving husband for being so wonderful at all times. I would like to thank my very best friends; my parents for their boundless support and love throughout my entire life.

Contents

1 Introduction

2 Background of the Problem

3 Distribution-Free Test Based on Rank Spacings

4 Modified Cornish-Fisher t Test for Asymmetrical Populations

5 Standardized Median Test (SMT) Using A Working Dispersion

6 Generalized SMT for Combined Sample

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation of the Problem

Standing more than 220 meters high in the Atlantic Ocean off the coast of Newfoundland, the Hibernia platform began the oil drilling process on November 17th, 1997. More than 80 wells will be required to tap the Hibernia reservoirs in order to reach the oil reserves deep beneath the ocean floor. The Topsides facilities on the Hibernia platform accommodate all drilling, producing and utility equipment, as well as provide living quarters for platform workers. These facilities were designed to produce up to 150,000 barrels of oil a day, as a major economic step for the province of Newfoundland and Labrador and Canada as a whole.

The Topsides is supported by a concrete pedestal called the The Gravity Base Structure (GBS) which was constructed in Bull Arm, Newfoundland. The GBS consists of high strength concrete reinforced with steel rods. The GBS is located less than 150kms from St. John's, it is 111 meters high and it has storage capacity for 1.3 million barrels of crude oil. The GBS is designed to withstand the impact of sea ice and icebergs to allow for year-round production (Platform Support, 1999).

Drilling operations occur within the Newfoundland-built Wellhead Module. The platform is designed to drill two wells at one time and these wells reach depths of more than 3,700 meters below sea level into the oil reservoirs. A drill bit is used to drill the wells into the ocean floor by spiraling its way down through the ground, crushing rock and stone as it goes. The crushed rock and stone produced by the drill bit are removed from the well by drilling mud, a compound of drilling chemicals, water and clay that is produced inside a Mud Module. Drilling muds are continuously pumped down the drill pipe and through holes in the drill bit to cool the bit, and to prevent the hole from collapsing and wash the drilled cuttings away from the bottom of the hole. The drilling mud flows constantly, and is eventually re-circulated back to the surface where the cuttings are separated (Drilling & Processing, 1999).

Note that in any major oil drilling operation, it is likely that there will be some negative effects on the environment such as on the flora and fauna surrounding the drilling platform. If such environmental effects are more than acceptable national standards then it would be necessary to control the associated factors responsible for the environmental damage. Failure to do this may result in the closure of the operations. From this point of view, environmental agencies are employed to collect necessary data before and after the drilling process. More specifically, measurements of organic and inorganic chemicals were taken in 1995 and 1998. Some statistical analysis and summary of the changes in sediment chemistry data collected from the Hibernia Gravity Based Structure (GBS) were released by Jacques Whitford Environment Ltd. St. John's, Newfoundland. The Hibernia statistics report is titled as Environmental Effects Monitoring (Lye and Bartholomew, 1999). In this original report, conclusions were made from the initial examination of plots and statistical tests including bubble plots and paired t-tests. It will be shown, however, that the samples extracted for certain chemicals in 1995 and 1998 actually follow asymmetrical distributions, which raises a concern about the validity of a normality based paired t-test.

Further, the chemical data collected in 1995 and 1998 are naturally longitudinally correlated. As testing procedures for independent groups are different than those for correlated groups, some authors Lye and Bartholomew, 1999 conducted paired t-tests. As the paired t-test is constructed by using the differences between the 1995 and 1998 chemical concentrations, and because the differences adjust for these correlations, the paired t-test would have been a valid test provided the two groups have the same number of observations at each distance and the chemical concentrations are symmetric. However, this is not the situation with the Hibernia data set, especially when one is interested in the environmental changes at different distances from the drilling platform. This consequently raises concerns about the application of the paired t-test.

In this practicum, we look at each of these twelve distances separately and test for sediment changes at each point radiating out from the Hibernia platform. As the chemical concentrations may be asymmetrical at a given distance and because responses for 1995 and 1998 are longitudinally correlated, we develop a number of suitable tests after taking the asymmetric nature of the data as well as the longitudinal correlations into account. Finally, we will combine the chemical data for all distances similar to Lye and Bartholomew, 1999. We will, however, account for the spatial correlation existing within these data because of their origin at different distances from the drilling platform.

1.2 Objective of the Practicum

The objective of this practicum is to thoroughly examine the chemical data collected before and after the Hibernia drilling process began and to determine whether the introduction of the GBS platform resulted in changes in the concentration of various chemicals in the surrounding sediment. This procedure will involve the modification of existing statistical tests to accommodate the asymmetrical nature of the data as well as longitudinal and spatial correlation between the chemical test sites in 1995 and 1998.

Statistical tests previously released in the Hibernia statistics report (1999) revealed that chemical concentrations for barium, iron and lead were significantly different from samples taken in 1995 and 1998. Attenuation tests showed that both barium and lead concentrations were significantly higher closer to the drilling platform. Overall, there is a significant increase in many of the chemicals after the start of the drilling process that is generally confined within 500m of the GBS. These existing tests on the combined distances did not however accommodate for the asymmetrical nature of the data as well as the spatial or longitudinal correlation existing between the test periods and sites. As a remedy, an extensive analysis will be completed as part of this practicum that will involve examination of the chemical data at each of the twelve test sites radiating from the GBS platform. Finally, additional tests will be performed on the combined data from all the locations accounting for the naturally occurring spatial correlation.

Chapter 2

Background

2.1 The Data

Sediment chemistry data were collected from the vicinity of the Hibernia GBS drill site before the drilling began in 1995 and one year after the start of drilling operations in 1998. Specific samples ranging in both distance and radial direction from Hibernia platform were gathered. A total of 46 locations were tested with at least three replicate sediment samples collected at each of these areas. All chemical measurements were reported in mg/kg and concentration levels were determined in an onshore laboratory.

Figure 1 shows the web-like collection pattern spreading out from the GBS drilling site. Each location is marked by distance in meters and radial direction from the Hibernia platform. Samples were collected at 12 distances from the site including 250m, 500m, 750m, 1000m, 1500m, 2000m, 2500m, 3000m, 4000m, 6000m, 8000m and 16,000m. The sampling plan also included eight radials spaced equal distances apart and surrounding the drill site. Each collection area was marked by distance and radial direction from the GBS for test both years.

A limit of quantification (LOQ) was determined (10 mg/kg) and any samples found with chemical amounts lower than this value did not undergo any further analysis. In some cases data were not available for certain elements in both test years or there may have been a change in the LOQ status. Higher levels of chemical concentration near the GBS in 1998 would suggest an effect due to the drilling process.

2.2 Exploratory Analysis

An extensive exploratory analysis was completed on the 1995 and 1998 chemical data sets including concentration levels for barium (Ba), lead (Pb), iron (Fe) and chromium (Cr) at each of the 12 distances measured from the GBS platform. Tables 1, 2 and 3 summarize the descriptive statistics for each of these chemicals at three selected distances. These distances include 250 meters, 1000 meters and 6000 meters from the Hibernia drilling site.

Table 1. Descriptive Statistics at 250m from the GBS Platform

Chemical/Date	*Mean*	σ	*Median*	*Minimum*	*Maximum*
*Barium*1998	409	374	240	160	1400
*Barium*1995	154	29	150	120	210
*Iron*1998	3655	3527	2100	1600	13700
*Iron*1995	2467	1244	1850	1300	5100
*Lead*1998	2.83	0.49	2.70	2.30	4.00
*Lead*1995	2.73	0.55	2.55	2.10	4.00
*Chromium*1998	9	9	6	4	36
*Chromium*1995	4	1	4	3	8

Table 2. Descriptive Statistics at 1000m from the GBS Platform

Chemical/Date	*Mean*	σ	*Median*	*Minimum*	*Maximum*
*Barium*1998	133	21	140	100	170
*Barium*1995	106	28	110	64	150
*Iron*1998	1314	435	1200	680	2100
*Iron*1995	1683	730	1350	1100	3100
*Lead*1998	2.02	0.30	2.15	1.30	2.30
*Lead*1995	2.23	0.69	2.15	1.20	3.80
*Chromium*1998	3.25	0.75	3	2	4
*Chromium*1995	4.06	1.58	3.8	2.30	8.40

Table 3. Descriptive Statistics at 6000m from the GBS Platform

Chemical/Date	*Mean*	σ	*Median*	*Minimum*	*Maximum*
*Barium*1998	93	48	83	58	240
*Barium*1995	106	47	94	63	220
*Iron*1998	1286	832	1100	640	3700
*Iron*1995	2183	1081	2000	700	3900
*Lead*1998	1.69	0.61	1.6	1.10	3.40
*Lead*1995	2.11	0.74	1.85	1.30	3.70
*Chromium*1998	2.83	1.69	2	2	8
*Chromium*1995	4.22	1.86	4	2.30	9.10

Tables 1 to 3 show that the concentration of each chemical decreases as the distance from the GBS platform increases in 1998, whereas in 1995 these data do not show any specific pattern in general. The average barium concentration was 409mg/kg at 250m, 133mg/kg at 1000m and 93mg/kg at 6000m test site, respectively. Similar trends are evident for iron (3655mg/kg to 1286mg/kg), lead (2.83mg/kg to 1.69mg/kg) and chromium (9.00mg/kg to 2.83mg/kg). In addition, these summaries also show cases where the chemical mean and median are widely separated. For example, at the 250m test site, there are substantial differences between the mean and median for both iron and barium in 1995. The gap between these descriptive statistics indicated the data may be asymmetrical and therefore additional examination was necessary.

Next, we created various bar charts to help summarize the chemical data. Figures 2 and 3 show the mean concentration levels for barium and iron in 1998 at increasing distances from the Hibernia platform. In both of these graphs the amount of sediment contamination is evident close to the drilling operations. A similar picture is seen in figures 4 and 5 for lead and chromium as the mean concentration decreases in test sites further away from the GBS in 1998.

In addition, similar to Lye and Bartholomew (1999) we combined the chemical data for all distances. In later chapters we will perform statistical tests on this combined sample to identify any significant changes in sediment chemistry from 1995 to 1998 after accommodating for spatial and longitudinal correlations. Here, as part of our exploratory analysis, we calculated descriptive statistics for the combined data set. These results are summarized in Table 4.

Table 4. Descriptive Statistics at Combined Distances from the GBS Platform

Chemical/Date	*Mean*	σ	*Median*	*Minimum*	*Maximum*
*Barium*1998	169	139	140	51	1400
*Barium*1995	143	60	140	60	450
*Iron*1998	1889	1497	1600	530	13700
*Iron*1995	2046	955	1800	700	6100
*Lead*1998	2.35	0.76	2.2	1.10	4.90
*Lead*1995	2.46	0.67	2.40	1.20	4.80
*Chromium*1998	4.22	3.08	4	2	36
*Chromium*1995	4.23	1.22	4	2	9.10

For the combined sample, Table 4 shows some common trends among the chemical data as well as some of the potential problems in the overall distributions. Outliers are visible within chromium and barium and are most apparent for the Iron data set. The spread between the overall mean and median is also the most severe in the Iron data set.

The combined results show that only the mean concentration of barium is greater in 1998 as compared to that of 1995. This does not agree with tables 1 to 3 which show significant differences in concentration of other variables between 1995 and 1998 closer to the GBS drilling site. This shows the importance of a more extensive analysis for these data at all distances.

Mean concentration comparisons were also examined for each chemical from 1995 to 1998. Figures 6 and 7 show these mean comparisons for all chemicals at 250m from the drilling site. In all cases at this distance, there are higher concentrations in 1998 when compared to 1995. Figures 8 and 9 show the same information for 1000m from the GBS and Figures 10 and 11 graphically compare mean concentration differences at 6000m from the platform. These sets do not show a clear pattern due to outliers existing in the data, however, it is evident that as the distance increases from the GBS the change in mean concentration decreases.

Finally, a graphical analysis was completed including sets of box plots that were produced for each chemical at each distance to determine the asymmetrical nature of the data. Figures 12 to 17 are some examples of these graphs showing the possible asymmetrical nature of the data and the presence of outliers.

Figures 12 and 13 show the distributions for iron and barium at the 250m test site. iron was shown to be highly asymmetrical and many outliers are identified. Barium, at this same distance in 1995 was shown to be symmetrical and no outliers were detected. Similarly, Figures 14 and 15 show both lead and chromium concentrations to be asymmetrical at the 1000m test site. Outliers are also identified for lead in 1998. Finally, Figures 16 and 17 show the sediment chemistry at 6000m from the GBS drilling site. In 1998, the barium sample is seen to be highly asymmetrical while the 1995 iron sample is seen to be symmetrical.

As many chemicals were found to have outliers existing in several test samples, statistical testing in this practicum will include analysis with and without the inclusion of these outliers. Table 5 summarizes the asymmetrical nature of the chemical data sampled in 1995 and 1998.

Table 5. Distribution Pattern of Sediment Data at each Distance (meters) in 1995 and 1998. (Note: A ≡ Asymmetrical; S ≡ Symmetrical).

Distance	Ba 98	Ba 95	Fe 98	Fe 95	Pb 98	Pb 95	Cr 98	Cr 95
250m	A	S	A	A	S	S	A	S
500m	A	S	A	A	A	A	S	S
750m	A	S	A	A	A	S	A	A
1000m	A	A	A	A	A	S	A	S
1500m	S	A	S	A	A	S	A	A
2000m	A	S	S	S	A	S	A	S
2500m	A	S	A	S	A	A	A	S
3000m	S	A	A	A	S	A	A	A
4000m	A	A	A	A	S	A	A	A
6000m	S	A	S	A	A	A	A	S
8000m	A	A	A	S	A	A	A	A
16000m	A	A	S	S	A	A	A	S

Thorough examination of the samples collected in 1995 and 1998 showed that a large number of samples followed asymmetrical distributions and that outliers exist in the data. For example, the data collected for iron was shown to have the highest amount of outliers and as seen in Table 5, iron has the most cases where both the 1995 and 1998 data sets are asymmetrical. Also, for many other variables, the distributions are mainly asymmetric either at 1995 or 1998. As a remedy, alternative tests that accommodate the asymmetrical nature of the data as well as longitudinal and spatial correlations are suggested in this report to determine any significant differences in chemical deposits before and after the Hibernia drilling process began.

2.3 Notation

In this sub-section, before we formally develop appropriate tests in the chapters to follow, we introduce some notations for such confirmatory analysis. First, let y_{tij} denote the j^{th} ($j = 1,...,n_{ti}$) response (concentration) for a selected chemical at the i^{th} ($i = 1,...,K$; $K=12$) distance group, collected at time points t ($t = 1,...,T$; $T = 2$). Here, n_{ti} denotes the number of observations at time t for the i^{th} distance for the selected chemical.

We recognize that for given i and j, y_{1ij} and y_{2ij} will be longitudinally correlated. Similarly, for a given t and for $i \neq l$, y_{tij} and y_{tlj} will be spatially correlated as we compare the observations in each of the distances away from the Hibernia platform.

Next, as shown in previous sections, the distributions of $y_{ti1}, \ldots, y_{tin_{ti}}$ in some cases were normal and in most of the other cases they were found to be asymmetrical.

Finally, we let $\mu_{ti} = E\left(y_{tij}\right)$ for all j and we also define m_{1i} and m_{2i} as the medians of the i^{th} group for time 1 and 2 respectively. In this practicum, our main objective will be to study the possible change in the chemical concentration levels surrounding the Hibernia platform. For this purpose, we plan to test either:

$$H_o : \mu_{1i} = \mu_{2i} \text{ or } H_o : m_{1i} = m_{2i} \text{ or both as appropriate.}$$

Next, we define μ_1 and μ_2 as the means of the responses for the combined data set at times 1 and 2 respectively. Similarly, we define m_1 and m_2 as the corresponding medians for the combined data. Another objective of this practicum is to test whether an overall change occurred in the sediment chemistry from 1995 to 1998 at all distances. That is, we test either:

$$H_o : \mu_1 = \mu_2 \text{ or } H_o : m_1 = m_2 \text{ or both as appropriate.}$$

Chapter 3

Distribution-Free Test Based on Rank Spacings

3.1 Introduction

As shown in Chapter 2, many of the chemical samples follow asymmetrical distributions. In this chapter, we use a distribution free approach to test the equality of the distributions of two samples collected at times 1 and 2. This we do for the samples at a given distance. Note that we realize that these samples are longitudinally correlated. However, to begin with, we perform a distribution free test under a working assumption that the samples are independent. The correlation existing between these samples will be accounted for in later chapters. Further note that the equality of the distributions will imply that there was no significant change in chemical concentration in 1998 as compared to 1995.

To be specific, we follow the distribution-free test based on Rank Spacings as suggested by Kaigh (1994). In notation, at a given distance i (i=1,...,12), we define our test to be:

$$H_o : F_{y1i} = F_{y2i}$$

where F_{y1i} is the continuous cumulative distribution function of y_{1i}. We note that y_{1i} is the sample chemical concentration of size n_{1i} for the year 1998 ($t = 1$) at a given distance i (i = 1,...,12). Similarly, F_{y2i} is the continuous cumulative distribution function of y_{2i}, where y_{2i} is the sample chemical concentration of size n_{2i} for the year 1995 ($t = 2$) at a given distance i (i = 1,...,12).

For convenience, however, we will now drop the subscript i as it is our intention in this chapter to test the equality of two distributions at a given distance. For example, consider the distance $i = 1$ corresponding to the first distance, 250 meters from the GBS drilling site, and the y_1 Barium sample observations for 1998 are as follows:

$$y_1 \equiv [380, 930, 220, 230, 190, 250, 160, 220, 400, 1400, 310, 220]$$

Similarly, the 1995 ($t = 2$) Barium sample y_2 at 250 meters is given as follows:

$$y_2 \equiv [160, 130, 180, 190, 210, 120, 130, 140, 140, 170, 120, 160]$$

Although we recognize here that y_1 and y_2 are longitudinally correlated, we use a working independence in order to test the equality of the distributions. We use the distribution-free test as suggested by Kaigh (1994) as the distributions of y_1 and y_2 are mostly asymmetrical as summarized in the previous chapter (see Table 5, section 2.2).

3.2 Computation of the Test Statistic

In order to construct the test statistic, for convenience, we follow the notation used by Kaigh, (1994, section 1.2, pages 159-161). We thus order the observations of the y_1 sample and write $y_{1_{1:n_{1i}}} < \ldots < y_{1_{n_{1i}:n_{1i}}}$ and the combined-sample ordered ranks as $R^{y_1}_{j:n_{1i}} = rank\left(y_{1_{1:n_{1i}}}\right) = j + \sum_{1 \le l \le n_{2i}} I\left(y_{2_l} < y_{1_{1:n_{1i}}}\right), 1 \le i \le n_{1i}$. Here, $I(\cdot)$ is an indicator variable that takes the value 1 for $y_{2_l} < y_{1_{1:n_{1i}}}$ and otherwise 0. Under the null hypothesis H_o: $F_{y1}=F_{y2}$, entries of $R^{y_1} = \left[R^{y_1}_{1:n_{1i}}, \ldots, R^{y_1}_{n_{1i}:n_{1i}}\right]^T$ are distributed as the order statistics in a simple random sample of n_{1i} from integers $(1, \ldots n_i)$ with $0 < 1 \le R^{y_1}_{1:n_{1i}} < \ldots < R^{y_1}_{n_{1i}:n_{1i}} \le n_i < n_i + 1$. The components of the $(n_{1i}+1)$ dimensional y_1 rank spacings vector $D^{y_1} = \left[D^{y_1}_1, \ldots, D^{y_1}_{n_{1i}+1}\right]$ are as follows:

$$D^{y_1}_1 = R^{y_1}_{1:n_{1i}};\, D^{y_1}_j = R^{y_1}_{j:n_{1i}} - R^{y_1}_{j-1:n_{1i}} \text{ for } 2 \le j \le n_{1i}$$

and

$$D^{y_1}_{n_{1i}+1} = n_i + 1 - R^{y_1}_{n_{1i}:n_{1i}}$$

At a given distance, the y_1 and y_2 samples for each chemical were first combined, sorted and ranked and the rank spacings D^{y_1} were produced. We now define y_1 rank spacing components as follows:

$$Z^{y_1}_{s;n_i} = -\left[\frac{(n_{1i}+1)(n_{1i}+2)}{n_{2i}(n_i+1)}\right]^{\frac{1}{2}} \times \sum_{1\leq j\leq n_{1i}+1} \pi_{s,n_{1i}}(j) D^{y_1}_j \tag{3.1}$$

where $\pi_{s,n_{1i}}(j) = \frac{p_{s,n_{1i}}(j)}{\left\|p_{s,n_{1i}}\right\|}$ are well known Hahn Polynomials with:

$$p_{1,n1i}(j) = 2j-(n_{1i}+2) \tag{3.2}$$

$$p_{2,n1i}(j) = 6j^2-6(n_{1i}+2)j+\left(n_{1i}^2+5n_{1i}+6\right) \tag{3.3}$$

$$p_{3,n1i}(j) = 20j^3-30(n_{1i}+2)j^2+\left(12n_{1i}^2+54n_{1i}+64\right)j-\left(n_{1i}^3+9n_{1i}^2+26n_{1i}+24\right) \tag{3.4}$$

$p_{4,n1i}(j) = 70j^4 - 140(n_{1i}+2)j^3 + (90n_{1i}^2 + 390n_{1i} + 470)j^2 - (20n_{1i}^3 +$
$150n_{1i}^2$
$+410n_{1i}+380)j+(n_{1i}^4+14n_{1i}^3+71n_{1i}^2+154n_{1i}+120)$

with $\left\|p_{s,n_{1i}}\right\| = \sqrt{\sum_{1\leq j\leq n_{1i}+1} p^2_{s,n_{1i}}(j)}$.

Similarly, we construct $Z_{s;n_i}^{y_2}$ for s = 1,...q. Next, we average $\sum\limits_{1 \leq s \leq q} \left(Z_{s;n_i}^{y_1}\right)^2$ and $\sum\limits_{1 \leq s \leq q} \left(Z_{s;n_i}^{Y_2}\right)^2$ and define the test statistic for the combined sample as follows:

$$\Psi_q^{y_1 y_2 2} = \sum_{1 \leq s \leq q} \frac{\left(\left(Z_{s;n_i}^{y_1}\right)^2 + \left(Z_{s;n_i}^{y_2}\right)^2\right)}{2} \tag{3.6}$$

see Kaigh, (1994, section 1.3 page 161). Under the null hypothesis, this test statistic follows a χ^2 distribution with ($q = 4$) degrees of freedom.

3.3 Analysis of the Test Results

In this section we summarize the results for the rank component procedure for the sediment data at each test site. Table 6 gives the value for the $\Psi_q^{y_1 y_2 2}$ test statistic (3.6) that we compare to $\chi^2_{(q=4)} = 9.49$ at level $\alpha = 0.05$.

Table 6. $\Psi_q^{y_1 y_2 2}$ test values for Sediment Chemistry using Rank Spacings Procedure

Distance	Barium 95/98	Iron 95/98	Lead 95/98	Chromium 95/98
$250m$	37.8	2.14	1.59	11.45
$500m$	16.1	9.68	12.96	3.54
$750m$	6.37	17.56	28.59	20.18
$1000m$	20.06	45.5	40.46	53.83
$1500m$	9.03	9.70	11.09	16.52
$2000m$	8.17	2.41	12.15	6.13
$2500m$	6.97	5.51	15.25	18.92
$3000m$	6.88	16.98	10.54	13.47
$4000m$	3.50	1.25	2.63	6.68
$6000m$	6.21	5.90	5.07	11.28
$8000m$	19.56	16.67	15.92	20.95
$16000m$	6.58	2.61	4.65	15.92

This table shows the change in the distribution of the chemical concentrations of each element before and after the drilling process began off the coast of Newfoundland in 1997. Barium and Iron distributions are significantly different in sites closest to the platform (1000m and less). The resulting patterns for Lead and Chromium are not as clear, however, the null hypothesis is rejected in the majority of the distances tested for these elements.

To better understand these results, a second analysis was performed after the removal of all outliers in each of the sediment data sets. Table A1 shows these results for the $\Psi_q^{y_1 y_2 2}$ test statistic. When the outliers are removed we see the same trend in the distribution for Barium which are significantly different in those areas closest to the platform. The results for Iron show a different trend once the outliers are removed as these data were found to have many outliers and to be highly asymmetrical. There does not seem to be significant differences in the Iron sample distributions based on the rank components procedure closer to the drilling area. The patterns for both Lead and Chromium become more clear as the null is rejected within predominantly smaller distances from the GBS.

3.4 Conclusion

The Rank Spacings Components method (Kaigh (1994)) accommodates for the asymmetrical nature of the sediment data to test for chemical concentration differences at each of the collection locations. However, this procedure does not take longitudinal correlations into consideration. Nevertheless, the application of the $\Psi_q^{y_1 y_2 2}$ test revealed some useful patterns for the changes in chemical concentrations. This test, in particular, revealed the importance of the distance factor that one should consider in understanding the possible changes in sediment chemistry from 1995 to 1998. It also reveals that outliers should be considered in the decision making process.

Chapter 4

Modified Cornish-Fisher t test for Asymmetrical Populations

4.1 Introduction

In the last chapter, we have tested for equality of the distributions for y_1 and y_2, where y_1 and y_2 refer to the chemical concentration samples in 1998 and 1995 respectively. In this chapter, we concentrate on testing the mean level of chemical concentrations in the populations corresponding to y_1 and y_2. Once again, as these samples are asymmetrically distributed we can not apply the traditional two sample t test. Some authors, such as Johnson (1978) and Chen (1995) have developed a modified Cornish-Fisher t test to test the mean concentration level for an asymmetrical population.

We follow these authors and develop a modified two sample Cornish-Fisher t test to test for significant differences in sediment chemistry before and after the oil drilling process began in 1995. Note however that although these samples are longitudinally correlated, in the spirit of the last chapter, we assume a working independence between the 1995 and 1998 chemical concentration samples.

Many outliers were detected in these sediment samples as shown in the exploratory analysis (section 2.2). For this reason, we conduct the modified Cornish-Fisher t test with and without these outliers included. We also recognize that by removing these outliers we are losing information. As a result, in the next chapter we will perform a sample median test where it is not necessary to remove moderate outliers.

4.2 Computation of the Test Statistic

Following Johnson (1978, section 2, page 537) and Chen (1995), we first write a one sample modified Cornish-Fisher test statistic T* for testing H$_o$: $\mu_i = \mu_i(0)$ where $\mu_i = E(y_{ij})$ with y_{ij} as the jth chemical concentration in the ith group or distance from the GBS drilling platform. The T* statistic is given by:

$$T^* = \left[(\overline{y_i} - \mu_i) + \lambda + \gamma\left[(\overline{y_i} - \mu_i)^2 - \left(\frac{\sigma_i^2}{n_i}\right)\right]\right] \cdot \left[\frac{s_i^2}{n_i}\right]^{-\frac{1}{2}} \qquad (4.1)$$

where $\overline{y_i}$ is the mean concentration of the sample of size n$_i$ for the chemical under consideration for the ith distance, s_i^2 is the variance of the corresponding sample and σ_i^2 is also estimated by s_i^2. We recognize that μ_i is the population mean of y_i and μ_{3_i} is the third central moment of y_i. Also in 4.1 we define γ *and* λ as follows:

$$\gamma = \frac{\mu_{3_{1i}}}{3\sigma_{1i}^4} \text{ and } \lambda = \frac{\mu_{3_{1i}}}{2n_i\sigma_{1i}^2} \qquad (4.2)$$

In our study, we need to generalize the one sample T^* test statistic to the case of two samples as it is our intention to test for changes in sediment chemistry between both test years, 1995 and 1998. We generalize the modified one sample Cornish-Fisher t test to produce the following two sample modified Cornish-Fisher t test given by T^{**}:

$$T^{**} = \frac{\left[(\overline{y_{1i}} - \overline{y_{2i}}) - (\mu_{1i} - \mu_{2i}) + \hat{\lambda} + \hat{\gamma}\left[((\overline{y_{1i}} - \overline{y_{2i}}) - (\mu_{1i} - \mu_{2i}))^2 - \left(\frac{\sigma_{1i}^2}{n_{1i}} + \frac{\sigma_{2i}^2}{n_{2i}}\right)\right]\right]}{\sqrt{\left(\frac{s_{1i}^2}{n_{1i}} + \frac{s_{2i}^2}{n_{2i}}\right)}} \tag{4.3}$$

where $\overline{y_{1i}}$ and $\overline{y_{2i}}$ are the mean concentrations of the samples of size n_{1i} and n_{2i} for the chemical under consideration. These two samples correspond to 1998 and 1995 respectively. For the ith distance, μ_{1i} and μ_{2i} are the population mean concentrations of the chemical under consideration in 1998 and 1995 respectively. Further, s_{1i}^2 and s_{2i}^2 are the variances of the corresponding samples y_{1i} and y_{2i} and σ_{1i}^2 and σ_{2i}^2 are also estimated by s_{1i}^2 and s_{2i}^2 respectively.

The definitions for $\hat{\lambda}$ and $\hat{\gamma}$ (generalizations of (4.2)) are as follows:

$$\hat{\lambda} = \left(\frac{\hat{\mu}_{3_{1i}}}{6n_{1i}\hat{\sigma}_{1i}^2} - \frac{\hat{\mu}_{3_{2i}}}{6n_{2i}\hat{\sigma}_{2i}^2}\right) - \hat{\gamma}\left(\frac{\hat{\mu}_{3_{1i}}}{6n_{1i}\hat{\sigma}_{1i}^2} - \frac{\hat{\mu}_{3_{2i}}}{6n_{2i}\hat{\sigma}_{2i}^2}\right)^2 + \hat{\gamma}\left(\frac{\hat{\sigma}_{1i}^2}{n_{1i}} + \frac{\hat{\sigma}_{2i}^2}{n_{2i}}\right) \ , \tag{4.4}$$

where

$$\hat{\gamma} = \left[\frac{a\left(\frac{\hat{\sigma}_{1i}^2}{n_{1i}} + \frac{\hat{\sigma}_{2i}^2}{n_{2i}}\right)^{-1} d_2^{-1} - 2d_1^{-1}}{a\left(\frac{\hat{\sigma}_{1i}^2}{n_{1i}} + \frac{\hat{\sigma}_{2i}^2}{n_{2i}}\right) + 2d_1 d_2^{-1} - 4d_1^{-1} d_2}\right] \tag{4.5}$$

In 4.5 we define d_1, d_2 and a as follows:

$$d_1 = \left(\frac{\hat{\sigma}_{1i}}{\sqrt{n_{1i}}} - \frac{\hat{\sigma}_{2i}}{\sqrt{n_{2i}}} \right) \tag{4.6}$$

$$d_2 = \left(\frac{\hat{\mu}_{3_{1i}}}{6 n_{1i} \hat{\sigma}_{1i}^2} - \frac{\hat{\mu}_{3_{2i}}}{6 n_{2i} \hat{\sigma}_{2i}^2} \right) \tag{4.7}$$

$$a = \left[\left(\frac{\hat{\sigma}_{1i}^2}{n_{1i}} \sqrt{\left(\frac{\hat{\mu}_{4_{1i}} - \hat{\sigma}_{1i}^4}{n_{1i} \hat{\sigma}_{1i}^4} \right)} \frac{\hat{\mu}_{3_{1i}}}{\hat{\sigma}_{1i}^2 \left(\hat{\mu}_{4_{1i}} - \hat{\sigma}_{1i}^4 \right)} \right) + \left(\frac{\hat{\sigma}_{2i}^2}{n_{2i}} \sqrt{\left(\frac{\hat{\mu}_{4_{2i}} - \hat{\sigma}_{2i}^4}{n_{2i} \hat{\sigma}_{2i}^4} \right)} \frac{\hat{\mu}_{3_{2i}}}{\hat{\sigma}_{2i}^2 \left(\hat{\mu}_{4_{2i}} - \hat{\sigma}_{2i}^4 \right)} \right) \right] \tag{4.8}$$

This new Cornish-Fisher two sample t statistic was used to test the sediment concentration differences in the 1995 and 1998. We also calculated the regular two sample t statistic to compare values for both methods (Table A2). Results for all tests will be discussed in the following sub-section.

4.3 Analysis of the Test Results

The regular two sample t statistic as well as modified t test statistic (4.3) were calculated for each chemical at each distance from the Hibernia GBS. Once again, we reject the null hypothesis that there is no difference in chemical concentration from 1995 to 1998 at level $\alpha = 0.05$ and $(n_{1i} + n_{2i} - 2)$ degrees of freedom. Table 7 summarizes the statistical results for the modified two sample t test.

Table 7. Test Results for Sediment Chemistry using Cornish-Fisher Modified Two Sample t Procedure

Distance	Barium 95/98	Iron 95/98	Lead 95/98	Chromium 95/98
250m	2.55	1.23	−30.6	1.83
500m	3.47	1.47	−499.5	0.35
750m	1.31	−2.06	−381.1	155.24
1000m	1.52	−1.64	−52.71	−13.95
1500m	−8.36	−2.31	11.46	−25.93
2000m	0.61	−0.52	−2.9	−280.73
2500m	−1.29	−2.21	−134.1	33.58
3000m	−0.39	−3.78	4.29	−1.66
4000m	1.15	−0.68	−2.23	−0.14
6000m	1.02	−0.62	1.01	−0.37
8000m	−0.71	0.56	4.84	−0.32
16000m	−0.8	−2.94	8.48	−71.08

This table shows that within 500m of the Hibernia drilling platform, both Barium and Lead have significantly different concentration levels in 1998 compared to 1995. At more distant collection sites, tests for both Iron and Barium do not show significant changes in the concentration amounts (with the exception of iron at 16,000m). Chromium levels are significantly different in areas between 750 and 2500 meters and no differences are detected between 3000 and 8000 meters. It is also noticed that the results for the Cornish-Fisher modified T and the regular T (Table A2) agree in the majority of cases.

Once again, to understand the effect of possible outliers present in the data, we removed the outliers and re-tested the samples. Table A3 shows the final results for the Cornish-Fisher modified t Test and Table A4 shows the regular T Test, both with these outliers removed. These figures clearly show the same trend for Barium both close to and further away from the GBS. The output for Iron, Lead and Chromium is different when the outliers are removed and a clear pattern is more difficult to determine. It is also noted that there is less agreement between the regular and the two sample Cornish-Fisher t statistics when these outliers are excluded.

4.4 Conclusion

The Cornish-Fisher modified t Test method accommodates for the asymmetrical nature of the sediment data to test for chemical concentration differences from 1995 to 1998. However, similar to the Rank Spacings test, this procedure also does not take longitudinal correlations into consideration. Thus, we assumed a working independence between the samples although we recognize that a longitudinal correlation exists between the 1995 and 1998 samples.

Using the modified two sample t test statistic we continue to see trends within the chemical data as the distance increases from the GBS platform. Once again, we clearly see the significance of the outliers in these data and how they affect the results of different statistical tests.

Chapter 5

Standardized Median Test (SMT) Using A Working Dispersion

5.1 Introduction

In the previous two chapters, we tested the two asymmetrical chemical samples using a working independence assumption, whereas these samples collected in the same locations in both 1995 and 1998 are in fact longitudinally correlated. To accommodate for this longitudinal correlation, we now follow the paired t-test rationale. To be specific, we take the differences between the observations in both time points so that the resulting differences become time independent. This leads to a one sample test for testing that the mean and median of the new population is zero.

However, we test for the median concentration levels as these samples are generally asymmetric.

Note that unlike the paired t-test, in our situation we have unequal sample sizes. Consequently, we were unable to use a one sample Cornish-Fisher t-test modified for asymmetrical data. As a remedy, we use the well known permutation approach to determine if there is a significant difference in chemical concentrations before and after the GBS drilling operations. This approach includes a Standardized Median Test (SMT) using a Working Dispersion (Hoyles, Sutradhar and Friel (2000)) and determines if the median is significantly different from zero indicating that the difference in the sediment chemical concentration between 1995 and 1998 will be significant. The null hypothesis assumes that the median is zero and the test statistic has a asymptotically standard normal distribution. This test performs well for moderately asymmetrical data. In the following sections we will define the SMT approach and summarize the results for each chemical and test site.

5.2 Computation of the Test Statistic

In this section, we follow the notation as suggested by Hoyles, Sutradhar and Friel (2000) . In the first step, the $n_i^* = n_{1i} \times n_{2i}$ differences namely, $\delta_{ji}= y_{1ij_1} - y_{2ij_2}$ for $j = 1,...,n_i^*$ with $j_1 = 1,...,n_{1i}$ and $j_2 = 1,....,n_{2i}$ are calculated. Here, y_{1ij} is the j^{th} observation in the 1998 chemical data set at the i^{th} distance and y_{2ij} is the j^{th} observation in the 1995 chemical data set also at the i^{th} distance. If the median of the n^* differences significantly differs from zero, then there is a significant difference in the sediment chemistry of the ocean floor near the GBS from 1995 to 1998. This approach consequently takes care of both longitudinal correlations and the asymmetrical nature of the data, especially when the test is computed for a given distance.

For simplicity, we denote these n_i^* differences by $\left(\delta_1, ...\delta_{n_i^*}\right)$belonging to the i^{th} group or distance from the Hibernia drilling area. Let m_d denote the median of the n_i^* differences and let s_i^* denote the dispersion of these differences.

In notation,

$$s_i^* = \frac{Median\,|\delta_j - m_{di}|}{0.6745}, \; for\ j = 1, ..., n_i^* \qquad (5.1)$$

We can construct a normal test statistic as

$$z_i^* = \sqrt{n_i^*}\,(m_{di} - M_{di})\,/s_i^*\ , \qquad (5.2)$$

and refer to this as the Standardized Median Test (SMT).

Note that this test statistic (5.2) may be treated as a working normal test statistic. This is because it is constructed following a regular z test by replacing the mean by the median and the standard deviation by dispersion, appropriate for these longitudinal and asymmetrical data.

5.3 Analysis of the Test Results

The differences between the sediment concentrations at each of the twelve distances from the GBS drilling area were tested using this SMT z_i^* test statistic. These results are summarized in Table 8 where the null hypothesis was rejected at $\alpha = 0.05$ level for Z $\leq$ -1.96 and Z $\geq$ 1.96.

Table 8. $\mathbf{z}_i^*$ Test Statistic Results for Sediment Chemistry using SMT Procedure

Distance	Barium 95/98	Iron 95/98	Lead 95/98	Chromium 95/98
$250m$	11.6	2.43	2.02	8.72
$500m$	10.41	6.24	−3.47	8.32
$750m$	0.00	−4.28	−10.71	−4.02
$1000m$	4.96	−3.97	−4.97	−7.43
$1500m$	−3.36	−7.29	−6.56	−8.74
$2000m$	0.00	−2.03	−1.04	−2.40
$2500m$	−8.34	−8.28	−16.21	−13.90
$3000m$	−3.63	−10.88	−6.53	−8.55
$4000m$	2.11	−1.68	2.11	2.11
$6000m$	1.01	−2.74	−1.56	−3.12
$8000m$	−3.57	−4.28	−1.07	−4.59
$16000m$	−4.37	0.00	0.00	−6.56

This table shows that within 500m of the Hibernia drilling platform, all chemicals have significantly different concentration levels in 1998 compared to 1995. At collection sites further away, tests for lead do not show significant changes in the concentration amounts. Chromium levels are significantly different in all areas including up to 16,000 meters. We also see the appearance of zeros indicating the median difference between the two test years was zero.

We removed the outliers and re-examined the samples to see their effect within these data. Table A5 shows the final results for the SMT Test with these outliers removed. Barium, lead and chromium all show significant differences in concentration levels close to the drill site. The output for iron is also erratic when the outliers are removed and a clear pattern is difficult to define. It is also noted however, that in the majority of cases the test results are similar to those concluded using procedures from previous chapters when the outliers are excluded.

5.4 Conclusion

Similar to the Cornish-Fisher modified t test method, the SMT accommodates the asymmetrical nature of the sediment data to test for differences in sediment chemistry from 1995 to 1998. This test, unlike the Cornish-Fisher modified t test, also takes care of longitudinal correlations. This is because the test was developed based on the computed differences of concentrations between 1995 and 1998. In the next chapter, we combine the data from all locations and generalize this SMT to test for change in chemical levels after taking the additional spatial correlations between observations collected from different distances into account.

Chapter 6

Generalized SMT for Combined Sample

6.1 Introduction

In previous chapters we performed several tests comparing the y_{1i} and y_{2i} samples where y_{1i} and y_{2i} refer to the chemical concentration data collected at the i^{th} distance (i = 1,...,12) in 1998 and 1995 respectively. In this chapter, we now combine all these sediment data collected at each of the twelve distances from the Hibernia drilling area for the two time points 1995 and 1998. Again, we wish to test if any overall significant change occurred in chemical contamination after the drilling process began in 1995.

In the spirit of the last chapter, we accommodate for the longitudinal correlation existing between the y_{1i} and y_{2i} samples by following the paired t-test rationale.

For this, we take the differences between the observations in both time points so that the resulting differences become time independent. This leads to a one sample test for testing the hypothesis that the mean or median of the new population is zero.

In section 6.2, we compute a z test statistic based on the mean of these time independent differences. We test these combined data using a sample mean and working variance estimate. Next, in section 6.3, similar to the previous chapter, we apply this same rational and test the combined sample using a modified Standardized Median Test with a working variance (Hoyles, Sutradhar and Friel, 2000). In the following sub-sections we show the derivation of these formulas in brief for these procedures and present the test results for the combined distances for all chemicals.

6.2 Computation of the Z Test Statistic Based on the Mean

As in the previous chapter, we first find the differences between the 1995 and 1998 chemical concentrations. In this chapter, however, we compute these differences for all distances combined. Recall that n_{1i} and n_{2i} denote the number of observations in 1998 and 1995 respectively for the i^{th} distance. Let

$$n_1 = \sum_{i=1}^{k} n_{1i} \text{ and } n_2 = \sum_{i=1}^{k} n_{2i},$$

with k=12 being the number of groups based on their distances from the drilling operations. Consequently, we have $n_1 \times n_2 = n^*$ differences. For j = 1,...,n*, $j_1 = 1,...,n_1$ and $j_2 = 1,...,n_2$, the n* differences can be computed as $\delta_j = y_{1j_1} - y_{2j_2}$ where y_{1j_1} is the j_1^{th} observation in the 1998 chemical data set and y_{2j_2} is the j_2^{th} observation in the 1995 chemical data set. To test for n* differences, we use a pooling technique and pool the information from all k distance groups as follows.

For the i^{th} distance we have $\delta_{j1} = y_{1j_{1i}} - y_{2j_{2i}}$ for $j_1 = 1,....n_{1i}$ and $j_2 = 1,....n_{2i}$ with $n_{1i} \times n_{2i}$ difference values. Let

$$n_d = \sum_{i=1}^{k} n_{1i}n_{2i} = \sum_{i=1}^{k} n_i^*$$

The purpose of this section is to test $\overline{D} = 0$, where $\overline{D}$ is the population mean corresponding to $\overline{\delta}$ computed from all the n_d differences. Note that the present test is developed in such a way that the spatial correlation between observations at different distances is taken into account. Thus, both longitudinal and spatial correlation will be taken into account while testing for a possible change in mean chemical concentration level.

Let $\overline{\delta_1}......\overline{\delta}_k$ denote the sample means of the differences for all k = 12 distances. We combine these means of differences to create an overall mean $\overline{\delta}$ which we define as follows:

$$\overline{\delta} = \frac{(n_{11} \cdot n_{21})\,\overline{\delta_1} + ...\,(n_{1k} \cdot n_{2k})\,\overline{\delta}_k}{n_d} \tag{6.1}$$

where $\overline{\delta}_i$ is the mean of the difference values for the ith group.

Under the null hypothesis that there is no change in the mean chemical concentration level, we have $\mathrm{E}\left(\overline{d}\right) = 0$ and the variance of $\overline{d}$ can be estimated as follows:

$$\hat{V}\left(\overline{d}\right) = \sum_{i=1}^{k} \frac{n_i^{*2}}{n_d^2} \cdot \frac{\hat{\sigma}_i^2}{n_i^*} + 2\sum\sum_{i<l} \frac{n_i^* n_l^*}{n_d^2} \cdot cov\left(\overline{\delta_i}, \overline{\delta_l}\right), \tag{6.2}$$

where $cov\left(\overline{\delta_i}, \overline{\delta_l}\right)$ is given by:

$$cov\left(\overline{\delta_i}, \overline{\delta_l}\right) = cov\left(\frac{\sum\limits_{j_1=1}^{n_i^*} \delta_{ij_1}}{n_i^*}, \frac{\sum\limits_{j_2=1}^{n_l^*} \delta_{lj_2}}{n_l^*}\right) = \frac{1}{n_i^* n_l^*} \sum_{j_1=1}^{n_i^*} \sum_{j_2=1}^{n_l^*} cov\left(\delta_{ij_1}, \delta_{lj_2}\right) \tag{6.3}$$

Note that to compute $cov\left(\sigma_{ij_1}, \sigma_{lj_2}\right)$we first express this as the following:

$$cov\left(\sigma_{ij_1}, \sigma_{lj_2}\right) = corr\left(\sigma_{ij_1}, \sigma_{lj_2}\right) \sigma_i \sigma_l.$$

Since σ_{ij_1} and σ_{lj_2} refer to the values in groups i and l, one requires an appropriate function to accommodate for this spatial relationship. There exists some work dealing with this type of spatial correlation. For convenience, we follow Jones and Vecchia (1993, page 953), and use a Gaussian correlation function: $corr\left(\sigma_{ij_1}, \sigma_{lj_2}\right) \leq e^{-b^2 r^2_{|i-l|}}$ where $\mathrm{r}_{|i-l|}$ is a suitable function to represent the proportional distances between observations in the i^{th} and l^{th} groups. To be specific, we consider $\mathrm{r}_{|i-l|} = |i-l|$. It then follows that:

$$cov\left(\overline{\delta_i}, \overline{\delta_l}\right) = \frac{1}{n_i^* n_l^*} \sum_{j_1=1}^{n_i^*} \sum_{j_2=1}^{n_l^*} \sigma_i \left[e^{-b^2 r^2_{|i-l|}}\right] \sigma_l \qquad (6.4)$$

We calculate individual $\sigma's$ for each k = 1,... 12 distances and we set b=1.00 for simplicity. Therefore, (6.2) reduces to:

$$\hat{V}\left(\overline{d}\right) = \sum_{i=1}^{k} \frac{n_i^*}{n_d^2} \cdot \hat{\sigma}_i^2 + 2 \sum\sum_{i<l} \frac{n_i^* n_l^*}{n_d^2} \hat{\sigma} i \hat{\sigma}_l e^{-b^2 r^2_{|i-l|}} \qquad (6.5)$$

Now, as we have constructed a combined sample mean $\overline{d}$ and its appropriate variance $\hat{V}\left(\overline{d}\right)$, we may now construct an asymptotically normal test statistic given by:

$$Z = \frac{\overline{d} - 0}{\sqrt{V\left(\overline{d}\right)}}, \qquad (6.6)$$

for testing $\overline{D} = 0$.

6.3 Computation of the Z Test Statistic Based on the Median

The test statistic in section 6.2 is based on the mean which is not an appropriate function for asymmetrical data. In this section, we extend the combined sample rationale to test about the median which is a more suitable function when dealing with asymmetrical data. For this procedure we once again find the differences between the 1995 and 1998 chemical concentrations. As it is our purpose to combine the sample data, therefore we recall that n_{1i} and n_{2i} are the number of observations in 1998 and 1995 respectively for the i^{th} distance. For the i^{th} distance we have $\delta_{j1} = y_{1j_{1i}} - y_{2j_{2i}}$ for $j_1 = 1, \ldots n_{1i}$ and $j_2 = 1, \ldots n_{2i}$ with $n_{1i} \times n_{2i}$ difference values. Let $n_d = \sum_{i=1}^{k} n_{1i} n_{2i}$, as in section 6.2.

The purpose of this section is to test $M = 0$, where M is the population median corresponding to the median δ_m computed from all the n_d differences. If the median of the n^* differences significantly differs from zero, then there is a significant difference in the sediment chemistry of the ocean floor near the GBS from 1995 to 1998. We note that both longitudinal and spatial correlation will be taken into account while testing possible change in median chemical concentration level.

Let $\delta_{m1}......\delta_{mk}$ denote the sample medians of the differences for all k = 12 distances. We combine these medians of differences to create an overall median δ_m which we define as follows:

$$\delta_m = \frac{(n_{11} \cdot n_{21})\, \delta_{m1} + ...\, (n_{1k} \cdot n_{2k})\, \delta_{mk}}{n_d} \tag{6.7}$$

where δ_{mi} is the median of the difference values for the ith group.

We apply the same procedure as seen in section 6.2 for the variance. In the present case, we compute the dispersion instead of variance, which is given by:

$$\text{Dispersion}\,(\delta_m) = \sum_{i=1}^{k} \frac{n_i^*}{n_d^2} \cdot \sigma_i^2 + 2 \sum_{i<l}\sum \frac{n_i^* n_l^*}{n_d^2} \cdot \sigma_i \left[e^{-b^2 r^2_{|i-l|}}\right] \sigma_l, \tag{6.8}$$

where by analogy as in section 5.2 we define σ_i as:

$$\sigma_i = \frac{Median\,|\delta_{mi} - \delta_m|}{0.6745}, \tag{6.9}$$

where δ_m is the median of the sample medians of the differences for all k = 12 distances denoted $\delta_{mi.}$. Now, once again, as we have constructed a combined sample median δ_m and its appropriate variance or Dispersion(δ_m), we may now construct an asymptotically normal test statistic given by:

$$Z = \frac{\delta_m - 0}{\sqrt{Dispersion\,(\delta_m)}} \tag{6.10}$$

for testing M = 0.

6.4 Analysis of the Test Results

The mean and median z tests with working dispersion were completed for the combined samples including concentration levels for all chemicals at each of the twelve distances from the Hibernia oil drilling platform. Table 9 reveals the test results for each of these procedures at level $\alpha = 0.05$. where the null hypothesis was rejected for z $\leq$ -1.96 and z $\geq$ 1.96.

Table 9. Z Test Statistic for Combined Sample Sediment Chemistry using Mean and Median Estimates with Working Dispersion.

Combined Sample	Barium 95/98	Iron 95/98	Lead 95/98	Chromium 95/98
Mean Test	1.04	-0.604	-0.74	-0.224
Median Test	1.50	-10.98	-0.82	-1.06

As table 9 shows, with the exception of Iron, all chemicals are not significantly different from 1995 to 1998. It is noted, also with the exception of Iron, that both the mean and median tests agree with the decision of not rejecting the null hypothesis.

6.5 Conclusion

These tests show that overall, when the data is combined, the amounts of chemical contamination after the drilling process is not significant. However as seen in previous chapters, when the data is tested at each distance, elements closer to the drill site are detected to have significantly higher concentrations. By combining the data into one large set, valuable information is lost due the spatial correlation present among the test sites. Areas closer to the drilling platform show significantly higher concentrations in 1998.

Chapter 7

Conclusion

In this practicum, we examined sediment chemistry data collected from the Hibernia Gravity Based Structure (GBS). The measurement of these elements occurred in both 1995 and 1998 to determine if this drilling process resulted in a significant contamination of the sediment on the ocean floor surrounding the GBS drill site. Statistical analysis and summary of the sediment chemistry data were released in 1999 by Jacques Whitford Environment Ltd. St. John's, Newfoundland. The objective of this practicum was to more extensively examine these data and suggest alternative tests accommodating for the fact that the samples extracted for certain chemicals actually followed asymmetrical distributions. We also accounted for the naturally occurring longitudinal and spatial correlation existing in the data. After a thorough exploratory analysis was complete, this project involved the modification of existing statistical methods to test for significant differences at each distance from the GBS platform.

Statistical tests previously released in the Hibernia statistics report revealed that chemical concentrations for Barium, Iron, and Lead were significantly different from samples taken in 1995 and 1998. It was reported that there was a significant increase in many of the chemicals after the start of the oil drilling process within 500m of the GBS. These tests did not accommodate for the asymmetrical nature of the data and they also did not accommodate for spatial correlation between the test sites. As a remedy, an extensive analysis was completed as part of this practicum. The tests included a Rank Components Test, a modified T Test using Cornish-Fisher Expansion, a Median Test with a Working Dispersion, and additional Z tests based on the mean and the median for the combined data from all the distances.

Our exploratory analysis summarized in Chapter 2 showed that many of the chemical samples followed asymmetrical distributions. As a result we first used a distribution free approach to test the equality of the distributions for each distance in 1995 and 1998. As these samples are longitudinally correlated we followed a Rank Spacings Components procedure, using the $\Psi_q^{y_1 y_2 2}$ test statistic with $q = 4$, under a working assumption that the samples were independent. This test revealed common trends within the chemical concentration data. Closer to the Hibernia drilling platform the chemical concentration levels are significantly different when compared to 1995. This result is the most clear for the Barium data set.

Next, we concentrated on testing the mean level of chemical concentrations in the populations corresponding to the 1995 and 1998 samples. As these samples were determined to be asymmetrically distributed we were not able to apply the traditional two sample t test. We followed Johnson (1978) and Chen (1995) and developed a modified Cornish-Fisher t test in chapter 4 to test the mean concentration level for an asymmetrical population. Although these samples are also naturally longitudinally correlated, we once again assumed a working independence between the 1995 and 1998 chemical concentration samples. Results from Cornish-Fisher modified t test statistic continue to display trends within the element data as the distance increases from the GBS platform.

In the next procedure in chapter 5, we accommodated for the longitudinal correlation between the 1995 and 1998 samples by following the paired t-test rationale. We calculated the differences between the observations in both time points so that the resulting differences become time independent. This lead to a one sample test for testing that the mean or median of the new population is zero. As we were dealing with unequal sample sizes and asymmetrical data, we used a well known permutation approach to determine if there is a significant difference in median chemical concentrations before and after the GBS drilling operations.

This approach included a Standardized Median Test (SMT) using a Working Dispersion and determines if the median is significantly different from zero indicating that the difference in the sediment chemical concentration between 1995 and 1998 will be significant (Hoyles, Sutradhar and Friel, 2000).

The SMT test statistic does not reflect similar trends within the element data as compared to these other test procedures. In all cases, the significance of outliers in the data was observed once these outliers were removed for further analysis.

In chapter 6, we combined all the data from all twelve distances and computed a sample mean and working variance estimate. Next, we applied the same rationale and tested the combined sample using a modified Standardized Median Test with a working variance (Hoyles, Sutradhar and Friel, 2000). This test also accommodates spatial correlations that may be present among the responses at different distances from the Hibernia platform. The modified SMT tests for the combined data showed that the amounts of chemical contamination after the drilling process is not significant. However, as seen in previous procedures outlined in this practicum, when the data is tested at each distance, elements closer to the drill site are detected to have significantly higher concentrations. By combining the data, important information is lost due the spatial correlation present among the collection sites.

In conclusion, it is the opinion of the author that areas closer to the drilling platform show significantly higher concentrations in 1998. Summaries for Barium and Lead clearly show this trend, while due to several outliers this trend was not as clearly defined for Iron and Chromium. Combined data test results of this practicum do not in fact differ significantly from those reported in 1999 by Jacques Whitford Environment Ltd.

Bibliography

[1] Chen, Ling (1995) , Testing the Mean of skewed Distributions, Journal of the American Statistical Association, 90, 767-772..

[2] Drilling & Processing (1999) . About Hibernia [On-line] Available: www.hibernia.ca

[3] Hoyles, M. and Sutradhar, B. and Friel, J. (2000) , Longitudinal Changes in Mineral Contents in Breast Milk from Lactating Mothers of Premature and Full term Infants, Statistical Methods 2(1), 1-28.

[4] Johnson, Norman J. (1978) , Modified t Tests and Confidence Intervals for Asymmetrical Populations, Journal of the American Statistical Association,73, 536-544.

[5] Jones, R. and Vecchia, A. (1993) , Fitting Continuous ARIMA Models to Unequally Spaced Spatial Data, Journal of the American Statistical Association, 88, 947-953.

[6] Kaigh, W.D. (1994) , Distribution-Free Two Sample Tests Based on Rank Spacings, Journal of the American Statistical Association,89, 159-167.

[7] Lye, L. and Bartholomew, L. (1999) , Hibernia Development Project, Jaques Whitford Environment, St. John's Newfoundland.

[8] Platform Support (1999) . About Hibernia [On-line] Available: www.hibernia.ca

Appendix

Table A1. Statistical Results for Rank Spacings Test ψ_q^{y1y22} - Outliers Excluded

Chemical/Distance	Barium 98/95	Iron 98/95	Lead 98/95	Chromium 98/95
250m	22.46	5.99	1.79	11.45
500m	16.30	7.25	12.96	3.54
750m	38.82	28.02	30.98	14.02
1000m	20.06	45.48	40.46	40.48
1500m	9.03	3.31	11.09	16.52
2000m	8.17	2.41	12.15	6.93
2500m	7.24	6.98	33.79	33.33
3000m	9.83	19.84	11.18	28.96
4000m	2.83	1.07	2.63	11.87
6000m	5.15	5.38	4.04	12.60
8000m	19.56	40.89	15.92	20.95
16,000m	6.58	1.91	4.65	15.92

Table A2. Statistical Results for Regular T Test - $T_{(n1i + n2i - 2)}$

Chemical/Distance	Barium 98/95	Iron 98/95	Lead 98/95	Chromium 98/95
250m	2.46	1.15	0.53	1.84
500m	3.49	1.46	-1.58	2.24
750m	1.27	-2.05	-3.09	-2.12
1000m	1.62	-1.58	-1.7	-2.22
1500m	-1.05	-2.25	-1.96	-2.42
2000m	0.61	-0.5	-0.29	-0.37
2500m	-1.61	-2.15	-3.29	-3.71
3000m	-0.39	-3.78	-1.77	-2.43
4000m	1.15	-0.66	0.29	-0.39
6000m	1.03	-0.65	0.22	-0.20
8000m	-0.88	0.5	3.28	-0.46
16,000m	-0.79	-2.72	0.55	-1.41

Table A3. Results for Cornish-Fisher Modified $T_{(n1i + n2i -2)}$ Test - Outliers Excluded

Chemical/Distance	Barium 98/95	Iron 98/95	Lead 98/95	Chromium 98/95
250m	4.14	-0.41	-140.24	1.83
500m	3.33	2.15	-499.5	0.35
750m	-0.82	-2.61	1256.05	4.67
1000m	1.52	-1.64	-52.71	5.01
1500m	-8.36	-1.66	11.46	-25.93
2000m	0.61	-0.52	-2.90	0.12
2500m	-1.93	-1.48	3355.84	-20.13
3000m	1.67	-4.07	-23.09	-1.18
4000m	0.62	-0.07	-2.23	-5.00
6000m	1.24	-1.10	-3.85	-1.47
8000m	-0.71	-1.37	4.84	-0.32
16,000m	-0.80	-2.94	-215.43	-71.08

Table A4. Statistical Results for Regular T Test - Outliers Excluded

Chemical/Distance	Barium 98/95	Iron 98/95	Lead 98/95	Chromium 98/95
250m	4.11	-0.40	0.77	1.84
500m	3.33	2.11	-1.58	2.24
750m	-0.90	-2.58	-3.63	-1.98
1000m	1.62	-1.58	-1.70	-2.04
1500m	-1.05	-1.73	-1.96	-2.42
2000m	0.61	-0.50	-0.29	0.14
2500m	-2.09	-1.60	-5.38	-4.56
3000m	1.60	-4.08	-0.94	-3.17
4000m	0.60	-4.08	0.29	-1.01
6000m	1.23	-1.06	-0.25	-7.07
8000m	-0.88	-1.63	3.28	-0.46
16,000m	-0.79	-2.72	-0.96	-1.41

Table A5. Results for Standardized Median Test - Z ($\alpha = 0.05$) - Outliers Excluded

Chemical/Distance	Barium 98/95	Iron 98/95	Lead 98/95	Chromium 98/95
250m	13.30	1.05	2.47	8.72
500m	10.09	6.06	-3.47	8.32
750m	-9.49	-6.19	-15.04	-4.89
1000m	4.96	-3.97	-4.97	-8.25
1500m	-3.36	-6.48	-6.56	-8.74
2000m	0.00	-2.03	-1.04	0.00
2500m	-9.42	-7.26	-21.86	-17.13
3000m	0.00	-11.66	-5.00	-10.90
4000m	2.70	0.00	2.11	0.00
6000m	1.24	-4.31	-3.62	-3.71
8000m	-3.57	-5.95	-1.07	-4.59
16,000m	-4.37	0.0	-3.69	-6.56

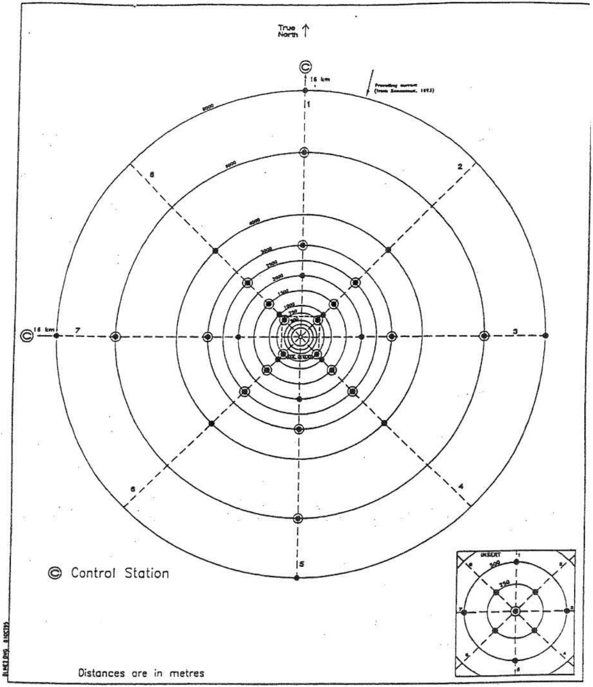

Figure 1: Baseline Sample Station Net For Hibernia EEM Program

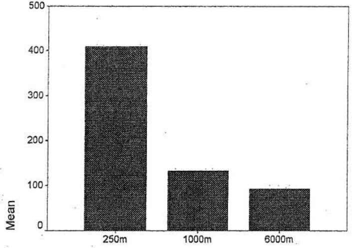

Figure 2. Sediment Chemistry 1998

Mean Barium Concentration By Distance

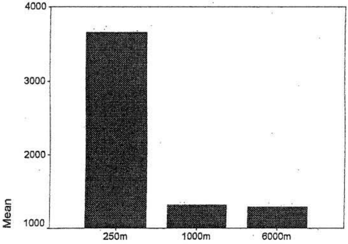

Figure 3. Sediment Chemistry 1998

Mean Iron Concentration By Distance

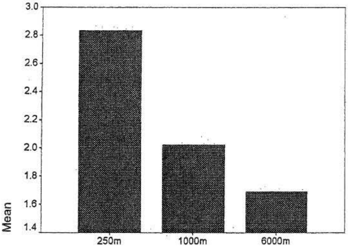

Figure 4. Sediment Chemistry 1998

Mean Lead Concentration By Distance

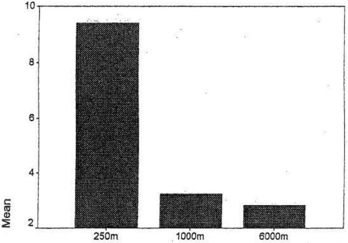

Figure 5. Sediment Chemistry 1998

Mean Chromium Concentration By Distance

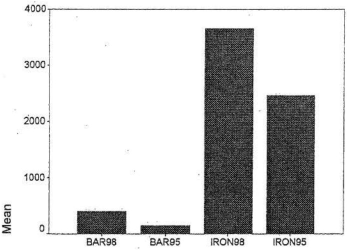

Figure 6. Mean Barium and Iron Concentrations 1995 & 1998

Distance = 250m From GBS

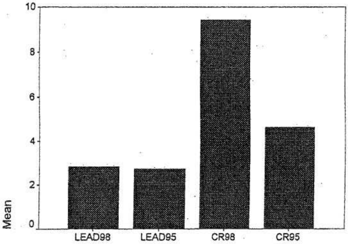

Figure 7. Mean Lead and Chromium Concentrations 1995 & 1998

Distance = 250m From GBS

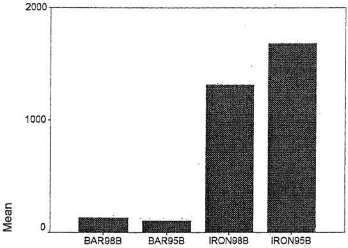

Figure 8. Mean Barium and Iron Concentrations 1995 & 1998

Distance = 1000m From GBS

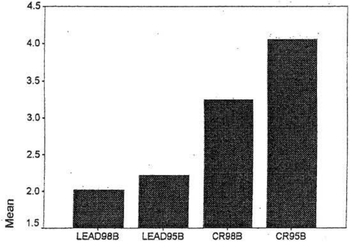

Figure 9. Mean Lead and Chromium Concentrations 1995 & 1998

Distance = 1000m From GBS

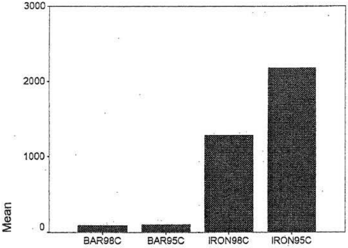

Figure 10. Mean Barium and Iron Concentrations 1995 & 1998

Distance = 6000m From GBS

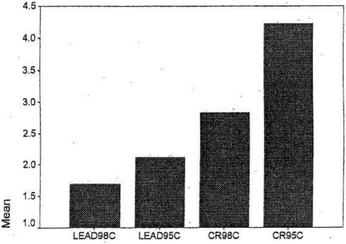

Figure 11. Mean Lead and Chromium Concentrations 1995 & 1998

Distance = 6000m From GBS

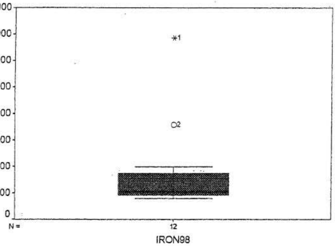

Figure 12. Iron 1998 - 250m From GBS

Conclusion: Asymmetrical

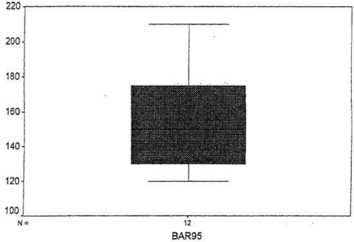

Figure 13. Barium 1995 - 250m From GBS

Conclusion: Symmetrical

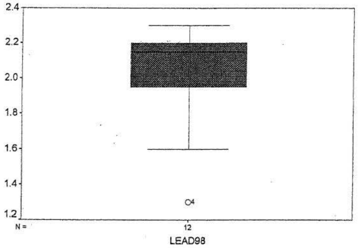

Figure 14. Lead 1998 - 1000m From GBS.

Conclusion: Asymmetrical

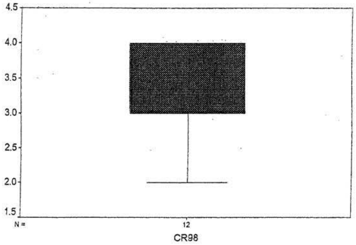

Figure 15. Chromium 1998 - 1000m From GBS

Conclusion: Asymmetrical

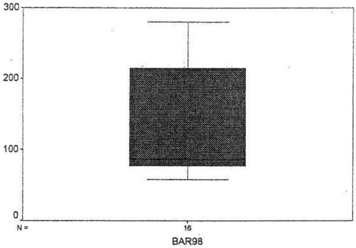

Figure 16. Barium 1998 - 6000m From GBS

Conclusion: Asymmetrical

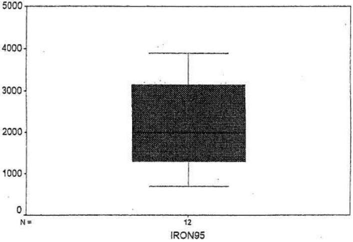

Figure 17. Iron 1995 - 6000m From GBS

Conclusion: Symmetrical