A TEST FOR DETECTING SPACE-TIME CLUSTERING AND A COMPARISON WITH SOME EXISTING METHODS

James C. Gear

A dissertation to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the Doctor of Public Health degree in the School of Public Health (Biostatistics)

Chapel Hill
2006

Approved by:

Advisor: Dana Quade

Reader: Shrikant I. Bangdiwala

Reader: Gerardo Heiss

Reader: Gary Koch

Reader: Michael J. Symons

ABSTRACT

JAMES C. GEAR: A Test For Detecting Space-Time Clustering
And A Comparison With Some Existing Methods
(Under the direction of Dana Quade)

This research introduces a new statistical test for evaluating space-time clustering in data where exact location and time information are available for the points of interest (cases). The test statistic, DP, is defined as the length of the path from $X_{[1]}$ to $X_{[n]}$ when the n cases are ordered by time of occurrence. Significance of the test is most appropriately determined by comparing the directed path length of the data to the empirical distribution of lengths obtained from all possible orderings of the n cases, or a random subset of those orderings when n is large.

The first three moments of DP are developed, and its properties are investigated using simulation on clustered and unclustered data. DP is then compared with Knox's test and Mantel's Generalized Regression using simulation on clustered and unclustered data. Two data sets from the literature (fifteen years of Burkitt's Lymphoma data from Uganda, and three years of birth defect data from California) are then used to compare the performance of these tests on actual data.

This work was largely done as this author's dissertation under the competent, inspirational, and greatly appreciated advisement of Professor Emeritus Dana Quade, Department of Biostatistics, UNC-Chapel Hill.

ACKNOWLEDGMENTS

I must first thank the Department of Biostatistics for its support during my extended stint as a student, and Drs. James Grizzle, Regina C. Elandt-Johnson, David Kleinbaum, Dennis Gillings and Roger Grimson for their advice during my graduate study. Thanks also to Dr. Chirayath Suchindran, the Director of Graduate Admissions, for supporting my continuing candidacy for this degree. Gratitude is also expressed to the members of my advisory committee; Drs. Shrikant Bangdawala, Gerardo Heiss, Gary Koch, and Michael Symons. Special appreciation is given to my dissertation advisor, Dr. Dana Quade, for consenting to work with me (even after retirement), and for his initial suggestion of this topic, his encouragement and support during the course of this research, his truly infinite patience with an over-committed graduate student, and the inspirational example of his tremendous scholarship.

I want to recognize professional colleagues who offered support and help. Thanks to the 'old' Bell Laboratories and to all of the statisticians in Depts. 1214 and 1215. Their initial support and encouragement changed my desire for a terminal degree from a dream to a real possibility, and to Ramanathan Gnanadesikan, Lorraine Denby, Jim Landwehr and Jon Kettering, for their personal concern. I want to recognize my current employer, Midwest Employers Casualty Company and Peter Shaw, for their support, and all of my past employers, who inadvertently "encouraged" me to complete this degree by refusing to give me a promotion without it. Thanks to Dr. Geoffrey Jacquez for letting me use his beta version of ***STATS!***, and to Dr. Tim Aldrich, who supplied the initial data for analysis in this work. Also, thanks to the late Dr. Joel Swartz for allowing me to consult with him in space-time clustering, and to Dr. Martin Kulldorff for helping me to stay current in the area of space-time clustering, and for offering invaluable advice at crucial moments during this work. And I greatly appreciate the word processing expertise of Kathy Fish, who created a coherent document from my disjointed chapters in different versions of Word.

Now, I must acknowledge by name a few of the host of family and friends that continually encouraged and motivated me in this pursuit, including Bettie Nelson, Jerome Wilson, Roberta Clark, Magdalene Johnson, Angela Hayden, Agnes Boozer, Dorothy Batts and Renee Gear (my sisters) and Jesse Lee Baines (my deceased brother). Special thanks to Pamela Jean Williams, for pushing me to be the best I could be, by challenging me to be as good as she is. I appreciate the support of Grace Community Bible Church (my home church), and of my pastor, Rev. John Sykes. Thanks also to the members of Greater Emmanuel Pentecostal Temple, Shiloh Baptist Church, Spring Creek Baptist Church, New Bethel Pentecostal Holiness Church, and all of the 'saints' who have borne me up in their prayers.

I must thank my wife, Kunita. My ADD-driven spastic efforts to complete this work sorely challenged her limited reserves of patience. She met that challenge by learning enough statistics and SAS to be a real help. Without the undergirding support of her love, I would have given this pursuit up years ago. I love her immeasurably. I also thank my four children, Benjamin, Jewel, Jennifer and Joann. Much of what I am today is due to their unconditional love and acceptance.

Finally, I must acknowledge the four people who have immeasurably shaped my life: the late Edward B. Fowlkes, who taught me to love and respect data; Rev. Vernon Hodelin, who taught me to love and respect God; the late Rev. Corrie Anders, who taught me to love and respect others; and my 'adopted' mother, Marjorie Hampton, who taught me to love and respect myself. And I continually thank God for His mercy and grace, without which, this work would not have even been attempted.

"Without God, I could do nothing, Without Him, I would fail,"
"Without God, I would be drifting, Like a ship, without a sail."
Fannie J. Crosby

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION AND REVIEW OF THE LITERATURE

1.0 Introduction.

In various disciplines it is often of interest to examine the pattern of occurrence of some event. As an example, in botany one might be interested in the growth pattern of some specific type of tree in the forest, or plant in a field (Strauss, 1975). In astronomy one might be interested in the patterns exhibited by constellations and galaxies in the sky (Peebles, 1974). Other examples can be found in disciplines as diverse as geography, sociology, biology, ecology, and archaeology, as well as the various health sciences, where the purpose of study is to examine the pattern of occurrence of some event of interest. If there is no pattern, the events are said to be distributed at random. However, it is usually the alternatives to randomness that are of most interest.

There are several alternatives to randomness in the pattern of occurrence of events of interest. Events could occur in regular or irregular clusters. Events could occur in cycles, or there could be trends in their occurrence. In fact, when events are too evenly spaced, that also represents a departure from randomness. Randomness (and alternatives) can occur in any of several 'dimensions' of measurement; physical location in space and/or the occurrence in time are by far the most common. Randomness has, however, been considered in other 'dimensions', such as interpersonal relationships (Mantel, 1967) and occupations (Whittemore et al., 1986). The departure from randomness that is of the most interest is clustering. In a very general sense, events are thought to be clustered if they occur more closely together than would be expected if they were randomly

distributed[1]. Most common tests are those for clustering in space alone, clustering in time alone, or clustering in space and time simultaneously. In the last case, called space-time clustering or space-time interaction, clustering is considered present if events that are near neighbors in space are also near neighbors in time.

Tests for clustering fall into two broad categories: *general* and *focused* (Besag and Newell, 1991). *General clustering tests* do not partition or group observations into 'clusters' in the manner that hierarchical clustering methods do, while *Focused clustering tests* are concerned with identifying disease clusters. General clustering tests look for the *presence* of clustering in data, usually over a large geographic region. For these tests, clustering is a characteristic of the data, rather than a construct of the data, and detecting the presence of clustering identifies a characteristic that is related to the mechanism that generated the data. The presence of general clustering can provide valuable clues to understanding the underlying mechanism that generated the data under analysis, provide support for theories about that mechanism, and may occasionally suggest other factors that may be important to the etiology of the disease (Besag and Newell, 1991).

As an example, consider the etiology of leukemia. It is conjectured that leukemia may be caused by some infectious agent, since a virus has been found to be the cause of feline leukemia (Hardy and McClelland, 1977). If one visually inspects the spatial distribution of the location of leukemia cases in some geographic area, a "clumping together" of cases may not be apparent. However, a positive indication from a test of general clustering, while not proving anything in itself, would certainly lend credence to the aforementioned theory of causation, and has been used as evidence to support further investigation in that vein (Kulldorff and Hjalmars, 1999). This was indeed the case in the investigation of Burkitt's lymphoma. Attention to and observation of the spatial

[1] This implies a decrease in the mean distance from a randomly chosen point and its nearest neighbor. That is, if nearest neighbors are closer together than would be expected if the points were randomly distributed, clustering exists.

distribution of the disease was vital to hypothesizing and demonstrating that the disease is caused by Epstein-Barr virus (de-The', 1979).

Focused clustering tests are concerned with identifying disease clusters, often by grouping disease cases in a smaller region or regions, because of some factor (e.g. nuclear installation, toxic waste dumps) previously hypothesized to be associated with the disease. The primary goal of focused clustering tests is to identify smaller areas (those containing the disease clusters) that warrant further investigation and/or more sophisticated study.

Both general and focused clustering tests are most appropriately considered as a part of 'pre-epidemiology' studies (Wartenberg and Greenberg 1993). These are analytic investigations that precede more traditional, time-consuming and costly epidemiologic studies. Cluster analyses are among the few useful tools available to epidemiologists for screening and surveillance. Hence, these tests may be used to generate hypotheses and ideas, provide demographic explanations for unusual observations, prioritize cluster reports for field investigation, and sometimes to address public concern and fear when the lack of clustering can be demonstrated.

1.1 Conducting Cluster Analyses.

Because cluster analyses are preliminary 'pre-epidemiology' studies, they often are conducted less rigorously than are 'real' epidemiologic studies. This can be an advantage, in that more ideas and potential hypotheses may be generated for further consideration and study. However, some guidelines are needed for selecting cluster tests. Many tests exist for detecting clustering, and for identifying disease clusters. In an unpublished review article, Kulldorff (2002) has documented and characterized over 100 cluster tests from various disciplines. With such a wide range of choices, guidelines are needed to select an appropriate cluster test, and to appropriately conduct cluster analyses. Wartenberg and Greenberg (1993) suggest that consideration be given to four areas, in order to conduct an appropriate cluster analysis. These are:

1. Determining the domain (or cluster type) to be considered.
2. Characterizing the data available for analysis.
3. Specifying the null hypothesis.
4. Specifying the alternative hypothesis.

Determining the domain refers to the type of clustering that is appropriate for the disease or other problem being studied. Are we looking for temporal clustering, spatial clustering, or space-time clustering?[2] The etiology of the disease under study and the specific null hypothesis are factors that impact on this choice. Grimson (1983) used time clustering methods to compare the incidence of birth defects over several areas in England to several areas in the United States, because his specific null hypothesis was that the series of cases in each location are independent of each other, as opposed to all the data being a part of, or belonging to, the same epidemic. Spatial clustering methods may be chosen if the time component is so small that it becomes irrelevant in the analysis (Lloyd and Roberts, 1973) or if long and/or variable time periods exist between exposure and the onset of disease (Whittemore et al., 1986). However, nearly every such analysis actually involves space-time clustering, for clustering in space usually refers to data collected over some time period, and clustering in time to data over some geographic area. Space-time techniques can be used in almost every application, but should be applied when concurrent clustering in both time and space is of interest and supported by the (presumed) etiology of the disease under study.

Characterizing the data available for analysis involves understanding what the data are, what data are available, and how the data may be aggregated. The actual event of interest may be an individual (case), or a household, block, or county containing a case or cases, etc. Also, some tests will accept individual events of interest with time and location recorded for each as data, while some require that these observations be aggregated into larger space and/or time sub-units, with the analysis being performed on

[2] Note that the domain may be something other than space and/or time, such as occupational cohort, school cohort, etc. (Wartenberg and Greenberg, 1993)

these sub-units (i.e., the counts within each sub-unit), while other tests must be performed on rates for regions or sub-regions. Other tests require data for both cases and controls; these data may be case and control counts, case and control locations, etc. Regional confounding data and/or exposure data may also be available that can be utilized in the cluster analysis.

Next, it is important to specify and understand the *null hypothesis*. This is often the hypothesis that the events occur randomly, or with uniformly distributed probability or risk, throughout the study area, given the domain being considered, e.g. the space-time dimensions of the study area. The presence of confounding factors in the study area may have an impact on this. For cluster methods that can utilize confounding information, the null hypothesis is typically that the events occur randomly throughout the study area after adjusting for these factors (Wartenberg and Greenberg, 1993).

Finally, the *alternative hypothesis* must be specified and understood. In what way is departure from randomness expected? For cluster tests, what kind of clustering is expected to be found in the data? Because cluster tests are usually designed to detect specific departures from randomness, the better we can specify the kind of clustering that we expect to find, the more powerful the cluster analysis is likely to be. A cluster test that is designed to detect one type of clustering may be sensitive to others, but it will certainly be less powerful at detecting them. Similarly, cluster tests intended to detect general departures from randomness may do so, but typically at a loss of power as compared to those designed to detect a specific type of clustering.

Careful attention to these areas of consideration will produce a more structured approach to cluster studies, one that is likely to yield more consistent, reliable and useful results.

1.2 The Effect of Scale in Clustering Studies.

Optimal detection of disease clusters also involves the judicious choice of *scale* when using the test. In a generic sense, *scale* refers to the precision of the elements of the analysis. This precision may impact on several of the elements of the analysis, and it may impact on them in several ways.

First, an appropriately sized geographic area and/or time period must be chosen for analysis, since one too small would clearly not detect clustering, while one too large may mask or obscure any indication of clustering by including many events unrelated to the mechanism under consideration. In some instances the area under analysis may be predetermined by the nature of the problem. For example, clustering tests are routinely used as proactive surveillance tools for state public health departments, to signal when an increase in localized disease rates warrant a response. County or census tract may be the default area for analysis in such cases (Waller and Turnbull, 1993). When the selection of the area for analysis is up to the researcher, this selection must be made with care.

Scale also enters in when considering the data involved in the analysis. The actual event of interest (or unit of analysis) may be an individual (case), or a household, block, or county containing a case or cases, etc. Further, some tests will accept as data the individual events of interest with time and location recorded for each, while some require that these observations be aggregated into larger space and/or time sub-units, with the analysis being performed on these sub-units (i.e., the counts within each sub-unit). As data are aggregated, not only is information lost (Clark and Avery, 1976), but the power of the test is adversely affected (Waller and Turnbull, 1993). Further, the actual event of interest (or unit of analysis) may be an individual (case), or a household, block, or county containing a case or cases, etc.

And all of this assumes that scale remains constant throughout the study period. There are instances in which the scale of the data is aggregated differently for different sub-regions in the study area, or for different time segments of the overall time period

under consideration (King, 1979). This usually forces a re-aggregation of the study data, and that can impact significantly on the power of the test.

Finally, many tests require that some parameter be specified, such as the choice of a 'critical' distance in time or space. Determining these critical parameters is often heuristic, with only general guidelines provided for their selection. Because cluster tests are notoriously non-robust with regard to the values of these parameters (Roberson and Fisher, 1983), improperly specifying critical parameters can have a profound impact on the power of cluster tests (Kulldorff and Hjalmars, 1999). Hence these critical parameters must be chosen with care.

The upshot of this is that the universally applicable, general and robust clustering test cannot exist. After considering factors such as the etiology of the disease under study, the hypothesis to be tested, available data and scale, a good test for the specific analysis can be chosen.

1.2.1 Categorization of Clustering Tests.

If a 'good' test for a particular analysis is to be chosen, comparisons must be made between the available space-time clustering tests. Valid comparisons are difficult to make for several reasons. First, several factors must be considered when choosing an appropriate clustering test (i.e., disease etiology, scale, null hypothesis, data, etc.). Also, not much work has been done to group or classify space-time clustering tests, or to clarify which situations are most appropriate for which kinds of space-time clustering tests (Waller and Turnbull, 1993). Finally, while a framework has been established so that tests for clustering in space alone can be compared with regard to statistical power (Kulldorff, Tango and Park, 2003), almost no work has been done to compare space-time clustering tests in this regard (Waller and Turnbull, 1993). Hence, reducing the scope of any of these factors can only simplify the task of comparing (and choosing among) the available clustering tests.

A great reduction in the scope of the null hypotheses to be considered can be achieved by organizing clustering tests according to some categorization of their null hypotheses. Then, if a specific analysis implies a certain category of null hypothesis, only those tests which address that type of null hypothesis need be considered.

Many areas of statistical analyses have been grouped into three categories, according to the null hypotheses of the analyses. In basic statistics, chi-squared tests are generally categorized in the following way: the test for homogeneity, the goodness of fit test, and the test of independence (Remington and Schork, 1985). Applied statistics courses tend to be organized around a similar trichotomy: goodness-of-fit tests and other one-sample problems, two-sample problems (or homogeneity tests, for more than two samples), and association or independence tests (Hollander and Wolfe, 1973).

Clustering tests seem to naturally fall into three analogous groups. Tests in the first group, which we call “goodness-of-fit” tests, specify that the distribution of the points of interest is the same as some theoretical distribution. Often the null hypothesis is randomness, meaning that the underlying theoretical distribution is uniform (in space alone, or in time alone, or in space and time jointly), and departures from a uniform distribution are considered departures from randomness. Hence, these tests approach the clustering problem from a "goodness-of-fit" perspective.

Tests in the second group compare the distribution of the events of interest to the distribution of events in some actual population. These tests tend to be data-driven and consider the clustering problem as a "two-sample" problem. Often, such tests evaluate significance using a randomization technique such as repeatedly testing samples chosen from the comparison population and using the empirical distribution of the results from these tests to determine an empirical p-value for the test that the events of interest are distributed in the test population in the same manner as they are distributed in the comparison population. We call these “two-sample” tests.

Tests in the third group consider space-time clustering from an "independence" perspective, for they assess the marginal distributions of the data in space and in time, and evaluate the independence of these marginal distributions. Knox's test does this by a simple binary categorization of the data in space and time. Barton and David's test compares within-group mean squared pairwise distances (a function of the temporal distribution) to overall mean squared spatial distances (a function of the spatial distribution). We call tests of this type “independence” tests.

This categorization of clustering tests clarifies the problem in at least two ways. First, when choosing a clustering test, once the null hypothesis under consideration has been appropriately categorized, the number of valid tests is greatly reduced, simplifying the choice of test. Secondly, when evaluating clustering tests, this categorization facilitates comparison of tests within the same category, so that tests similar to each other are being compared. Occasionally in the literature one test is implied to be better than certain others, when actually the difference is one of type of test rather than of power of the test (Whittemore et al., 1986). In effect, such comparisons are analogous to 'comparing apples and oranges'; while the tests in question may not necessarily be better than the others, they are **different** from the others.

This introduction continues with a review of some of the more widely used tests in the literature for clustering in space and/or time. The tests are identified according to our type categorization.

1.3 Literature Review: Description of Specific Methods.

This section contains a review of selected specific methods for detecting space-time clustering. These are general methods, in that they detect the presence of clustering in the data rather than identifying clusters. Additionally, the methods reviewed here are not adjusted for lack of homogeneity in the underlying population (population shift bias).

1.3.1 Knox's Tests (Independence)

Knox (1963) first approached this problem from a contingency table standpoint. He created a cross-classification by dividing distances in space and time into r and c categories, respectively (see Table 1.1). Each pair of cases was then assigned to a specific space-time cell according to the spatial and temporal distances between the two cases. Knox analyzed the table using a X^2 test with (r-1)(c-1) df, though he acknowledged that the test might not be appropriate due to the dependence of the pairs. The appropriate test for this kind of table was derived by Abe (1973).

Table 1.1

DATA FORMAT FOR SPACE-TIME TESTS AND PROCEDURES

		Time Interval							Total
		1	2	3	...	j	...	c	
Space Units	1	n_{11}	n_{12}	n_{13}	...	n_{1j}	...	n_{1c}	n_1
	2	n_{21}	n_{22}	n_{23}	...	n_{2j}	...	n_{2c}	n_2
	3	n_{31}	n_{32}	n_{33}	...	n_{3j}	...	n_{3c}	n_3
	i	n_{i1}	n_{i2}	n_{i3}	...	n_{ij}	...	n_{ic}	n_i
		.	.	.		.		.	.
	r	n_{r1}	n_{r2}	n_{r3}	...	n_{rj}	...	n_{rc}	n_r
Total		n_1	n_2	n_3	...	n_j	...	n_c	n

n_{ij} is the number of events in the ij^{th} space-time unit;
i=1, 2, ..., r, j=1, 2, ..., c.

Refining his original idea, Knox (1964a) proposed the use of a 2x2 table to categorize the pairs of cases. A 'critical' time and distance are chosen, and the pairs of cases are classified as to which pairs are close in space, or in time, or both. The number

of pairs that are close in both space and time, **a**, is the test statistic. Knox felt that 'close pairs' could reasonably be considered as rare events, so that the null distribution of **a** could be assumed Poisson, with parameter λ equal to the expected value of **a** computed from the marginals of the table. The test of significance using this procedure has come to be known as Knox's test.

Table 1.2

KNOX'S TEST

		SPACE: close	SPACE: far	
TIME	close	a	b	a+b
	far	c	d	c+d
		a+c	b+d	N=n(n-1)/2

The distributional assumptions of Knox's test were later questioned in the literature. Barton and David (1966), using a graph theoretic approach, identified conditions under which the Poisson approximation is appropriate, specifically, when the first two moments of **a** are sufficiently large, in comparison to the other moments. Mantel (1967) suggested that the permutational distribution of **a**, or the appropriate normal distribution, might have wider application than the Poisson approximation, and outlined the methodology for obtaining the exact permutational distribution.

Knox's test has been widely utilized in the literature, because it is intuitive and elegant, and calculation of the Knox test statistic is simple and straightforward. At least 59 studies using Knox's test appear in the literature between 1960 and 1990 (Kulldorff and Hjalmars, 1999). Many of these were leukemia studies, (including Knox, 1964b), and the results of these studies have been used as evidence of a viral etiology of leukemia (Kulldorff and Hjalmars, 1999).

Knox's test has been criticized because of the arbitrariness involved in selecting a critical distance and time. It has been shown (Roberson, 1979) that the choice of a critical distance and time greatly affects the power of the test. Baker (1996) has proposed a modification to Knox's test that requires only that ranges of critical distances and times be specified to conduct the test. Also, a weakness of many space-time clustering tests is that they do not take into account possible geographic shifts or temporal trends in the underlying population. Kulldorff and Hjalmars (1999) have proposed a modification to such tests that can account for population shifts, if the background population and its temporal shifts are known. If random replications of cases generated under the null hypothesis are obtained, these replications can be used for hypotheses testing using the Monte Carlo procedure. Because randomization is conducted proportionate to population size at each time and place, population shifts are adjusted for.

1.3.2 Barton and David's Test (Independence)

Barton et al., (1965) and David and Barton (1966) proposed for use in detecting space-time clustering a test adapted from one originally used for evaluating randomness of points on a line (Barton and David, 1962). In this test, the events of interest are divided into temporal clusters by grouping together cases that occur within some critical time span of each other. The test statistic is then defined as:

$$Q = \left(\frac{n - 1}{n - h}\right)\left[1 - \sum_{t=1}^{h} \frac{r_t\left((\overline{X}_t - \overline{X})^2 + (\overline{Y}_t - \overline{Y})^2\right)}{n\left(\mathrm{Var}(X) + \mathrm{Var}(Y)\right)}\right]$$

where h is the number of clusters, and r_t the number of events in the t^{th} cluster. This is the ratio of the average squared within-group spatial distance between pairs of events to the squared overall spatial distance between pairs of events. Essentially, this is the ratio of within-group variance to overall variance and, as such, is analogous to traditional analysis of variance. Under the null hypothesis there is no difference between the

within-group variance and the overall variance, and the expected value of **Q** is 1. When clustering exists, **Q** should be less than 1. Significance can be assessed using a randomization test to determine the exact distribution of **Q**. When this is not feasible (and it often is not), various approximations are suggested. For small numbers of events, Barton and David suggest a Beta approximation is appropriate. When the number of events and/or the number of clusters is large, **Q** is approximately normally distributed.

The Barton and David test uses the actual times and distances so there is no loss of information, and the only arbitrariness is in choosing the critical time span. However, this test magnifies the effect of large distances, while small distances are of most interest. This results in loss of power of this test and, in some instances, existing clustering may not be detected at all because of the presence of large pairwise distances in the data.

1.3.3 The EMM Test[3]

Ederer, Myers and Mantel (1964) proposed a test of temporal clustering that can easily be expanded to accommodate space-time clustering. The test is based on dividing the time period under consideration into disjoint sub-intervals, each containing the same number of time units. Their original example used fifteen years of leukemia data, divided into three 5-year intervals. Under the null hypothesis of random distribution of cases throughout the time periods, the probability of any given arrangement of cases among the time intervals is multinomial. The test statistic is based on **a**, the maximum number of cases in a time unit and n_+, the total number of cases in a sub-interval. Tables of the distribution of **a**, given n_+, appear in their paper and extended tables are given in Mantel, Kryscio and Myers (1976).

In order to assess space-time interaction, this test can be performed on several spatial locations (or on subdivisions of one spatial location). Then the significance of the overall test can be assessed by:

[3] This is a goodness-of-fit test for time alone, and an independence test for space-time clustering.

$$X^2 = \frac{\left(\left|\sum m_1 - \sum E(m_1)\right| - 0.5\right)^2}{\sum Var(m_1)}, \text{ with 1 df}$$

This test was used to examine clustering in Down's syndrome by Stark and Mantel (1967a), and childhood leukemia by Stark and Mantel (1967b), but did not detect significant clustering for either. Disadvantages of this test include its sensitivity to marginal clustering in space or time alone, as well as its sensitivity to space-time clustering. Additionally, dividing the time period into sub-intervals, or the spatial location into sub-divisions, requires the subjective determination of critical criteria.

1.3.4 Mantel's Generalized Regression

Mantel felt that since small temporal and spatial interpoint (pairwise) distances were of greatest interest in determining space-time clustering, statistics to detect space-time clustering would be more powerful if they emphasized small pairwise distances. After considering the regression of temporal pairwise distance on spatial pairwise distance (time vs. space), Mantel (1967) proposed the following:

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} g(X_{ij} + a)\, h(Y_{ij} + b)$$

where x_{ij} and y_{ij} are the temporal and spatial distances between points i and j respectively. g(.) and h(.) are functions chosen to emphasize closeness, and a and b are arbitrarily small constants (spatial and temporal "offset constants") to preclude operation by the functions on zero distances. Significance of **Z** would be determined by obtaining its exact randomization distribution for small sample sizes, Monte Carlo simulations to approximate its distribution for larger sample sizes, or by calculating a normal deviate under the assumption of asymptotic normality.

Since **Z** has the form of a U-statistic, asymptotic normality was felt to be a reasonable assumption. However, simulation studies have shown that the distribution of **Z** is highly skewed and use of the normal approximation is not reasonable when trying to assess borderline significance. In borderline cases, Klauber (1971) and Siemiatycki (1971) were unable to consistently determine significance using the normal approximation, even with sample sizes as large as 250.

The 'closeness' function recommended by Mantel is the reciprocal transform (hence, the need for constants a and b). However, the choice of these constants does affect the power of the test: the larger they are, the smaller the value of **Z** (Glass et al, 1971 and Siemiatycki, 1971). Mantel suggested that the constants be close to the expected distance between pairs and that finding the 'best' constants might involve trial and error, affecting the validity of the test. The reciprocal transform is not the only appropriate choice as a 'closeness' function. In fact, if the 'closeness' functions are selected to be the following indicator functions:

$$g(x) = I\{x_{ij} < t^*\}$$

$$h(x) = I\{y_{ij} < d^*\},$$

where t* and d* are critical time and distance respectively, then **Z** reduces to Knox's test. The effect of different 'closeness' functions on **Z** and its corresponding normal deviate was examined by Siemiatycki (1971) and Siemiatycki and McDonald (1972), and found to be minimal in the Pearson percentage points. But Mantel felt (and Siemiatycki confirmed in 1978) that this test has low power when no transform is used. Mantel's Generalized Regression has been widely used to examine clustering in such diverse areas as typhoid and salmonellosis data (Siemiatycki, 1971) and childhood leukemia (Glass et al, 1971).

1.3.5 Zero-One Matrix Tests[4]

Dat (1982) has derived three tests, one test for clustering in space alone, one for clustering in time alone, and one for space-time interaction. When the data are organized as in Figure l.l, the statistic is:

$$A = \sum_{i}^{r} \sum_{j}^{c} a_{ij}$$

For time clustering: $a_{ij} = I\{n_{ij} >= n_{i.}/c\}$;

For space clustering: $a_{ij} = I\{n_{ij} >= n_{.j}/r\}$;

For space-time interaction: $a_{ij} = I\{n_{ij} >= n_{i.}n_{.j}/n_{..}\}$.

Given a random distribution of cases over the intervals of interest (space, time or both), one would expect the same number of cases in each cell. The 'comparison ratio' in each test is the appropriate expected number of cases. These statistics are approximately binomially distributed, with p=l/2 and n=rc. When n is large (greater than l0, say), the distribution of **A** is approximated by a normal distribution with mean equal to rc/2 and variance equal to rc/4. From simulation results, Dat (1982) found that the distribution of **A** in his three tests is closer to the binomial distribution if the 'comparison ratio' is adjusted by subtracting 0.5, then rounding the result up to the next integer. So, in his test for space-time interaction the definition of a_{ij} becomes:

$$a_{ij} = I\{n_{ij} >= [(n_{i.}n_{.j}/n_{..})-0.5]\},$$

where the expression in the square brackets [X] indicates the smallest integer greater than X.

The Zero-One Matrix tests are interesting in that, since the null value of **A** is n/2 and its range is [0, n], **A** can test for departures from randomness at both extremes of its range

[4]These are goodness-of-fit tests for space or time alone, and independence tests for space-time interaction.

of values. Small values of **A** indicate the presence of clustering, while large values indicate cluster avoidance. A particular advantage of the Zero-One Matrix test for space-time clustering is that it is sensitive only to joint clustering in both space and time.

l.3.6 Nonrandom Concordant Clustering (Two-Sample)

Occasionally one has incidence or morbidity data collected at two locations over the same period of time. Nonrandom concordant clustering exists if some degree of temporal clustering exists for each spatial location, and if the pattern of occurrence of the data is similar across the spatial locations (concordance). For data organized as in Table l.l, where the spatial sub-units are the two locations of collection and the time period is divided into (equal) sub-units, Ingram (1983) proposed the following test of nonrandom concordant clustering, where the statistic of interest is:

$$X = \sum_{i}^{n_1} \sum_{j}^{n_2} x_{ij}$$

where X_{ij} = I{events i and j occur in the same time cell}.

A computational formula for **X** is:

$$X = \sum_{j=1}^{c} n_{1j} n_{2j}$$

where n_{1i} and n_{2i} are the numbers of events from the two locations that fall into the j^{th} time sub-unit or cell.

Under the null hypothesis of no concordance between the two series and no temporal clustering, the probability of an event falling into any given time cell is l/C, for the $n_{1.}$ and $n_{2.}$ events are mutually independent and distributed randomly amongst the C time cells. **X**, then, is the sum of N dependent binomial variables. But the dependencies among the variables affect only the higher moments of **X** (Grimson and Ingram, 1982). Hence, the first two moments of **X**:

$$E[\mathbf{X}] = N/c,$$

$$\mathrm{Var}[\mathbf{X}] = (N/c)(c-1/c)$$

are identical to those of a binomial with parameters $n=n_1.n_2.$ and p = l/C, and the asymptotic distribution of X is closely approximated by the standard normal variate

$$Z = \frac{[X - E[X]]}{\sqrt{\mathrm{Var}(X)}}$$

Results of simulations by Grimson and Ingram have shown that using the normal approximation for X yields a slightly conservative test for $p > 0.05$, and a slightly anti-conservative test for $p <= 0.05$. This test has been used to examine shigellosis morbidity in urban North Carolina, and cancer mortality in counties around and including Cherokee County, N.C. (Ingram, 1983).

1.4 Conclusions.

This review surveys some of the more popular general tests in the literature for detecting clustering in space, or time, or both. This work continues in Chapter 2, with the introduction of the directed path statistic DP, and an examination of its moments and theoretical properties.

Chapter II

DESCRIPTION AND THEORETICAL PROPERTIES OF DP

2.0 Introduction

Many of the existing tests for joint clustering in space and time have characteristics that make them inherently unsuitable for wide application, or generally unwieldy and difficult to use. Some tests require the choice of a critical value or constant that introduces arbitrariness into the application of the test. Tests for joint space/time clustering are often inappropriately more sensitive to marginal clustering in space or time. Some of the tests are difficult to apply, or to understand and use properly, and for many, levels of significance are difficult to compute.

This research proposes a statistic for space-time clustering that attempts to address some of these shortcomings. In this chapter we will define the directed path length statistic, DP, and examine its characteristics.

2.1 Definition of the Statistic

First, let us define for two points (cases) in space, c_1 and c_2, the ordinary Euclidean distance between them as $d(c_1,c_2)$. Then, for a set of N points, $\{c_1,...,c_N\}$, we denote as C an ordered arrangement of them, $\{c_{[1]},...,c_{[N]}\}$. Then we define the path length function, PL(C), such that:

$$\mathrm{PL}(\mathbf{C}) = \sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})$$

Now, for a set of N points in space, with associated times of occurrence $t_1,...t_N$, if we denote by C_t the set of points ordered by their times of occurrence, we define the directed path length of the points, DP, as:

$$\mathrm{PL}(\mathbf{C_t}) = \sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})$$

which is simply the length of the path from $c_{[1]}$ to $c_{[N]}$, where the N points are ordered by their times of occurrence.

The directed path statistic is conceptually straightforward and easy to understand. The statistic is based on the classical model of contagion with direct case-to-case transmission (Benenson, 1975). While DP directly models the presumed path of the disease through the population under study under the assumption that each case causes only one other, it is a valid assessment of joint clustering in space and time even when the above assumption does not hold. If there is joint clustering in space and time in the data, the length of this path will be relatively short, compared to the lengths of other paths through the data, computed with the cases permuted. Hence, DP can easily be utilized in a Monte Carlo test as a randomization statistic.

As a randomization statistic, DP is compared with its empirical distribution from the actual data under study. This makes DP useful in situations where the disease has a long or variable incubation period, for significance is based on the actual data, and not on distributional assumptions.

2.2 The Significance Level of DP

Given some regularity assumptions on the underlying population of points and assuming complete spatial randomness of the cases of interest, the exact theoretical distribution of DP under the null hypothesis (of complete spatial randomness) can be specified or approximated. Under these assumptions, the first three moments of DP have

been derived, and appear in a later section. However, using the theory to determine the significance level of DP is impractical for several reasons. The assumption of regularity on the base population is quite severe and unrealistic. Also, expressions for the moments of DP contain symbols that represent the mean path length for connected points (two, three, four, and so on) in that population. These values are almost never known. They can be derived from the density of points in the area under study, but that derivation depends on regularity assumptions already acknowledged as unrealistic.

The significance level of DP can also be determined using Monte Carlo procedures, so that is the recommended method. Specifically, the empirical permutational distribution of DP can be determined by tabulating the path length for all possible permutations of the N cases (N! permutations). The value obtained for DP is then compared to this empirical distribution. If DP is less than or equal to M of the N! values in the empirical distribution, then the exact significance is P=M/N!.[5] The number of permutations, N!, rapidly increases as N increases, so that tabulating the path length for all possible permutations of the N cases quickly becomes computationally prohibitive[6]. For large N, the empirical distribution can be estimated by tabulating the path length for an arbitrarily large number (say, N_p) of random permutations of the N cases. Then, the exact significance is estimated by $P=M/N_p$.[7]

---

[5]Note that the path length is the same regardless of the direction the path is traversed. Hence, there are actually N!/2 distinct paths, and if M' is the number of those N!/2 distinct paths that are **less** than DP, the exact empirical significance is P'=M'/(N!/2). However, since each distinct path appears twice, M'=M/2, and P'=M'/(N!/2)=(M/2)/(N!/2)=M/N!=P.

[6]On one IBM mainframe used during the course of this research, integers could not be represented with precision with more than twelve significant digits. Because of this, tabulating the path length for all possible permutations of the N cases became computationally prohibitive for N greater than 14.

[7]When the permutations of the N cases are randomly generated, they should yield a tabulation of path lengths that is proportionate to the full empirical distribution of DP, so that the number of permutations greater than DP in the tabulated distribution is proportionate to the number greater than DP in the complete distribution. Hence, this is a valid estimate of the exact significance of DP.

In general, the question of whether significance should be determined using classical distribution theory or Monte Carlo testing is a philosophical one, and has been the subject of some debate. Diggle (1983) made the following observation in his book, Statistical Analysis of Spatial Point Patterns:

> "When asymptotic distribution theory is available, Monte Carlo testing provides an exact alternative for small samples and a useful check on the applicability of the asymptotic theory. If the results of classical and Monte Carlo tests are in substantial agreement, little or nothing has been lost; if not, the explanation is usually that the classical test uses inappropriate distributional assumptions."

The general experience of this researcher is in agreement with this observation. Also, while theoretical results are presented for this statistic, DP is defined so that the comparison of its value to other path lengths generated from permutations of the data is quite natural and intuitive. Hence, DP is most appropriately utilized in a Monte Carlo test.

It is clear from its definition that DP is a spatial statistic, but constrained temporally so that significance only exists if the cases that are close together in space are also close together in time. Since temporal information is utilized simply to order the pairwise distances that make up DP, only the ranks of the times influence the ordering of the points (and the value of DP). This serves the purpose of making DP insensitive to temporal clustering alone, for the ordering of the points is not affected by the spacing between them.

2.3 Illustrative Example

Let us look at an application of DP in the following illustrative example. Suppose that cases of a disease (or some other points of interest) appear on a 'map' of some geographic area as in Figure 2.1 below:

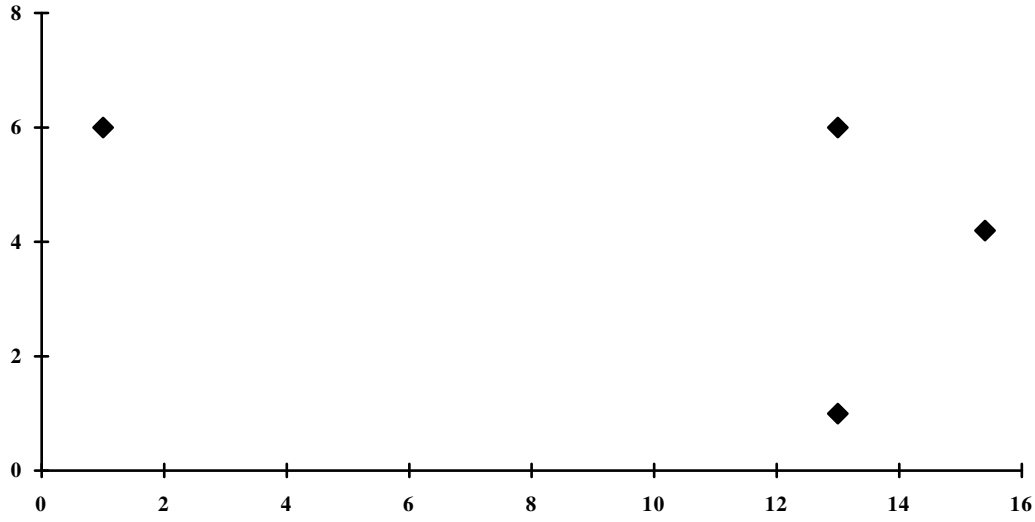
Figure 2.1

DP Example – Cases of Interest

We will impose an arbitrary coordinate system on the map as in Figure 2.2, so that we can quantitatively locate points and compute distances.

Figure 2.2

DP Example - Coordinates



If we number the points according to their times of occurrence (Table 2.1)

Table 2.1

CASES OF INTEREST

Point	Coordinates
1	(1,6)
2	(13,6)
3	(15.4,4.2)
4	(13,1)

Then, the pairwise Euclidean distances between the points can be computed as:

Table 2.2

PAIRWISE DISTANCES

	1	**2**	**3**	**4**
1	-	12	14.512	13
2	12	-	3	5
3	14.512	3	-	4
4	13	5	4	-

The value of DP for these data is the length of the path from point 1 to point 4, or the sum of the pairwise distances $d(1,2)$, $d(2,3)$, and $d(3,4)$, which is (12+3+4) = 19 units. The order of the cases is permuted, and DP is computed. This is repeated for each of the twenty-four possible permutations of the cases, and tabulated in Table 2.3. These values are ordered, to generate the permutational distribution shown in the right-most column of Table 2.3, shown below.

Table 2.3

PERMUTATIONAL DISTRIBUTION

Permutation	Pairwise Distances	Path Length	Ordered Path Length
1,2,3,4	**12+3+4**	**19**	**19.0**
1,2,4,3	12+5+4	21	19.0
1,3,2,4	14.5+3+5	22.5	20.0
1,3,4,2	14.5+4+5	23.5	20.0
1,4,2,3	13+5+3	21	21.0
1,4,3,2	13+4+3	20	21.0
2,1,3,4	12+14.5+4	30.5	21.0
2,1,4,3	12+13+4	29	21.0
2,3,1,4	3+14.5+13	30.5	22.5
2,3,4,1	3+4+13	20	22.5
2,4,1,3	5+13+14.5	32.5	23.5
2,4,3,1	5+4+14.5	23.5	23.5
3,1,2,4	14.5+12+5	31.5	28.0
3,1,4,2	14.5+13+5	32.5	28.0
3,2,1,4	3+12+13	28	29.0
3,2,4,1	3+5+13	21	29.0
3,4,1,2	4+13+12	29	30.5
3,4,2,1	4+5+12	21	30.5
4,1,2,3	13+12+3	28	30.5
4,1,3,2	13+14.5+3	30.5	30.5
4,2,1,3	5+12+14.5	31.5	31.5
4,2,3,1	5+3+14.5	22.5	31.5
4,3,1,2	4+14.5+12	30.5	32.5
4,3,2,1	4+3+12	19	32.5

There are two (out of all possible) permutations that have path lengths less than or equal to DP. So, the empirical p-value for these data is 2(1/24) = 2(0.04167) = 0.083.

2.4 The Moments of DP

In order to document the distributional characteristics of DP, its first three moments are derived here. The moments could be used to develop a parametric test using DP, so that its significance can be determined directly. However, these moments depend on unknown parameters, which are the expected value of the lengths of various

combinations of path segments. These have been derived, and appear below. These parameters could conceivably be estimated from the data, should such a test be desired.

2.4.1 The First Moment of DP.

DP is defined as:

$$\mathrm{PL}(\mathbf{C_t}) = \sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})$$

The first moment of DP is:

$$\mathbf{E}\{\mathrm{DP}\} = \mathbf{E}(\mathrm{PL}(\mathbf{C_t})) = \mathbf{E}\left\{\left\{= \sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})\right\}\right\}$$

$$= \mathbf{E}\left\{d(c_{[1]}, c_{[2]}) + d(c_{[2]}, c_{[3]}) + \ldots + d(c_{[N-1]}, c_{[N]})\right\} \text{ (a total of N-1 terms)}$$

$$= (\mathrm{N\text{-}1})\ \mathbf{E}[\mathrm{d}_{12}],$$

where $\mathbf{E}[\mathrm{d}_{12}]$ is the mean interpoint distance (or the average distance between two points) in a random sample of size N from the population under consideration. Note that the subscript notation used for defining terms is determined by the subscripts on the points of the *first occurrence* of a typical segment of a particular type. In this instance, the first occurrence of a pairwise distance is $d(c_1,c_2)$, so the subscript notation becomes d_{12}.

Because this is the mean value of DP for random samples of size N from the population under consideration, we will also denote this quantity as μ_{DP}. That is:

$$\mathbf{E}\{\mathrm{DP}\} = (\mathrm{N\text{-}1})\ \mathbf{E}[\mathrm{d}_{12}] = \mu_{DP}$$

2.4.2 The Second Moment of DP.

The second moment of DP is:

$$\mathbf{E}\{\mathrm{DP}^2\} = \mathbf{E}\{\mathrm{PL}(\mathbf{C_t})^2\} = \mathbf{E}\left\{\left(\sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})\right)^2\right\}$$

$$= \mathbf{E}\left\{\sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]}) \sum_{j=1}^{N-1} d(c_{[j]}, c_{[j+1]})\right\}$$

$$= \mathbf{E}\left\{\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} d(c_{[i]}, c_{[i+1]}) d(c_{[j]}, c_{[j+1]})\right\}$$

Equation **E2.1:**

$$= \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} \mathbf{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[j]}, c_{[j+1]})\right\}$$

Evaluating the second moment of DP involves evaluating the sum of the square of the (N-1) terms of DP, where each term is the product of two line segments. To assist in understanding the $(N-1)^2$ terms that make up this expression, we will lay these terms out in the following two-dimensional *product matrix*:

Table 2.4

PRODUCT MATRIX

	j = 1	2	…	j	…	n-1
1	$d(c_{[1]}, c_{[2]})$ $d(c_{[1]}, c_{[2]})$	$d(c_{[1]}, c_{[2]})$ $d(c_{[2]}, c_{[3]})$	…	$d(c_{[1]}, c_{[2]})$ $d(c_{[j]}, c_{[j+1]})$	…	$d(c_{[1]}, c_{[2]})$ $d(c_{[n-1]}, c_{[n]})$
2	$d(c_{[2]}, c_{[3]})$ $d(c_{[1]}, c_{[2]})$	$d(c_{[2]}, c_{[3]})$ $d(c_{[2]}, c_{[3]})$	…	$d(c_{[2]}, c_{[3]})$ $d(c_{[j]}, c_{[j+1]})$	…	$d(c_{[2]}, c_{[3]})$ $d(c_{[n-1]}, c_{[n]})$
i = …	…	…	…	…	…	…
i	$d(c_{[i]}, c_{[i+1]})$ $d(c_{[1]}, c_{[2]})$	$d(c_{[i]}, c_{[i+1]})$ $d(c_{[2]}, c_{[3]})$	…	$d(c_{[i]}, c_{[i+1]})$ $d(c_{[j]}, c_{[j+1]})$	…	$d(c_{[i]}, c_{[i+1]})$ $d(c_{[n-1]}, c_{[n]})$
…	…	…	…	…	…	…
n-1	$d(c_{[n-1]}, c_{[n]})$ $d(c_{[1]}, c_{[2]})$	$d(c_{[n-1]}, c_{[n]})$ $d(c_{[2]}, c_{[3]})$	…	$d(c_{[n-1]}, c_{[n]})$ $d(c_{[j]}, c_{[j+1]})$	…	$d(c_{[n-1]}, c_{[n]})$ $d(c_{[n-1]}, c_{[n]})$

The $(N-1)^2$ terms that comprise this double sum fall into three groups. The first contains the terms where the segments in the product are identical (when i=j). The second group contains the terms where the two segments are connected, that is, when they share a common point (when i=j+1 or j=i+1). The third group contains the terms where the segments in the product are disjoint (all other terms). Evaluating these groups separately, we have:

Group 1: Identical Segments.

When i=j, the terms fall on the major diagonal of the product matrix. In this case Equation **E2.1** above reduces to:

$$= \sum_{i=1}^{N-1} \mathbf{E}\left\{ d(c_{[i]}, c_{[i+1]})^2 \right\}$$

$$= \sum_{i=1}^{N-1} \mathbf{E}\left\{ d(c_{[i]}, c_{[i+1]}) \right\}^2$$

$$= \sum_{i=1}^{N-1} \mathbf{E}\left\{ (d_{12})^2 \right\}$$

$$= (N\text{-}1)\, d_{12} d_{12}$$

$$= (N\text{-}1)\, \mathbf{E}[d_{12}^2]$$

where $\mathbf{E}[d_{12}^2]$ represents the expectation of the square of the mean interpoint distance for samples of size N from the population under consideration.

Group 2: Connected Segments.

When i=j+1 or j=i+1, the terms fall on the two minor diagonals just above and below the major diagonal of the product matrix. To select these terms, Equation **E2.1** above becomes:

$$= \sum_{i=1}^{N-1} \mathbf{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[i+1]}, c_{[i+2]}) \right\}$$

$$= \sum_{j=1}^{N-1} \mathbf{E}\left\{ d(c_{[j]}, c_{[j+1]}) d(c_{[j+1]}, c_{[j+2]}) \right\}$$

If we make the following definition:

$$\mathbf{E}[d_{12}d_{23}] = \mathbf{E}\left\{ d(c_{[j]}, c_{[j+1]}) d(c_{[j+1]}, c_{[j+2]}) \right\}$$

then $\mathbf{E}[d_{12}d_{23}]$ is the expectation of the product of two connected segments, and the above expression becomes:

$$= \sum_{i=1}^{N-2} \mathrm{E}[d_{12}d_{23}] + \sum_{j=1}^{N-2} \mathrm{E}[d_{12}d_{23}]$$

$$= 2\,(\mathrm{N}\text{-}2)\,\mathbf{E}[d_{12}d_{23}]$$

Group 3: Disjoint Segments.

There are $(\mathrm{N}\text{-}1)^2 - (\mathrm{N}\text{-}1) - (2(\mathrm{N}\text{-}2)) = (\mathrm{N}\text{-}2)(\mathrm{N}\text{-}3)$ terms left in this summation. Under the assumption that the segments in the terms are independent, the expectation in Equation **E2.1** distributes over the product segments, yielding (N-2)(N-3) terms of the form

$$= \mathrm{E}\left\{d(c_{[i]}, c_{[i+1]})\right\}\mathrm{E}\left\{d(c_{[j]}, c_{[j+1]})\right\}$$

$$= \mathbf{E}[d_{12}]\;\mathbf{E}[d_{12}] = \mathbf{E}^2[d_{12}],$$

$\mathbf{E}^2[d_{12}]$ is the square of the expectation of the mean interpoint distance for samples of size N from the population under consideration, and the sum of these independent terms is:

$$(\mathrm{N}\text{-}2)(\mathrm{N}\text{-}3)\;\mathbf{E}^2[d_{12}]$$

Note that the binomial expansion can be used to enumerate the terms in these groups. This double summation represents the square of an expression containing (N-1) terms, e.g. $(x_1 + x_2 + x_3 + \ldots + x_{N-1})^2$. From the binomial expansion, we know that this expression contains (N-1) squared terms, and (N-1)(N-2) linear terms, for a total of $(\mathrm{N}\text{-}1)^2$ terms. The (N-1) squared terms are those terms that contain identical segments. Both the connected and disjoint terms are linear terms, so the (N-1)(N-2) linear terms can be partitioned into connected and disjoint terms. So, the sum of the 2(N-2) connected terms

and the (N-2)(N-3) disjoint terms yields the (N-1)(N-2) linear terms of the binomial expansion of this expression.

Now, combining the partial sums from the three groups above, we have:

$$\mathbf{E}\{DP^2\} = (N\text{-}1)\ \mathbf{E}[d_{12}{}^2] + 2\ (N\text{-}2)\ \mathbf{E}[d_{12}d_{23}] + (N\text{-}2)(N\text{-}3)\ \mathbf{E}^2[d_{12}]$$

2.4.3 The Third Moment of DP.

In order to compute the third central moment of DP, we need to evaluate the expectation of DP^3. This is:

$$\mathbf{E}\{DP^3\} = \mathbf{E}\{PL(\mathbf{C_t})^3\} = \mathbf{E}\left\{\left(\sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})\right)^3\right\}$$

$$= \mathbf{E}\left\{\sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]}) \sum_{j=1}^{N-1} d(c_{[j]}, c_{[j+1]}) \sum_{k=1}^{N-1} d(c_{[k]}, c_{[k+1]})\right\}$$

Equation **E2.2:** $$= \sum_{i=1}^{N-1}\sum_{j=1}^{N-1}\sum_{k=1}^{N-1} \mathrm{E}\left\{d(c_{[i]}, c_{[i+1]}) d(c_{[j]}, c_{[j+1]}) d(c_{[k]}, c_{[k+1]})\right\}$$

As before, the $(N\text{-}1)^3$ terms that comprise this triple sum will be organized into groups to facilitate evaluating this expectation. The first of these six groups contains the terms where the three segments in the product are identical (when i=j=k). The second group contains terms with two identical segments, and one connected segment (when i=j=k+1, i=j=k-1, j=k=i+1, j=k=i-1, k=i=j+1 or k=i=j-1). The third group contains terms made up of two identical segments, and one disjoint segment (when i=j, i=k or j=k, and excluding terms in groups 1 and 2). The fourth group contains terms with three connected segments (when i=j+1=k+2, i=k+1=j+2, j=k+1=i+2, j=i+1=k+2, k=j+1=i+2, or k=i+1=j+2, excluding terms in groups 1, 2 and 3). The fifth group contains terms with

two segments that share a common point (connected), and a third disjoint segment (when i=j+1, i=k+1, j=i+1, j=k+1, k=i+1. or k=j+1) excluding terms appearing in any preceding groups). The sixth group contains the terms with three disjoint segments. First we will enumerate the terms in these groups, and then we will evaluate these groups separately, as before.

Group 1: Three Identical Segments:

When i=j=k, the terms in the 3-tuple are identical, and fall on the major diagonal of the three-dimensional product matrix. In this case there are exactly N-1 of these terms, so Equation **E2.2** above reduces to:

$$= \sum_{i=1}^{N-1} \mathbf{E}\left\{ d(c_{[i]}, c_{[i+1]})^3 \right\}$$

$$= \sum_{i=1}^{N-1} \mathbf{E}\left\{ d(c_{[i]}, c_{[i+1]}) \right\}^3$$

$$= \sum_{i=1}^{N-1} \mathrm{E}[\mathrm{d}_{12}\, \mathrm{d}_{12}\, \mathrm{d}_{12}]$$

$$= (\mathrm{N}\text{-}1)\ \mathbf{E}[\mathrm{d}_{12}{}^3]$$

where $\mathbf{E}[\mathrm{d}_{12}{}^3]$ represents the expectation of the cube of the mean interpoint distance for samples of size N from the population under consideration.

Group 2: Two Identical and One Connected Segment.

The terms that contain two identical segments and one connected segment are of one of these following forms: i=j=k+1, i=j=k-1, j=k=i+1, j=k=i-1, k=i=j+1 or k=i=j-1. For any of these six forms, we have from Equation **E2.2**:

$$\sum_{i=1}^{N-2} \mathrm{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[i]}, c_{[i+1]}) d(c_{[i+1]}, c_{[i+2]}) \right\}$$

If we make the following definition:

$$\mathbf{E}[d_{12}{}^2 d_{23}] = \mathrm{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[i]}, c_{[i+1]}) d(c_{[i+1]}, c_{[i+2]}) \right\},$$

then $\mathbf{E}[d_{12}{}^2 d_{23}]$ is the expectation of a product that contain two identical segments, and one connected segment. Then we have:

$$(\mathrm{N}\text{-}2)\ \mathbf{E}[d_{12}{}^2 d_{23}]$$

And, accounting for all six cases, we have:

$$6\ (\mathrm{N}\text{-}2)\ \mathbf{E}[d_{12}{}^2 d_{23}]$$

These terms may be enumerated more succinctly using combinatorics. First we count the number of ways to choose a pair of connected segments (for example, a,b). There are N-2 ways to do that. Then, note that there are 6 ways to obtain a 3-tuple that contains two identical segments and one connected segment: (a,a,b), (a,b,a), (b,a,a), (b,b,a), (b,a,b) and (a,b,b). So, there are 6(N-2) terms with two identical segments and one connected segment, and given the above definition, for this group we have:

$$6\ (\mathrm{N}\text{-}2)\ \mathbf{E}[d_{12}{}^2 d_{23}]$$

Group 3: Two Identical and One Disjoint Segment.

All terms that contain two identical segments are of the following form:

$$\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}\sum_{k=1}^{N-1} \mathrm{E}\left\{d(c_{[i]},c_{[i+1]})d(c_{[j]},c_{[j+1]})d(c_{[k]},c_{[k+1]})\right\}$$

where i=j, i=k or j=k. As an example, we have for any one of these three functionally identical cases:

$$\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} \mathrm{E}\left\{d(c_{[i]},c_{[i+1]})d(c_{[i]},c_{[i+1]})d(c_{[j]},c_{[j+1]})\right\}$$

The subscripts on the first summation is unconstrained, and the subscript on the second summation can be anything except the (single) value of the first summation, so there are (N-1)(N-2) terms of this form in each dimension, for a total of 3(N-1)(N-2) of these terms to account for. 6 (N-2) of these terms are accounted for in Group 2, the terms with two identical segments and one connected segment. Hence, there are:

$$3(N-1)(N-2) - (6(N-2))$$

$$= 3(N-2)[(N-1) - (2)]$$

$$= 3(N-2)(N-3)$$

terms of this form. Evaluating the expectation for one term, we have:

$$= \mathrm{E}\left\{d(c_{[i]},c_{[i+1]})d(c_{[i]},c_{[i+1]})d(c_{[j]},c_{[j+1]})\right\}$$

$$= \mathbf{E}[d_{12}d_{12}d_{34}]$$

$$= \mathbf{E}[d_{12}^{2}]\mathbf{E}[d_{34}] = \mathbf{E}[d_{12}^{2}]\ \mathbf{E}[d_{12}]$$

Accounting for all terms in this group, we have:

$$3(N-2)(N-3)\ \mathbf{E}[d_{12}^{2}]\ \mathbf{E}[d_{12}]$$

Again, let us confirm this result by using combinatorics to numerate these terms. First we count the number of ways to choose a pair of disjoint segments. There are ${}_{N-1}C_2$ ways to choose a pair of segments, but (from Group 2 above) N-2 of these pairs are adjacent. Because N-2 = ${}_{N-2}C_1$, the number of ways to choose a pair of disjoint segments from N-1 segments is

$$ {}_{N-1}C_2 - {}_{N-2}C_1 $$

Now, a basic relationship in combinatorics is:

$$ {}_{n}C_k = {}_{(n-1)}C_{(k-1)} + {}_{(n-1)}C_k $$

(see, for example, Charalambides 2002). Substituting N-1 for n, and 2 for k, we have

$$ {}_{N-1}C_2 = {}_{N-2}C_1 + {}_{(N-2)}C_2 \ . $$

And the number of ways to choose a pair of disjoint segments is

$$ {}_{N-1}C_2 - {}_{N-2}C_1 = {}_{(N-2)}C_2 = \frac{(N-2)(N-3)}{2} \ . $$

Then (as demonstrated above), note that there are 6 ways to obtain a 3-tuple that contains two identical segments and one disconnected segment. So, the total number of terms that contain two identical segments and one disconnected segment are

$$ 6\ \frac{(N-2)(N-3)}{2} = 3(N-2)(N-3) \ , $$

and given the above definition, for this group we have:

$$ 3(N-2)(N-3)\ \mathbf{E}[d_{12}{}^2]\ \mathbf{E}[d_{12}] $$

Group 4: Three Connected Segments.

The terms that contain three connected segments are of the form:

$$\sum_{i=1}^{N-3} \mathrm{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[i+1]}, c_{[i+2]}) d(c_{[i+2]}, c_{[i+3]}) \right\}$$

Note that there are $\binom{3}{2}$ distinct (and functionally identical) orderings of the three summands: (i, j, k), (i, k, j), (j, i, k), (j, k, i), (k, i, j) and (k, j, i), so there are 6 ways to choose an appropriate 3-tuple. Now, how many such 3-tuples are there? Note that there are N-3 ways to select the last term. However, once that term is selected, the first two terms are already determined, so there are N-3 ways to choose 3-tuples with three connected terms, for a total of 6(N-3) terms containing three connected segments. These terms are of the form:

$$\mathrm{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[i+1]}, c_{[i+2]}) d(c_{[i+2]}, c_{[i+3]}) \right\}$$

$$= \mathbf{E}[\mathrm{d}_{12}\mathrm{d}_{23}\mathrm{d}_{34}]$$

And, accounting for all terms in this group, we have:

$$6(\mathrm{N}\text{-}3)\ \mathbf{E}[\mathrm{d}_{12}\mathrm{d}_{23}\mathrm{d}_{34}]$$

Group 5: Two Connected and One Disjoint Segment.

The terms that contain two connected segments and one disjoint segment are of the form:

$$\sum_{\substack{i=1 \\ j \neq [i-1, i, i+1, i+2]}}^{N-2} \sum_{j=1}^{N-1} \mathrm{E}\left\{ d(c_{[i]}, c_{[i+1]}) d(c_{[i+1]}, c_{[i+2]}) d(c_{[j]}, c_{[j+1]}) \right\}$$

First we will count the number of ways to obtain 3-tuples of this form. From Group 2 we know that there are N-2 ways to choose a pair of adjacent terms. Now, we must select a term that is not adjacent to either of the terms in the pair. If the pair of adjacent terms is either the first two terms or the last two terms, there are n-4 possible terms to choose. If the pair of adjacent terms is neither the first two terms nor the last two terms, there are n-5 possible terms to choose. Therefore, there are a total of:

$$2(N-4) + ((N-2)-2)(N-5) = 2(N-4) + (N-4)(N-5) = (N-3)(N-4)$$

ways to select the terms that contain two connected segments and one disjoint segment. We have shown earlier that there are 6 ways to choose an appropriate 3-tuple, so there are a total of 6(N-3)(N-4) terms in this group. Now, evaluating the expectation for one term, we have:

$$\mathrm{E}\left\{d(c_{[i]}, c_{[i+1]})d(c_{[i+1]}, c_{[i+2]})d(c_{[j]}, c_{[j+1]})\right\}$$

$$= \mathrm{E}\left\{d(c_{[i]}, c_{[i+1]})d(c_{[i+1]}, c_{[i+2]})\right\} \mathrm{E}\left\{d(c_{[j]}, c_{[j+1]})\right\}$$

$$= \mathbf{E}[\mathrm{d}_{12}\mathrm{d}_{23}]\mathbf{E}[\mathrm{d}_{45}]$$

$$= \mathbf{E}[\mathrm{d}_{12}\mathrm{d}_{23}]\mathbf{E}[\mathrm{d}_{12}]$$

because $\mathbf{E}[\mathrm{d}_{45}]$ is the same as $\mathbf{E}[\mathrm{d}_{12}]$. Accounting for all terms in this group, we have:

$$6(N-3)(N-4)\ \mathbf{E}[\mathrm{d}_{12}\mathrm{d}_{23}]\mathbf{E}[\mathrm{d}_{12}]$$

Group 6: Three Disjoint Segments.

The terms that contain three disjoint segments are of the form:

$$\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}\sum_{k=1}^{N-1} \mathrm{E}\left\{d(c_{[i]}, c_{[i+1]})d(c_{[j]}, c_{[j+1]})d(c_{[k]}, c_{[k+1]})\right\}$$

Note that there are six distinct (and functionally identical) orderings of the subscripts; *i*,*j*,*k*; *i*,*k*,*j*; *j*,*i*,*k*; *j*,*k*,*i*; *k*,*i*,*j*; and *k*,*j*,*i*. This corresponds to the six ways to obtain appropriate 3-tuples of this form. Now we will determine the number of ways to choose three disjoint terms, the terms in this group will be 6 times this number.

We will use an inductive argument to demonstrate that there are ${}_{N-3}C_3$ ways to choose three disjoint terms. It is clear by observation that N-3 must be greater than 4. First we count the number of ways to choose three disjoint segments where none of the segments is the last one. This is simply choosing three disjoint segments from the first N-2 segments, and by my inductive hypothesis there are ${}_{(N-2-3)}C_3$ or ${}_{(N-5)}C_3$ ways to do this. Now, we determine the remaining ways to choose three disjoint segments, that is, the ways where one of the segments is the last one. In this case, the next to the last segment cannot be selected, and two disjoint segments must be selected from the remaining N-3 objects. From Group 3 we already know that the number of ways to choose a pair of disjoint segments from N-1 segments is

$$ {}_{N-2}C_2 = \frac{(N-2)(N-3)}{2} . $$

So the number of ways to choose a pair of disjoint segments from N-3 segments is

$$ {}_{N-4}C_2 = \frac{(N-4)(N-5)}{2} . $$

Therefore, the total number of ways to select three disjoint segments is

$$ {}_{N-4}C_3 + {}_{N-4}C_2 $$

However, using the basic combinatorics relationship from Group 3, we know that

$$ {}_{N-3}C_3 = {}_{N-4}C_3 + {}_{N-4}C_2 $$

Hence, the total number of terms that contain three disjoint segments is

$$6({}_{N-3}C_3) = 6\left(\frac{(N-3)!}{3!(N-6)!}\right) = 6\left(\frac{(N-3)(N-4)(N-5)(N-6)!}{6\,(N-6)!}\right) = (N-3)(N-4)(N-5).$$

So, for all six cases there are a total of (N-3)(N-4)(N-5) terms.

Evaluating the expectation for one term, we have:

$$E\left\{d(c_{[i]}, c_{[i+1]})d(c_{[j]}, c_{[j+1]})d(c_{[k]}, c_{[k+1]})\right\}$$

$$= \mathbf{E}[d_{12}d_{34}d_{56}]$$

$$= \mathbf{E}[d_{12}]\ \mathbf{E}[d_{34}]\ \mathbf{E}[d_{56}]$$

$$= \mathbf{E}[d_{12}]\ \mathbf{E}[d_{12}]\ \mathbf{E}[d_{12}] = \mathbf{E}^3[d_{12}]$$

And, accounting for all terms with three disjoint segments, we have:

$$[(N-3)(N-4)(N-5)]\ \mathbf{E}^3[d_{12}]$$

Note that we have now enumerated all of the terms that make up these six cases. As these cases represent the cube of (N-1) terms, the total number of terms in these six cases should be $(N-1)^3$. As a check, we demonstrate that:

$$(N-1)^3 \stackrel{?}{=} (N-1) + 6\,(N-2) + 3(N-2)(N-3) + 6(N-3) + 6(N-3)(N-4) + [(N-3)(N-4)(N-5)]$$

Note first the left side of the equation, and $(N-1)^3 = N^3 - 3N^2 + -3N - 1$. Now, expanding the right-hand side of the above equation, we have:

$$(N-1) + (6N-12) + (3N^2-13N+18) + (6N-18) + (6N^2-42N+72) + (N^3-12N^2+47N-60)$$

And collecting like terms, we have:

$$N^3 + (3+6-12)N^2 + (1+6-15+6-42+47)N + (-1-12+18-18+72-60)$$

$$N^3 - 3N^2 + -3N - 1$$

The expectation of DP^3 is the sum of all of the expressions derived in the preceding six cases. Hence, the expectation of DP^3 is:

$$\mathbf{E}\{DP^3\} = \mathbf{E}\{PL(\mathbf{C_t})^3\} = \mathbf{E}\left\{\left(\sum_{i=1}^{N-1} d(c_{[i]}, c_{[i+1]})\right)^3\right\}$$

$$= (N\text{-}1)\,\mathbf{E}[d_{12}{}^3] + 6(N\text{-}2)\,\mathbf{E}[d_{12}{}^2 d_{23}] + 3(N\text{-}2)(N\text{-}3)\,\mathbf{E}[d_{12}{}^2]\mathbf{E}[d_{12}]$$
$$+ 6(N\text{-}3)\,\mathbf{E}[d_{12}d_{23}d_{34}] + 6(N\text{-}3)(N\text{-}4)\,\mathbf{E}[d_{12}d_{23}]\mathbf{E}[d_{12}] + [(N\text{-}3)(N\text{-}4)(N\text{-}5)]\,\mathbf{E}^3[d_{12}].$$

Now, the third central moment is

$$E(DP-\mu_{DP})^3 = E(DP^3 - 3\mu_{DP}DP^2 + 3\mu_{DP}{}^2 DP - \mu_{DP}{}^3)$$
$$= E(DP^3) - 3\,\mu_{DP}E(DP^2) + 3\,\mu_{DP}{}^2 E(DP) - \mu_{DP}{}^3$$
$$= E(DP^3) - 3\,\mu_{DP}E(DP^2) + 3\,\mu_{DP}{}^2\mu_{DP} - \mu_{DP}{}^3$$
$$= E(DP^3) - 3\,\mu_{DP}E(DP^2) + 2\mu_{DP}{}^3$$

Substituting the previously derived expressions for $E(DP^3)$ and $E(DP^2)$ in the equation above, we have for the third central moment:

$$E(DP-\mu_{DP})^3 = (N\text{-}1)\,\mathbf{E}[d_{12}{}^3] + 6(N\text{-}2)\,\mathbf{E}[d_{12}{}^2 d_{23}] + 3(N\text{-}2)(N\text{-}3)\,\mathbf{E}[d_{12}{}^2]\mathbf{E}[d_{12}]$$
$$+ 6(N\text{-}3)\,\mathbf{E}[d_{12}d_{23}d_{34}] + 6(N\text{-}3)(N\text{-}4)\,\mathbf{E}[d_{12}d_{23}]\mathbf{E}[d_{12}] + [(N\text{-}3)(N\text{-}4)(N\text{-}5)]\,\mathbf{E}^3[d_{12}]$$
$$- 3\,\mu_{DP}\{(N\text{-}1)\,\mathbf{E}[d_{12}{}^2] + 2\,(N\text{-}2)\,\mathbf{E}[d_{12}d_{23}] + (N\text{-}2)(N\text{-}3)\,\mathbf{E}^2[d_{12}]\} + 2\mu_{DP}{}^3$$

or

$$= (N\text{-}1)\,\mathbf{E}[d_{12}{}^3] + 6(N\text{-}2)\,\mathbf{E}[d_{12}{}^2 d_{23}] + 3(N\text{-}2)(N\text{-}3)\,\mathbf{E}[d_{12}{}^2]\mathbf{E}[d_{12}]$$
$$+ 6(N\text{-}3)\,\mathbf{E}[d_{12}d_{23}d_{34}] + 6(N\text{-}3)(N\text{-}4)\,\mathbf{E}[d_{12}d_{23}]\mathbf{E}[d_{12}] + [(N\text{-}3)(N\text{-}4)(N\text{-}5)]\,\mathbf{E}^3[d_{12}]$$
$$- 3\,(N\text{-}1)\,\mu_{DP}\,\mathbf{E}[d_{12}{}^2] - 6\,(N\text{-}2)\,\mu_{DP}\,\mathbf{E}[d_{12}d_{23}] - 3\,(N\text{-}2)(N\text{-}3)\,\mu_{DP}\,\mathbf{E}^2[d_{12}] + 2\,\mu_{DP}{}^3$$

2.4.4 The Mean, Variance, Standard Deviation and Skewness of DP

From the moments of DP, the mean, variance and skewness can easily be derived. The mean is simply the first (central) moment, so:

$$\mu_{DP} = (N\text{-}1)\,\mathbf{E}[d_{12}]$$

The variance of DP is:

$$\mathbf{E}\{DP^2\} - \mathbf{E}^2\{DP\}$$

$$= (N\text{-}1)\ \mathbf{E}[d_{12}^2] + 2\ (N\text{-}2)\ \mathbf{E}[d_{12}d_{23}] + (N\text{-}2)(N\text{-}3)\ \mu_{DP}^2 - \mu_{DP}^2$$

$$= (N\text{-}1)\ \mathbf{E}[d_{12}^2] + 2\ (N\text{-}2)\ \mathbf{E}[d_{12}d_{23}] + ((N\text{-}2)(N\text{-}3)\text{-}1)\ \mu_{DP}^2$$

The standard deviation of DP is square root of the variance, or

$$\{(N\text{-}1)\ \mathbf{E}[d_{12}^2] + 2\ (N\text{-}2)\ \mathbf{E}[d_{12}d_{23}] + ((N\text{-}2)(N\text{-}3)\text{-}1)\ \mu_{DP}^2\}^{½}$$

The skewness of DP is the third central moment divided by the cube of the standard deviation, or

$$\begin{gathered}(N\text{-}1)\ \mathbf{E}[d_{12}^3] + 6(N\text{-}2)\ \mathbf{E}[d_{12}^2 d_{23}] + 3(N\text{-}2)(N\text{-}3)\ \mathbf{E}[d_{12}^2]\mu_{DP} \\ + 6(N\text{-}3)\ \mathbf{E}[d_{12}d_{23}d_{34}] + 6(N\text{-}3)(N\text{-}4)\ \mathbf{E}[d_{12}d_{23}]\mu_{DP} \\ - 3(N\text{-}1)\mathbf{E}[d_{12}^2] + 6(N\text{-}2)\mathbf{E}[d_{12}d_{23}] + 3(N\text{-}2)(N\text{-}3)\mu_{DP}^2 + (2+(N\text{-}3)(N\text{-}4)(N\text{-}5))\ \mu_{DP}^3\end{gathered}$$

divided by

$$(\{(N\text{-}1)\ \mathbf{E}[d_{12}^2] + 2\ (N\text{-}2)\ \mathbf{E}[d_{12}d_{23}] + ((N\text{-}2)(N\text{-}3)\text{-}1)\ \mu_{DP}^2\}^{½})^3$$

or,

$$\begin{gathered}(N\text{-}1)\ \mathbf{E}[d_{12}^3] + 6(N\text{-}2)\ \mathbf{E}[d_{12}^2 d_{23}] + 3(N\text{-}2)(N\text{-}3)\ \mathbf{E}[d_{12}^2]\ \mu_{DP} \\ + 6(N\text{-}3)\ \mathbf{E}[d_{12}d_{23}d_{34}] + 6(N\text{-}3)(N\text{-}4)\ \mathbf{E}[d_{12}d_{23}]\ \mu_{DP} \\ - 3(N\text{-}1)\mathbf{E}[d_{12}^2] + 6(N\text{-}2)\mathbf{E}[d_{12}d_{23}] + 3(N\text{-}2)(N\text{-}3)\ \mu_{DP}^2 + (2+(N\text{-}3)(N\text{-}4)(N\text{-}5))\ \mu_{DP}^3\end{gathered}$$

divided by

$$\{(N\text{-}1)\ \mathbf{E}[d_{12}^2] + 2\ (N\text{-}2)\ \mathbf{E}[d_{12}d_{23}] + ((N\text{-}2)(N\text{-}3)\text{-}1)\ \mu_{DP}^2\}^{1.5}$$

2.5 Conclusions

The space-time clustering statistic DP was formally presented in this chapter, with some discussion of its properties. The first three moments of DP were derived here, along with its mean, variance and skewness. The next chapter describes the simulation studies that were carried out to further investigate DP and its properties.

Chapter III

METHODOLOGY AND RESULTS OF SIMULATION COMPARISON STUDIES

3.0 Introduction

Simulation studies were carried out on DP, Knox's Test (KNOX), and Mantel's Generalized Regression (MGR), to investigate the properties of DP under various conditions, and to compare the performance of DP to that of KNOX and MGR, using data of various sample sizes and degrees of clustering. In this chapter, the methods and rationale for generating the data for use in the simulation studies are described, and the procedure for carrying out these simulation studies is outlined.

3.1 Generating Clustered Data.

There are three generally recognized models for disease clustering. *Point-source clustering* occurs when the observed clustered data all arise from a single point source. *Separate-source clustering* occurs when the observed clustered data is made up of several separate point source clusters, which may or may not be complete clusters. *Contagion clustering* occurs when the observed clustered data may be made up of several separate clusters, but successive data points within a cluster arise from other data points (instead of from point sources). The observed contagion clustering data may or may not be complete.

To investigate the statistics in this research, data were generated using a separate-source clustering model in the following manner. Data were generated using the unit square as the spatial area and the unit interval as the time interval. So, a case is denoted

by the ordered triplet (X, Y, T), where $0 \leq X \leq 1$, $0 \leq Y \leq 1$, and $0 \leq T \leq 1$. Let **n** denote the number of observed cases, that is, the total sample size in this set of clustered data.

The sizes of the clusters in the observed set of data will be determined by the Poisson distribution. Select $n_1 \sim$ Poisson (λ) as the size of the first cluster. If $n_1 \geq \mathbf{n}$, then set $n_1 = \mathbf{n}$, and there is only one cluster in the observed set of data. If $n_1 \leq \mathbf{n}$, select $n_2 \sim$ Poisson (λ) as the size of the second cluster. If $n_1+n_2 \geq \mathbf{n}$, then set $n_2 = \mathbf{n} - n_1$, and there are two clusters in the observed set of data. If $n_1+n_2 \leq \mathbf{n}$, select $n_3 \sim$ Poisson (λ) as the size of the third cluster. If $n_1+n_2+n_3 \geq \mathbf{n}$, then set $n_3 = \mathbf{n}-(n_1+n_2)$, and there are three clusters in the observed set of data. This continues until enough cluster sizes are determined that sum to **n**, the sample size of the observed data set.

Data points (cases) within each of the clusters in the observed set of data will be determined as follows. The first cluster source (X_{S1}, Y_{S1}, T_{S1}) is selected at random in the unit square and on the unit interval. The first case in the first cluster is located in a random direction θ_1 away from the first cluster source, at a distance d_1 from the first cluster source, where $d_1 \sim$ exponential (δ). The Cartesian coordinates X_1 and Y_1 of this point are $X_1 = X_{S1} + \cos(\theta_1)*d_1$ and $Y_1 = Y_{S1} + \sin(\theta_1)*d_1$. T_1 is set equal to t_1, where $t_1 \sim$ exponential (τ). This produces the first point (X_1, Y_1, T_1). The next case (X_2, Y_2, T_2) will be located in a random direction θ_1 away from the cluster source, at a distance d_2 from the cluster source, where $d_2 \sim$ exponential (δ). The Cartesian coordinates X_2 and Y_2 of this point are $X_2 = X_{S1} + \cos(\theta_2)*d_2$ and $Y_2 = Y_{S1} + \sin(\theta_2)*d_2$. T_2 will be set at a time interval t_2 from the cluster source, where $t_{12} \sim$ exponential (τ). So, $T_2 = T_1 + t_1$, which produces the second point (X_2, Y_2, T_2). Each subsequent case in the first cluster will be determined in the same manner: This continues until all of the points in the cluster have been generated. The entire process repeats for each cluster in the observed set of data, until all of the **n** data cases in the observed set of data have been generated.

Note that this process may generate cases that fall outside the unit square in space, or outside the unit interval in time. Cases that fall outside the unit square or the unit interval

are discarded. This emulates what happens in real life, for an investigation of disease clustering typically is bounded in space or time. Hence, there may be cases that are a part of the disease cluster but are unobservable, because they fall outside the geographic region or time interval of interest. Because of this, the actual number of data cases analyzed may be slightly less than **n**, the target number of data cases in the observed set of data.

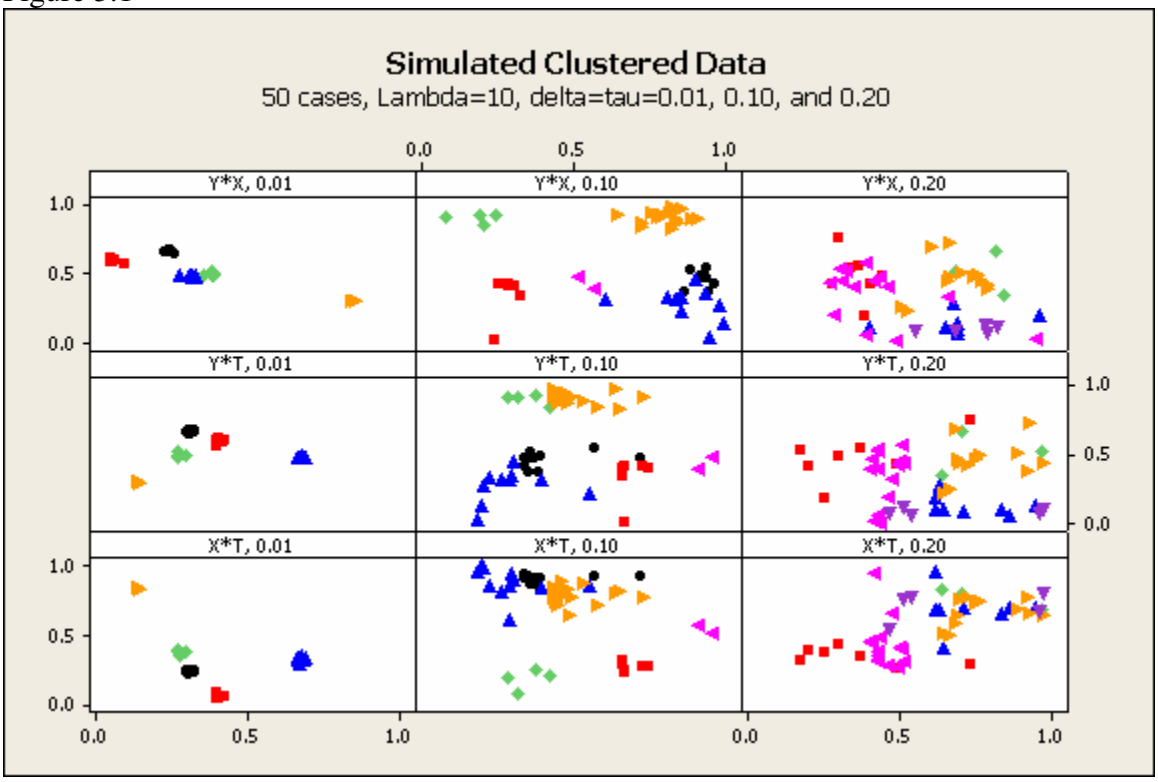
3.2 Simulation Plan.

The following parameters govern the simulation of clustered data cases:

n	The number of cases in the complete data set.
λ	The Poisson parameter, which is the average cluster size in the complete data set.
δ	The exponential parameter, which is the average distance from a case to its point source.
τ	The exponential parameter, which is the average time from a case to the next case (if any) in the same cluster..

The observed number of cases and average cluster size have expectations slightly smaller than **n** and λ, respectively. A preliminary study was conducted to determine how these parameters influence the degree of clustering in the simulated data cases. In this study, clustered data were generated by varying δ and τ through 12 values ranging from 0.01 to 0.5. Plots were made of x vs. y, x vs. t, and y vs. t, for the data cases generated at each of these 12 parameter values, to examine the degree of clustering that these values produced. Examples of these plots appear on the following page; the different colors in the plots represent different clusters. Based on an evaluation of these results, δ and τ values of 0.01 were used to simulate data with strong clustering, values of 0.10 were used to simulate data with moderate clustering, and values of 0.20 were used to simulate data with weak clustering. To account for all of the possible combinations of the three values of δ and τ, nine simulation runs must be conducted.

Figure 3.1



Observed data cases were generated for **n** = 25, 50, 100, 250 and 500. Cases were also generated for λ = 5, 10, 25, and 50. Because it is reasonable to expect λ to be small as compared to **n**, simulated data cases were generated only when $\lambda \leq$ **n**/5. So data cases were generated for λ = 5 when **n**=25, λ = 5 and 10 when **n**=50 and when **n**=100, λ = 5, 10 and 25 when **n**=250, and λ = 5, 10, 25, and 50 when **n**=500. At each of these 12 combinations of λ and **n**, a simulation run was conducted for data generated at each of the nine combinations of the values of δ and τ, for a total of 108 simulation runs.

In addition to the simulations conducted using data cases generated according to the separate-source clustering model, simulations were also conducted for data cases generated randomly in both space and time. Simulations on data that are completely random in both space and time serve to validate the computer programming in this research, for there should be no unusual patterns in the results of these statistics on random data. These simulations were conducted for **n**=25, **n**=50, **n**=100, **n**=250, and **n**=500, for a total of 5 simulations on completely random data. Altogether, 113

simulations were conducted to investigate the statistics in this research. This is depicted in the following table:

Table 3.1: **Number of Simulations Conducted to Examine the Performance of DP, Knox's Test, and Mantel's Generalized Regression**

lambda	delta	tau	**n** = 25	**n** = 50	**n** = 100	**n** = 250	**n** = 500
5	3 values	3 values	9 simulations	9 simulations	9 simulations	9 simulations	9 simulations
10	3 values	3 values		9 simulations	9 simulations	9 simulations	9 simulations
25	3 values	3 values				9 simulations	9 simulations
50	3 values	3 values					9 simulations
Random	in space	and time	1 simulation	1 simulation	1 simulation	1 simulation	1 simulation

A simulation run consisted of 500 simulated data sets. For unclustered data, the 500 data sets in a simulation run are made up of data cases generated randomly in both space and time. For clustered data, the 500 data sets in a simulation run are made up of data cases generated for the specified values of **n**, λ, δ and τ, according to the separate-source clustering model. DP, KNOX and MGR are all computed on each of the simulated data sets that are generated. In fact, the three cluster statistics are computed on exactly the same data sets in these simulation studies, as opposed to each statistic being computed on 500 data sets generated the same way, using the same parameters and the same clustering model. This greatly enhances the ability to compare the performance of these statistics.

3.3 Computational Details for these Cluster Statistics.

Various assumptions must be made in order to determine the significance for each of these statistics based on their distributional properties. To avoid any impact of the validity of the assumptions on significance, it was decided to compute significance for each of these statistics from their empirical distribution using randomization techniques as follows. The empirical permutational distribution of a statistic is determined by tabulating the value of the statistic for all possible permutations of the times of occurrence of the N cases (N! permutations). The value obtained for the statistic is then

compared to this empirical distribution. An estimate[8] of the full empirical distribution can be determined by randomly re-assigning the times of occurrence to the cases some number of times, and re-computing the statistic for each of these random reassignments. If this is done (N-1) times, the resulting set of values (including the one calculated from the original assignment) produces an estimated empirical permutational distribution of size N. The value obtained for the statistic is then compared to this empirical distribution. If the statistic is more extreme than or equal to M of the N values in the empirical distribution, then the exact significance is P=M/N.

For these simulation studies, estimated empirical distributions of size N=500 were used to compute significance. Because of this, the smallest possible p-value is 0.002, which will occur when the observed value of the statistic is the smallest value in the empirical distribution, so that p=1/500. In these simulations, $0.002 \leq p \leq 1.000$.

3.3.1 Computational Details for DP.

Because DP is defined as the length of the path from the first case to the last case when the cases are ordered by their times of occurrence, the simulated data are first ordered by time before the value of DP is computed. On these time-sorted data, the value of DP is the path length from the first to the last case. The p-value is then computed as described above, and in Chapter 2.

3.3.2 Computational Details for Knox's Test.

Knox's Test requires that a 'critical' distance and time be specified, in order to compute and evaluate the statistic. In practice, these are determined by considering the etiology of the disease under investigation. In simulation studies, there is no disease

[8] Randomly generated permutations of the **n** cases should yield an unbiased tabulation of path lengths that is proportionate to the full empirical distribution of DP, such that the number of permutations greater than DP in this tabulated distribution is proportionate to the number greater than DP in the complete empirical distribution. Hence, this is a valid estimate of the exact significance of DP.

etiology to provide clues regarding what values for the critical distance and time might be appropriate, so it is important to identify a way to determine these 'critical parameters' based on the generated simulated data.

Mantel (1967) argues that there should be relatively few close pairs when clustering is not present, and provides some discussion of K, the expected number of close pairs under the null hypothesis. With that in mind, expressions for T, the critical time difference and R, the critical distance difference, have been derived as follows:
If points are chosen at random, a second point is close in time if it is no farther than T from the first point. Ignoring any "edge effects", we have

$$\Pr\{\text{a pair is close in time}\} = 2T$$

Similarly, given one randomly chosen point, a second point is close in space if it is no farther than R from the first point. Again ignoring "edge effects", we have

$$\Pr\{\text{a pair is close in space}\} = \pi R^2.$$

Under the null hypothesis that space and time are independent,

$$\Pr\{\text{a pair is close in both time and space simultaneously}\} = 2T\pi R^2.$$

In a set of **n** cases, there are **n** (**n** -1)/2 pairs. The expected number of pairs that are close in both time and space simultaneously is then

$$K = T\pi R^2 n(n-1).$$

This yields one equation with two unknowns, so an additional restriction is needed to arrive at a solution. It is reasonable to require that the probability that a pair of points is close in space is equal to the probability that a pair of points is close in time. Given this assumption, we obtain

$$2T = \pi R^2 \text{ or } T = \pi R^2/2.$$

Expressions for T and R may now be derived given any choice of K, based on the expected number of pairs that are close in both time and space simultaneously. For time, we have

$$K = T\pi R^2 n(n-1) \text{ and } 2T = \pi R^2$$

$$\Rightarrow \quad K = T(2T)n(n-1)$$

$$\Rightarrow \quad T = SQRT(K/2n(n-1))$$

For space, we have

$$K = T\pi R^2 n(n-1) \text{ and } T = \pi R^2/2$$

$$\Rightarrow \quad K = (\pi R^2/2)\pi R^2 n(n-1)$$

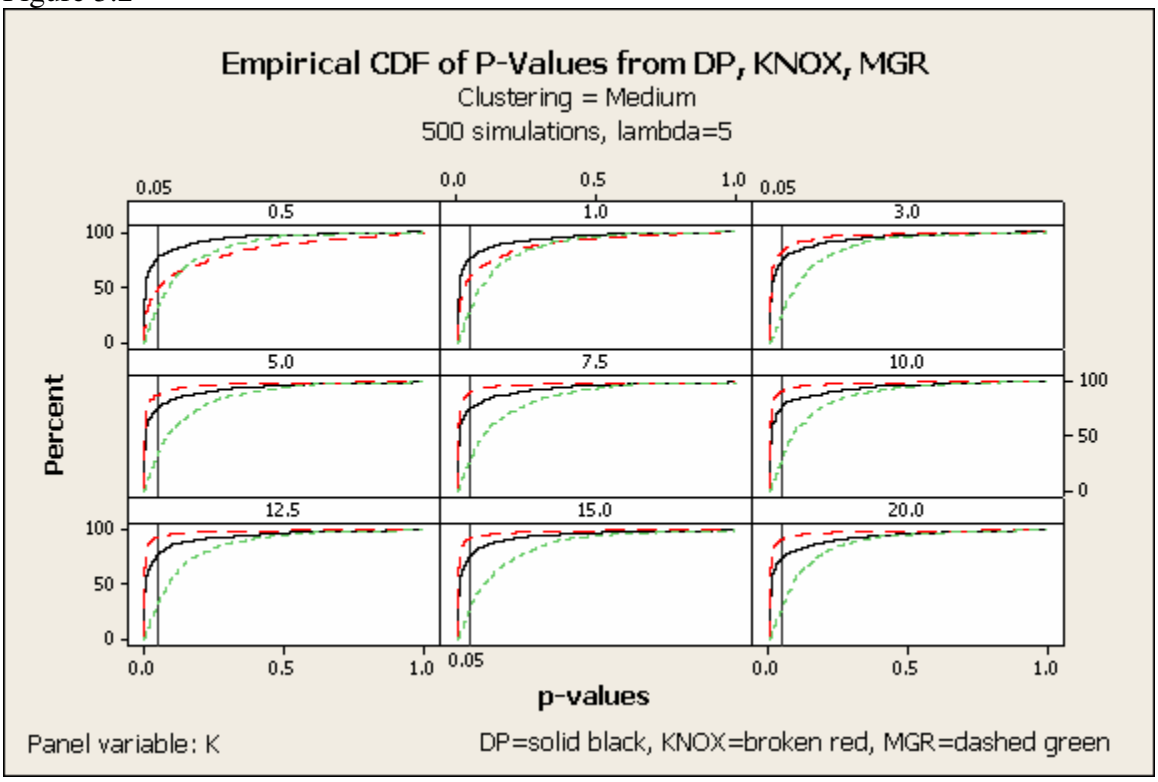
$$\Rightarrow \quad 2K = R^4(\pi^2 n(n-1))$$

$$\Rightarrow \quad R^4 = 2K/\pi^2 n(n-1)$$

These expressions allow appropriate values for critical distance and time to be computed directly during a simulation run by simply specifying K, the expected number of pairs close in both time and space simultaneously. But how large should K be? Mantel determined that Knox used a K of 0.8 for **n**=96, given the data in Knox's original paper (Mantel, 1967). Also, Mantel's hypothetical data in the same paper had 50 close pairs for **n**=100, which would have been highly significant for K of up to 15 or so.

To investigate this, simulations were conducted to help determine an appropriate value of K for our investigations. For **n**=50 and λ=5, simulations were conducted for values of K $\in$ (0.5, 1.0, 3.0, 5.0, 7.5, 10.0, 12.5, 15.0, 20.0). In each simulation run, 500 data sets were generated for the specified values of K, with weak clustering (δ=τ=0.20), medium clustering (δ=τ=0.10), and strong clustering (δ=τ=0.01). DP, KNOX and MGR were all computed on each of the simulated data sets. This was repeated for **n**=50 and λ=10. The following graph is a typical example of the results from these simulations.

Figure 3.2



In general, the results of these 54 simulation runs indicated that Knox's Test performed much better when K was at least 5, and was less than optimal for values of K less than 5. Note that the results from Mantel's hypothetical data indicated that K should be no more than 15 or so. After considering this, K was set to a value of 5.0 for these simulation studies. While the critical distances and times from this process have no meaning from the perspective of disease etiology, they do produce a Knox's Test for these simulations that is close to optimal from a statistical sense, in terms of identifying clustering.

3.3.3 Computational Details for Mantel's Generalized Regression.

Because Mantel's Generalized Regression (MGR) is essentially the product of the reciprocals of the pairwise distances in space and time, the possibility of encountering values of 0 while computing the test statistic is of great concern. MGR is defined by adding a "small constant" to each value, to preclude operation on 0 values. The value of

this constant can greatly impact the power of the statistic; so determining this constant is an important issue.

However, in a simulation study there is almost no chance of encountering values of 0 while computing the test statistic. Real data are often rounded, so that the precision of measurement is by month or week with regards to time, or by county, mile, or block with regards to space. In a simulation study, data are all from continuous distributions, and it is virtually impossible to get two identical points at random. Hence, MGR was computed without adding the recommended constant in these simulations, avoiding any impact on the performance of the statistic from this source.

3.4 Comparing Cluster Statistics.

In a given simulation run, 500 data sets are generated. These data are generated either completely at random, or clustered based on the specified values of $\mathbf{n}$, λ, δ and τ, according to the separate-source clustering model. The results of these simulation studies are sets of 500 p-values that represent the performance of DP, Knox's Test (KNOX), and Mantel's Generalized Regression (MGR) under the conditions determined by the simulated data. These sets of p-values are the empirical distributions of the p-values for DP, KNOX, and MGR under the specified conditions. In any meaningful discussion of the performance of these statistics, these empirical p-value distributions must be compared and assessed. The Kolmogorov-Smirnov Test was used when comparing these empirical distributions, and the expected significance level (ESL) was used to assess and compare the power of these statistics. The following discusses these methods.

3.4.1 The Kolmogorov-Smirnov Test.

The Kolmogorov-Smirnov test (also known as the Smirnov Test, see Conover, 1999) may be used to test for any kind of difference between two empirical distributions. If we denote the empirical distribution function of sample 1 as $F_1(x_j)$, and the empirical

distribution function of sample 2 as $F_2(x_j)$, the Kolmogorov-Smirnov test is the maximum absolute difference between the two empirical distributions, and is expressed as:

$$\boldsymbol{D} = \max_j |F_1(x_j) - F_2(x_j)| \quad \text{where } j = 1, 2, 3, \ldots, n$$

$\boldsymbol{D}$ and its asymptotic p-value were computed by the NPAR1WAY procedure in SAS, and are reported here for all empirical distribution comparisons (*SAS/STAT User's Guide, Version 8*, 1999).

3.4.2 The Expected Significance Level, ESL.

Alternative tests are often compared in terms of statistical power, or in terms of asymptotic relative efficiency. However, using a power function requires that a specific level of significance be decided upon in advance. If this cannot be done, power functions must be considered at several levels of significance, which may be problematic. Also, when the test statistic is a discrete random variable, it may not be possible to obtain a specific level of significance directly. Further, if the asymptotic relative efficiency is used, some other criterion must be used for small n. Joiner (1969) suggests the use of the expected significance level for Monte Carlo studies. Also known as the expected p-value, the expected significance level was introduced by Dempster and Schatzoff (1965), and further discussed more recently by Hung, O'Neill, Bauer and Köhne (1997), and by Sackrowitz and Samuel-Cahn (1999). This alternative criterion for comparing tests is derived by considering the p-value as a random variable, deriving its expectation, and understanding some of its properties.

Given a random test statistic T, distributed under H_0 as F_0, and under H_θ as F_θ, the p-value is the random variable X such that

$$X = 1 - F_0(T)$$

Under H_0, the p-value is uniformly distributed on [0, 1]. The power β of a test of size α is the probability under H_0 that the p-value will be less than or equal to α. So, for a given alternative θ, the power β of a test of size α is:

$$\beta = P_\theta (X \leq \alpha) \text{ or}$$

(Eqn. 3.1) $$P_\theta (X \leq \alpha) = 1 - F_\theta (F^{-1}{}_0 (1-\alpha))$$

If α takes on all values $0 < \alpha < 1$, this yields the distribution function of the *p*-value under H_θ. Under H_θ, the *p*-value is right-skewed, so that the *p*-value is always smaller under H_θ than under H_0. The expectation of the *p*-value under H_θ, ESL(θ), can be derived from Eqn.1. Note that expected value of the *p*-value under H_θ will be less than ½, the expected *p*-value under H_0, and the smaller ESL(θ) is, the better the test is suited for distinguishing between H_0 and H_θ.

Alternately, if T is distributed according to $F_\theta(\cdot)$, and T^* is independently distributed according to $F_0(\cdot)$,

(Eqn. 3.2) $$\text{ESL}(\theta) = \text{P}(\text{T}^* \geq \text{T})$$

As derived in Eqn. 2, ESL(θ) can easily be used in simulation studies and direct evaluation. For each simulation in this study, the empirical p-value was computed in precisely this way. In simulation studies such as this one, ESL(θ) may be estimated by computing the average of these empirical p-values. That is, if m is the number of simulations conducted:

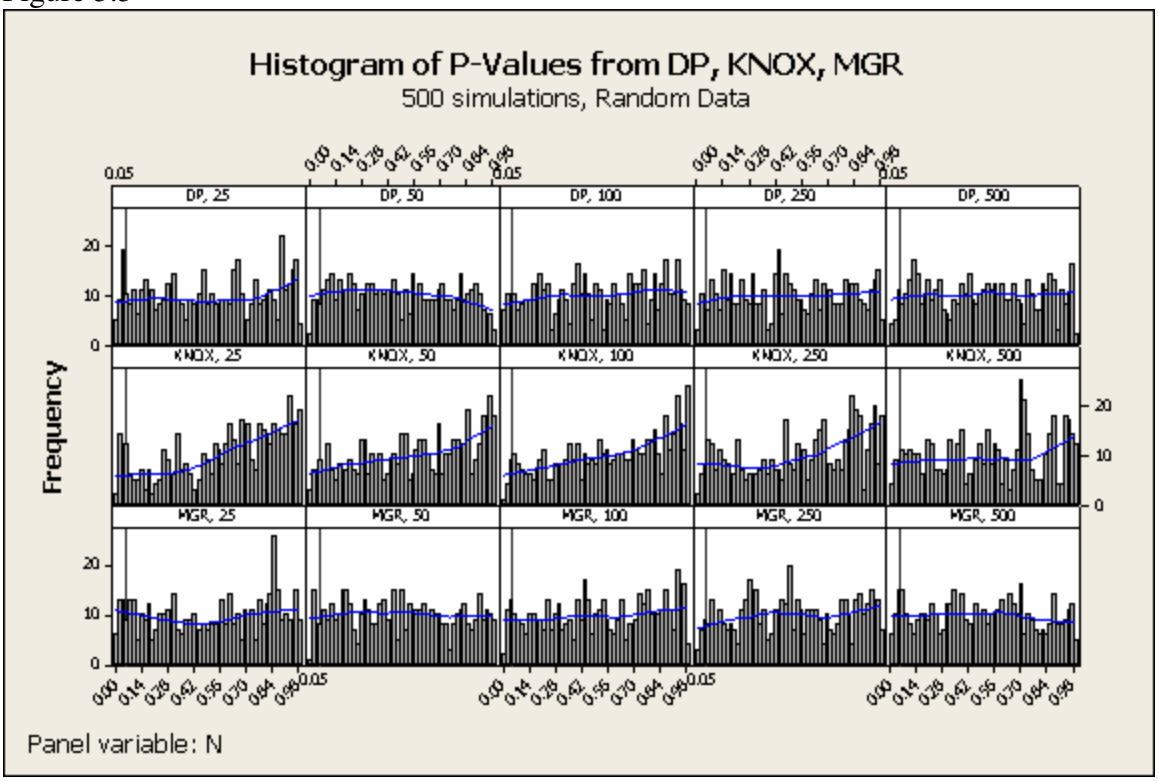
$$\text{ESL}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{1} I(T^* \geq T) \ ,$$

The arithmetic average of these empirical p-values, is an unbiased estimator of ESL(θ), with variance of ESL(1-ESL)/m. Comparisons of two ESL's are conducted using a normal test based on the standard error of the difference of the two ESL's. This yields an approximate p-value for the comparison that is liberal, for it is always less than the theoretical p-value. Because of this, some significant ESL comparisons may not be truly significant. The test with the smaller ESL(θ) is the one better suited for distinguishing between H_0 and H_θ.

3.5 Simulation Studies on Random Data.

Simulations were conducted for data cases generated randomly in both space and time to validate the computer programming in this research. No unusual patterns should appear in the results of these statistics on random data. These simulations were conducted for **n**=25, **n**=50, **n**=100, **n**=250, and **n**=500, and the histograms below were produced to assess the results of these statistics on random data.

Figure 3.3

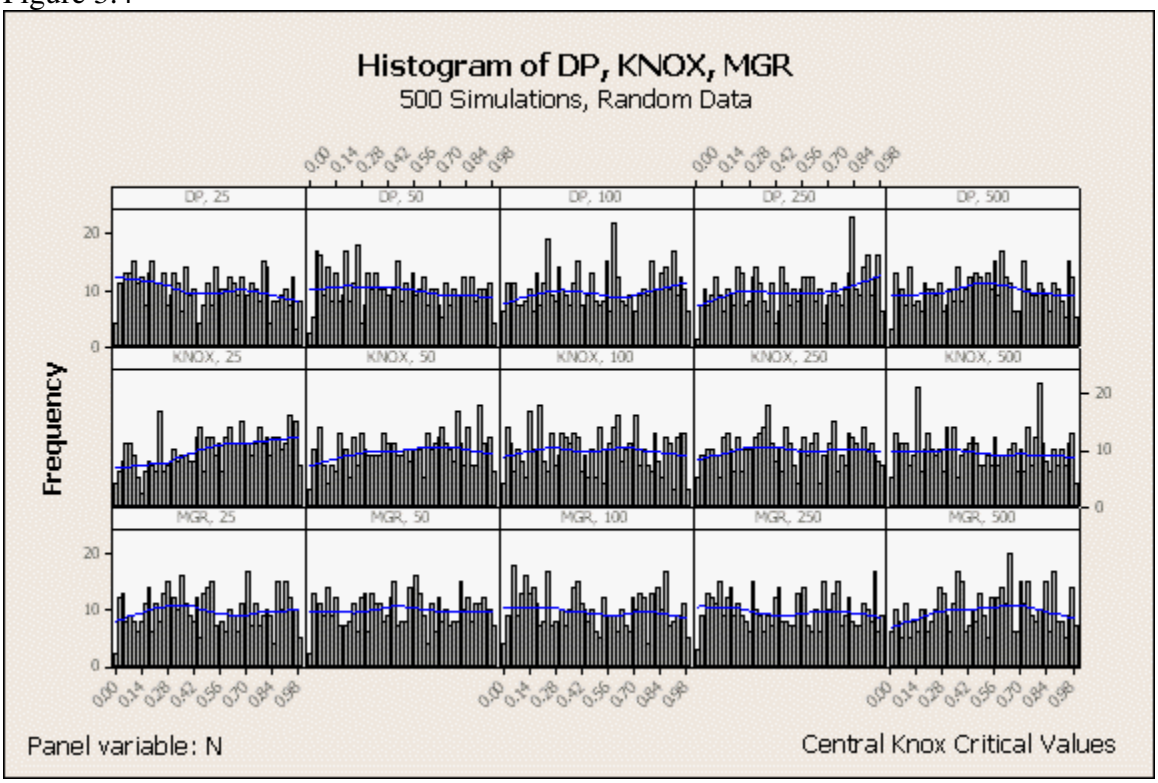


Notice the unanticipated strong bias towards large p-values for the Knox Test results. This bias exists for all sample sizes, indicating some unaddressed issue with the manner the Knox Test was calculated. After some thought, it became apparent that a fundamental assumption was no longer valid. Knox's Test requires that a 'critical' distance and time be specified, in order to compute and evaluate the statistic. In this work, these 'critical parameters' were based on K, the expected number of close pairs under the null hypothesis. When testing for clustering, Mantel argues that K should be relatively small,

as it avoids the detection of spurious clustering. However, it is clear that by avoiding the detection of spurious clustering, a bias towards large p-values is introduced.

When data are completely random, one expects a relatively uniform distribution of p-values. Considering each axis separately, setting the critical value for distance and time at their respective medians will yield a K value of ¼ of the total number of pairs. So, setting K to ¼ of the total number of pairs should remove the systematic bias, and correct the propensity towards large p-values. These simulations on random data were repeated for **n**=25, **n**=50, **n**=100, **n**=250, and **n**=500, with K= ¼ (**n**(**n**-1)/2), at each sample size. Results from these simulations yielded the set of histograms below:

Figure 3.4

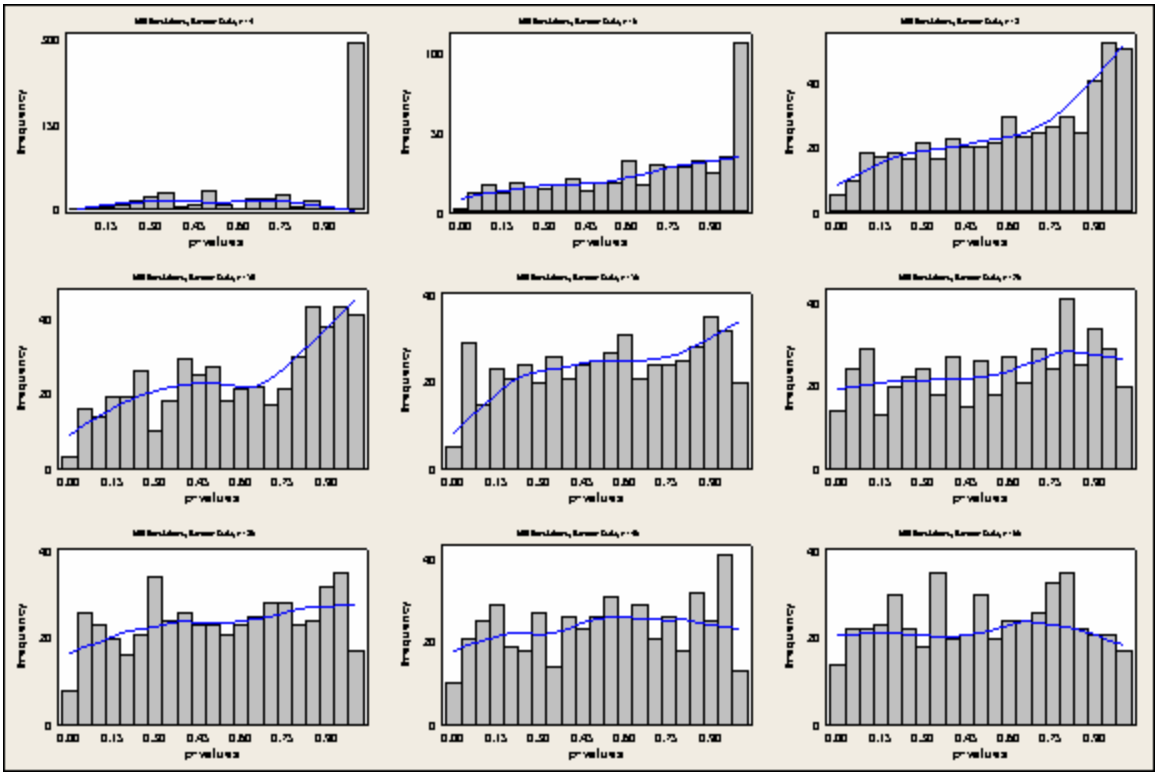


The strong bias towards large p-values in the Knox Test results is virtually eliminated in all cases except for **n**=25. The presence of the bias in the smallest sample size is a clue that it is related to the discrete nature of Knox's Test. To investigate this, additional sets of simulations were carried out using Knox's test for a variety of small sample sizes, with K= ¼ (**n**(**n**-1)/2), at each sample size. The relationship between the

strength of this bias and the smallness of the sample size was confirmed through these simulations. Results from the simulations on smaller sample sizes (where **n**=4, **n**=6, **n**=8, **n**=10, **n**=15, **n**=25, **n**=35, **n**=45 and **n**=55) appear in the following histograms:

Figure 3.5

Histograms of KNOX Test Results on Random Data of Various Sample Sizes



Note that apart from random variation, as **n** increases the bias becomes less and less. This indicates that this bias is, in fact, related to the discrete nature of Knox's Test. The p-value is the probability of obtaining the observed result, or a result more extreme than the observed result. This usually implies that the p-value is an upper tail probability. This definition poses no problems when the distribution is continuous, because a single observed value has no weight of probability in a continuous distribution. However, with a discrete distribution, each value represents some probability weight. The p-value is too large (biased) because the entire weight of probability of the observed value goes into the upper tail.

There are two implications to this result. First, some adjustment can be made to the way the p-value is computed for Knox's test, to compensate for this bias. However, it is unclear if the bias due to the discrete nature of Knox's test is known. The significance of Knox's test is typically determined either by using the Poisson approximation, or by using the permutational distribution of **a**, the number of close pairs, or by considering **a** to be approximately normal and using the appropriate normal distribution. Neither of these methods is adjusted for discreteness in the literature. Because we want our results to be as general as possible, we have chosen not to adjust the p-values from Knox's test in this work for discreteness.

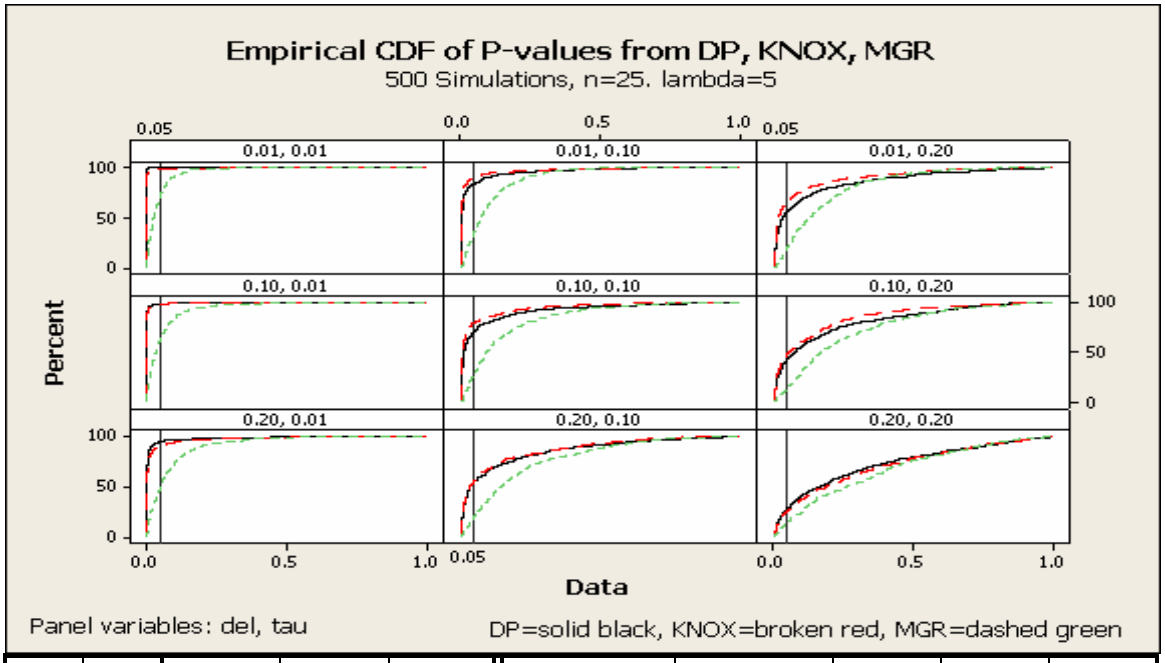
This also highlights the fact that these tests should not be conducted on data with small sample sizes. While we have conducted our simulations starting with **n**=25, that number is likely too small for practical work. In practice, n should probably be at least 50 or so, to avoid any untoward influence by the bias due the discreteness of the test.

3.6 Simulation Studies on Clustered Data.

Simulations were conducted for clustered data cases generated according to the plan detailed earlier in this chapter. Briefly, for 12 combinations of λ and **n**, a simulation run consisting of 500 simulated data sets was conducted at each of the nine combinations of the values of δ and τ, for a total of 108 simulation runs. These simulation results are displayed graphically on cumulative distribution function plots (CDF plots). These graphs are organized so that for each λ-**n** combination, nine CDF plot panels are produced that represent each of the nine combinations of the values of δ and τ. The CDFs for DP, KNOX, and MGR are overlaid on each panel to facilitate comparisons between them. The nine plot panels are arranged together on the same plot, to facilitate panel-to-panel comparisons.

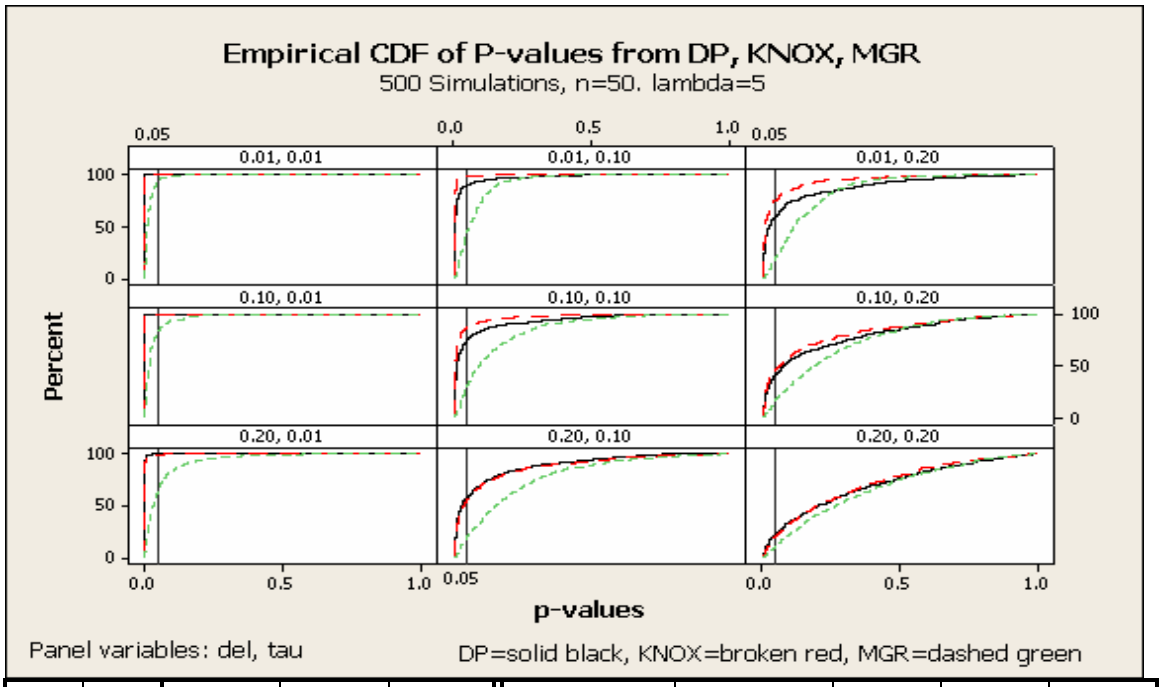
The table accompanying each set of plots contains the expected significance level (ESL) for each statistic and its standard deviation, and the results from all pairwise comparisons of ESLs among the three statistics. The table also contains the results from all pairwise comparisons of the EDFs among the three statistics, using the Kolmogorov-Smirnov test. These results follow.

Figure 3.6 **Results from 500 Simulations, n=25, λ=5**



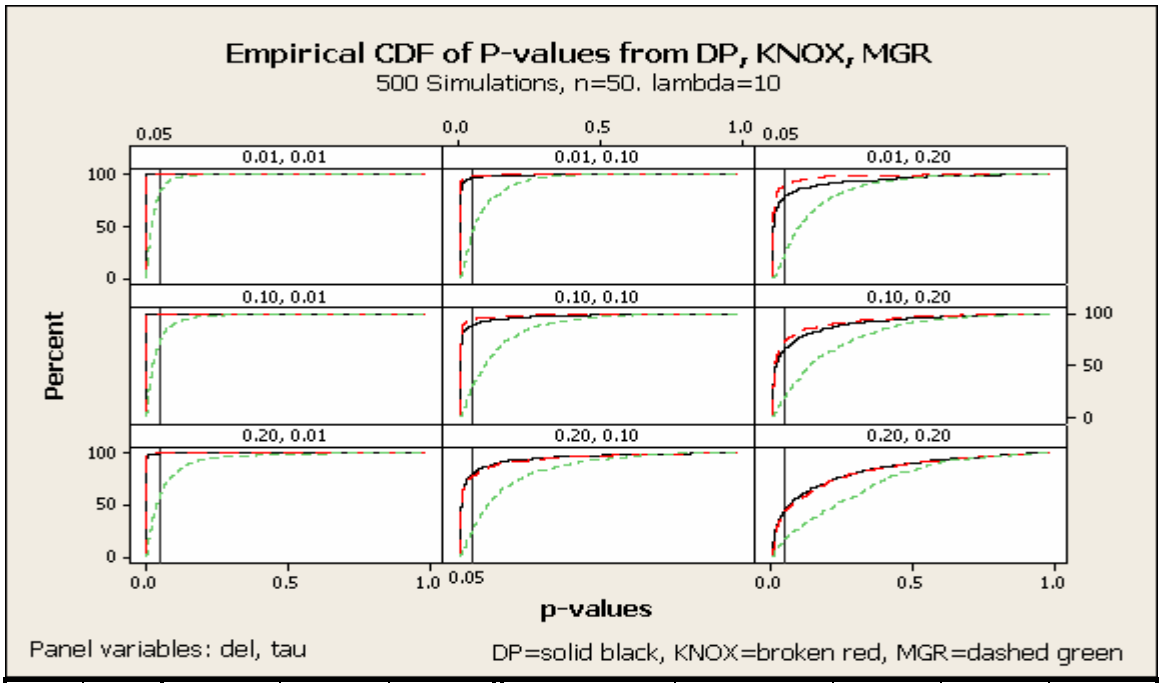
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	$p\text{-val}_{(5)}$ (6)	K-S ***D*** (7)	$p\text{-val}_{(7)}$ (8)
0.2	0.2	DP	0.2637	0.0197	DP vs KNOX	-0.0143	0.6105	0.0500	0.5596
0.2	0.2	KNOX	0.2780	0.0200	DP vs MGR	-0.0520	0.0694	0.1620	<.0001
0.2	0.2	MGR	0.3157	0.0208	KNOX vs MGR	-0.0377	0.1917	0.1260	0.0007
0.2	0.1	DP	0.1338	0.0152	DP vs KNOX	0.0141	0.5023	0.0420	0.7699
0.2	0.1	KNOX	0.1197	0.0145	DP vs MGR	-0.0765	0.0013	0.3680	<.0001
0.2	0.1	MGR	0.2103	0.0182	KNOX vs MGR	-0.0906	0.0001	0.3780	<.0001
0.2	0.01	DP	0.0152	0.0055	DP vs KNOX	-0.0077	0.3715	0.0800	0.0815
0.2	0.01	KNOX	0.0230	0.0067	DP vs MGR	-0.0677	0.0000	0.7300	<.0001
0.2	0.01	MGR	0.0829	0.0123	KNOX vs MGR	-0.0599	0.0000	0.6640	<.0001
0.1	0.2	DP	0.1711	0.0168	DP vs KNOX	0.0353	0.1211	0.0760	0.1114
0.1	0.2	KNOX	0.1358	0.0153	DP vs MGR	-0.0712	0.0053	0.2940	<.0001
0.1	0.2	MGR	0.2423	0.0192	KNOX vs MGR	-0.1065	0.0000	0.3420	<.0001
0.1	0.1	DP	0.0736	0.0117	DP vs KNOX	0.0244	0.1071	0.1120	0.0038
0.1	0.1	KNOX	0.0492	0.0097	DP vs MGR	-0.0763	0.0001	0.4920	<.0001
0.1	0.1	MGR	0.1499	0.0160	KNOX vs MGR	-0.1007	0.0000	0.5880	<.0001
0.1	0.01	DP	0.0041	0.0029	DP vs KNOX	-0.0014	0.7441	0.0200	1.0000
0.1	0.01	KNOX	0.0055	0.0033	DP vs MGR	-0.0526	0.0000	0.8260	<.0001
0.1	0.01	MGR	0.0567	0.0103	KNOX vs MGR	-0.0512	0.0000	0.8140	<.0001
0.01	0.2	DP	0.1268	0.0149	DP vs KNOX	0.0376	0.0551	0.1300	0.0004
0.01	0.2	KNOX	0.0892	0.0127	DP vs MGR	-0.0458	0.0421	0.3800	<.0001
0.01	0.2	MGR	0.1725	0.0169	KNOX vs MGR	-0.0833	0.0001	0.4940	<.0001
0.01	0.1	DP	0.0347	0.0082	DP vs KNOX	0.0092	0.3963	0.0860	0.0495
0.01	0.1	KNOX	0.0256	0.0071	DP vs MGR	-0.0653	0.0000	0.6320	<.0001
0.01	0.1	MGR	0.1001	0.0134	KNOX vs MGR	-0.0745	0.0000	0.7120	<.0001
0.01	0.01	DP	0.0022	0.0021	DP vs KNOX	-0.0039	0.3332	0.0420	0.7699
0.01	0.01	KNOX	0.0062	0.0035	DP vs MGR	-0.0376	0.0000	0.8820	<.0001
0.01	0.01	MGR	0.0398	0.0087	KNOX vs MGR	-0.0336	0.0004	0.8400	<.0001

Figure 3.7 **Results from 500 Simulations, n=50, λ=5**



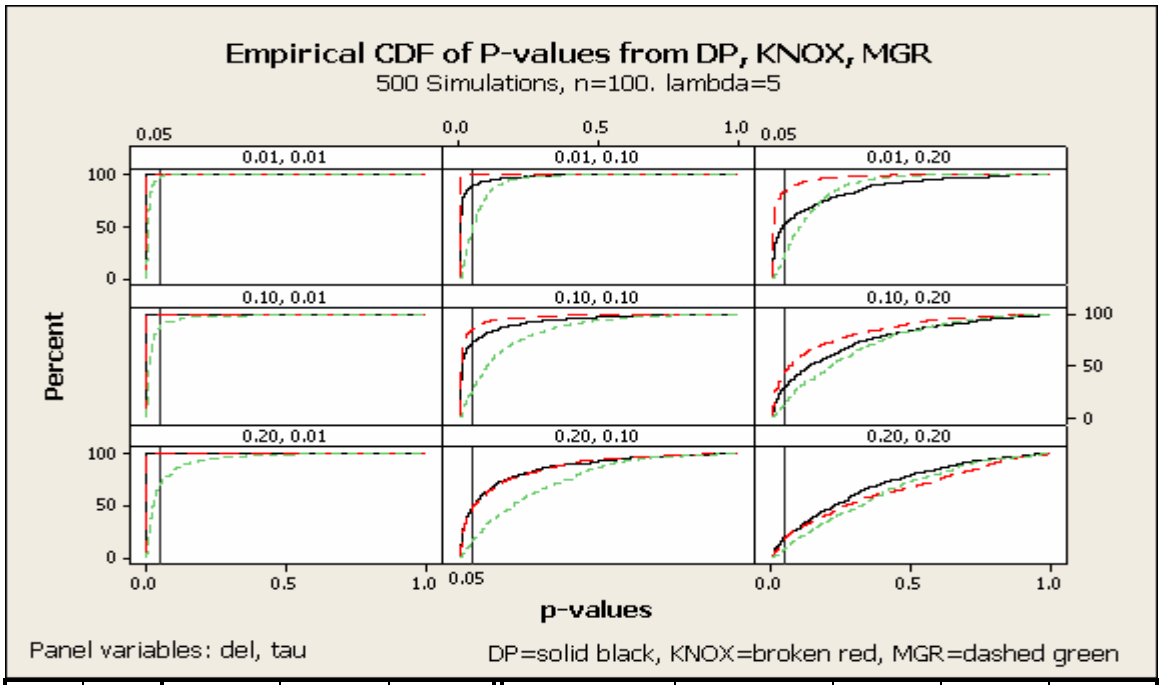
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	ESL_A - ESL_B (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.2916	0.0203	DP vs KNOX	0.0124	0.6630	0.0380	0.8632
0.2	0.2	KNOX	0.2791	0.0201	DP vs MGR	-0.0343	0.2398	0.1280	0.0006
0.2	0.2	MGR	0.3259	0.0210	KNOX vs MGR	-0.0468	0.1070	0.1140	0.0030
0.2	0.1	DP	0.1059	0.0138	DP vs KNOX	-0.0117	0.5582	0.0500	0.5596
0.2	0.1	KNOX	0.1176	0.0144	DP vs MGR	-0.1077	0.0000	0.3940	<.0001
0.2	0.1	MGR	0.2136	0.0183	KNOX vs MGR	-0.0961	0.0000	0.3660	<.0001
0.2	0.01	DP	0.0025	0.0023	DP vs KNOX	-0.0010	0.7693	0.0200	1.0000
0.2	0.01	KNOX	0.0036	0.0027	DP vs MGR	-0.0588	0.0000	0.8780	<.0001
0.2	0.01	MGR	0.0614	0.0107	KNOX vs MGR	-0.0578	0.0000	0.8580	<.0001
0.1	0.2	DP	0.1924	0.0176	DP vs KNOX	0.0274	0.2577	0.0680	0.1979
0.1	0.2	KNOX	0.1650	0.0166	DP vs MGR	-0.0544	0.0373	0.2620	<.0001
0.1	0.2	MGR	0.2468	0.0193	KNOX vs MGR	-0.0818	0.0013	0.3120	<.0001
0.1	0.1	DP	0.0599	0.0106	DP vs KNOX	0.0357	0.0047	0.2220	<.0001
0.1	0.1	KNOX	0.0242	0.0069	DP vs MGR	-0.0825	0.0000	0.5260	<.0001
0.1	0.1	MGR	0.1424	0.0156	KNOX vs MGR	-0.1182	0.0000	0.6880	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	0.9966	0.0040	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0291	0.0003	0.9040	<.0001
0.1	0.01	MGR	0.0311	0.0078	KNOX vs MGR	-0.0291	0.0003	0.9080	<.0001
0.01	0.2	DP	0.1104	0.0140	DP vs KNOX	0.0583	0.0007	0.1600	<.0001
0.01	0.2	KNOX	0.0521	0.0099	DP vs MGR	-0.0400	0.0597	0.4320	<.0001
0.01	0.2	MGR	0.1505	0.0160	KNOX vs MGR	-0.0983	0.0000	0.5860	<.0001
0.01	0.1	DP	0.0234	0.0068	DP vs KNOX	0.0186	0.0124	0.2460	<.0001
0.01	0.1	KNOX	0.0048	0.0031	DP vs MGR	-0.0504	0.0002	0.6720	<.0001
0.01	0.1	MGR	0.0738	0.0117	KNOX vs MGR	-0.0690	0.0000	0.8540	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	0.9989	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0157	0.0117	0.8600	<.0001
0.01	0.01	MGR	0.0177	0.0059	KNOX vs MGR	-0.0157	0.0118	0.8580	<.0001

Figure 3.8 **Results from 500 Simulations, n=50, λ=10**



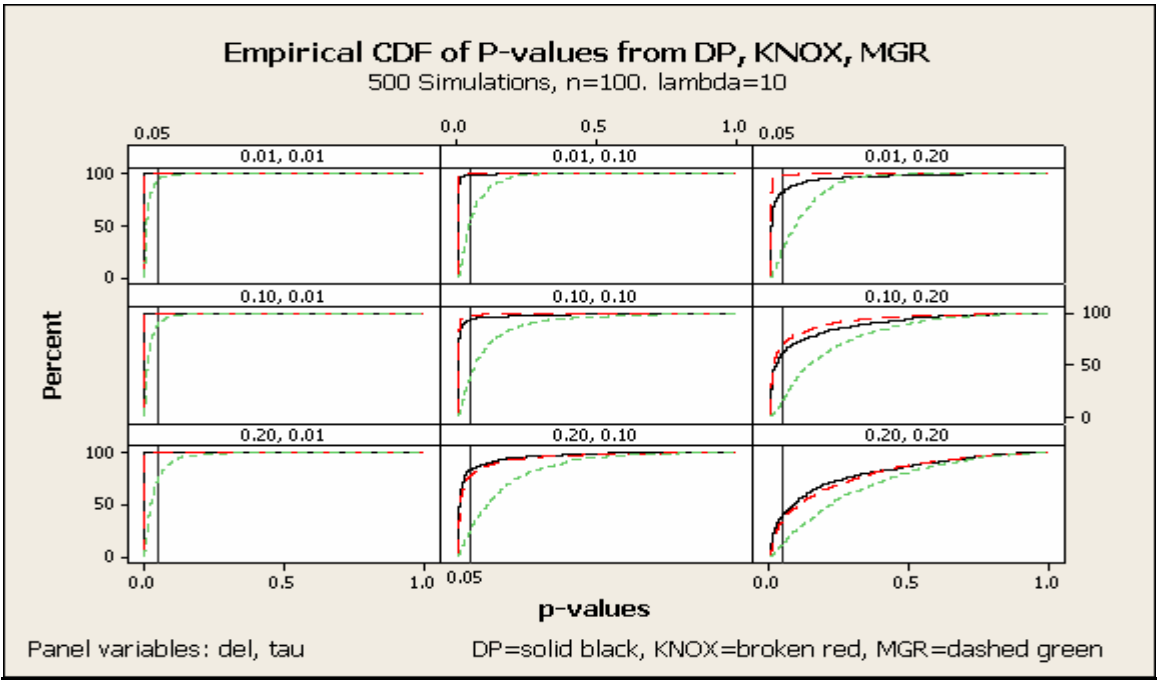
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.1587	0.0163	DP vs KNOX	-0.0069	0.7667	0.0720	0.1497
0.2	0.2	KNOX	0.1656	0.0166	DP vs MGR	-0.1116	0.0000	0.3060	<.0001
0.2	0.2	MGR	0.2703	0.0199	KNOX vs MGR	-0.1047	0.0001	0.2800	<.0001
0.2	0.1	DP	0.0473	0.0095	DP vs KNOX	-0.0072	0.6061	0.0320	0.9600
0.2	0.1	KNOX	0.0545	0.0102	DP vs MGR	-0.1104	0.0000	0.6120	<.0001
0.2	0.1	MGR	0.1578	0.0163	KNOX vs MGR	-0.1033	0.0000	0.5960	<.0001
0.2	0.01	DP	0.0036	0.0027	DP vs KNOX	0.0011	0.7590	0.0080	1.0000
0.2	0.01	KNOX	0.0025	0.0022	DP vs MGR	-0.0653	0.0000	0.9400	<.0001
0.2	0.01	MGR	0.0689	0.0113	KNOX vs MGR	-0.0663	0.0000	0.9440	<.0001
0.1	0.2	DP	0.0847	0.0125	DP vs KNOX	0.0200	0.2297	0.0900	0.0348
0.1	0.2	KNOX	0.0647	0.0110	DP vs MGR	-0.1150	0.0000	0.4820	<.0001
0.1	0.2	MGR	0.1997	0.0179	KNOX vs MGR	-0.1350	0.0000	0.5600	<.0001
0.1	0.1	DP	0.0258	0.0071	DP vs KNOX	0.0121	0.1676	0.0760	0.1114
0.1	0.1	KNOX	0.0137	0.0052	DP vs MGR	-0.1050	0.0000	0.7660	<.0001
0.1	0.1	MGR	0.1308	0.0151	KNOX vs MGR	-0.1171	0.0000	0.8340	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	0.9944	0.0020	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0404	0.0000	0.9540	<.0001
0.1	0.01	MGR	0.0424	0.0090	KNOX vs MGR	-0.0404	0.0000	0.9540	<.0001
0.01	0.2	DP	0.0574	0.0104	DP vs KNOX	0.0356	0.0037	0.1560	<.0001
0.01	0.2	KNOX	0.0218	0.0065	DP vs MGR	-0.0877	0.0000	0.6300	<.0001
0.01	0.2	MGR	0.1451	0.0158	KNOX vs MGR	-0.1234	0.0000	0.7840	<.0001
0.01	0.1	DP	0.0084	0.0041	DP vs KNOX	0.0038	0.4574	0.0840	0.0587
0.01	0.1	KNOX	0.0046	0.0030	DP vs MGR	-0.0717	0.0000	0.8480	<.0001
0.01	0.1	MGR	0.0800	0.0121	KNOX vs MGR	-0.0754	0.0000	0.9160	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0276	0.0004	0.9280	<.0001
0.01	0.01	MGR	0.0296	0.0076	KNOX vs MGR	-0.0276	0.0004	0.9280	<.0001

Figure 3.9 **Results from 500 Simulations, n=100, λ=5**



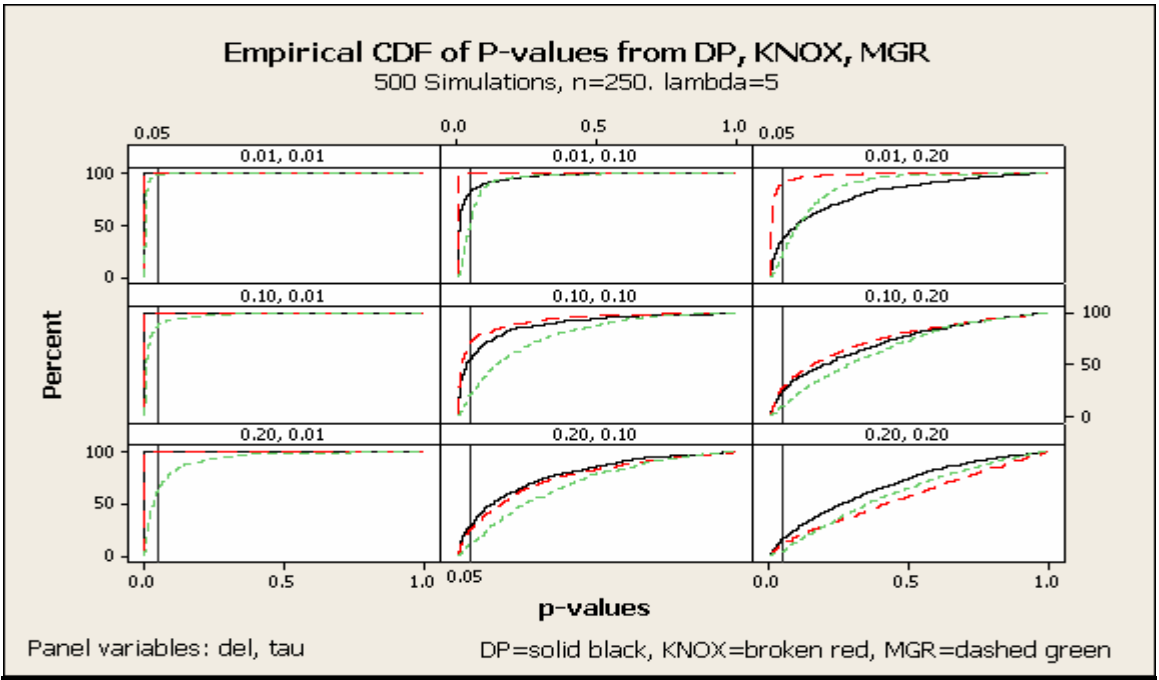
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.2857	0.0202	DP vs KNOX	-0.0605	0.0392	0.1240	0.0009
0.2	0.2	KNOX	0.3462	0.0213	DP vs MGR	-0.0629	0.0321	0.1420	<.0001
0.2	0.2	MGR	0.3486	0.0213	KNOX vs MGR	-0.0024	0.9359	0.1140	0.0030
0.2	0.1	DP	0.1329	0.0152	DP vs KNOX	0.0008	0.9710	0.0320	0.9600
0.2	0.1	KNOX	0.1321	0.0151	DP vs MGR	-0.1134	0.0000	0.3440	<.0001
0.2	0.1	MGR	0.2463	0.0193	KNOX vs MGR	-0.1142	0.0000	0.3200	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	0.9899	0.0140	1.0000
0.2	0.01	KNOX	0.0021	0.0020	DP vs MGR	-0.0569	0.0000	0.9560	<.0001
0.2	0.01	MGR	0.0589	0.0105	KNOX vs MGR	-0.0568	0.0000	0.9420	<.0001
0.1	0.2	DP	0.2273	0.0187	DP vs KNOX	0.0726	0.0033	0.1740	<.0001
0.1	0.2	KNOX	0.1547	0.0162	DP vs MGR	-0.0374	0.1695	0.1720	<.0001
0.1	0.2	MGR	0.2647	0.0197	KNOX vs MGR	-0.1100	0.0000	0.3180	<.0001
0.1	0.1	DP	0.0616	0.0108	DP vs KNOX	0.0357	0.0057	0.1720	<.0001
0.1	0.1	KNOX	0.0260	0.0071	DP vs MGR	-0.0940	0.0000	0.5100	<.0001
0.1	0.1	MGR	0.1556	0.0162	KNOX vs MGR	-0.1296	0.0000	0.6540	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0266	0.0006	0.9040	<.0001
0.1	0.01	MGR	0.0286	0.0074	KNOX vs MGR	-0.0266	0.0006	0.9040	<.0001
0.01	0.2	DP	0.1330	0.0152	DP vs KNOX	0.1031	0.0000	0.3460	<.0001
0.01	0.2	KNOX	0.0299	0.0076	DP vs MGR	0.0040	0.8500	0.3360	<.0001
0.01	0.2	MGR	0.1290	0.0150	KNOX vs MGR	-0.0990	0.0000	0.6660	<.0001
0.01	0.1	DP	0.0229	0.0067	DP vs KNOX	0.0206	0.0033	0.3540	<.0001
0.01	0.1	KNOX	0.0023	0.0021	DP vs MGR	-0.0423	0.0010	0.6640	<.0001
0.01	0.1	MGR	0.0652	0.0110	KNOX vs MGR	-0.0629	0.0000	0.9580	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0092	0.0728	0.8260	<.0001
0.01	0.01	MGR	0.0112	0.0047	KNOX vs MGR	-0.0092	0.0728	0.8260	<.0001

Figure 3.10 **Results from 500 Simulations, n=100, λ=10**



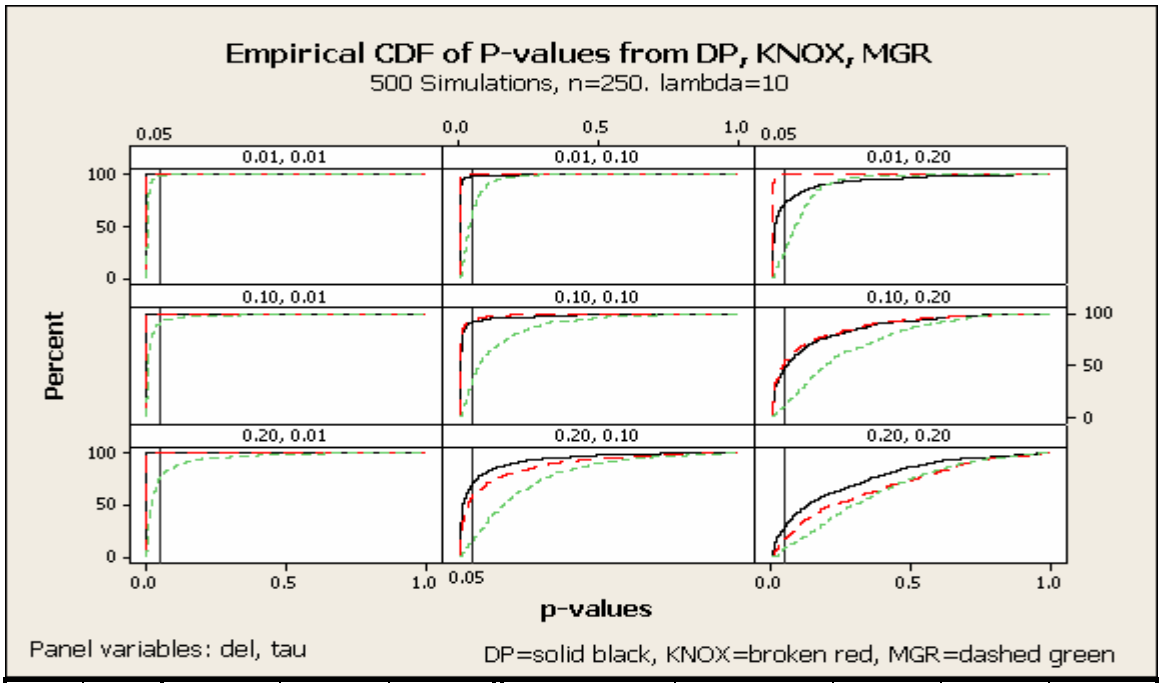
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	$p\text{-val}_{(5)}$ (6)	K-S ***D*** (7)	$p\text{-val}_{(7)}$ (8)
0.2	0.2	DP	0.1836	0.0173	DP vs KNOX	-0.0133	0.5933	0.0840	0.0587
0.2	0.2	KNOX	0.1969	0.0178	DP vs MGR	-0.0987	0.0002	0.2860	<.0001
0.2	0.2	MGR	0.2823	0.0201	KNOX vs MGR	-0.0854	0.0015	0.2380	<.0001
0.2	0.1	DP	0.0369	0.0084	DP vs KNOX	-0.0128	0.3183	0.1340	0.0003
0.2	0.1	KNOX	0.0498	0.0097	DP vs MGR	-0.1194	0.0000	0.6240	<.0001
0.2	0.1	MGR	0.1564	0.0162	KNOX vs MGR	-0.1066	0.0000	0.5560	<.0001
0.2	0.01	DP	0.0021	0.0020	DP vs KNOX	0.0001	0.9799	0.0020	1.0000
0.2	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0389	0.0000	0.9400	<.0001
0.2	0.01	MGR	0.0410	0.0089	KNOX vs MGR	-0.0390	0.0000	0.9420	<.0001
0.1	0.2	DP	0.1020	0.0135	DP vs KNOX	0.0330	0.0615	0.1040	0.0090
0.1	0.2	KNOX	0.0690	0.0113	DP vs MGR	-0.1039	0.0000	0.4820	<.0001
0.1	0.2	MGR	0.2059	0.0181	KNOX vs MGR	-0.1369	0.0000	0.5720	<.0001
0.1	0.1	DP	0.0167	0.0057	DP vs KNOX	0.0109	0.1022	0.0980	0.0164
0.1	0.1	KNOX	0.0058	0.0034	DP vs MGR	-0.0890	0.0000	0.7780	<.0001
0.1	0.1	MGR	0.1057	0.0137	KNOX vs MGR	-0.0999	0.0000	0.8660	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0201	0.0035	0.9260	<.0001
0.1	0.01	MGR	0.0221	0.0066	KNOX vs MGR	-0.0201	0.0035	0.9260	<.0001
0.01	0.2	DP	0.0411	0.0089	DP vs KNOX	0.0361	0.0001	0.3280	<.0001
0.01	0.2	KNOX	0.0049	0.0031	DP vs MGR	-0.0726	0.0000	0.6440	<.0001
0.01	0.2	MGR	0.1136	0.0142	KNOX vs MGR	-0.1087	0.0000	0.8940	<.0001
0.01	0.1	DP	0.0045	0.0030	DP vs KNOX	0.0024	0.4994	0.0940	0.0241
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0529	0.0000	0.8880	<.0001
0.01	0.1	MGR	0.0573	0.0104	KNOX vs MGR	-0.0553	0.0000	0.9760	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0149	0.0145	0.8500	<.0001
0.01	0.01	MGR	0.0169	0.0058	KNOX vs MGR	-0.0149	0.0145	0.8500	<.0001

Figure 3.11 **Results from 500 Simulations, n=250, λ=5**



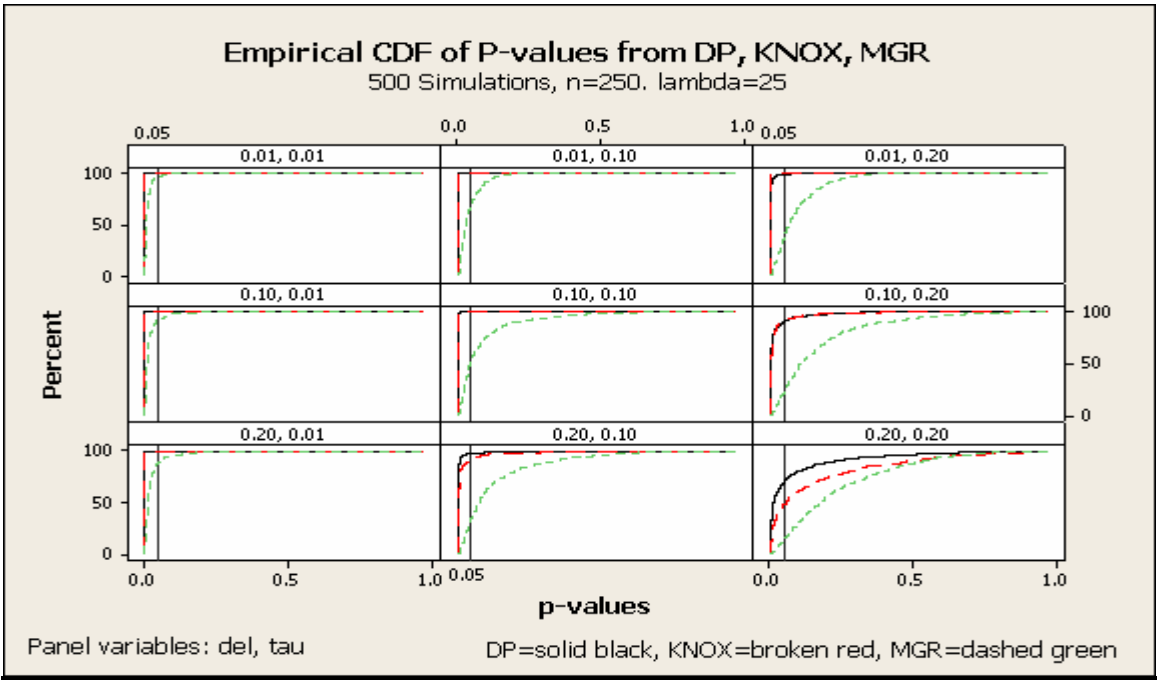
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.3184	0.0208	DP vs KNOX	-0.1183	0.0001	0.1860	<.0001
0.2	0.2	KNOX	0.4366	0.0222	DP vs MGR	-0.0801	0.0081	0.1480	<.0001
0.2	0.2	MGR	0.3985	0.0219	KNOX vs MGR	0.0382	0.2206	0.0940	0.0241
0.2	0.1	DP	0.2134	0.0183	DP vs KNOX	-0.0294	0.2675	0.0720	0.1497
0.2	0.1	KNOX	0.2428	0.0192	DP vs MGR	-0.0953	0.0006	0.2380	<.0001
0.2	0.1	MGR	0.3087	0.0207	KNOX vs MGR	-0.0659	0.0195	0.1800	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	-0.0001	0.9688	0.0080	1.0000
0.2	0.01	KNOX	0.0021	0.0021	DP vs MGR	-0.0700	0.0000	0.9640	<.0001
0.2	0.01	MGR	0.0720	0.0116	KNOX vs MGR	-0.0699	0.0000	0.9560	<.0001
0.1	0.2	DP	0.2793	0.0201	DP vs KNOX	0.0282	0.3118	0.0720	0.1497
0.1	0.2	KNOX	0.2510	0.0194	DP vs MGR	-0.0548	0.0599	0.1700	<.0001
0.1	0.2	MGR	0.3340	0.0211	KNOX vs MGR	-0.0830	0.0038	0.2100	<.0001
0.1	0.1	DP	0.1080	0.0139	DP vs KNOX	0.0389	0.0299	0.1660	<.0001
0.1	0.1	KNOX	0.0691	0.0113	DP vs MGR	-0.1032	0.0000	0.3580	<.0001
0.1	0.1	MGR	0.2111	0.0183	KNOX vs MGR	-0.1421	0.0000	0.5160	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0253	0.0008	0.8700	<.0001
0.1	0.01	MGR	0.0273	0.0073	KNOX vs MGR	-0.0253	0.0008	0.8700	<.0001
0.01	0.2	DP	0.1899	0.0175	DP vs KNOX	0.1709	0.0000	0.5940	<.0001
0.01	0.2	KNOX	0.0190	0.0061	DP vs MGR	0.0545	0.0192	0.1660	<.0001
0.01	0.2	MGR	0.1354	0.0153	KNOX vs MGR	-0.1164	0.0000	0.7400	<.0001
0.01	0.1	DP	0.0342	0.0081	DP vs KNOX	0.0321	0.0001	0.5480	<.0001
0.01	0.1	KNOX	0.0021	0.0020	DP vs MGR	-0.0263	0.0496	0.5420	<.0001
0.01	0.1	MGR	0.0605	0.0107	KNOX vs MGR	-0.0584	0.0000	0.9740	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0062	0.1685	0.7020	<.0001
0.01	0.01	MGR	0.0082	0.0040	KNOX vs MGR	-0.0062	0.1685	0.7020	<.0001

Figure 3.12 **Results from 500 Simulations, n=250, λ=10**



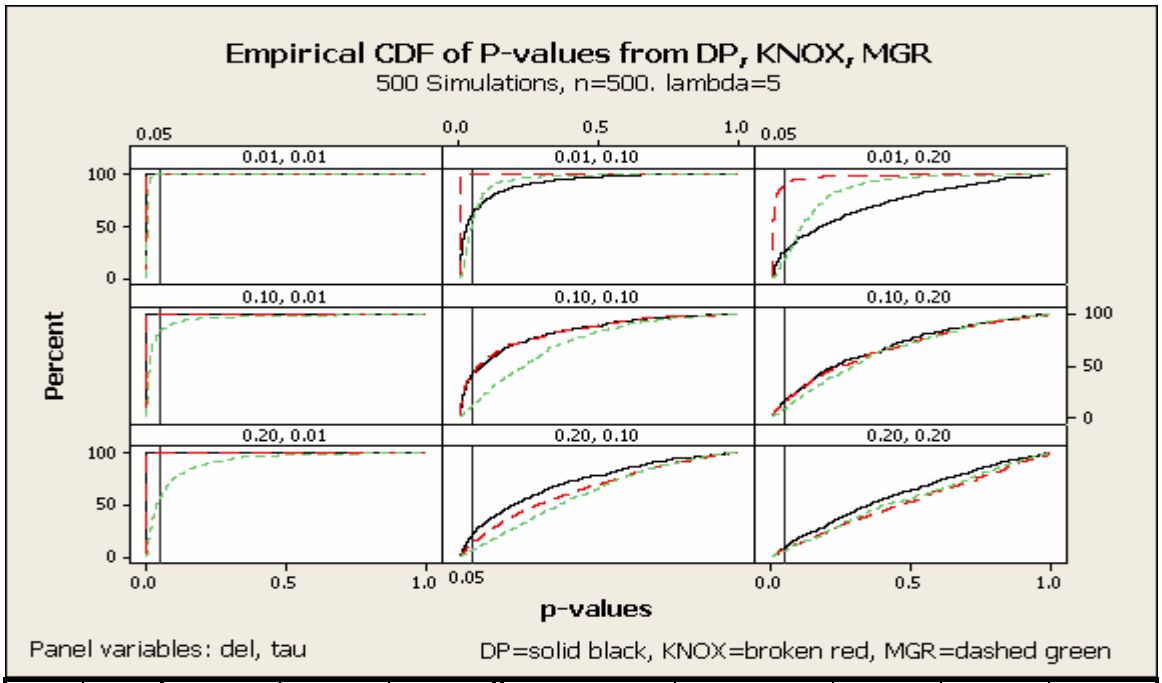
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.2199	0.0185	DP vs KNOX	-0.0845	0.0023	0.1500	<.0001
0.2	0.2	KNOX	0.3044	0.0206	DP vs MGR	-0.1203	0.0000	0.2900	<.0001
0.2	0.2	MGR	0.3402	0.0212	KNOX vs MGR	-0.0359	0.2248	0.1640	<.0001
0.2	0.1	DP	0.0633	0.0109	DP vs KNOX	-0.0470	0.0081	0.1900	<.0001
0.2	0.1	KNOX	0.1103	0.0140	DP vs MGR	-0.1535	0.0000	0.5560	<.0001
0.2	0.1	MGR	0.2168	0.0184	KNOX vs MGR	-0.1065	0.0000	0.4300	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.2	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0517	0.0000	0.9400	<.0001
0.2	0.01	MGR	0.0537	0.0101	KNOX vs MGR	-0.0517	0.0000	0.9400	<.0001
0.1	0.2	DP	0.1301	0.0150	DP vs KNOX	0.0120	0.5646	0.0740	0.1294
0.1	0.2	KNOX	0.1181	0.0144	DP vs MGR	-0.1224	0.0000	0.3800	<.0001
0.1	0.2	MGR	0.2524	0.0194	KNOX vs MGR	-0.1344	0.0000	0.4340	<.0001
0.1	0.1	DP	0.0215	0.0065	DP vs KNOX	0.0104	0.1933	0.0920	0.0290
0.1	0.1	KNOX	0.0111	0.0047	DP vs MGR	-0.1021	0.0000	0.7520	<.0001
0.1	0.1	MGR	0.1236	0.0147	KNOX vs MGR	-0.1125	0.0000	0.7860	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0219	0.0021	0.8140	<.0001
0.1	0.01	MGR	0.0239	0.0068	KNOX vs MGR	-0.0219	0.0021	0.8140	<.0001
0.01	0.2	DP	0.0646	0.0110	DP vs KNOX	0.0617	0.0000	0.5780	<.0001
0.01	0.2	KNOX	0.0029	0.0024	DP vs MGR	-0.0351	0.0428	0.4960	<.0001
0.01	0.2	MGR	0.0997	0.0134	KNOX vs MGR	-0.0968	0.0000	0.9340	<.0001
0.01	0.1	DP	0.0054	0.0033	DP vs KNOX	0.0034	0.3801	0.1060	0.0073
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0417	0.0000	0.8820	<.0001
0.01	0.1	MGR	0.0470	0.0095	KNOX vs MGR	-0.0450	0.0000	0.9780	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0058	0.1861	0.7040	<.0001
0.01	0.01	MGR	0.0078	0.0039	KNOX vs MGR	-0.0058	0.1861	0.7040	<.0001

Figure 3.13 **Results from 500 Simulations, n=250, λ=25**



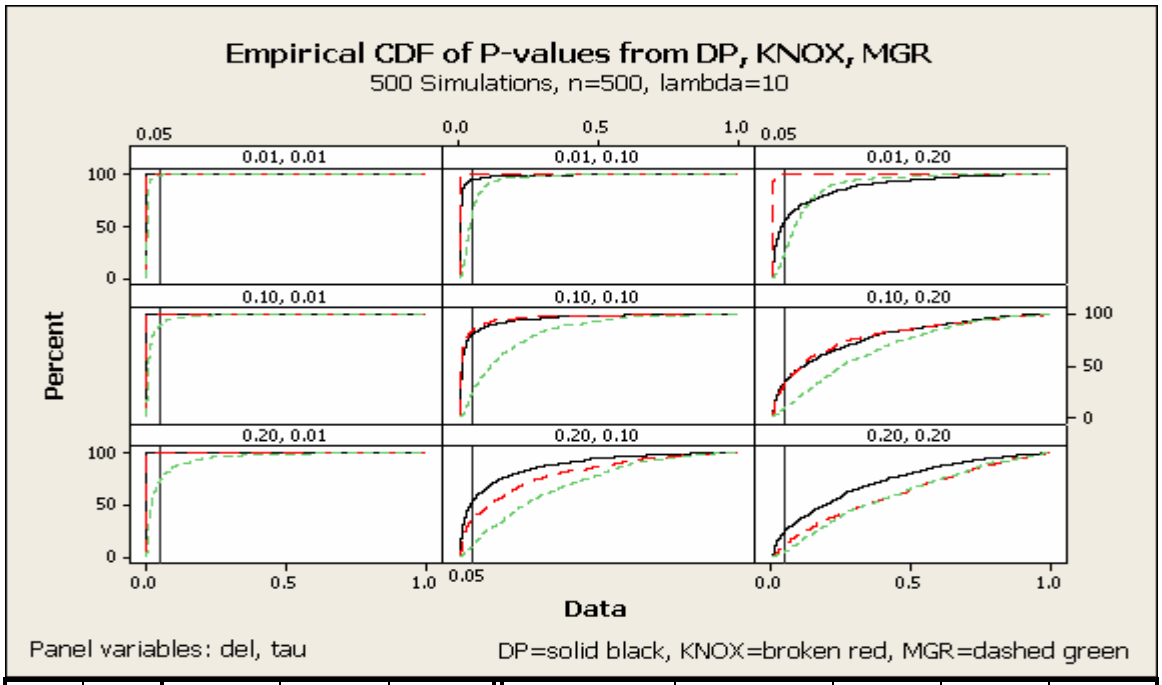
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.0699	0.0114	DP vs KNOX	-0.0784	0.0001	0.2340	<.0001
0.2	0.2	KNOX	0.1483	0.0159	DP vs MGR	-0.1601	0.0000	0.5620	<.0001
0.2	0.2	MGR	0.2301	0.0188	KNOX vs MGR	-0.0817	0.0009	0.3480	<.0001
0.2	0.1	DP	0.0057	0.0034	DP vs KNOX	-0.0092	0.1483	0.2340	<.0001
0.2	0.1	KNOX	0.0149	0.0054	DP vs MGR	-0.1182	0.0000	0.9040	<.0001
0.2	0.1	MGR	0.1239	0.0147	KNOX vs MGR	-0.1090	0.0000	0.7700	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.2	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0229	0.0016	0.9200	<.0001
0.2	0.01	MGR	0.0249	0.0070	KNOX vs MGR	-0.0229	0.0016	0.9200	<.0001
0.1	0.2	DP	0.0173	0.0058	DP vs KNOX	-0.0013	0.8793	0.0700	0.1725
0.1	0.2	KNOX	0.0186	0.0060	DP vs MGR	-0.1490	0.0000	0.7720	<.0001
0.1	0.2	MGR	0.1663	0.0167	KNOX vs MGR	-0.1477	0.0000	0.7660	<.0001
0.1	0.1	DP	0.0021	0.0020	DP vs KNOX	-0.0001	0.9694	0.0040	1.0000
0.1	0.1	KNOX	0.0022	0.0021	DP vs MGR	-0.0821	0.0000	0.9680	<.0001
0.1	0.1	MGR	0.0842	0.0124	KNOX vs MGR	-0.0820	0.0000	0.9680	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0171	0.0080	0.8800	<.0001
0.1	0.01	MGR	0.0191	0.0061	KNOX vs MGR	-0.0171	0.0080	0.8800	<.0001
0.01	0.2	DP	0.0046	0.0030	DP vs KNOX	0.0025	0.4872	0.1080	0.0059
0.01	0.2	KNOX	0.0021	0.0020	DP vs MGR	-0.0866	0.0000	0.8940	<.0001
0.01	0.2	MGR	0.0913	0.0129	KNOX vs MGR	-0.0892	0.0000	0.9900	<.0001
0.01	0.1	DP	0.0021	0.0020	DP vs KNOX	0.0001	0.9787	0.0040	1.0000
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0382	0.0000	0.9740	<.0001
0.01	0.1	MGR	0.0403	0.0088	KNOX vs MGR	-0.0383	0.0000	0.9780	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0101	0.0562	0.8260	<.0001
0.01	0.01	MGR	0.0121	0.0049	KNOX vs MGR	-0.0101	0.0562	0.8260	<.0001

Figure 3.14 **Results from 500 Simulations, n=500, λ=5**



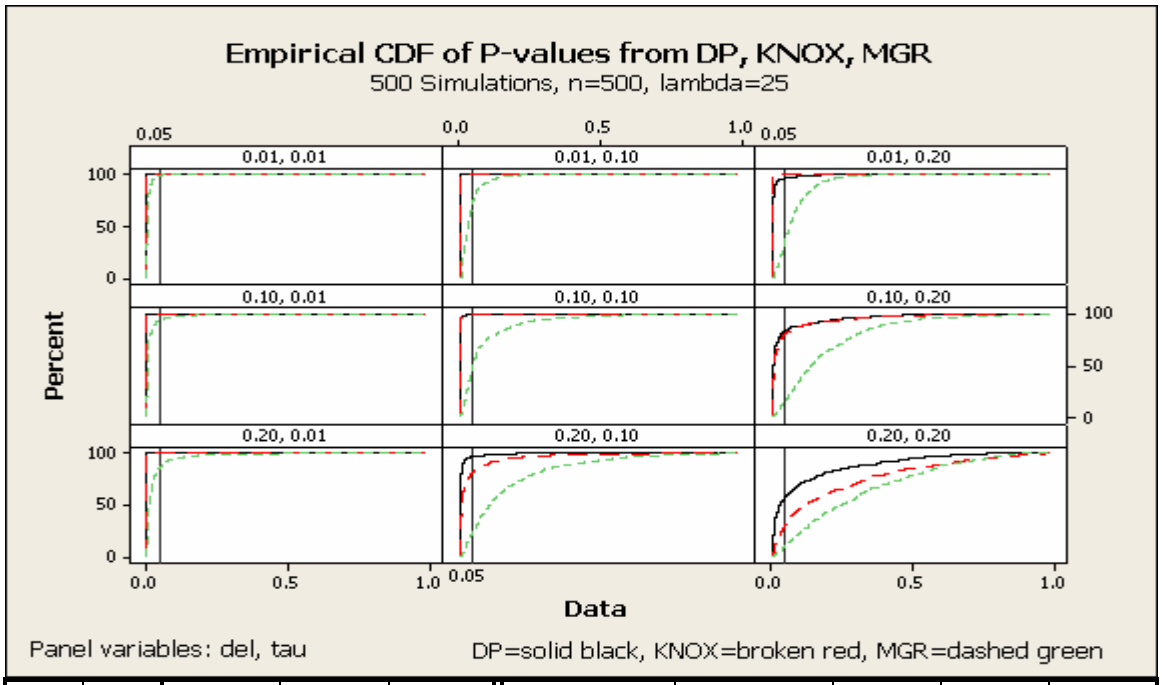
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.3957	0.0219	DP vs KNOX	-0.0728	0.0198	0.1200	0.0015
0.2	0.2	KNOX	0.4685	0.0223	DP vs MGR	-0.0557	0.0742	0.0920	0.0290
0.2	0.2	MGR	0.4514	0.0223	KNOX vs MGR	0.0171	0.5874	0.0580	0.3696
0.2	0.1	DP	0.2744	0.0200	DP vs KNOX	-0.0716	0.0141	0.1280	0.0006
0.2	0.1	KNOX	0.3460	0.0213	DP vs MGR	-0.1138	0.0001	0.2360	<.0001
0.2	0.1	MGR	0.3883	0.0218	KNOX vs MGR	-0.0423	0.1652	0.1260	0.0007
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	-0.0004	0.9039	0.0280	0.9895
0.2	0.01	KNOX	0.0024	0.0022	DP vs MGR	-0.0844	0.0000	0.9740	<.0001
0.2	0.01	MGR	0.0864	0.0126	KNOX vs MGR	-0.0840	0.0000	0.9460	<.0001
0.1	0.2	DP	0.3103	0.0207	DP vs KNOX	-0.0178	0.5467	0.0520	0.5085
0.1	0.2	KNOX	0.3281	0.0210	DP vs MGR	-0.0389	0.1900	0.1160	0.0024
0.1	0.2	MGR	0.3492	0.0213	KNOX vs MGR	-0.0212	0.4793	0.1060	0.0073
0.1	0.1	DP	0.1691	0.0168	DP vs KNOX	0.0050	0.8325	0.0540	0.4595
0.1	0.1	KNOX	0.1641	0.0166	DP vs MGR	-0.1092	0.0000	0.3300	<.0001
0.1	0.1	MGR	0.2783	0.0200	KNOX vs MGR	-0.1142	0.0000	0.3540	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0390	0.0000	0.9020	<.0001
0.1	0.01	MGR	0.0410	0.0089	KNOX vs MGR	-0.0390	0.0000	0.9020	<.0001
0.01	0.2	DP	0.2707	0.0199	DP vs KNOX	0.2500	0.0000	0.6620	<.0001
0.01	0.2	KNOX	0.0207	0.0064	DP vs MGR	0.1241	0.0000	0.2740	<.0001
0.01	0.2	MGR	0.1467	0.0158	KNOX vs MGR	-0.1260	0.0000	0.7640	<.0001
0.01	0.1	DP	0.0782	0.0120	DP vs KNOX	0.0762	0.0000	0.7760	<.0001
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	0.0164	0.3100	0.2800	<.0001
0.01	0.1	MGR	0.0618	0.0108	KNOX vs MGR	-0.0598	0.0000	0.9820	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0038	0.3368	0.6300	<.0001
0.01	0.01	MGR	0.0058	0.0034	KNOX vs MGR	-0.0038	0.3368	0.6300	<.0001

Figure 3.15 **Results from 500 Simulations, n=500, λ=10**



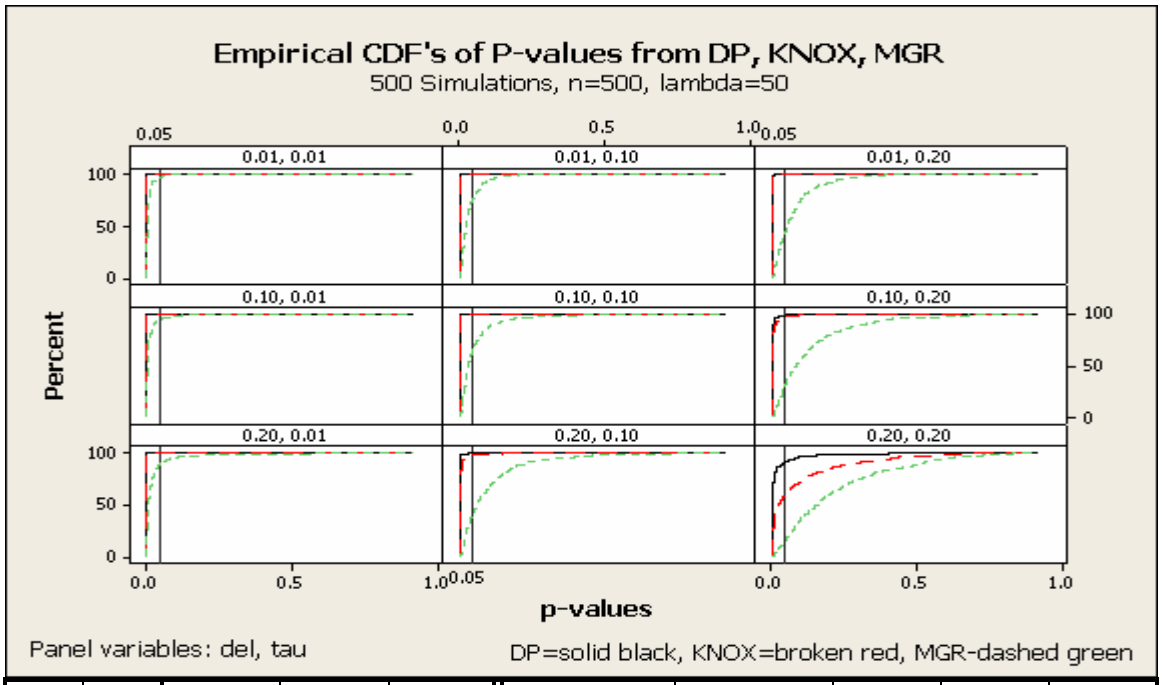
δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.2642	0.0197	DP vs KNOX	-0.1170	0.0001	0.1960	<.0001
0.2	0.2	KNOX	0.3812	0.0217	DP vs MGR	-0.1285	0.0000	0.2260	<.0001
0.2	0.2	MGR	0.3927	0.0218	KNOX vs MGR	-0.0115	0.7098	0.0920	0.0290
0.2	0.1	DP	0.1159	0.0143	DP vs KNOX	-0.0776	0.0006	0.1900	<.0001
0.2	0.1	KNOX	0.1935	0.0177	DP vs MGR	-0.1751	0.0000	0.4480	<.0001
0.2	0.1	MGR	0.2910	0.0203	KNOX vs MGR	-0.0975	0.0003	0.2600	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.2	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0514	0.0000	0.9620	<.0001
0.2	0.01	MGR	0.0534	0.0101	KNOX vs MGR	-0.0514	0.0000	0.9620	<.0001
0.1	0.2	DP	0.2108	0.0182	DP vs KNOX	0.0041	0.8738	0.0520	0.5085
0.1	0.2	KNOX	0.2067	0.0181	DP vs MGR	-0.1039	0.0002	0.2880	<.0001
0.1	0.2	MGR	0.3148	0.0208	KNOX vs MGR	-0.1080	0.0001	0.3040	<.0001
0.1	0.1	DP	0.0428	0.0091	DP vs KNOX	0.0104	0.3888	0.0920	0.0290
0.1	0.1	KNOX	0.0324	0.0079	DP vs MGR	-0.1258	0.0000	0.6240	<.0001
0.1	0.1	MGR	0.1686	0.0167	KNOX vs MGR	-0.1362	0.0000	0.6860	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0208	0.0029	0.8460	<.0001
0.1	0.01	MGR	0.0228	0.0067	KNOX vs MGR	-0.0208	0.0029	0.8460	<.0001
0.01	0.2	DP	0.1117	0.0141	DP vs KNOX	0.1093	0.0000	0.7720	<.0001
0.01	0.2	KNOX	0.0024	0.0022	DP vs MGR	-0.0001	0.9947	0.3380	<.0001
0.01	0.2	MGR	0.1118	0.0141	KNOX vs MGR	-0.1094	0.0000	0.9620	<.0001
0.01	0.1	DP	0.0127	0.0050	DP vs KNOX	0.0107	0.0475	0.3020	<.0001
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0371	0.0007	0.7560	<.0001
0.01	0.1	MGR	0.0498	0.0097	KNOX vs MGR	-0.0478	0.0000	0.9800	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0052	0.2266	0.6260	<.0001
0.01	0.01	MGR	0.0072	0.0038	KNOX vs MGR	-0.0052	0.2266	0.6260	<.0001

Figure 3.16 **Results from 500 Simulations, n=500, λ=25**



δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.1079	0.0139	DP vs KNOX	-0.1127	0.0000	0.2860	<.0001
0.2	0.2	KNOX	0.2206	0.0185	DP vs MGR	-0.1907	0.0000	0.4840	<.0001
0.2	0.2	MGR	0.2986	0.0205	KNOX vs MGR	-0.0780	0.0047	0.2540	<.0001
0.2	0.1	DP	0.0091	0.0043	DP vs KNOX	-0.0273	0.0037	0.4060	<.0001
0.2	0.1	KNOX	0.0364	0.0084	DP vs MGR	-0.1576	0.0000	0.8720	<.0001
0.2	0.1	MGR	0.1668	0.0167	KNOX vs MGR	-0.1303	0.0000	0.6080	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.2	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0281	0.0004	0.9060	<.0001
0.2	0.01	MGR	0.0301	0.0076	KNOX vs MGR	-0.0281	0.0004	0.9060	<.0001
0.1	0.2	DP	0.0377	0.0085	DP vs KNOX	-0.0072	0.5667	0.1460	<.0001
0.1	0.2	KNOX	0.0449	0.0093	DP vs MGR	-0.1504	0.0000	0.7020	<.0001
0.1	0.2	MGR	0.1881	0.0175	KNOX vs MGR	-0.1432	0.0000	0.6620	<.0001
0.1	0.1	DP	0.0025	0.0023	DP vs KNOX	0.0001	0.9839	0.0060	1.0000
0.1	0.1	KNOX	0.0025	0.0022	DP vs MGR	-0.0819	0.0000	0.9440	<.0001
0.1	0.1	MGR	0.0845	0.0124	KNOX vs MGR	-0.0820	0.0000	0.9420	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0134	0.0222	0.8020	<.0001
0.1	0.01	MGR	0.0154	0.0055	KNOX vs MGR	-0.0134	0.0222	0.8020	<.0001
0.01	0.2	DP	0.0085	0.0041	DP vs KNOX	0.0065	0.1556	0.2100	<.0001
0.01	0.2	KNOX	0.0020	0.0020	DP vs MGR	-0.0710	0.0000	0.8580	<.0001
0.01	0.2	MGR	0.0795	0.0121	KNOX vs MGR	-0.0775	0.0000	0.9940	<.0001
0.01	0.1	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0368	0.0000	0.9900	<.0001
0.01	0.1	MGR	0.0388	0.0086	KNOX vs MGR	-0.0368	0.0000	0.9900	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0060	0.1790	0.6700	<.0001
0.01	0.01	MGR	0.0080	0.0040	KNOX vs MGR	-0.0060	0.1790	0.6700	<.0001

Figure 3.17 **Results from 500 Simulations, n=500, λ=50**



δ	τ	Statistic (1)	ESL (2)	$sd_{(2)}$ (3)	$Stat_A$ vs $Stat_B$ (4)	$ESL_A - ESL_B$ (5)	p-$val_{(5)}$ (6)	K-S ***D*** (7)	p-$val_{(7)}$ (8)
0.2	0.2	DP	0.0170	0.0058	DP vs KNOX	-0.0744	0.0000	0.4400	<.0001
0.2	0.2	KNOX	0.0914	0.0129	DP vs MGR	-0.1980	0.0000	0.8040	<.0001
0.2	0.2	MGR	0.2150	0.0184	KNOX vs MGR	-0.1237	0.0000	0.4480	<.0001
0.2	0.1	DP	0.0027	0.0023	DP vs KNOX	-0.0049	0.2751	0.1960	<.0001
0.2	0.1	KNOX	0.0076	0.0039	DP vs MGR	-0.0990	0.0000	0.9720	<.0001
0.2	0.1	MGR	0.1017	0.0135	KNOX vs MGR	-0.0941	0.0000	0.8520	<.0001
0.2	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.2	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0258	0.0007	0.8900	<.0001
0.2	0.01	MGR	0.0278	0.0074	KNOX vs MGR	-0.0258	0.0007	0.8900	<.0001
0.1	0.2	DP	0.0040	0.0028	DP vs KNOX	-0.0029	0.5322	0.1360	0.0002
0.1	0.2	KNOX	0.0069	0.0037	DP vs MGR	-0.1272	0.0000	0.9200	<.0001
0.1	0.2	MGR	0.1312	0.0151	KNOX vs MGR	-0.1243	0.0000	0.8480	<.0001
0.1	0.1	DP	0.0020	0.0020	DP vs KNOX	0.0000	0.9989	0.0020	1.0000
0.1	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0523	0.0000	0.9780	<.0001
0.1	0.1	MGR	0.0543	0.0101	KNOX vs MGR	-0.0523	0.0000	0.9780	<.0001
0.1	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.1	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0125	0.0283	0.8360	<.0001
0.1	0.01	MGR	0.0145	0.0053	KNOX vs MGR	-0.0125	0.0283	0.8360	<.0001
0.01	0.2	DP	0.0022	0.0021	DP vs KNOX	0.0002	0.9364	0.0260	0.9959
0.01	0.2	KNOX	0.0020	0.0020	DP vs MGR	-0.0748	0.0000	0.9700	<.0001
0.01	0.2	MGR	0.0770	0.0119	KNOX vs MGR	-0.0750	0.0000	0.9960	<.0001
0.01	0.1	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.1	KNOX	0.0020	0.0020	DP vs MGR	-0.0327	0.0001	0.9740	<.0001
0.01	0.1	MGR	0.0347	0.0082	KNOX vs MGR	-0.0327	0.0001	0.9740	<.0001
0.01	0.01	DP	0.0020	0.0020	DP vs KNOX	0.0000	1.0000	0.0000	1.0000
0.01	0.01	KNOX	0.0020	0.0020	DP vs MGR	-0.0086	0.0845	0.7900	<.0001
0.01	0.01	MGR	0.0106	0.0046	KNOX vs MGR	-0.0086	0.0845	0.7900	<.0001

3.6.1 Results From Simulation Studies on Clustered Data.

An examination of the CDF plots reveals that, in every instance, Mantel's Generalized Regression (MGR) did not perform as well as Knox's Test or DP. Also, in most cases, KNOX performed at least slightly better than DP; but in a few cases DP was better. Upon closer examination, DP was better when the clustering in the data was weaker (δ or $\tau = 0.20$), and KNOX performed better when the clustering in the data was stronger (δ or $\tau = 0.01$). However, when the clustering in the data was very strong (both δ and $\tau = 0.01$), DP and KNOX performed equivalently, because they both identified every data set in the simulation run as highly significant. These findings are discussed in more detail in the sections that follow.

3.6.2 Comparisons Against Mantel's Generalized Regression.

The ESL for MGR is numerically larger than the ESL for KNOX and DP in every instance. This difference is nearly always significant (α=0.05), with KNOX and DP always having the smaller ESL. Also, Kolmogorov-Smirnov test results were nearly always highly significant (α<0.01) for the CDF comparisons against MGR. In general, MGR simply did not perform as well as DP or KNOX, in detecting the separate-source clustering examined in this study.

3.6.3 Comparisons Between DP And Knox's Test.

The comparisons between DP and Knox's Test are very interesting, because these tests performed quite similarly in this study. To facilitate examination of the numerical results, the p-values from comparisons between KNOX and DP have been extracted from the previous tables and put into a single table. When these p-values are less than 0.05, it is indicated in this table with an asterisk (*) for K-S test results, and with '**DP**' or '**KX**' for the ESL comparisons, to denote which test is superior.

Table 3.2 **All DP versus KNOX Comparison Results on Clustered Data**

n	λ	δ	τ	ESL p-val	K-S *D* p-val	n	λ	δ	τ	ESL p-val	K-S *D* p-val
25	5	0.2	0.2	0.6105	0.5596	250	10	0.2	0.2	DP 0.0023	* <.0001
25	5	0.2	0.1	0.5023	0.7699	250	10	0.2	0.1	DP 0.0081	* <.0001
25	5	0.2	0.01	0.3715	0.0815	250	10	0.2	0.01	1.0000	1.0000
25	5	0.1	0.2	0.1211	0.1114	250	10	0.1	0.2	0.5646	0.1294
25	5	0.1	0.1	0.1071	* 0.0038	250	10	0.1	0.1	0.1933	* 0.0290
25	5	0.1	0.01	0.7441	1.0000	250	10	0.1	0.01	1.0000	1.0000
25	5	0.01	0.2	0.0551	* 0.0004	250	10	0.01	0.2	KX <.0001	* <.0001
25	5	0.01	0.1	0.3963	* 0.0495	250	10	0.01	0.1	0.3801	* 0.0073
25	5	0.01	0.01	0.3332	0.7699	250	10	0.01	0.01	1.0000	1.0000
50	5	0.2	0.2	0.6630	0.8632	250	25	0.2	0.2	DP 0.0001	* <.0001
50	5	0.2	0.1	0.5582	0.5596	250	25	0.2	0.1	0.1483	* <.0001
50	5	0.2	0.01	0.7693	1.0000	250	25	0.2	0.01	1.0000	1.0000
50	5	0.1	0.2	0.2577	0.1979	250	25	0.1	0.2	0.8793	0.1725
50	5	0.1	0.1	0.0047	* <.0001	250	25	0.1	0.1	0.9694	1.0000
50	5	0.1	0.01	0.9966	1.0000	250	25	0.1	0.01	1.0000	1.0000
50	5	0.01	0.2	KX 0.0007	* <.0001	250	25	0.01	0.2	0.4872	* 0.0059
50	5	0.01	0.1	KX 0.0124	* <.0001	250	25	0.01	0.1	0.9787	1.0000
50	5	0.01	0.01	0.9989	1.0000	250	25	0.01	0.01	1.0000	1.0000
50	10	0.2	0.2	0.7667	0.1497	500	5	0.2	0.2	DP 0.0198	* 0.0015
50	10	0.2	0.1	0.6061	0.9600	500	5	0.2	0.1	DP 0.0141	* 0.0006
50	10	0.2	0.01	0.7590	1.0000	500	5	0.2	0.01	0.9039	0.9895
50	10	0.1	0.2	0.2297	* 0.0348	500	5	0.1	0.2	0.5467	0.5085
50	10	0.1	0.1	0.1676	0.1114	500	5	0.1	0.1	0.8325	0.4595
50	10	0.1	0.01	0.9944	1.0000	500	5	0.1	0.01	1.0000	1.0000
50	10	0.01	0.2	KX 0.0037	* <.0001	500	5	0.01	0.2	KX <.0001	* <.0001
50	10	0.01	0.1	0.4574	0.0587	500	5	0.01	0.1	KX <.0001	* <.0001
50	10	0.01	0.01	1.0000	1.0000	500	5	0.01	0.01	1.0000	1.0000
100	5	0.2	0.2	DP 0.0392	* 0.0009	500	10	0.2	0.2	DP 0.0001	* <.0001
100	5	0.2	0.1	0.9710	0.9600	500	10	0.2	0.1	DP 0.0006	* <.0001
100	5	0.2	0.01	0.9899	1.0000	500	10	0.2	0.01	1.0000	1.0000
100	5	0.1	0.2	KX 0.0033	* <.0001	500	10	0.1	0.2	0.8738	0.5085
100	5	0.1	0.1	KX 0.0057	* <.0001	500	10	0.1	0.1	0.3888	* 0.0290
100	5	0.1	0.01	1.0000	1.0000	500	10	0.1	0.01	1.0000	1.0000
100	5	0.01	0.2	KX <.0001	* <.0001	500	10	0.01	0.2	KX <.0001	* <.0001
100	5	0.01	0.1	KX 0.0033	* <.0001	500	10	0.01	0.1	KX 0.0475	* <.0001
100	5	0.01	0.01	1.0000	1.0000	500	10	0.01	0.01	1.0000	1.0000
100	10	0.2	0.2	0.5933	0.0587	500	25	0.2	0.2	DP <.0001	* <.0001
100	10	0.2	0.1	0.3183	* 0.0003	500	25	0.2	0.1	DP 0.0037	* <.0001
100	10	0.2	0.01	0.9799	1.0000	500	25	0.2	0.01	1.0000	1.0000
100	10	0.1	0.2	0.0615	* 0.0090	500	25	0.1	0.2	0.5667	* <.0001
100	10	0.1	0.1	0.1022	* 0.0164	500	25	0.1	0.1	0.9839	1.0000
100	10	0.1	0.01	1.0000	1.0000	500	25	0.1	0.01	1.0000	1.0000
100	10	0.01	0.2	KX 0.0001	* <.0001	500	25	0.01	0.2	0.1556	* <.0001
100	10	0.01	0.1	0.4994	* 0.0241	500	25	0.01	0.1	1.0000	1.0000
100	10	0.01	0.01	1.0000	1.0000	500	25	0.01	0.01	1.0000	1.0000
250	5	0.2	0.2	DP 0.0001	* <.0001	500	50	0.2	0.2	DP 0.0000	* <.0001
250	5	0.2	0.1	0.2675	0.1497	500	50	0.2	0.1	0.2751	* <.0001
250	5	0.2	0.01	0.9688	1.0000	500	50	0.2	0.01	1.0000	1.0000
250	5	0.1	0.2	0.3118	0.1497	500	50	0.1	0.2	0.5322	* 0.0002
250	5	0.1	0.1	KX 0.0299	* <.0001	500	50	0.1	0.1	0.9989	1.0000
250	5	0.1	0.01	1.0000	1.0000	500	50	0.1	0.01	1.0000	1.0000
250	5	0.01	0.2	KX <.0001	* <.0001	500	50	0.01	0.2	0.9364	0.9959
250	5	0.01	0.1	KX 0.0001	* <.0001	500	50	0.01	0.1	1.0000	1.0000
250	5	0.01	0.01	1.0000	1.0000	500	50	0.01	0.01	1.0000	1.0000

The CDF plots show that DP and KNOX are very close, even though KNOX is usually slightly better than DP. The numerical results bear this out. The K-S test results are often very non-significant for these comparisons. When these comparisons do produce significant K-S test results, some patterns can be seen. For smaller sample sizes, the CDFs tended to be significantly different when clustering was strong. As the sample sizes increased, the CDFs tended to be significantly different more often when clustering was weaker.

There were fewer significant ESL differences than were seen in the CDF comparisons, because the K-S test is sensitive to any kind of difference between two empirical distributions; however, the CDF comparison was always significant whenever the ESL comparison was significant. In the 108 comparisons that were conducted, the differences were significant 28 times. In 16 out of the 28 significant ESL comparisons, KNOX was superior, the other 12 times DP was superior. KNOX tended to be superior when clustering was moderate to strong; DP was superior when clustering was weaker. As $\mathbf{n}$ and λ increased, both tests detected strong clustering consistently, with a p-value of 0.002 all of the time. To summarize the ESL results, in the 108 DP-KNOX comparisons on simulated clustered data, 80 comparisons (74.1%) indicated that neither test was significantly different, 16 comparisons (14.8%) indicated that KNOX was significantly different, and 12 comparisons (11.1%) indicated that DP was significantly different.

3.7 Discussion.

Summarizing the results of these studies, DP and Knox's Test performed quite similarly, while Mantel's Generalized Regression did not perform as well. This was not expected, but it seems to be tacitly assumed in the literature. There were surprisingly few instances of Mantel's Generalized Regression being used to analyze data in the literature. In addition to the fact that Knox's test is more popular, perhaps practitioners in this area

are aware that other tests perform better, and have opted instead to use these other tests. This fact does lend some credence to our result.

Knox's test was the first test for space-time interaction, and is the best known test of this sort. It is quite popular because of its simplicity and effectiveness. It is easy to apply and use, and understanding and explaining Knox's test is quite intuitive (Kulldorff and Hjalmars, 1999). In part because of its popularity, the issues with the use of Knox's test are well documented. Like DP, Knox's test is a general, or *omnibus*, space-time interaction test. These omnibus tests are intended to detect any difference between the distribution of cases in space and the distribution of cases in time, regardless of the cause of that difference. It is known that Knox's test may be biased if the rate of population growth is not constant for all geographic sub-areas (population shift bias). Adjusting for this bias can be done either by applying the test within stable population time strata and combining the results across strata (Klauber and Mustacchi, 1970), or by generating random replications of the cases under the null hypothesis (Kulldorff and Hjalmars, 1999). However, the first method sacrifices the power of the test, while the second method requires prior knowledge of the underlying population distribution (Kulldorff and Hjalmars, 1999).

Neither method addresses one basic issue with Knox's test - the choice of a critical distance and time. In practice, these are determined by knowledge of the etiology of the disease under study. However, this is not always known with certainty. In fact, one method of applying Knox's test allows the results of several applications of the test to be combined to produce a single assessment of clustering in a "meta-analytical" fashion (Kulldorff and Hjalmars, 1999). It is even suggested in the literature that the practitioner, uncertain about the proper choice of critical distance and time, should conduct the test repeatedly, and combine the results to get a single assessment of clustering (Kulldorff and Hjalmars, 1999).

In this context, DP appears to be a viable alternative to Knox's test. Under the conditions when it is appropriate to use an omnibus test for space-time clustering, and for the kind of clustering examined in this study, the performance of DP is quite competitive to that of Knox's test, and it does not require the determination of a critical distance or time. In fact, because DP is superior to Knox's test when clustering is weak or mild, it may be a better choice for a clustering test when early detection of clustering is desired.

Note also that DP is competitive in this study, using simulated data with the critical parameters for Knox's test determined in a nearly optimal manner. In practice, this may be not so, for these parameters are determined by knowledge of the etiology of the disease under study. This knowledge is often spotty, weak, or faulty. Even when the knowledge of the disease etiology is good and strong, the critical parameters selected may still not be optimal (in a statistical sense) for identifying clustering. It is likely that Knox's test, as conducted in practice, is much less powerful than the Knox test as conducted in this study. Because of this, DP may be just as powerful as the Knox test conducted in practice is, if not more so.

Also note that while this study has mentioned the challenges surrounding the selection of critical parameters, it is clear that this is a necessary challenge inherent in the nature of Knox's test. If the practitioner knows that the power of Knox's test may be optimized by the judicious choice of the critical parameters, that is, by basing the selection of the critical parameters on the expected number of close pairs under the null hypothesis, it would be inappropriate to use this knowledge to optimize the performance of the test. In actuality, the best option is to select a different test that does not depend as heavily on the selection of critical parameters. We propose DP as that choice.

3.8 Conclusions.

The simulation studies indicated that Mantel's Generalized Regression did not perform as well as DP or Knox's test, and that DP and Knox's test are comparable in

performance against clustering of the type evaluated here (separate-source clustering). In the next chapter, these statistics are used to analyze fifteen years of Burkitt's lymphoma data from Uganda

Chapter IV

RESULTS OF ANALYSIS OF SPACE-TIME POINT DATA

4.0 Introduction

In order to compare and understand the performance of DP, Knox's test, and Mantel's Generalized Regression on actual data, these statistics are used to analyze two data sets. The first is fifteen years of Burkitt's Lymphoma (BL) data from Uganda. The second is a set of cardiac defects among newborn infants in Santa Clara County, CA for the three-year period 1981-1983. These analysis results are compared against the results of other analyses of these data sets in the literature.

4.1 Initial Cluster Analyses in the Uganda Region.

An early space-time clustering analysis of data from this region was conducted by Pike et al. (1967). They used both Knox's test and Barton and David's test to assess clustering, employing a range of critical distances and times for Knox's test, and using the average time interval between cases as the critical time parameter for Barton and David's test. Highly significant space-time clustering was found in cases during the period 1961-1965, for both Knox's test (over a wide range of the critical parameters used) and for Barton and David's test.

Cases from this region in the period 1966-1967 were assessed for space-time clustering in space and time by Williams, Spit and Pike (1969). Using Knox's test, they again found highly significant space-time clustering. In 1970 an important outbreak of BL was seen in Bwamba county (in the Toro District of southeastern Uganda) that would have demonstrated significant space-time clustering (Williams, Day and Geser, 1974).

Also, cases of new patients with BL in the West Nile District of Uganda demonstrated significant seasonal variation consistent with space-time clustering during the period 1966-1973 (Morrow et al. 1970).

These results all supported the conjecture that cases of BL cluster in space and time. However, some contradictory findings have also been reported. Morrow et al. (1976) examined data from the Mengo Districts of Uganda during the period 1959-1968. They were unable to find significant space-time clustering using both Knox's test and Barton and David's test on these data. Similarly, Brubaker, Geser and Pike (1973) were not able to find significant space-time clustering in the North Mara District of Tanzania in data from the period 1964-1970. Also, Smith (1974) reported an apparent reduction of the strength of space-time clustering in reported cases of BL in the West Nile district of Uganda. Because space-time clustering is believed to be related to the etiology of BL disease, it was troubling to find these inconsistent and somewhat contradictory results.

This was one impetus for the study conducted by Williams et al. (1978), in an attempt to confirm whether the space-time clustering detected in the West Nile district of Uganda represents a real biological phenomenon, or is simply spurious significance due to random variation in the data or some artifact of case selection. Data were collected on all 202 reported cases of BL from 1961 to 1975 in the West Nile District of Uganda. In this study, Knox's test was used to assess space-time clustering within the three 5-year groups spanned by the data (1961-1965, 1966-1970, and 1971-1975), and in the period 1972-1973. Five distances (2.5, 5.0, 10.0, 20.0 and 40.0 km), and six times (30, 60, 90, 120, 180 and 360 days) were used as critical parameters for Knox's test, so Knox's test was performed 30 times on data from each quinquennial group, and 30 times on data from the period 1972-1973. Significance was found 19 times in the 1961-1965 data, 2 times in the 1966-1970 data, and 3 times in the 1971-1975 data. However, significance was found 13 times in the 1972-1973 data, primarily using the larger space and time critical parameters.

Siemiatycki, Brubaker and Geser (1980) analyzed 40 BL cases in North Mara, Tanzania, and reanalyzed a subset of the West Nile data (cases from 1961 to 1965) using both Knox's test and Mantel's Generalized Regression. Their analyses with Mantel's Generalized Regression confirmed the detection of space-time clustering in the 1961 to 1965 data, but found weak to no clustering in the latter time periods.

4.2 Description of the Burkitt's Lymphoma Data.

Assessing the presence and degree of space-time clustering has been historically important in understanding the etiology of Burkitt's Lymphoma. To accomplish this, epidemiologic data were collected on all 202 reported cases of BL from 1961 to 1975 in the West Nile District of Uganda (see Williams et al. 1978 for additional details). These data included for each case the date of onset of the disease, and the exact location of the cases as determined by a 1 kilometer grid superimposed on the map of this district in Uganda. This yielded spatial units of 1 kilometer and temporal units of 1 day, in these analyses. Because the location information was not available for nine of the reported cases, and the exact date of onset was not available for five additional cases, only 188 of these cases were available for analysis.

4.3 Cluster Analyses of the Burkitt's Lymphoma Data.

These data were previously analyzed by Williams et al. (1978), in the three 5-year groups 1961-1965, 1966-1970, and 1971-1975, and in the period 1972-1973. We replicated those analyses, and assessed clustering in the full data set, using Knox's test, Mantel's test, and DP. Knox's test was conducted setting K=5, as was done in the simulation studies, to produce a test that is close to optimal from a statistical sense, in terms of identifying clustering. Because ties exist in these data, spatial and temporal offset constants must be set, to preclude taking the reciprocal of 0. It is known that the power of Mantel's test is related to the size of these offset constants (Glass et al, 1971 and

Siemiatycki, 1971), but only general guidelines exist for selecting these constants. Mantel (1967) indicated that these constants should be "arbitrarily small", and suggested that the constants be close to the expected distance between pairs. Robertson (1979) observed an important impact on the power of Mantel's test when the offset constants were larger than the expected distance between pairs. The expected spatial and temporal distance between pairs in these data is larger than 1 kilometer or 1 day, so Mantel's test was conducted using spatial and temporal offsets of 0.1.

P-values were also determined as was done in the simulation studies, using the estimated empirical distributions of these statistics. Because estimates of the empirical distribution are used, the observed individual p-values for actual data may be somewhat variable for estimated empirical distributions with relatively few simulations. To address this, the analyses were conducted using 10,000 simulations to estimate the empirical distributions of each of the three statistics. Also, standard errors are reported to give some sense of the variability in these observed p-values.

Table 4.1

Analysis Results For Burkitt's Lymphoma Data

Data Group	# Cases	DP (SE)	Knox (SE)	MGR (SE)
1961-75 data	35	0.0103 (0.001010)	0.0031 (0.000556)	0.0323 (0.001768)
1966-70 data	72	0.6796 (0.004666)	0.2416 (0.004281)	0.8651 (0.003416)
1971-75 data	81	0.2955 (0.004563)	0.1388 (0.003457)	0.0023 (0.000479)
1972-73 data	37	0.1702 (0.003758)	0.0146 (0.001199)	0.0005 (0.000224)
All BL data	188	0.0627 (0.002424)	0.2074 (0.004054)	0.0023 (0.000479)

All three tests found strong clustering in the first 5-year period, and neither test found significant clustering in the second 5-year period (although the Knox's test p-value was relatively low (p=0.2416). DP was not significant for any of the remaining time periods, or for the overall analysis of the data. Mantel's test was highly significant for all

of the remaining time periods, and for the overall analysis of the data, while Knox's test was significant only for the 1972-73 period. In fact, Mantel's test seems to be particularly sensitive to the type of clustering present in these data, as it was highly significant for every analysis except for the 1966-70 period.

4.4 Description and Prior Analyses of the Cardiac Defects Data.

In 1981, the drinking water supplied to part of Santa Clara County, California was found to be contaminated with chemical solvents. It was felt that this contamination was caused by leakage from solvent storage tanks that are associated with the prevailing electronics manufacturing industry. An increase in the number of cases of infants born with cardiac defects was also noticed during this period. Selvin (1996) presents these data as time of diagnosis in months (from January, 1981), and the location in space (given as coordinates in kilometers, modified to protect confidentiality). He then details a space-time clustering analysis of these data using the correlation between the pairwise distances and the pairwise time intervals as the test statistic. No evidence of clustering was found ($p=0.889$) using this method.

Baker (1996) reanalyzed these data using Knox's test. A search through several possible critical values (as is customary when Knox's test is used in practice) yielded significant clustering ($p=0.0204$), using a critical distance of 2 km and a critical time of 1 month. However, suitable critical values for distance and time are not known in advance, so this p-value is inappropriately small. Baker then used a proposed modification of Knox's test to adjust for the fact that the critical parameters are unknown, and could detect only marginal clustering ($p=0.18$).

4.5 Cluster Analyses of the Cardiac Defects Data.

We used Knox's test, Mantel's test, and DP to assess clustering in the cardiac defects data set. Knox's test was conducted setting $K=5$, as was done in the simulation studies,

to again produce a test that is close to optimal from a statistical sense, in terms of identifying clustering. Mantel's test was conducted using spatial and temporal offsets of 0.1, due to the presence of ties in these data. P-values were also determined as was done in the simulation studies, using estimated empirical distributions based on 10,000 simulations for each of the three statistics. Standard errors are again reported to give some sense of the variability in these observed p-values.

Table 4.2

Analysis Results For Cardiac Defect Data

Data	# Cases	DP (SE)	Knox (SE)	MGR (SE)
Cardiac Defect data Saint Clara County, CA	48	0.0976 (0.002968)	0.0474 (0.002125)	0.6157 (0.004864)

DP found marginal clustering in these data, corresponding to the conclusion that Baker reached from her analysis of these data using the proposed modification of Knox's test. Knox's test (using 'optimal' critical parameters) found significant clustering in these data, corresponding to the conclusion that Baker reached from her analysis of these data after searching for 'optimal' critical parameters. Mantel's test did not seem to be sensitive to the type of clustering present in these data, as it did not find significant clustering in these data.

4.6 Discussion.

The results from these analyses indicate that Mantel's test was better at identifying clustering in the Ugandan Burkitt's Lymphoma data, followed by Knox's test and DP. Knox's test and DP performed similarly on these data. The same conclusions regarding clustering in these data would be reached using either Knox's test or DP, except for the 1972-73 time period.

The results from the analyses of the Cardiac Defect data indicate that the performance of DP is similar to that of a Knox's test that has been modified to adjust for

the fact that the critical parameters are unknown. Mantel's test did not indicate clustering in these data as the other statistics did.

Chapter V

CONCLUSIONS AND SUGGESTIONS FOR ADDITIONAL RESEARCH

5.0 Introduction

The purpose of this research is to introduce and investigate the directed path statistic (DP), a new statistical test for evaluating space-time clustering in data where exact location and time information are available for the disease cases or other points of interest. Disease clustering is discussed in a general manner in Chapter 1, and other important selected statistics in this area are reviewed. DP is formally defined in Chapter 2, and its distributional properties are developed there. Chapter 3 describes and presents results from the simulation studies that were carried out on DP, Mantel's Generalized Regression, and Knox's test to better understand these statistics and to compare their performance. Finally, the classic Burkitt's Lymphoma data set from Uganda and three years of birth defect data from California were used in Chapter 4 to compare the performance of these tests on actual data.

In this chapter we will further discuss the challenge of assessing disease clustering, to help clarify the setting in which DP makes a novel contribution. Then, some opportunities for additional research with DP will be identified

5.1 The Role of Disease Cluster Analyses

Historically, disease cluster analyses were initially conducted primarily to provide clues to the cause of diseases with unknown etiology. Cluster analysis results were used to confirm or validate hypotheses that were postulated about the mechanism of causation. More recently, these analyses have been used as screening tools by public health agencies.

Routinely applied to health data, a positive indication may alert health officials when action should be taken to address some health issue. Purposefully applied to health data, a negative indication may be used to dispel public concerns about some health concern. There statistics are also being considered for use to provide warning signals in bioterrorism attacks.

The common factor in most of these applications is that cluster analyses are rarely an end in themselves. Their results are generally precursors to additional, more rigorous and more refined studies. This observation motivated Wartenberg and Greenberg (1993) to promote the use of cluster analyses as preliminary 'pre-epidemiology' studies. Because of this, it is a distinct advantage when these analyses are simple, easy to carry out and understand, and have relatively meager data requirements. This reserves time, money and other resources for the more rigorous epidemiologic studies that should follow.

5.2 Comparing DP, Knox's Test, and Mantel's Generalized Regression.

The results of this research indicate that DP is a reasonable, viable alternative to Knox's test. General, or omnibus, tests for space-time clustering require a non-differential population growth rate for all geographic sub-areas, or they will be affected by population shift bias. When the assumptions for the valid use of a general space-time interaction test apply, DP may be used without having to specify a critical distance and time, which is an important drawback to Knox's test. Mantel's Generalized Regression emphasizes small interpoint distances, so its performance for the kind of clustering examined in this study was disappointing[9], as compared against DP and Knox's test. Because the power of DP is comparable to that of Knox's test, DP is a reasonable alternative to Knox's test for separate source clustering which was investigated in this research.

[9] The performance of Mantel's Generalized Regression on different types of clustering may be an area for further research.

The major drawbacks to Knox's test have been addressed in the recent literature. Baker (1996) has proposed a modification to Knox's test that requires only ranges of critical distances and times to be specified, while Kulldorff and Hjalmars (1999) have proposed a modification which accounts for population shifts, given that knowledge of temporal shifts in the background population exists. These modifications may be effective, but they greatly increase the complexity of the test, and the burden of applying it in actual practice. Arguably, this may simply be too much for a 'pre-epidemiology' analysis. Also, data on the temporal shifts of the background population often is not available, or is not available given the available time, money, and other resources. When the resources are available to acquire the data on the temporal shifts of the background population, one should probably conduct a more rigorous, 'true' epidemiology study. Further, for a more rigorous study, a space-time interaction test specifically designed to account for possible geographic shifts or temporal trends in the underlying population should probably be considered before using a modified Knox test. So in this context, given the purpose of the test and for this kind of 'pre-epidemiology' analysis, DP is a viable alternative to Knox's test.

5.3 Suggestions for Additional Research.

It is of interest to better understand the performance of these statistics for data generated for other clustering models. Simulation studies using separate-source clustering were conducted to investigate the performance of DP, Knox's test, and Mantel's Generalized Regression in this research. Additional work using the other disease clustering models (point-source clustering and contagion clustering) would be helpful in characterizing the performance of these statistics, and in determining which statistics are superior for which disease clustering models.

Geographic features (such as a lake in the middle of the area under consideration) can make some interpoint distances impossible, thus excluding some possible paths

through the data. The way that the significance of DP is calculated can be adjusted to account for this. Some work investigating this adjustment could enhance the usefulness of DP.

As a general or omnibus space-time interaction test, DP is subject to population shift bias if there are geographic or temporal population shifts. Kulldorff and Hjalmars (1999) have proposed a method for adjusting space-time interaction tests for geographic or temporal shifts in the underlying population. If this method were applied to DP, it would be more robust to shifts in the underlying population. Some work investigating and applying this method could enhance the usefulness of DP.

Finally, DP actually looks at the relationship between a univariate variable (time) and a multivariate variable (space). In some sense, DP actually represents a kind of nonparametric 'multiple correlation' between a single univariate and a multivariate variable, and could possibly be used to assess the relative superiority of competing summary statistics. An example of this might be in psychometrics, where several test instruments are used to assess some multi-dimensional concept and characterize it with a single summary score (such as IQ). Suppose there are competing ways to generate this summary score from the component test scores. How might one determine which of these competing summary scores is 'best'? DP could provide a way to assess and rank these summary scores, and facilitate identifying the best one. Some work investigating this area of application could enhance the usefulness of DP in other settings.

APPENDIX A –SIMULATION DRIVER PROGRAM

This is the SAS program used to conduct the simulation studies. The simulations are performed by the **SIM** macro; each execution of the SIM macro produces one simulation run. As written, this driver program executes the macro 9 times, once for each of the 9 combinations of δ and τ considered at each of the 12 combinations of λ and **n**. This code executes successfully in SAS Version 8.2 or higher.

==========

```
*OPTIONS LS=116 PS=500 NOCENTER FORMCHAR="|----|+|---";

options  ls=106 ps=56 pageno=1  FORMCHAR="|----|+|---"; /* 9 point portrait */

*OPTIONS MLOGIC MPRINT SYMBOLGEN;
LIBNAME USER "C:\Documents and Settings\James\My Documents\SAS_D\Temp";

/*
%LET N=50 ;  %LET M=500 ;  %LET SEED=0 ;  %LET K=5 ;  %LET W=100 ;
%LET AVCLUSIZ=10 ;  %LET AVDIST=0.1 ;  %LET AVTIME=0.1 ;  %LET
PI=3.14159265358979 ;
*/

* N is the number of cases;
* M the number of random permutations to estimate P-values
  (This is the size of the empirical distribution);
* AVCLUSIZ = parameter LAMBDA or average cluster size
  (including cases OUT of bounds or discarded after N have been found);
* AVDIST = parameter DELTA, average spatial distance from point source to case;
* AVTIME = parameter TAU, average time from point source and case;
* SEED is the seed for the random number generator, 0 for production runs;
* K is the expected number of close pairs for the Knox test under the null
hypothesis;
* W is the number of simulations for given parameters;

%MACRO SIM(N=, AVCLUSIZ=, AVDIST=, AVTIME=) ;

%LET M=500 ;  %LET SEED=0 ;  %LET K=5 ;  %LET W=500 ;
%LET PI=3.14159265358979 ;

%LET NC2 = %EVAL(&N*(&N-1)/2);
%LET CT  = %SYSFUNC(SQRT(%SYSEVALF(&K/&NC2/4)));
%LET CD  = %SYSFUNC(SQRT(%SYSEVALF(2*&CT/&PI)));

%DO I = 1 %TO &W;

TITLE1 "COMPARISON OF SPACE-TIME INTERACTION TESTS BASED ON PAIR DIFFERENCES";
TITLE2 "&W SIMULATIONS OF &N CASES USING &M PERMUTATIONS";
TITLE3 "LAM=&AVCLUSIZ, DEL=&AVDIST, TAU=&AVTIME, SEED=&SEED";
TITLE4 "EXPECTED NUMBER OF CLOSE PAIRS FOR KNOX TEST IS &K";
TITLE5 "CRITICAL DISTANCE AND TIME FOR KNOX TEST ARE &CD AND &CT";


DATA GOOD ALL; N = &N; LAM = &AVCLUSIZ; DEL = &AVDIST; TAU = &AVTIME; SEED =
&SEED;
  NN = 0; * NN is the ID# of the current case, not counting cases that are "out
of bounds";
```

```
  NC = 0; * NC is the ID# of the current cluster; * C is the size of the current
cluster;
  START: CC = 1; * CC is the ID# of the case within the cluster;
  NC + 1; IF CC = 1 THEN DO; GETC: C = RANPOI(0,LAM); IF C = 0 THEN GOTO GETC;
             PX = RANUNI(SEED); PY = RANUNI(SEED); PT = RANUNI(SEED); STATUS =
"OUT";
             * PX, PY, AND PT ARE THE COORDINATES OF THE POINT SOURCE; END;
  GETDIST: DIST = RANEXP(SEED)*DEL; * DIST is the random distance from source to
case;
  ANGL = 2*&PI*RANUNI(SEED); * ANGL is the random angle (in radians);
  XX = PX + COS(ANGL)*DIST; YY = PY + SIN(ANGL)*DIST; TIME = PT +
RANEXP(SEED)*TAU;
  * XX, YY, and TIME are the coordinates of the current case;
  IF (0 < XX < 1) AND (0 < YY < 1) AND (0 < TIME < 1) THEN DO;
     NN + 1; STATUS = "IN"; OUTPUT GOOD; END; OUTPUT ALL;
  IF NN = N THEN STOP; IF CC = C THEN GOTO START; CC + 1; GOTO GETDIST; RUN;
*PROC PRINT DATA=ALL;
*PROC MEANS DATA=GOOD N MIN MEAN MAX STD;
*PROC PLOT DATA=GOOD;
*     PLOT YY*XX=NC / HAXIS=0 TO 1 BY .2 VAXIS=0 TO 1 BY .2; RUN; OPTIONS
PS=500;

PROC SORT DATA=GOOD(KEEP = XX YY TIME); BY TIME; *OPTIONS NODATE NONUMBER;
TITLE; RUN;

DATA RESULTS&I; SET GOOD END=LAST; FILE PRINT;
* READING IN THE GENERATED DATA AND CALCULATING THE SPATIAL AND TEMPORAL
DIFFERENCES;
ARRAY X(&N) _TEMPORARY_; ARRAY Y(&N) _TEMPORARY_; ARRAY T(&N) _TEMPORARY_;
ARRAY DD(&N,&N) _TEMPORARY_; ARRAY TT(&N,&N) _TEMPORARY_;
X(_N_) = XX; Y(_N_) = YY; T(_N_) = TIME; DD(_N_,_N_) = 0; TT(_N_,_N_) = 0;
IF _N_ > 1 THEN DO N = 1 TO _N_ - 1;
   I = SQRT((XX-X(N))**2 + (YY-Y(N))**2); DD(_N_,N) = I; DD(N,_N_) = I;
   J = ABS((TIME-T(N)));                  TT(_N_,N) = J; TT(N,_N_) = J;
   IF I*J = 0 THEN DO; PUT "THE IMPOSSIBLE HAS OCCURRED!" / _N_= N= I= J=;
ABORT; END;
   END; IF NOT LAST THEN RETURN;
/*
PUT "SPATIAL "; DO I = 1 TO &N; PUT I 2. %DO J = 1 %TO &N; DD(I,&J) 6.3 %END;;
END;
PUT "TEMPORAL"; DO I = 1 TO &N; PUT I 2. %DO J = 1 %TO &N; TT(I,&J) 6.3 %END;;
END;
*/
* COMPUTING THE VALUES OF DP, KX, AND MR (OF WHICH THE FIRST IS THE "OBSERVED"
VALUE);
ARRAY R(&N) _TEMPORARY_; DO I = 1 TO &N; R(I) = I; END;
* Vector R holds random permutations of (1,2,...,N), starting with (1,2,...,N);
DO M = 1 TO &M; * The DO loop in the next 2 lines randomly permutes the vector
R;
   IF M > 1 THEN DO I = N TO 2 BY -1; J = CEIL(I*RANUNI(&SEED));
                    Q = R(J); R(J) = R(I); R(I) = Q; END;
   DP = 0; KX = 0; MR = 0;
   DO I = 1 TO N-1; DP + DD(R(I),R(I+1));
      DO J = I+1 TO N; DIJ = DD(R(I),R(J)); TIJ = TT(I,J);
         IF DIJ < &CD AND TIJ < &CT THEN KX + 1; MR + 1/DIJ/TIJ; END; END;
   IF M = 1 THEN DO; DPOBSRVD = DP; KXOBSRVD = KX; MROBSRVD = MR; END;
*PUT M 2.0 DP 9.3 KX 6.0 MR 11.1;
   IF DP LE DPOBSRVD THEN DPSIG + 1;
   IF KX GE KXOBSRVD THEN KXSIG + 1;
   IF MR GE MROBSRVD THEN MRSIG + 1; END;
DPPVAL = DPSIG/&M; KXPVAL = KXSIG/&M; MRPVAL = MRSIG/&M;
* Note that all P-values are estimated based on the same permutations of times;
OUTPUT; KEEP DPOBSRVD--MRPVAL; RUN;

%END;

DATA SUMMARY; SET %DO I = 1 %TO &W; RESULTS&I %END;;
/*PROC PRINT; VAR DPOBSRVD KXOBSRVD MROBSRVD DPPVAL KXPVAL MRPVAL;
```

```
PROC MEANS DATA=SUMMARY N MIN MEAN MAX STD;*/
PROC CORR; VAR DPPVAL KXPVAL MRPVAL;

%MEND SIM;

* Send LOG window contents to a file ;
proc printto log='C:\Documents and Settings\James\My
Documents\SAS_D\Temp\sim_log.txt' ;
    run ;

%SIM(N=500,AVCLUSIZ=25,AVDIST=0.2,AVTIME=0.2);

data p25_20_20 ;    set summary ;
    lam=25 ;    del=0.20 ;    tau=0.20 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;

%SIM(N=500,AVCLUSIZ=25,AVDIST=0.2,AVTIME=0.1);

data p25_20_10 ;    set summary ;
    lam=25 ;    del=0.20 ;    tau=0.10 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;

%SIM(N=500,AVCLUSIZ=25,AVDIST=0.2,AVTIME=0.01);

data p25_20_01 ;    set summary ;
    lam=25 ;    del=0.20 ;    tau=0.01 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;


%SIM(N=500,AVCLUSIZ=25,AVDIST=0.1,AVTIME=0.2);

data p25_10_20 ;    set summary ;
    lam=25 ;    del=0.10 ;    tau=0.20 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;

%SIM(N=500,AVCLUSIZ=25,AVDIST=0.1,AVTIME=0.1);

data p25_10_10 ;    set summary ;
    lam=25 ;    del=0.10 ;    tau=0.10 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;

%SIM(N=500,AVCLUSIZ=25,AVDIST=0.1,AVTIME=0.01);

data p25_10_01 ;    set summary ;
    lam=25 ;    del=0.10 ;    tau=0.01 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;


%SIM(N=500,AVCLUSIZ=25,AVDIST=0.01,AVTIME=0.2);

data p25_01_20 ;    set summary ;
    lam=25 ;    del=0.01 ;    tau=0.20 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;

%SIM(N=500,AVCLUSIZ=25,AVDIST=0.01,AVTIME=0.1);

data p25_01_10 ;    set summary ;
    lam=25 ;    del=0.01 ;    tau=0.10 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;
```

```
%SIM(N=500,AVCLUSIZ=25,AVDIST=0.01,AVTIME=0.01);

data p25_01_01 ;    set summary ;
    lam=25 ;    del=0.01 ;    tau=0.01 ;
    keep  lam del tau DPPVAL KXPVAL MRPVAL ;
        run ;


proc printto  log=log ;
run ;


data pv25_500 ;    set  p25_20_20  p25_20_10  p25_20_01  p25_10_20  p25_10_10
                        p25_10_01  p25_01_20  p25_01_10  p25_01_01 ;
    run ;

data pv25_500 ;    set  pv25_500 ;
    rename  DPPVAL=DP  KXPVAL=KNOX  MRPVAL=MGR ;
    keep  lam del tau DP KNOXL MGR ;
    run ;

proc sort ;    by  lam del tau ;
    run ;

proc print  data=pv25_500 ;
    title "P-values For N=500 and Lambda=25" ;
    run ;

PROC EXPORT DATA=user.pv25_500
    OUTFILE='C:\Documents and Settings\James\My
Documents\SAS_D\Results\pv25_500.xls'
    DBMS=excel2000 REPLACE;
    RUN;


                                    quit ;
```

List of References

Abe, O. (1973): A note on the methodology of Knox's tests of "time and space interaction". *Biometrics* 29: 67-77

Baker, Rose D. (1996): Testing for space-time clusters of unknown size. *Journal of Applied Statistics*, 23(5): 543-554

Barton, D.E., and David, F.N. (l962): Randomization basis for multivariate tests. I. The bivariate case: randomness of n points in a plane. *Bulletin of the International Statistical Institute* l39: 455- 467

Barton, D.E., and David, F.N. (l966): The random intersection of two graphs. Research Papers in Statistics (Neyman Festschrift), F.N. David, (ed.), pp. 445-459, New York: John Wiley & Sons Inc.

Barton, D.E., and David, F.N., and Merrington, Maxine (l965): A criterion for testing contagion in time and space. *Annals of Human Genetics* 29: 97-l02

Benenson, Abram S. (l975): Control of Communicable Disease in Man, 12th Ed, AS Benenson, Ed. APHA, Washington, DC, l975

Besag, Julian, and Newell, James (1991): The detection of clusters in rare diseases. *Journal of the Royal Statistical Society*. A 154. Part 1, pp 143-155.

Brubaker, G., Geser, A. and Pike, M.C. (1978): Burkitt's lymphoma in the North Mara district of Tanzania 1964-70: Failure to find evidence of time-space clustering in a high-risk isolated rural area. *British Journal of Cancer*. 28:409.

Charalambides, Charles A. (2000): Enumerative Combinatorics. Boca Raton, FL Chapman and Hall

Clark, W.A.V. and Avery, Karen L. (1976): The effects of data aggregation in statistical analysis. *Geographical Analysis* 8: 428-438

Conover, W.J. (1999): Practical Nonparametric Statistics, 3rd Edition. New York: John Wiley & Sons Inc.

Dat, Nguyen V. (l982): Tests for time-space clustering of disease. Institute of Statistics Mimeo Series No. 1409, Chapel Hill, North Carolina

de-The', Guy (l979): The epidemiology of Burkitt's lymphoma: Evidence for a causal association with Epstein-Barr virus. *Epidemiologic Review* l: 32-54

David, F.N. and Barton, D.E. (1966): Two space-time interaction tests for epidemicity. *British Journal of Preventive and Social Medicine* 20, 44-48

Dempster, A.P. and Schatzoff, M. (1965): Expected significance level as a sensitivity index for test statistics. *Journal of the American Statistical Association* 60: 420-436

Diggle, Peter J (1983) Statistical Analysis of Spatial Point Patterns, Academic Press Inc., New York, NY

Ederer, Fred, Myers, Max H., and Mantel, Nathan (l964): A statistical problem in space and time: Do leukemia cases come in clusters? *Biometrics* 20: 626-638

Glass, A.G., Mantel, Nathan, Gunz, F.W., and Spears, G.F.S. (1971): Time-space clustering of childhood leukemia in New Zealand. *Journal of the National Cancer Institute* 47: 329-336

Grimson, Roger C. and Ingram, Deborah D. (1982): A test for concordant epidemic patterns. Unpublished manuscript.

Grimson, Roger C. (l983): A test for epidemicity. Unpublished manuscript

Grimson, Roger C. (1993): Disease clusters, exact distributions of maxima, and *P*-values. *Statistics In Medicine* 13: 1773-1794

Hardy, WD Jr. and McClelland, AJ (1977): Feline leukemia virus. Its related diseases and control. *The Veterinary Clinics of North America*. 7(1): 93-103

Hollander, Myles and Wolfe, Douglas A. (1973): Nonparametric Statistical Methods, John Wiley and Sons, Inc., New York

Hung, H.M., O'Neill, R.T., Bauer, P., and Köhne, K. (1997): The behavior of the p-value when the alternative hypothesis is true. *Biometrics* 53: 11-22

Ingram, Deborah D. (1983): A test for concordant nonrandom patterns among series with epidemiologic applications. Institute of Statistics Mimeo Series No. 1430, Chapel Hill, North Carolina

Jacquez, Geoffrey M. (1996): A k nearest neighbour test for space-time interaction. *Statistics in Medicine* 15-30;15(17-18):1935-49.

Joiner, B.L. (1969): The median significance level and other small sample measures of test efficacy. *Journal of the American Statistical Association* 64: 971-985

King, Paul E. (1979): Problems of spatial analysis in geographical epidemiology. *Social Science and Medicine* 13D: 249-252

Klauber, Melville R. (1971): Two-sample randomization tests for space-time clustering. *Biometrics* 27: 129-142

Klauber, Melville R. and Mustacchi, P (1970): Space-time clustering of childhood leukemia in San Francisco. *Cancer Research* 30: 1965-1973

Knox, G (1963): Detection of low-intensity epidemicity. *British Journal of Preventive and Social Medicine* l7: 121-127

Knox, G (1964): The detection of space-time interaction. *Applied Statistics* 13: 25-30

Knox, G (1964): Epidemiology of childhood leukemia in Northumberland and Durham. *British Journal of Preventive and Social Medicine* 18: 17-24

Knuckles, Bettie N. (1983): Strategies for investigating health and outcome patterns and an extension of Mantel's generalized regression approach: Application to homicide and suicide data. DrPH dissertation, University of North Carolina at Chapel Hill

Kulldorff, Martin. (2002): Tests for spatial randomness adjusted for an inhomogeneity: A general framework. Unpublished manuscript.

Kulldorff, Martin and Hjalmars, Ulf. (1999): The Knox method and other tests for space-time interaction. *Biometrics* 55: 544-552

Kulldorff, Martin, Tango, Toshiro and Park, Peter J. (2003): Power comparisons for disease clustering tests. *Computational Statistics and Data Analysis* 24: 665-684

Lloyd, Setsuko and Roberts, C.J. (1973): A test for space clustering and its application to congenital birth defects. *British Journal of Preventive Social Medicine* 27: 188-191

Mantel, Nathan (1967): The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220

Mantel, Nathan, Kryscio, Richard J., and Myers, Max H. (1976): Tables and formulas for extended use of the Ederer-Myers-Mantel disease clustering procedure. *American Journal of Epidemiology* 104: 576-584

Marriott, F.H.C. (1978): Barnard's monte carlo tests: How many simulations? *Applied Statistics* 28: 75-77

Morrow, R.H., Kisuule, A., Pike, M.C., Smith, P.G.. (1976): Burkitt's lymphoma in the Mengo districts of Uganda: Epidemiologic features and their relationship to malaria. *Journal of the National Cancer Institute*, 56:479

Morrow R.H., Pike, M.C. and Smith, P.G. (1977): Further studies of space-time clustering of Burkitt's lymphoma in Uganda. *British Journal of Cancer*. 35(5): 668-73.

Morrow R.H., Pike, M.C., Smith, P.G., Ziegler, J.L., and Kisuule, A. (1976): Burkitt's lymphoma: A time-space cluster in Bwamba county of Uganda. *British Medical Journal* 2: 491

Peebles, P.J.E. (1974): The nature of the distribution of galaxies. *Astronomical Astrophysics* 32: 197-202

Pike, M.C., Williams, Edward H. and Barbara Wright (1967): Burkitt's tumor in the West Nile district of Uganda 1961-5. *British Medical Journal* 2: 395-399

Remington, Richard D. and Schork, M. Anthony (1985): Statistics with Applications to the Biological and Health Sciences, 2nd Ed. Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Roberson, Paula K. (1979): Distributional and robustness problems in time-space clustering. Ph.D. dissertation, University of Washington, Seattle.

Roberson, Paula K. and Fisher, Lloyd (1983): Lack of robustness in time-space disease clustering. *Communications In Statistics – Simulation And Computation* 12(1): 11-22

SAS Institute, Inc., SAS/STAT User's Guide, Version 8. Cary, NC: SAS Institute, Inc., 1999. 3884 pp.

Sackrowitz, Harold and Samuel-Cahn, Ester (1999): *P*-values as random variables – Expected *P* values. *The American Statistician* 53:326-331

Schatzoff, N. (1956): Sensitivity comparison among tests in the general linear hypothesis. *Journal of the American Statistical Association* 61: 415-435

Selvin, Steve (1996): Statistical Analysis of Epidemiologic Data, 2nd Edition. New York, Oxford University Press.

Siemiatycki, Jack (1971): Space-time clustering: Finding the distribution of a correlation-type statistic. Master's Thesis, McGill University.

Siemiatycki, Jack (1978): Mantel's space-time clustering statistic: Computing higher moments and a comparison of various data transforms. *Journal of Statistical Computation and Simulation* 7: 13-32

Siemiatycki, Jack and McDonald, AD (1972): Neural tube defects in Quebec: A search for evidence of clustering in time and space. *British Journal of Preventive and Social Medicine* 26: 40-44

Siemiatycki Jack, Brubaker, G, and Geser, A. (1980): Space-time clustering of Burkitt's lymphoma in East Africa: analysis of recent data and a new look at old data. *International Journal of Cancer*. 25(2): 197-203.

Silva, Claudio and Quade, Dana (1980): Analysis of weighted rankings using expected significance level. *Communications In Statistics - Theory And Methods* A9(10): 1087-1096

Smith PG (1974). Burkitt's lymphoma in Africa. *Cancer Research*. 34, 1211:349

Stark, Charles R. and Mantel, Nathan (1967): Lack of seasonal or temporal-spatial clustering of Down's syndrome births in Michigan. *American Journal of Epidemiology* 86: 199-213

Stark, Charles R. and Mantel, Nathan (1967): Temporal-spatial distributions of birth dates for Michigan children with leukemia. *Cancer Research* 27: 1749-1755

Strauss, D.J. (1975): A model for clustering. *Biometrika* 62: 467-475

Waller, L.A. and Turnbull, B.W. (1993): The effects of scale on tests for disease clusters. *Statistics In Medicine* 13: 1869-1884

Wartenberg, Daniel, and Greenberg, Michael (1993): Solving the clusters puzzle: Clues to follow and pitfalls to avoid. *Statistics In Medicine* 12: 1763-1770

Whittemore, Alice S., Friend, Nina, Brown, Byron W., and Holly, Elizabeth A. (1986): *A test to detect clusters of disease.* Presented at ENAR-86, Atlanta, GA

Williams, Edward H., Day, Nicholas E., and Geser, Anton G. (1974): Seasonal variation in onset of Burkitt's lymphoma in the West Nile district of Uganda. *Lancet* 2(19):19-22

Williams, Edward H., Smith PG, Day, Nicholas E., Geser, Anton G., Ellice J, Tukei P (1978). Space-time clustering of Burkitt's lymphoma in the West Nile district of Uganda: 1961-1975. *British Journal of Cancer*. 37(1):109-22.

Williams, Edward H., Spit, P., and Pike, M.C. (1969). Further evidence of space-time clustering of Burkitt's lymphoma patients in the West Nile District of Uganda. *British Journal of Cancer*. 23(2):235-46.