2017

Longitudinal And Geographical Modeling Of Circular Data With An Application To Sudden Infant Death Syndrome

Xinyan Cai
University of South Carolina

Recommended Citation

Cai, X.(2017). *Longitudinal And Geographical Modeling Of Circular Data With An Application To Sudden Infant Death Syndrome.* (Master's thesis). Retrieved from http://scholarcommons.sc.edu/etd/4084

LONGITUDINAL AND GEOGRAPHICAL MODELING OF CIRCULAR DATA WITH AN APPLICATION TO SUDDEN INFANT DEATH SYNDROME

by

Xinyan Cai

Bachelor of Arts
East Tennessee State University 2013

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Public Health in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2017

Accepted by:

Marco Geraci, Director of Thesis

James Hardin, Reader

Nansi Boghossian, Reader

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

ACKNOWLEDGMENTS

This thesis is a summary of my research work during my master’s program under the supervision of my advisor Dr. Marco Geraci. I am deeply grateful for his patient and helpful guidance. His suggestions greatly enlightened me during my research. I appreciate his insight and support during this process. I would not have made it without his guidance.

I would like to thank Dr. Nansi Boghossian for her enlightening comments on my thesis work. She has gone through the writing of my thesis in detail and my thesis is polished with her help. Dr. Nansi Boghossian also enhanced my work by contributing her knowledge of epidemiology.

I also want to express my gratitude to Dr. James Hardin. He gave me helpful suggestions on my work, which greatly motivated me as I began my thesis. I also have learned a lot about biostatistical concepts and methodology during his class, which added to my research work.

ABSTRACT

The aim of this thesis is to study seasonality of death in U.S. infants who died from SIDS. We also propose to investigate secular trends and geographical patterns of seasonal patterns of mortality. The application of circular statistics is used to describe the seasonality of the month of death in infants who died from SIDS in 1990, 2000 and 2010. The secular trends of seasonal patterns of SIDS mortality are investigated using a circular linear regression model after adjusting for potential confounders. The geographical variation in seasonal patterns of SIDS mortality is explored from the U.S. map and quantified by using Moran’s I auto correlation. We conclude insignificant correlation between geographical region and the mean direction of month of death in infants who died from SIDS and/or the residuals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

Infant Mortality, Seasonal and Geographical Patterns

Infant mortality is an important indicator to measure the health and well-being of a country's population. According to the National Center for Health Statistics (NCHS) – Centers for Disease Control and Prevention (CDC), over 23,000 infants died in the United States (U.S.) in 2014. The leading causes of infant mortality were birth defects, extreme low birthweight and sudden infant death syndrome (SIDS). According to the National Vital Statistics Report (NVSR), of the 28,035 U.S. infants who died in 2000, 5,743 (20.5%) died from congenital malformations and deformation, 4,397 (15.7%) died from disorders related to short gestation and low birth weight, and 2,523 (9%) died from sudden infant death syndrome (Minino et al., 2002). From 2000 to 2010, the number of infant deaths decreased to 24,586, but the three leading causes of infant mortality were still congenital malformation and deformation (20.8%), disorders related to short gestation and low birth weight (16.9%), and sudden infant birth syndrome (8.4%) (Mathews & MacDorman, 2013).

Pregnancy outcomes including neonatal mortality are greatly influenced by a woman's health and demographic factors. Significant differences exist in infant mortality by race; among black infants the mortality rate is more than twice the rate among white infants (Centers for Disease Control and Prevention, 2016). Maternal smoking during pregnancy is independently associated with infant mortality; 10% of infant deaths would be reduced if women refrained from smoking while pregnant, according to Kleinman et al. (1988). Lower levels of maternal education increase the risk of neonatal and post neonatal infant death (Luo et al., 2006). A significantly greater rate of infant mortality occurs in those born to a mother aged less than 20 years (Babson & Clarke, 1983).

Gestational age of an infant plays a fundamental role in infant mortality. For example, according to Kramer et al. (2000), the relative risk of all causes of mortality in the US for infants born at 32-33 gestational weeks was 6.6 (95% confidence interval [CI], 6.1-7.0) higher compared to infants born at 28-31 gestational weeks, while the risk for those born at 34-36 gestational weeks was 2.9 (95% CI, 2.8-3.0) times higher compared to infants born at 28-31 gestational weeks.

Other neonatal characteristics that impact neonatal mortality include multiple births. Multiples deliveries are five times more likely to result in infant mortality than singleton births (Mathews & MacDorman, 2013).

There are also significant seasonal patterns in infant mortality. According to a cohort study conducted in Minnesota by Keller & Nugent (1983), perinatal death increases in July, August and September. McMullen (1983) also concluded that the seasonal climatic variation has an independent effect on infant mortality after controlling for race, metropolitan residence, and socioeconomic status.

SIDS is one of the causes of death that has been often studied in relation to its seasonality patterns. The seasonality patterns are more evident in outcomes for babies older than four months as compared to those who die at younger ages (Douglas et al, 2014). The peak of SIDS mortality typically occurs in January (Valdeś-Dapena, 1980; Adam et al, 1998), with an overall pattern that resembles that of viral respiratory infections. Other studies showed similar patterns. The proportion of infant deaths in each season does not vary by maternal race (Valdeś-Dapena, 1980).

Finally, infant mortality has been seen to vary based on geography. For example, in the U.S., infant mortality rates vary considerably by state, ranging from the highest age-adjusted rate in Mississippi to the lowest rate in Hawaii. The infant mortality variation among states is partly due to differences in race/ethnicity and socioeconomic status with higher infant mortality rates in the South and Midwest of the U.S. and lower elsewhere (Mathews & MacDorman, 2013).

In this project, we propose to study seasonality of birth and death among U.S. infants who died from SIDS. We also propose to investigate circular trends and geographical patterns of seasonal patterns of mortality. In the following sections, we briefly introduce the methods and the data. In Chapter 2, we discuss the methods at greater length, while in Chapter 3 we present the results of our study. In Chapter 4, we conclude with some final remarks.

Circular Statistics

Circular statistics refer to statistics that deal with directional data. This type of data is supported by the unit circle, as opposed to linear data that are supported by the real line. Circular data usually are recorded as angles, expressed in degrees or radians, and are measured clockwise from an (arbitrary) origin. The most fundamental property of circular data is that the beginning and end scales coincide. For observations within the circular scale, angles can be converted to a point on the circumference. Time can be represented as circular data, whether it is the hour of the day, the day of the week or the calendar month (Pewsey et al., 2013).

The application of statistical methods appropriate for linear data to circular data is generally inappropriate and may lead to misleading results. For example, if a drug reaches the peak of its effectiveness between 11:00AM and 1:00PM, the “average” would be 12:00pm. However, the arithmetic mean of these directions is 6:00, which points in the opposite direction.

Circular statistics have been used in different areas of epidemiologic studies. Mooney et al. (2002) combined the von Mises distribution, a circular distribution, and uniform distribution, a linear distribution, to model the data on infant death from SIDS to assess the seasonal pattern of infant mortality. The result shows that SIDS data can be adequately fitted by the von Mises distribution. Circular statistics are also used as a tool to detect periodontal disease, a common cause of tooth loss in adults. The study (Maitra & Braun, 2012) visualized the mouth as a circle and teeth as points located on the circumference of the circle. Circular statistics were used to determine the average location of periodontal disease. Barnett et al. (2012) applied circular statistics to the monthly number of cardiovascular disease deaths in people aged greater than 75 years old in Los Angeles from the years 1987 to 2000. By visualizing the circular pattern of the month in which cardiovascular death occurred, they found an unevenly distributed seasonal pattern of cardiovascular death and concluded that mortality does not start in earnest until October. Serfling (1963) examined the circular trend and its extrapolation on pneumonia influenza mortality, and also estimated the seasonal variation. The circular trend of the seasonal change was described by combining the sine and cosine mathematical functions into a linear regression model to investigate the trend of seasonal change by years.

Data Description

The data used in our project were abstracted from the NVSS (NCHS). NVSS data are released in two formats: *period* data and birth *cohort* data. The numerator for the period-linked file consists of all infant deaths occurring in a given data year linked to their corresponding birth certificates, whether the birth occurred in that year or the previous year. The numerator for the birth cohort-linked file consists of deaths of infants born in a given year. In both cases, the denominator is all births occurring in the year. Infant mortality rates are calculated from the cohort data.

The NVSS only records births and infant deaths occurring in the U. S. for infants born to either residents or nonresidents. U.S. infants who died outside the United States and non-U.S. infants who die in the United States are excluded. Both birth records and linked death records were compiled in a national database through the Vital Statistics Cooperative Program of the CDC and Prevention’s National Center for Health Statistics (Murphy et al., 2013).

In this project, we applied circular statistics to describe the seasonality of the month of death in infants who died from SIDS in 1990, 2000 and 2010. Using circular linear regression, we then investigated circular trends of seasonal patterns of SIDS mortality after adjusting for potential confounders. Finally, we explore geographical variation in seasonal patterns of SIDS mortality by using Moran’s I auto correlation.

CHAPTER TWO: METHODS

Circular Statistics

According to Pewsey et al. (2013), the circular observation represented by the angle, θ, is subtended by the arc around the circle, with a clockwise or counterclockwise rotation from the origin. In our analyses, the circular observation θ is measured in radians, thus θ corresponds to the same point for $\theta+2\pi p$, where $p=\pm 1, \pm 2, \ldots$. We define the unit vector **x** in the real plane which is related to θ through **x**=(cos θ, sin θ)^T. Using the complex plane, the circular observation with unit vector **x** can be also represented as complex number $z = e^{i\theta} = cos\theta + isin\theta$, where *i* is the imaginary unit. We denote *n* circular observations with $x_1, \ldots, x_n$, and $z_1, \ldots, z_n$. The first sample trigonometric moment in the zero direction is given by

$$\mathrm{t}_{1,0} = \frac{1}{n}\sum_{j=1}^{n} z_j = \frac{1}{n}\sum_{j=1}^{n} e^{i\theta_j} = \frac{1}{n}\sum_{i=1}^{n}(\cos\theta_j + i\sin\theta_j) = C + iS, \tag{1}$$

where

$$C = \sum_{i=1}^{n} \cos\theta_j \,, S = \sum_{i=1}^{n} \sin\theta_j. \tag{2}$$

The circular mean, $\bar{\theta}$, measures the location of the circular data, which is calculated by

$$\bar{\theta}=\begin{cases} \tan^{-1}\left(\frac{S}{C}\right) & S > 0, C > 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + \pi & C < 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + 2\pi & S < 0, C > 0, \end{cases} \tag{3}$$

where the $\tan^{-1}$ function is in the range of value between $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ in radians.

The direction $\bar{\theta}$ of the vector resultant of $\theta_1, \theta_2, \ldots, \theta_n$ is given by

$$\cos\bar{\theta} = \frac{C}{R}, \sin\bar{\theta} = \frac{S}{R} \tag{4}$$

where R is the resultant length of the vector resultant and it lies in the range between 0 to n. The mean resultant $\bar{R}$, is defined as R/n, which measures the dispersion on the circle. The length of the mean resultant vector in the complex plane can be represented as

$$\bar{R}=(C^2 + S^2)^{\frac{1}{2}}, \tag{5}$$

which lies between 0 and 1. When $\bar{R}$ is close to 0, the distribution is uniform, i.e. the data is spread evenly across the circle; when $\bar{R}$ is close to 1, the data are clustered around the mean direction.

The standard deviation of the circular statistics is defined as $\sigma = \{-2\log(1 - V)\}^{1/2}$, where V is the variance of the circular data, and the relationship between V and $\bar{R}$ can be defined as $V = 1 - \bar{R}$. It follows that V, too, ranges from 0 to 1. Similar to its counterpart used for linear data, the smaller the value of V or σ, the more concentrated the distribution (Fisher, 1995).

The circular median direction, denoted by $\tilde{\theta}$, provides a robust alternative estimation to the mean direction, $\bar{\theta}$. It is more efficient when the sample distribution is skewed. If the data are mostly distributed along the circle instead of being concentrated on the arc of the circle, estimation of the circular median can be carried out using methods for linear data. Otherwise, it can be calculated by minimizing the function

$$\mathrm{d}(\tilde{\theta}) = \pi - \sum_{i=1}^{n} |\pi - |\theta_i - \theta|| \tag{6}$$

Circular Linear Regression Model

The circular linear model is a regression model where the response variable is circular and its mean is a function of explanatory variables. Typically, it is assumed that the distribution of the response is von Mises. The von Mises distribution plays a similar role as the normal distribution of linear data with its unimodality and reflective symmetry of the mean. The von Mises distribution with parameters μ and k, denoted by $\mathrm{VM}(\mu, k)$ has density function

$$f(\theta) = [2\pi I_0(k)]^{-1}\exp[k\cos(\theta - \mu)], \tag{7}$$

where $0 \leq \theta < 2\pi$, $0 \leq k < \infty$; k is the concentration parameter, and

$$I_p(k)=\frac{1}{2\pi}\int_0^{2\pi} \cos p\theta e^{k\cos\theta} d\theta \tag{8}$$

is called the Bessel function. When $k = 0$, the distribution is centered on μ; when k is large, the distribution is unimodal and relatively symmetric, which leads to $\mu = \tilde{\mu}$. The relationship between the concentration parameter k and the mean resultant length $\bar{R}$ can be expressed as

$$\bar{R} = A_1(k) = \frac{I_{1(k)}}{I_{0(k)}} = \frac{1}{2\pi I_0(k)}\int_0^{2\pi} \cos(\theta - \mu)e^{k\cos(\theta-\mu)}d\theta. \tag{9}$$

For $0 \leq \theta \leq 2\pi$, the distribution function is given by

$$F(\theta) = \frac{1}{2\pi I_0(k)}\int_0^{\theta} e^{k\cos(\emptyset-\mu)}d\emptyset. \tag{10}$$

The mean direction μ is related to the explanatory variable X_j by the equation

$$\mu_j=\mu+\mathrm{g}(\gamma_1 x_1 + \gamma_2 x_2 + \cdots \gamma_n x_n), \tag{11}$$

where g is the link function, $\mathrm{g}(\mu) = 2\tan^{-1}(\mu)$, and $\mathrm{g}(\mu)$ ranges between $-\pi$ and π. Maximum likelihood estimation has been used to estimate the unknown parameter γ in the linear circular model. The linear predictors may have great influence on the concentration and the mean direction of the response (Fisher, 1995).

We used circular-linear regression to model the mean direction of month of death in infants who died from SIDS. In the model, we included the following variables: age at death (days), year of death (1990, 2000, 2010), the interaction between age at death and year of death, maternal prenatal care, maternal age, alcohol consumption, smoking status, plurality, gestational age (weeks), marital status, education level and race.

Maternal prenatal care (yes vs no), plurality (single vs multiple births), smoking status during pregnancy (yes vs no), race (black vs white), marital status (married/living with a partner vs no), age (≤19 years vs >19 years), and prenatal care (prenatal care in the first month of pregnancy vs later/no care) were treated as binary variables. The numbers of years of education years were divided into five categories: 0 to 9 (reference group), 9 to 11, 12, 13 to 15 and more than 16 years. Age at death was grouped into 4 intervals: 0 to 27, 28 to 120 (reference), 121 to 180, and 181 to 365 days.

For calculation purposes, each month was assigned a direction on the 12-hour clock template, i.e. $j - 0.5$, $j = 1, \ldots, 12$ and then converted to radians (Table 2.1)

Table 2.1 *Month with Corresponding Direction and Radians*

Month	Direction	Radians
January	0.5	0.262
February	1.5	0.785
March	2.5	1.309
April	3.5	1.833
May	4.5	2.356
June	5.5	2.880
July	6.5	3.403
August	7.5	3.927
September	8.5	4.451
October	9.5	4.974
November	10.5	5.498
December	11.5	6.021

In Equation (11), the regression coefficients can be interpreted as follows: as compared to the baseline, the average month of death occurs earlier when the coefficient is negative; if the coefficient is positive, the average month of death occurs later.

Under the assumption of von Mises residuals, the P-P (probability-probability) plot and Q-Q (quantile-quantile) plot were used to check the goodness of fit of the model. The residual for each month, $r_j, j = 1, \dots, 12$, is defined as the difference between the observed radian and predicted radian for that month.

For all circular statistics analyses, we used the R package “circular” (Agostinelli & Lund, 2013). P-values were based on normal approximations. The significance level was set as 5% in all analyses.

Geographical Residual Pattern

The Moran’s autocorrelation coefficient (Moran’s I) is a statistic used to assess whether there is correlation in a signal among close locations (Paradis, 2010). The formula for Moran’s autocorrelation coefficient is

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{12}$$

where w_{ij} is a weight for observations *i* and *j* and S_0 is the sum of all w_{ij}’s

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}. \tag{13}$$

In our study, the weight w_{ij} takes the value 1 if the states *i* and *j* are geographically contiguous (“neighbors”), or the value 0 otherwise. The null hypothesis of the test is that there is no spatial correlation. Under this assumption, the expected value of the statistic *I* is equal to $I_0 = -1/(n-1)$. If the observed *I* is significantly greater than I_0, then there is indication of a positive autocorrelation. If the observed *I* is significantly smaller than I_0, then there is indication of a negative autocorrelation. We used the spdep package (Bivand et al, 2015) to define the spatial weights matrix.

CHAPTER THREE: RESULTS

Summary Statistics

Table 3.1 shows summary circular statistics for the three leading causes of death in infants in 1990, 2000, and 2010. The mean direction of month of death from SIDS was January in 1990, shifting to mid- and late February in 2000 and 2010. The mean direction of month of death for extremely low birthweight or immaturity was between late April (1990) and mid-July (2010). Finally, the mean direction of the month of death from heart malformations was between late September (1990) and mid-January (2000 and 2010). The median direction was in general consistent with the mean, except for the median direction of SIDS in 2010, which was substantially lower than the mean.

Table 3.1 *Summary Statistics of Three Main Causes of Infant Death Month*

	SIDS			Extremely low birthweight or immaturity			Heart malformations		
	1990	2000	2010	1990	2000	2010	1990	2000	2010
Number of deaths	5332	2499	2093	3866	3248	3142	2520	1625	1135
Mean direction month of death	0.46	1.9	1.4	4.03	5.23	6.43	8.87	0.45	0.64
Standard deviation	3.77	4.97	4.51	5.19	6.02	5.21	5.51	5.68	5.04
Median direction month of death	0.5	2.5	0.5	4.5	5.5	6.5	9.5	0.5	0.5

Summary circular statistics stratified by infant age are reported in Table 3.2. Most infants died from SIDS at ages between 28 and 120 days, and most of the infants died in the winter months regardless of age. However, in 2000, the average death month was the beginning of May for infants who died between 7 and 27 days of age, and the average death month was late July for infants who died between 121 – 180 days of age. The median direction of death month among infants who died from SIDS was close to the mean direction for each category of the age of infant death, indicating symmetry in the distribution. The standard deviations were in general close to 4 (except in 2000 at ages 121-180), which is relatively large based on the 12-hour clock template and thus indicates a relatively flat distribution.

Table 1.2 *Summary Statistics of the Direction of Death Month in Infants who Died from SIDS by Year and Age at Death*

Year of death			1990					2000					2010		
Age at infant deaths	0 - 6	7 – 27	28 - 120	121 - 180	181 - 365	0 - 6	7 - 27	28 - 120	121 - 180	181 - 365	0 - 6	7 - 27	28 - 120	121 - 180	181 - 365
Number of deaths	58	305	3686	855	428	29	168	1661	410	231	26	169	1235	372	237
Mean direction	11.43	10.77	0.45	1.29	11.31	1.24	4.08	1.31	6.91	2.18	11.24	0.50	1.75	0.05	1.90
Standard deviation	3.51	4.57	4.51	3.79	3.95	4.17	3.95	4.83	5.97	4.63	4.90	4.57	4.56	4.48	3.91
Median	10.50	11.50	0.50	1.50	11.50	1.50	4.50	2.50	6.00	1.50	10.50	0.50	1.50	0.50	1.50

Circular Linear Regression Model

The output for the circular linear model is shown in Table 3.3, where the baseline direction μ, the regression coefficients γ and the concentration parameter k from the output are measured in radians. Based on the output, the baseline direction μ is 5.85, and k is 1.534. The baseline direction of infant death month can be interpreted as the mean direction of death month for black infants who died in 1990 aged between 28 and 120 days, with mothers who had multiple births and did not smoke during pregnancy, lived alone, received late prenatal care, were older than 19 years at childbirth, and had less than 9 years of education.

As compared to baseline (infants who died in 1990 at 28 to 120 days after birth), infants who died at similar ages had statistically significant earlier (2000) or later (2010) average month of death. Infants who died in 1990 at 0 to 27 days or 181 to 365 days had significantly earlier average month of death as compared to those who died at 28 to 120. In contrast, infants who died at 121 to 180 days had similar average month of death as compared to baseline. There were significant interactions between year of death and age at death. Notably, infants who died in 2000 aged 0 to 27 days had similar average month of death (net effect $0.095 - 0.053 - 0.034 = 0.008$ radians) as compared to baseline, while those who died aged 181 to 365 days had later (net effect 0.026 radians) average month of death. Finally, the average death month for infants who died in 2010 aged 121-180 days was 0.026 radians earlier than that for infants in the baseline group.

Smoking during pregnancy, early prenatal care, maternal age and maternal education were significantly associated with infant death month. On average, infants whose mothers smoked during pregnancy died 0.023 radians earlier compared to infants whose mothers did not smoke, while infants whose mothers received early prenatal care died 0.025 radians earlier compared with infants whose mothers received late or no prenatal care. For each additional week of gestational age, the mean infant death month increased by 0.003 radians (corresponding to about 0.024 radians difference between a very preterm infant and an infant born at term). Also, for infants born to mothers with more than 9 years of education, the average death month occurred earlier as compared to those born to mothers with less than 9 years of education. However, the magnitude of the effect was noticeable only for > 16 years of education. No statistically significant associations were found for the remaining variables.

Figure 3.1 shows the P-P and Q-Q plots of the residuals. There are some departures between the estimated empirical function and the reference line, which indicates our model does not fit the von Mises distribution very well. We hypothesized that the lack of fit of the P-P and Q-Q plots are due to some unexplained factors: for example, geographic factors. For the next part, the paper will investigate the geographical pattern of the direction of death month among infants who died from SIDS.

Table 3.3 *Circular Linear Regression for Infant Death Month from SIDS*

Circular-Linear Regression for Infant Death Month for SIDS				
log-likelihood:2556				
Exposure	Estimation	Std.Error	t-value	P-value
Intercept	-0.071	0.082	0.866	0.193
Year of death				
1990	ref			
2000	-0.034	0.018	1.933	0.027
2010	0.106	0.030	3.490	0.000
Age at death				
0 to 27	-0.053	0.034	1.542	0.062
28 to 120	ref			
121 to 180	0.012	0.024	0.485	0.314
181 to 365	-0.072	0.032	2.244	0.012
Interaction between year of death and age at death				
2000 * 0 to 27	0.095	0.055	1.748	0.040
2000 * 121 to 180	-0.025	0.040	0.640	0.261
2000 * 181 to 365	0.132	0.052	2.555	0.005
2010 * 0 to 27	-0.070	0.089	0.785	0.216
2010 * 121 to 180	-0.144	0.068	2.111	0.017
2010 * 181 to 365	0.093	0.080	1.161	0.123
Plurality				
Single	0.010	0.031	0.308	0.379

twin or more	ref			
smoking status during pregnancy				
Smoking	-0.023	0.014	1.625	0.052
no smoking	ref			
mother's race				
White	0.008	0.016	0.524	0.300
Black	ref			
month of prenatal care began				
early (1st Trimester)	-0.025	0.144	1.743	0.041
late or no	ref			
maternal marital status				
married/live together	0.019	0.015	1.234	0.109
single	ref			
gestational age				
(week)	0.003	0.002	1.425	0.077
maternal age				
less or equal to 19	-0.046	0.017	2.689	0.004
older than 19	ref			
maternal education years				
0 to 9	ref			
9 to 11	-0.006	0.029	0.207	0.418
12	-0.029	0.029	0.996	0.160
13 to 15	-0.015	0.033	0.456	0.324
more than 16	-0.056	0.039	1.452	0.073
mu: 5.849 (0.01319)		kappa: 1.534 (0.02584)		

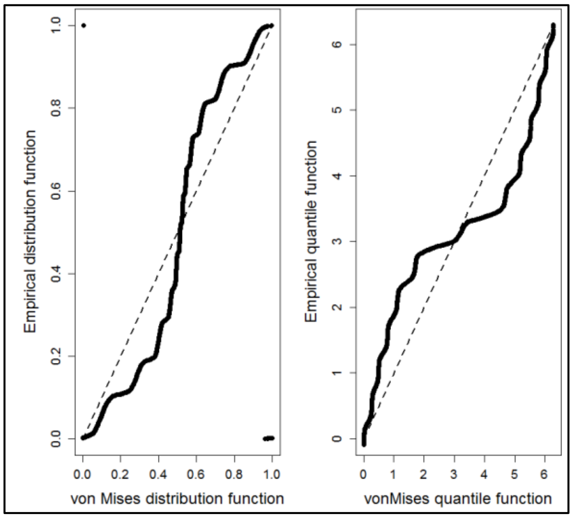

Figure 3.1 *P-P and Q-Q plots of Residuals from the Circular-Linear Model*

Geographical Patterns

We focus on infant deaths occurred after the "Back to Sleep" campaign launched in 1994. Since geographical information was not available for deaths occurred in 2010, we consider those in 2000 only.

Out of 2499 SIDS occurrences, only 1,382 had information on county of occurrence as deaths occurred in counties with a population of less than 250,000were not geo-referenced. Figure 2 shows a map of the 190 counties (out of 3144) for which SIDS information was available. Since an analysis at the county level was not feasible, we considered the state level. Each state was assigned the mean direction calculated from all counties where information was available. Thus, some states had to be excluded, while other states had very little information available.

Figure 3.2 *Map of the counties (black areas) for which information on SIDS was available. Data restricted to year 2000.*

Figure 3.3 shows a colored map of the mean direction of month of death where winter months are in grey, spring months in green, summer months in red, and fall months in yellow. We can see that for states on the west coast the mean direction was sometime in the spring, while for states in the middle of the country and the southeast, the mean direction was sometime between the fall and the winter. However, overall there does not seem to be any geographical correlation. This was supported by the large p-value (0.69) of the Moran's test.

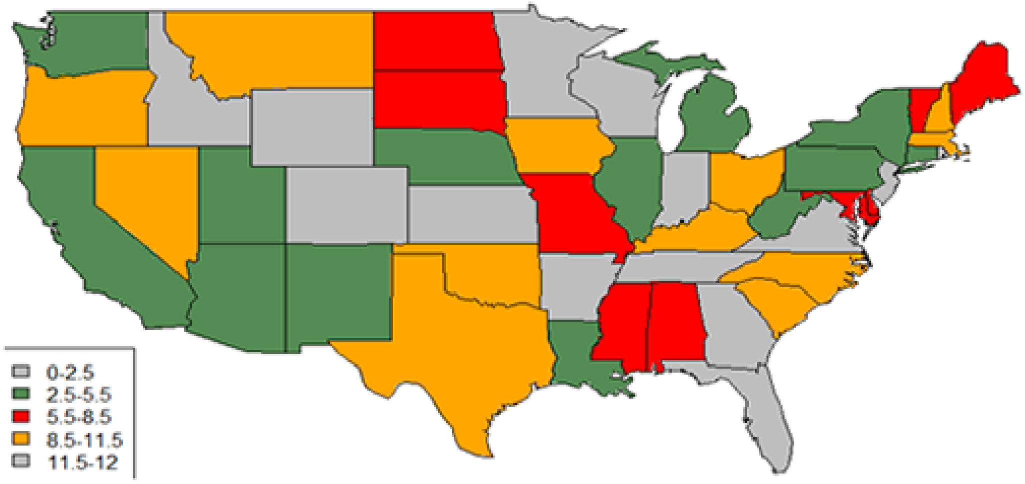

Figure 3.3 *The seasonal pattern of the direction of death month in infants who died from SIDS*

Figure 3.4 shows the map of the residuals from the circular-linear model. The states that have the highest positive residuals are mostly located in the southwest and northeast regions of the U.S., which include Arizona, New Mexico, Pennsylvania, New York and Michigan; the highest negative residuals are found in Oregon, Montana, Maine, Mississippi and Alabama. Again, the Moran's I test did not reveal any spatial autocorrelation in the residuals (p-value 0.50).

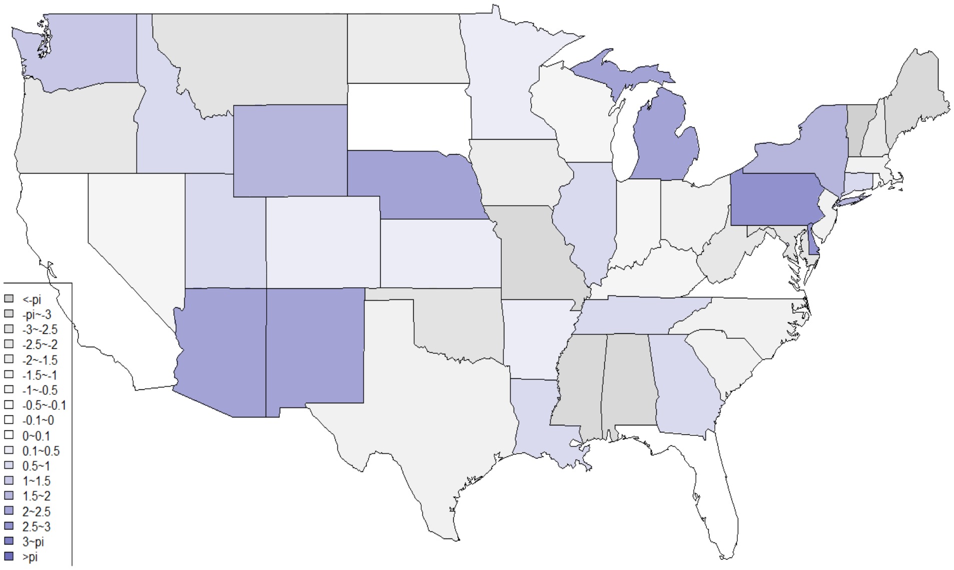

Figure 3.4 *The residuals pattern of the mean direction of death month in infants who died from SIDS*

CHAPTER FOUR: CONCLUSIONS

In this thesis, we applied circular statistics to the death month in infants who died from SIDS. We used a circular statistics package from R software to calculate the mean, standard deviation, and median direction of the death month among infants who died from SIDS. We also modeled the year of death and age of infant death by adjusting for potential risk factors. Additionally, we visualized, by state, the seasonal pattern and residuals pattern for SIDS deaths in months. Below, we conclude the summary statistics, model fitting and geographical residuals pattern, respectively.

In the first chapter, we reviewed the methods available for calculating the mean, median, and standard deviation of circular data. By applying the circular statistical package, we calculated the summary statistics of the direction of the death month among infants who died from SIDS, immaturity, and heart malformations stratified by the years 1990, 2000, and 2010. We also calculated the mean direction of the death month in infants who died from SIDS stratified by age at death in days for these three years. Based on the results, we conclude that most infants who died from SIDS died during the winter with months ranging between November and February. However, infants who died in 2000 at 7-27 days, the average death month was April, and for infants who died at ages 121-180, the average death month was July.

In the second chapter, we applied the circular linear model to our data to check if the mean direction of the death month among infants who died from SIDS is associated with death year and total lived days. The model is defined as the circular linear model;

with the circular statistics of infant death month as the outcome variable, the age of infant death and death years are explanatory variables and controlling for the potential confounders, which are the risk factors previously mentioned. The results show that the the mean direction of the death month in infants who died from SIDS changed significantly for infants who died after 0-27 days and 181-365 days in the year 2000, and those infants who died after 121-180 days in 2010 as compared to the infants who died in the year 1990. The residuals pattern for the QQ plot shows the residuals line does not closely align with the reference line, so the model is not good enough to predict the outcome. Some important factors not included in the model might explain this lack of fit.

In the third chapter, we checked states' geographical residuals pattern of the mean direction of the death month in infants who died from SIDS. The results show that few states have a high residuals pattern (for example, Pennsylvania, New York, Michigan and Maine). The result of Moran's autocorrelation shows that both predicted values and residuals for the direction of the death month in infants who died from SIDS are not correlated between states' geographical location. Moran's autocorrelation shows insignificant correlation between geographical region and the mean direction of month of death in infants who died from SIDS and/or the residuals.

REFERENCES

Agostinelli, C., & Lund, U. (2013). Circular: circular statistics, 2013. *URL http://CRAN. R-project. org/package= circular. R package version 0.4-7.[p 1]*.

Adams, E. J., Chavez, G. F., Steen, D., Shah, R., Iyasu, S., & Krous, H. F. (1998). Changes in the epidemiologic profile of sudden infant death syndrome as rates decline among California infants: 1990–1995. *Pediatrics*, *102*(6), 1445-1451.

Babson, S. G., & Clarke, N. G. (1983). Relationship between infant death and maternal age: comparison of sudden infant death incidence with other causes of infant mortality. *The Journal of Pediatrics*, *103*(3), 391-393.

Barnett, Adrian G., Baker, Peter, & Dobson, Annette (2012) Analysing Seasonal Data. *The R Journal*, 4(1), pp. 5-10

Bivand, R. S., & Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63(18), 1-36.

Bivand, R. S., Hauke, J., and Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, 45(2), 150-179.

Centers for Disease Control and Prevention (2016). Infant Mortality. Retrieved March 02, 2017, from https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm

Drew, T., & Doucet, S. (1991). Application of circular statistics to the study of neuronal discharge during locomotion. *Journal of Neuroscience Methods, 38*(2), 171-181.

Douglas, A.S., Helms, P.J., Jolli0e, I.T. (1998). Seasonality of sudden infant death syndrome in mainland Britain and Ireland 1985–1998. *Archives of Disease in Childhood*, 79, 269–270.

Fisher, N.I. (1995). *Statistical analysis of circular data.* Cambridge University Press.

Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics* (Vol. 5). World Scientific.

Keller, C. A., & Nugent, R. P. (1983). Seasonal patterns in perinatal mortality and preterm delivery. *American Journal of Epidemiology*, *118*(5), 689-698.

Kleinman, J. C., Pierre, M. B., Madans, J. H., LAND, G. H., & SCHRAMM, W. F. (1988). The effects of maternal smoking on fetal and infant mortality. *American Journal of Epidemiology*, *127*(2), 274-282.

Kramer, M. S., Demissie, K., Yang, H., Platt, R. W., Sauvé, R., Liston, R., & Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. (2000). The contribution of mild and moderate preterm birth to infant mortality. *Journal of the American Medical Association*, *284*(7), 843-849.

Kutil, R. (2012). Biased and unbiased estimation of the circular mean resultant length and its variance. *Statistics*, *46*(4), 549-561.

Lee, D. (2013). CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, *55*(13), 1-24.

Luo, Z. C., Wilkins, R., Kramer, M. S., & Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. (2006). Effect of neighbourhood income and maternal education on birth outcomes: a population-based study. *Canadian Medical Association Journal*, *174*(10), 1415-1420.

Maitra, S., & Braun, T. M. (2012). Analysis of Periodontal Data using Circular Statistics.

Mathews, T.J., & MacDorman, M.F. (2013). Infant Mortality Statistics from the 2010 Period Linked Birth/Infant Death Data Set. *National Vital Statistics Report*, 62(8).

McMullen, G. A. (1983). Climatic variation in seasonal patterns of infant mortality.

Mooney, J. A., Helms, P. J., & Jolliffe, I. T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics & Data Analysis*, *41*(3), 505-513.

Murphy SL, Xu JQ, Kochanek KD. Deaths: Final data for 2010. *National vital statistics reports*; vol 61 no 4. Hyattsville, MD: National Center for Health Statistics. 2013.

Paradis, E. (2010). Moran’s autocorrelation coefficient in comparative methods. 2010.

Pewsey, A., Neuhäuser, M., & Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press.

Serfling, R. E. (1963). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports, 78*(6), 494.

Tietjen, G. H., & Kripke, D. F. (1994). Suicides in California (1968–1977): absence of seasonality in Los Angeles and Sacramento counties. *Psychiatry Research*, *53*(2), 161-172.

Valdeś-Dapena, M. A. (1980). Sudden infant death syndrome: a review of the medical literature 1974-1979. *Pediatrics*, *66*(4), 597-614.