Psychological Distress and the Leading Cancers among American Adults: An Evidence from the 2013 National Health Interview Survey

A THESIS
SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF SCIENCE
BY
ABDULLAH MOHAMMED ALBALAWI
DR. MUNNI BEGUM - ADVISOR
BALL STATE UNIVERSITY
MUNCIE, INDIANA
MAY 2015
Psychological Distress and the Leading Cancers among American Adults:

An Evidence from the 2013 National Health Interview Survey

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

ABDULLAH MOHAMMED ALBALAWI

Committee Approval:

___________________________________ ___________________________
Committee Chairperson Date

___________________________________ ___________________________
Committee Member Date

___________________________________ ___________________________
Committee Member Date

Departmental Approval:

___________________________________ ___________________________
Departmental Chairperson Date

___________________________________ ___________________________
Dean of Graduate School Date

BALL STATE UNIVERSITY
MUNCIE, INDIANA
May, 2015
ACKNOWLEDGEMENTS

I would never have been able to finish my thesis project without the guidance of my committee members, help from friends, and support from my family and wife. I would like to express my deepest gratitude to my advisor, Dr. Munni Begum, for her excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research. I would like to thank Dr. Khubchandani, who helped me determining the research topic of studying and applying the statistics application in the field of the psychological distress and its relationship with cancer.

I would like to thank Manzur Rahman Farazi and Morshed Alam, who as good friends, were always willing to help and give their best suggestions. My research would not have been possible without their helps.

Finally, I warmly thank and appreciate my parents, sisters, and brothers. They were always supporting me and encouraging me with their best wishes. I would like to thank my lovely wife, Suha Manqarah. She was always there cheering me up and stood by me through the good and bad times. I understand it was difficult for you to do your master’s degree and take care of our son, Talal, in the same time. I can just say thanks for everything and may Allah give you all the best in return.

Abdullah Albalawi

May 02, 2015
ABSTRACT

THESIS: Psychological Distress and the Leading Cancers among American Adults: An Evidence from the 2013 National Health Interview Survey

STUDENT: Abdullah Mohammed Albalawi

DEGREE: Master of Science

COLLEGE: Sciences and Humanities

DATE: May, 2015

PAGES: 92

Aim: The objective of this study is to determine the prevalence of psychological distress (PD) among cancer patients and to investigate the association of PD with various socio-demographic factors

Methods and results: We consider the 2013 National Health Interview Survey, a large survey of the US non-institutionalized civilian population. PD is determined with a standardized questionnaire (K6). Cancer diagnoses are determined based on self-report. For the purpose of this study, four different types of cancer are selected based on the leading number of deaths caused by them. We fit three commonly used ordinal regression model for PD for both overall cancer patients as well as for patients with four sub-types of (breast, colon, lung, and prostate) cancer. According to the goodness of fit criteria, AIC and deviance, we select the adjacent category model as best model for PD. All the predictors along with afflicted by cancer that were found to be significant in bivariate analysis, are also found to be significant determinants of PD in the multivariate analysis. Subgroup analysis of PD among the subtypes of cancer (breast, colon, lung, and prostate) do not demonstrate any significant determinants of PD.

Conclusion: Psychological distress is found to be significantly prevalent among cancer patients that adds extra burden on them. An important finding is that differential psychological distress level exists across different race of overall cancer patients. However, among the different sub types of cancer (breast, colon, lung, and prostate) PD is not found to be different across race.
# Table of contents

ACKNOWLEDGEMENTS ................................................................................................................. ii

ABSTRACT ...................................................................................................................................... iii

Table of contents ............................................................................................................................. iv

List of Figures ................................................................................................................................. vi

List of Tables ................................................................................................................................. vii

1 Chapter One: Introduction ........................................................................................................ 1
   1.1 Background ......................................................................................................................... 1
   1.2 Motivation of the Research ............................................................................................... 2
   1.3 Objectives ........................................................................................................................... 3
   1.4 Importance from Public Health Point of View ................................................................. 4
   1.5 Outline of the Paper .......................................................................................................... 4

2 Chapter Two: Literature Review .............................................................................................. 6
   2.1 Introduction ......................................................................................................................... 6
   2.2 Review of Literature ......................................................................................................... 6
      2.2.1 Prevalence of Psychological Distress ......................................................................... 7
      2.2.2 Peak Periods of Psychological Distress ..................................................................... 8
      2.2.3 The Risk Factors Associated With Psychological Distress ..................................... 8
   2.3 Definition of Terms ............................................................................................................ 9
      2.3.1 Psychological Distress ............................................................................................... 9
      2.3.2 Conceptual Clarification ............................................................................................ 10
      2.3.3 Kessler Scale (K6) for Psychological Distress .......................................................... 10
      2.3.4 Cancer and the Most Leading New Cancer Cases and Deaths ................................ 11
   2.4 Concluding Remarks ......................................................................................................... 12

3 Chapter Three: Data and Research Methods ......................................................................... 13
   3.1 Introduction ........................................................................................................................ 13
3.2 Data source .................................................................................................................. 13
  3.2.1 Defining the target population.............................................................................. 14
  3.2.2 Description of the variables................................................................................. 14
  3.2.3 Description of the variables................................................................................. 15
  3.2.4 Socio-Demographic information of the study population.................................. 20
3.3 Statistical Method ........................................................................................................ 24
  3.3.1 The Cumulative Odds Model (CO)....................................................................... 25
  3.3.2 The Continuation Ratio Model (CR)..................................................................... 27
  3.3.3 The Adjacent Categories Model (AC)................................................................. 28
3.4 Assessing Model Fit...................................................................................................... 29
3.5 Computational tools..................................................................................................... 30
3.6 Concluding Remarks.................................................................................................... 30
4 Chapter Four: Data Analysis and Results....................................................................... 31
  4.1 Introduction................................................................................................................ 31
  4.2 The Sample Characteristic....................................................................................... 31
  4.3 Bivariate Analysis..................................................................................................... 32
  4.4 Multivariate Analysis (Ordinal Logistic Model).................................................... 37
    4.4.1 Model Selection.................................................................................................... 37
    4.4.2 Model Fitting and Interpretation........................................................................ 38
  4.5 Psychological Distress and Types of Cancer.......................................................... 41
  4.6 Psychological Distress and Race ............................................................................. 43
  4.7 Residual Analysis...................................................................................................... 44
  4.8 Discussion................................................................................................................ 44
5 Chapter Five: Conclusion and Discussion...................................................................... 46
6 The Appendixes.............................................................................................................. 52
List of Figures

Figure 4.1: The distribution of PD among Cancer patients ........................................... 33
Figure 4.2: The distribution of respondents’ gender and PD ...................................... 33
Figure 4.3: The Distribution of Cancer Types and PD ................................................. 41
Figure 4.4: The residual plots of the adjacent model ..................................................... 44
Figure 6.1: The psychological distress distribution ....................................................... 52
Figure 6.2: the distribution of the respondents region and psychological distress .......... 53
Figure 6.3: the distribution of the respondents’ race and psychological distress ........... 53
Figure 6.4: the distribution of the respondents’ age and psychological distress .......... 54
Figure 6.5: the distribution of the respondents’ Marital Status and psychological distress ... 54
Figure 6.6: the distribution of the respondents’ smoking status and psychological distress .... 55
Figure 6.7: the distribution of the respondents’ physical activity and psychological distress ..... 55
Figure 6.8: the distribution of the respondents’ BMI and psychological distress ............. 56
Figure 6.9: the distribution of the respondents’ insurance and psychological distress .......... 56
Figure 6.10: the distribution of the respondents’ Education level and psychological distress ..... 57
Figure 6.11: the distribution of the respondents’ income and psychological distress .......... 57
List of Tables

Table 3.1: The data adjustment table of the dependent variable ........................................... 16
Table 3.2: The data adjustment table of the independent variables ........................................ 17
Table 3.3: Descriptive statistics of the variables .................................................................... 21
Table 3.4: Descriptive statistics of the variables (Cont.) ....................................................... 22
Table 3.5: The descriptive statistics of the four types of cancer based on the gender .............. 23
Table 3.6: Category Comparisons of three Different Ordinal Regression Model Methods, Based on a 3-level Ordinal Outcome (j=1, 2,3) ........................................................................ 25
Table 4.1: Descriptive Statistics for All Variables (Cancer in general), N=17765 ................. 34
Table 4.2: The association between PD of the respondents and predictors for (the cancer sample) .................................................................................................................................................. 36
Table 4.3: The adequacy tool for model selection .................................................................... 38
Table 4.4: The results of the ordinal logistic regression models. .............................................. 39
Table 4.5: The results of the ordinal logistic regression models. (Con’t) ................................ 40
Table 4.6: The cross table between Type of Cancer and PD .................................................... 42
Table 4.7: The cross table between Race and PD .................................................................... 43
Table 6.1: The association between PD of the respondents and predictors for (the breast cancer sub-sample) .................................................................................................................. 58
Table 6.2: The association between PD of the respondents and predictors for (the colon cancer sub-sample) .................................................................................................................. 58
Table 6.3: The association between PD of the respondents and predictors for (the lung cancer sub-sample) .................................................................................................................. 59
Table 6.4: The association between PD of the respondents and predictors for (the prostate cancer sub-sample) ................................................................................................. 59
Table 6.5: Results of ordinal logistic models for the breast cancer sub-sample ....................... 60
Table 6.6: Con’t ....................................................................................................................... 61
Table 6.7: Results of ordinal logistic models for the colon cancer sub-sample ....................... 62
Table 6.8: Con’t ....................................................................................................................... 63
Table 6.9: Results of ordinal logistic models for the lung cancer sub-sample ......................... 64
Table 6.10: Con’t ..................................................................................................................... 65
Table 6.11: Results of ordinal logistic models for the prostate cancer sub-sample .................. 66
Table 6.12: Con’t ..................................................................................................................... 67
Chapter One: Introduction

1.1 Background


For example, a study conducted by Taylor and colleagues found association between psychological distress and the increase of incidence rate of breast cancer among black females in the United States (Alcálá, H. E. (2014)). In addition, other studies show that there are potential consequences of racial discrimination on psychological distress among the patients of selected cancer types (Taylor, T. (2007), and Alcálá, H. E. (2014)). The health outcome of cancer is of
interest because it is the second most common cause of death in the US (Krieger, N. (2011)). For the purpose of this paper, cancers types are selected according to the estimates of the leading new cancer cases and deaths in 2013 in the US. The most frequently diagnosed cancers that occurred in both males and females in 2013 were breast cancer (BC), colon cancer (CC), lung cancer (LC), and prostate cancer (PC) [4]. In particular, this study will estimate and compare the prevalence of overall cancer and that of four leading cancer subtypes namely, BC, CC, LC, and PC. Association with psychological distress among the black and white American adults. To be more precise, the research hypotheses of this study are:

1. The overall incidence of psychological distress in a large representative sample of self-reported cancer patients is at a nominal level in the US compared to rates published by the International Agency for Research on Cancer (IARC).

2. There are no differences in the levels of psychological distress in terms of selected types of cancer.

3. There is no difference in psychological distress in terms of race.

Considering the brief information in the background, we provide an overview of the relevant studies and the major components of their research. The research hypotheses stated in this section are broad goals of this study and we present our specific objectives in the subsection 1.3. Motivation of this study is discussed in the following subsection.

1.2 Motivation of the Research

Mental health is the study of the state of wellbeing of a person’s mind in all respects. Advances in science, medicine, and technology provide comfort and a better life for a handful of
people in the world, but a significant number of the world’s population still necessitates much needed attention to physical and mental health. The incidence of mental health problems and psychological disorders is growing at a steeper rate than ever before; this needs to be addressed with considerable attention. Among cancer patients, mental illness is even more problematic because it can contribute to new health problems or complicate existing ones. The most common method to assess psychological distress is with the Kessler 6-question screening scale (K6) or 10-question screening scale (K10) (Mitchell, C. M. (2011), and Kessler, R. C. (2002)). For the purposes of this paper, this type of measurement of mental health I used because of the availability of information in the data being used. This study is mainly conducted to examine the relationship between psychological distress and four different types of cancers. The specific objectives and goals of this research are discussed in the following subsection.

1.3 Objectives

The main objectives of this research are to determine the prevalence of psychological distress among cancer patients and to investigate the association of psychological distress with various socio-demographic factors. More specifically, the objectives are listed as below:

a) To identify the factors which influence psychological distress of cancer patients in a large representative sample of adults in the US.

b) To examine the variation of psychological distress among individuals with different types of cancer.

c) To make a comparison of the mental health impact of cancer patients with non-diseased people.
1.4 Importance from Public Health Point of View

Mental health is found to be an important health issue for cancer patients. Psychological distress is described as a measure of the mental health illness. Its symptoms are potentially treatable, so it is valuable to understand the possible population health impact of eliminating the distress symptoms in cancer patients.

The findings of this research are expected to provide the motive for future investigations that assess the role of the medical and mental health care professional and the functions of social support as drivers of reductions in psychological distress. Hence, public health awareness should be taken into consideration by using the available information to measure outcomes.

1.5 Outline of the Paper

The thesis is organized as follows. Chapter 1 discusses the background, motivation, and the objectives of the research. Chapter 2 presents the theoretical framework including terminologies and a literature review of the studies that have been conducted on psychological distress and the four most leading types of cancer. The results from the referenced papers are discussed thoroughly in chapter 2. In Chapter 3 we discuss the research methodology, including defining the target population, specification of the dependent and independent variables, study design, and a method to measure psychological distress. Data preparation and management process including missing values checking, editing, coding, transcribing, data cleaning and computation issues are also discussed. Chapter 4 presents results of our study. Results from the univariate, bivariate and multivariate analyses are presented in this chapter. Findings from the tests of our research hypotheses are presented in Chapter 4. We fit models for ordinal response psychological distress to determine the degree of relationship with a number of predictors. We considered
cumulative logit model, proportional odds model, continuation model, and adjacent logistic model. Based on model adequacy checking, we select the best model and fit our data. Results from extensive analysis on the distress level among different types of cancer are presented in this chapter. Chapter 5 presents conclusion about our findings and future direction of our research.
2 Chapter Two: Literature Review

2.1 Introduction

In this chapter, we present a brief review of available literature to assess the impact of psychological distress on two major diseases, such as cancer and heart disease. In particular we review literature on psychological distress and its predictive factors among people living with cancer. We also focus on the prevalence of psychological distress among the cancer patients, defines psychological distress, and provides an overview of the associated factors among people living with cancer in other existing studies. As the purpose of this study, we discuss the prevalence of psychological distress and the associated factors among individuals with different types of cancer in a large representative sample of adults in the US. Section 2.2 presents a discussion of several prior research works on psychological distress in general. Definitions of the important terms related to the study are given in section 2.3. Finally, section 2.5 concludes this chapter.

2.2 Review of Literature

Prior research has been conducted to determine the relationship between psychological distress (PD) and other chronic diseases, as well as its relationship with mortality. For instance, Ferketich and Binkley examined the burden of PD among individuals with different types of heart disease. They found that PD is a significant comorbidity of heart disease (Agresti, A. (2013)). In addition, another prospective study, conducted in London by (Stansfeld, S., Fuhrer, R., Shipley, M., & Marmot, M. (2002)) followed a group comprising of civil service employees in London for five years. In their study, they aimed to test whether there was an increased chronic heart disease (CHD) risk associated with PD. Their findings showed that the experience of psychological
distress increased CHD in males, but not consistently in females (Ferketich, A. K. & Binkley, P. F. (2005)).

A number of studies demonstrate a clear link between the affliction from cancer and PD measured by the symptoms of depression and anxiety (Mosher, C. (2012), Deimling, G. (2006), Pinquart, M. (2010), Yeh, M. (2014), Nakatani, Y. (2013), Liao, Y. (2011), Schulz, R. (1996), Honda, K. (2005), Zobora, J. (2001), Satin, J. (2009), and Massie, M. (2004)). Although past research has mainly focused on distress, in particular, demographic variables associated with PD (depression and anxiety), the issue of psychological screening has become increasingly important (Zabora, J. (2001)). A review of the psychological distress literature concludes that there are no simply identifiable characteristics of patients that can readily predict who has the potential need for psychosocial assessment and intervention (Mosher, C. (2012), and Deimling, G. (2006)). To explore the impact of PD, many research investigations have been carried out in many countries examining various factors that manipulate the psychological distress among cancer patients. The following sections review studies that have focused on the prevalence of PD and its associated factors among individuals with different types of cancer in a large representative sample of adults.

The most recent systematic review about PD among cancer patients was conducted by Yeh, M. and colleagues (2014). They found that the majority of cancer patients face significant psychological and emotional distress at some time during the course of the illness.

2.2.1 Prevalence of Psychological Distress

A recent study measuring the prevalence of PD in a large group of cancer patients (n=4496), revealed that 35.1% of the total sample had significant psychological distress as a result of cancer or cancer-related treatments. In this study, the rates varied from 43.4% in patients with
lung cancer to 29.6% in patients with gynecological cancers, with an overall average of 35.1% for all tumor site groups. Reported prevalence rates of PD also vary widely in research. PD prevalence rates of less than 5% to over 50% have been cited in the literature. There are many possible explanations for the wide variation in the PD prevalence rate (Ogawa et al., (2012)).

Prevalence rates vary according to the tumor site and extent of disease. In addition, prevalence rates would vary and be reported differently depending on which empirical tools were used to measure PD. Patnaik et al., (2011) reported that psychological distress is most frequent and severe among patients with a poorer prognosis and greater patient burden. The rates of psychological distress also vary as the concept is dynamic, and the levels of distress often change at various stages of the illness trajectory and treatment phase.

2.2.2 Peak Periods of Psychological Distress

According to Ogawa et al., (2012), common periods of crisis for cancer patients across the illness trajectory exist, and this can lead to significant PD. These critical periods of vulnerability include the following: while finding a suspicious symptom, during workup, at time of diagnosis, while awaiting the start of treatment, and during changes in treatment, post-treatment, medical follow-up, remission, time of recurrence, disease progression, and the transition to palliative care. Each period of vulnerability along the illness continuum provokes unique existential questions, requiring the use of different coping mechanisms and presenting particular obstacles.

2.2.3 The Risk Factors Associated With Psychological Distress

There are various risk factors that have been linked to mental illness. Numerous investigations were conducted to determine the factors that play a significant role on PD among cancer patients (Cukier, Y. (2013), Zabora, J. (2001), Mosher, C. (2012), Deimling, G. (2006),
Schwarts, M. (1995), Pinquart, M. (2010), Yeh, M. (2014), Forman-Hoffman, V. (2014), Nakatani, Y. (2013), Sunderland, M. (2012), Liao, Y. (2011), Schulz, R. (1996), Honda, K. (2005), Zabora, J. (2001)). For example, studies conducted in order to identify the risk factors of PD in females with breast cancer revealed that the age, marital status, and education of the patient were the significant factors that contributed to the level of PD (Sunderland, M. (2012), Schulz, R. (1996), and Maunsell, E. (1992)). Based on prior studies, the socio-demographic and clinical factors of health (i.e., BMI, insurance status, physical activity, smoking, drinking alcohol, etc.) may enhance distress. A common factor that is found to affect PD and cancer was race/ethnicity (Taylor, T. (2007), Alcalá, H. (2014), and Krieger, N. (2011)). Other studies showed there is correlation between all covariates under consideration and the PD being the response variable. In particular, older women with breast cancer had a significantly higher level of distress (Sunderland, M. (2012)). These factors can be used to identify individuals who may be at risk of experiencing PD.

2.3 Definition of Terms

This section introduces the scientific definitions of the major terms utilized in the study.

2.3.1 Psychological Distress

Psychological distress is a significant problem for patients with cancer at every stage of their disease. Although the concept of psychological distress is frequently used in the field of health science, it is seldom conceptually defined. There are many different manifestations of distress, with anxiety and depression being the most common. In addition, there are many signs of distress which can in turn negatively impact a patient's health status. Assessment and management of psychological distress are imperative in order to ease the burden on patients and to help them cope with their diagnosis and treatment (Bray, F., Jemal, A., Grey, N., Ferlay, J., & Forman, D. (2012)).
2.3.2 Conceptual Clarification

The concept of psychological distress is significant for many patients with cancer who experience some degree of emotional disturbance related to their diagnosis or treatment. Despite the fact that psychological distress is an important issue in the cancer population and is widely investigated in cancer research, the concept remains vague and not well defined. Psychological distress is often defined only by its empirical measurement tools. According to Bray et al., (2012), if a concept is unclear, then any work on which it is based is also unclear. The lack of conceptual clarity may result in unsuitable methodology that could threaten the internal validity of research and perhaps negatively impact patient care.

The term psychological distress is a concept that is frequently communicated in both lay and professional language, but it is seldom defined as a distinct concept. According to Siegel et al., (2014), the term psychological refers to a broad encompassing term: the study of the mind in all of its relationships. A definition of distress in the context of health and social sciences is referred to as "...a subjective response to internal or external stimuli that are threatening or perceived as threatening to the self.” Patnaik et al., (2011) defined the concept of PD as too much or not enough arousal resulting in harm to the mind. Stimuli become distressing only when perceived as such. The authors further conceptualized PD by describing it as an outcome of ongoing negative situational transactions.

2.3.3 Kessler Scale (K6) for Psychological Distress

K6, developed by Kessler et al. (2002), is considered one of the more extensively used measures for PD. In particular, using a 30-day reference period, respondents answered the questions of "how often they felt […] sad, nervous, restless, hopeless, everything was an effort,
and worthless." Therefore, the range of the combination of these six feelings is indicated on a scale between 0 and 24.

2.3.4 Cancer and the Most Leading New Cancer Cases and Deaths

Cancer is one of the significant reasons for death around the world. Around 10.9 million individuals worldwide are diagnosed with cancer and 6.7 million individuals die because of it every year. The World Health Organization (WHO) anticipates that death from disease will continue to rise, with an estimated 11.5 million deaths in 2030 (Siegel, R., Ma, J., Zou, Z., & Jemal, A. (2014)). There are several distinctive types of cancer cases. Cancer occurrence rate may differ in males and females. Lung malignancy is the primary reason for death in men, with a reported yearly mortality of 16% of the expected 6.6 million men diagnosed with lung tumor growth in 2007. Among females, breast cancer is a standout amongst the most habitually diagnosed situations where one out of four women worldwide is diagnosed with breast cancer growth. Figures that incline to cancer incorporate hereditary anomaly, tobacco, liquor, corpulence, dietary components, and natural and word-related dangers (Bray, F. (2012)). According to the National Cancer Institute (NCI), the term cancer has been used for diseases in which the abnormal cells separate without control and attack other tissues. This study investigates the effect of four different types of cancer on psychological distress. In particular, we consider the following four leading types of new cancer cases and deaths according to NCI:

1- Breast cancer (BC): A type of “cancer where it forms in tissues of the breast.” BC occurs more often in females than in males.

2- Lung cancer (LC): A type of “cancer where it conforms in tissues of the lung.”

3- Colon cancer (CC): A type of “cancer forming in tissues of the colon.”
4- Prostate cancer (PC): A type of “cancer that forms in tissues of the prostate (a gland in the male reproductive system found below the bladder and in front of the rectum). Prostate cancer usually occurs in older men.”

We selected the four highest types of cancer based on the estimates in the US in 2013. According to the American Cancer Society, There were 238,590 (28%) new cases of prostate cancer for males; 232,340 (29%) new cases of breast cancer for females; 118,080 (14%) and 110,110 (14%) new cases of lung cancer for males and females, respectively; and 73,680 (9%) and 69,140 (9%) new cases of colon cancer for males and females, respectively. In terms of deaths, there were 87,260 (28%) and 72,220 (26%) deaths from lung cancer for males and females, respectively; 29,720 (10%) deaths from prostate cancer for males; 39,620 (14%) deaths from breast cancer for females, and 26,300 (9%) and 24,530 (9%) deaths from colon cancer for males and females, respectively.

2.4 Concluding Remarks

Many studies have been carried out in different countries to find the impact of PD on chronic disease, such as cardiovascular diseases and cancer. In the previous sections we reviewed some important studies which are pertinent to our study and that help us meet our research objectives. In Chapter 3, we discuss the methodology to address our research hypotheses and specific objectives of our study.
3 Chapter Three: Data and Research Methods

3.1 Introduction

Psychological distress among cancer patients bears an important public health issue due to the both short term and long term impact on the wellbeing of these patients. The psychological distress adds additional burden on the cancer patients. This study attempts to identify the socio-demographic factors behind psychological distress among different types of cancer patients. We are also interested to assess if there is any differential statistical pattern in psychological distress across patients afflicted by four common types of cancer: breast, colon, prostate and lung. Data for this study has been considered from the National Health Interview Survey (NHIS). Due to the ordinal nature of our response variable, psychological distress, an ordinal regression modeling approach is suitable to address our specific research hypotheses. In this chapter we discuss the data source and a brief review on the available regression models for ordinal data.

3.2 Data source

Data are obtained from the 2013 National Health Interview Survey (NHIS), which is a large annual survey conducted on a random sample of individuals living in the United States. This data is a cross-sectional household survey of the US population conducted annually by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). Interviews were carried in respondents’ houses by face to face. But, for those who were not home a follow-up was performed over the telephone. This survey employed a randomly selected, stratified, multistage area design that is nationally representative sample of households. Data from the NHIS are organized into a number of different types of files. For the purpose of our study, we considered the sample adult module, which contains health-related information on a randomly
selected adult in the family. Information from the person file is used by merging the two files together based on the person number in each family. According to National Center for Health Statistics during the data collection period in 2013, there were 42,294 eligible individuals, of which 34,557 (81.7%) were agreed to be interviewed.

3.2.1 Defining the target population

We used the basic file for adult as well as the person file for extracting additional variables. The files have been merged based on Household Number, Family Number, and Person Number (within family). The target sample is the participants who are 18 years and older. The listwise deletion method used to omit randomly the missing values. The total number of individuals that completed the questions in the survey is 17765 (%51.40 are female).

3.2.2 Description of the variables

The dependent variable:

The dependent variable is psychological distress (PD) that has been widely used and well assessed by Kessler 6 (K6), which has been developed exactly for the NHIS (Mitchell, C. (2011), Kessler, R. (2002), Sunderland, M., & Andrewas, G. (2011), and Prochaska, J. (2012)). This measure K6 contains a six-item instrument, for each part the respondents were asked how frequently they experienced symptoms of psychological distress (sad, nervous, restless, hopeless, everything was an effort, and worthless) during the past 30 days. Each question has a 5-point scale with ranges from 0= ‘none of the time’ to 5=‘all of the time’ in the NHIS, but as established for K6 that the response should be scored between 0 and 4 on the Likert scale. As a result, the total of the response scores is ranged between 0 and 24. We group these scores into three groups as the following: 0-7 indicating a low level of PD, 8-12 indicating a moderate level of PD, and 13-24 indicating high PD.
The independent variables:

The predictors selected here are well established in the literature. There are various socio-demographic and other characteristics variables. Those variables examined in the current research are shown in (Table 3.3 & Table 3.4) with description and percentages, where all of the predictors were categorized. These include: region (coded 1=’northeast’, 2=’Midwest’, 3=’south’, and 4=’west’), sex (coded 1=’male’, and 0=‘female’), race categorized as (1=’white’, 2=’Black/African American’, 3=’Asian’, and 4=’others’), age in years (coded as 1=’18-30’, 2=’31-64’, and 3=’65+’), marital status (coded as 1=’Married’, and 2=’Unmarried’), education (coded as 1=’High school or below’, 2=’More than high school’), physical activities (coded as 1=’Yes’, and 0=’No’) [the participants been asked “How often do you do VIGOROUS leisure-time physical activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?”], alcohol (coded as 1=’Yes’, and 0=’No’) [the participants been asked “In ANY ONE YEAR, have you had at least 12 drinks of any type of alcoholic beverage?”], smoking (coded as 1=’Yes’, and 0=’No’) [the participants been asked “Do you NOW smoke cigarettes every day, some days or not at all?”], and insurance status (coded as 0= Not covered’, and 1=’Covered’) [the definition of uninsured matches that used in Health United States], and income (coded as 1=’ $0-$34,999’, 2=’ $35,000-$74,999’, 3=’ $>= 75,000’).

3.2.3 Description of the variables

In tables 3.1 & 3.2, we show the adjustment on the data coding. The original codes were done by National Health Interview Survey (NHIS).
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>The original code of the variables</th>
<th>Recode the variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>1= all of the time</td>
<td>0= None of the time</td>
</tr>
<tr>
<td></td>
<td>2= Most of the time</td>
<td>1= A little of the time</td>
</tr>
<tr>
<td></td>
<td>3= Some of the time</td>
<td>2= Some of the time</td>
</tr>
<tr>
<td>SAD, NERVOUS, RESTLESS, HOPELESS, EFFORT, WORTHLESS</td>
<td>4= A little of the time</td>
<td>3= Most of the time</td>
</tr>
<tr>
<td></td>
<td>5= None of the time</td>
<td>4= all of the time</td>
</tr>
<tr>
<td></td>
<td>7= Refused</td>
<td>7,8,9=missing</td>
</tr>
<tr>
<td></td>
<td>8= Not ascertained</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9= Don’t know</td>
<td></td>
</tr>
<tr>
<td>Psychological Distress (K6)</td>
<td>PD= SAD+NERVOUS+RESTLESS+HOPELESS+EFFORT+WORTHLESS</td>
<td>(0-7): 1=Low distress</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8-12): 2=Moderate distress</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(13-24): 3=High distress</td>
</tr>
</tbody>
</table>
Table 3.2: The data adjustment table of the independent variables

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>The original code of the variables</th>
<th>Recode the variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variables</td>
<td>1=Northeast</td>
<td>1=Northeast</td>
</tr>
<tr>
<td>REGION</td>
<td>2=Northeast</td>
<td>2=Northeast</td>
</tr>
<tr>
<td>REGION</td>
<td>3=South</td>
<td>3=South</td>
</tr>
<tr>
<td>REGION</td>
<td>4=West</td>
<td>4=West</td>
</tr>
<tr>
<td>REGION</td>
<td>1=male</td>
<td>1=Male</td>
</tr>
<tr>
<td>REGION</td>
<td>2=female</td>
<td>2=Female</td>
</tr>
<tr>
<td>REGION</td>
<td>01 =White only</td>
<td>1=White</td>
</tr>
<tr>
<td>REGION</td>
<td>02 =Black/African American only</td>
<td>2=Black/African American</td>
</tr>
<tr>
<td>REGION</td>
<td>03 =AIAN only</td>
<td>3=Asian</td>
</tr>
<tr>
<td>REGION</td>
<td>04 =Asian only</td>
<td>4=Others</td>
</tr>
<tr>
<td>REGION</td>
<td>05 =Race group not releasable*</td>
<td></td>
</tr>
<tr>
<td>REGION</td>
<td>06 =Multiple race</td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>00 Under 1 year</td>
<td>1=18-30</td>
</tr>
<tr>
<td>AGE</td>
<td>01-84 1-84 years</td>
<td>2=31-64</td>
</tr>
<tr>
<td>AGE</td>
<td>85 85+ years</td>
<td>3=65+</td>
</tr>
<tr>
<td>AGE</td>
<td>0 Under 14 years</td>
<td></td>
</tr>
<tr>
<td>MARITL</td>
<td>1 Married - spouse in household</td>
<td></td>
</tr>
<tr>
<td>MARITL</td>
<td>2 Married - spouse not in household</td>
<td></td>
</tr>
<tr>
<td>MARITL</td>
<td>3 Married - spouse in household unknown</td>
<td>(1,2,3): 1=married</td>
</tr>
<tr>
<td>MARITL</td>
<td>4 Widowed</td>
<td>(4,5,6,7,8) : 0=unmarried</td>
</tr>
<tr>
<td>MARITL</td>
<td>5 Divorced</td>
<td>9= missing</td>
</tr>
<tr>
<td>MARITL</td>
<td>6 Separated</td>
<td></td>
</tr>
<tr>
<td>MARITL</td>
<td>7 Never married</td>
<td></td>
</tr>
<tr>
<td>MARITL</td>
<td>8 Living with partner</td>
<td></td>
</tr>
</tbody>
</table>

17
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Less than/equal to 8th grade</td>
</tr>
<tr>
<td>02</td>
<td>9-12th grade, no high school diploma</td>
</tr>
<tr>
<td>03</td>
<td>High school graduate/GED recipient</td>
</tr>
<tr>
<td>04</td>
<td>Some college, no degree</td>
</tr>
<tr>
<td>05</td>
<td>AA degree, technical or vocational</td>
</tr>
<tr>
<td>06</td>
<td>AA degree, academic program</td>
</tr>
<tr>
<td>07</td>
<td>Bachelor's degree</td>
</tr>
<tr>
<td>08</td>
<td>Master's, professional, or doctoral degree</td>
</tr>
<tr>
<td>97</td>
<td>Refused</td>
</tr>
<tr>
<td>98</td>
<td>Not ascertained</td>
</tr>
<tr>
<td>99</td>
<td>Don't know</td>
</tr>
</tbody>
</table>

**EDUC**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>

**CANCER**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>

**SMOKE**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>ACTIVITY</td>
<td>0=Never</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALCOHOL</th>
<th>1=Yes</th>
<th>2=No</th>
<th>3=Refused</th>
<th>4=Not ascertained</th>
<th>5=Not ascertained</th>
<th>6=Don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI</th>
<th>0001-9994</th>
<th>0.01-99.94</th>
<th>9995 99.95+</th>
<th>9999 Unknown</th>
<th>&lt;1850: 1=Underweight (&lt;18.5)</th>
<th>(1850 &lt;= BMI &lt;2499): 2=Normal (18.5-24.99)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>2500 &lt;= BMI &lt;3000: 3=Overweight (25-29.99)</td>
<td>BMI &gt;=3000: 4=Obese (&gt;=30)</td>
<td>9999=missing</td>
<td>1=Not covered</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INSURANCE</th>
<th>1=Covered</th>
<th>2=Covered</th>
<th>3=Refused</th>
<th>4=Not ascertained</th>
<th>5=Not ascertained</th>
<th>6=Don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 3.2.4 Socio-Demographic information of the study population

The descriptive statistics of the variables used in this study is displayed in Tables 3.3 and 3.4. There were 90% of the respondents having low distress level, followed by 7% and 3% for moderate level and high level, respectively. 36% of them were from the South region while just 16% were from the Northeast whereas 21% and 27% were form the Midwest and the West. Most of the participants are white American by 76% of the sample. There were 12086 (68%) of the respondents in the aged 31-64 years old. More than 65% of the whole sample are having high school or above. The most important factor here is the cancer where people who have cancer are
923 (5%). Smokers were about 38% of the participants while 71% of them are alcoholic. Approximately, 20% of the respondents said they don’t have insurance. Finally, more than half of them were in the group income of 0$-34,999 annually.

Table 3.3: Descriptive statistics of the variables

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Categories</th>
<th>N= 17765</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>Psychological Distress (K6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low distress</td>
<td>16018</td>
<td>90.17</td>
</tr>
<tr>
<td></td>
<td>Moderate distress</td>
<td>1288</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>High distress</td>
<td>459</td>
<td>2.58</td>
</tr>
<tr>
<td>Independent variables</td>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Northeast</td>
<td>2838</td>
<td>15.98</td>
</tr>
<tr>
<td></td>
<td>Midwest</td>
<td>3693</td>
<td>20.79</td>
</tr>
<tr>
<td></td>
<td>South</td>
<td>6450</td>
<td>36.31</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>4784</td>
<td>26.93</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>8634</td>
<td>48.60</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>9131</td>
<td>51.40</td>
</tr>
<tr>
<td></td>
<td>Race/ethnic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>13527</td>
<td>76.14</td>
</tr>
<tr>
<td></td>
<td>Black/African American</td>
<td>2548</td>
<td>14.34</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>1107</td>
<td>6.23</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>583</td>
<td>3.28</td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characteristics</td>
<td>Categories</td>
<td>N= 17765</td>
<td>%</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>Education</td>
<td>High School or below</td>
<td>5723</td>
<td>32.22</td>
</tr>
<tr>
<td></td>
<td>More than High school</td>
<td>12042</td>
<td>67.78</td>
</tr>
<tr>
<td>Cancer</td>
<td>Yes</td>
<td>923</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>16842</td>
<td>94.80</td>
</tr>
<tr>
<td>Smoke</td>
<td>Yes</td>
<td>6684</td>
<td>37.62</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>11081</td>
<td>62.38</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Yes</td>
<td>12691</td>
<td>71.44</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>5074</td>
<td>28.56</td>
</tr>
<tr>
<td>Physical activity</td>
<td>No</td>
<td>8065</td>
<td>45.40</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>9700</td>
<td>54.60</td>
</tr>
<tr>
<td>BMI1</td>
<td>Underweight (&lt;18.5)</td>
<td>240</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>Normal (18.5-24.99)</td>
<td>5963</td>
<td>33.57</td>
</tr>
<tr>
<td></td>
<td>Overweight (25-29.99)</td>
<td>6142</td>
<td>34.57</td>
</tr>
<tr>
<td></td>
<td>Obese (&gt;=30)</td>
<td>5420</td>
<td>30.51</td>
</tr>
</tbody>
</table>

Insurance status
Body Mass Index (BMI) is a simple index of weight-for-height that is commonly used to classify underweight, overweight and obesity in adults. It is defined as the weight in kilograms divided by the square of the height in meters (kg/m²).

In addition to these, four different types of cancer condition included for the participants who reported having cancer; breast cancer (coded 1='Yes', 0='No'), colon cancer (coded 1='Yes', 0='No'), lung cancer (coded 1='Yes', 0='No'), prostate cancer (coded 1='Yes', 0='No'), and other types of cancer (coded 1='Yes', 0='No'), all are shown in (Table 3.5).

Table 3.5: The descriptive statistics of the four types of cancer based on the gender.

<table>
<thead>
<tr>
<th>Types of cancer</th>
<th>Categories</th>
<th>Gender</th>
<th>N=17765</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Yes</td>
<td>4</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>357</td>
<td>413</td>
</tr>
<tr>
<td>colon cancer</td>
<td>Yes</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>347</td>
<td>547</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Yes</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>353</td>
<td>554</td>
</tr>
<tr>
<td>prostate cancer</td>
<td>Yes</td>
<td>82</td>
<td>0</td>
</tr>
</tbody>
</table>
The ordinal logistic regression model is an extension of the logistic regression model of binary response (Dobson, A. J. (2001)). Logistic regression, also called a logit model, is applied to model dichotomous outcome variables. In the logistic regression model, the log odds of the outcome are modeled as a linear combination of the predictor variables. Epidemiological and most of the health-related studies often depend on the ordered outcomes. Since we have an obvious natural order among the response categories [psychological distress (PD) = (1=Low, 2=Moderate, 3=High)], the ordinal logistic model will be taken into consideration.

Although ordinal response can be simple and meaningful, many researchers are challenged to handle ordinal responses in terms of choosing the appropriate method. Agresti (2013) and O’Connell (2006) reviewed the most commonly used methods to model ordinal responses in detail. For the purpose of this study, we discuss the three common methods for analyzing ordinal responses. These methods can be selected based on the research question. These include: the Cumulative Odds Model (CO), sometimes referred by the proportional odds model, the Continuation Ratio Model (CR), and the Adjacent Categories Model (AC). Table 3.6 displays the comparison between the three methods based on a 3 level ordinal outcome for our study.
Table 3.6: Category Comparisons of three Different Ordinal Regression Model Methods, Based on a 3-level Ordinal Outcome (j=1, 2, 3)

<table>
<thead>
<tr>
<th>Cumulative Odds (ascending) $P(Y \leq j)$</th>
<th>Cumulative Odds (descending) $P(Y \geq j)$</th>
<th>Continuation Ratio $P(Y &gt; j/Y \geq j)$</th>
<th>Adjacent Categories $P(Y = j + 1/Y = j$ or $Y = j + 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1 vs. all above</td>
<td>Category 3 vs. all below</td>
<td>Categories 2 and 3 vs. 1</td>
<td>Category 2 vs. 1</td>
</tr>
<tr>
<td>Categories 1 and 2 combined vs. 3</td>
<td>Categories 3 and 2 combined vs. 1</td>
<td>Categories 3 vs. 1</td>
<td>Category 3 vs. 2</td>
</tr>
</tbody>
</table>

The following sections give a brief overview of each method:

3.3.1 The Cumulative Odds Model (CO)

The Cumulative odds model (CO), also known as the proportional odds model (PO), is indicated when an originally continuous response variable is later grouped. The CO model is the most frequently used ordinal regression model, mostly in the educational sciences. Anath, C. & Klenbaum (1997) discussed that the CO model was first proposed by Walker and Duncan (1967) and later developed by McCullagh (1980) and called the proportional odds models. The CO model is the default setting for ordinal regression model by most of the statistical software, i.e. SAS in our case. Anath, C. & Klenbaum (1997) also reviewed several other statistical models for ordinal response. Six different models were considered for analyzing ordinal response. In their work, examples were given to illustrate the fit of these models to large data from a prenatal health registry. However, they suggested that the CO is the ideal choice for the epidemiological and biomedical applications. For testing the CO or PO assumption, chi-square score test is used to assess the assumption. If the p-value is not significant at $\alpha=0.05$ with respect to the chi-square distribution, this implies the model fits the data well (Christensen, R. (2013). In this method, the CO with $J$ categories is divided to $J-1$ logit equations. We use the category ordering by forming logits of cumulative probabilities,
\[ P(Y \leq j/x) = \pi_1(x) + \cdots + \pi_j(x), \quad j = 1, \ldots, J. \]

The cumulative logits are defined as:

\[ \text{logit}[P(Y \leq j/x)] = \log \frac{P(Y \leq j/x)}{1-P(Y \leq j/x)} = \log \frac{\pi_1(x) + \cdots + \pi_j(x)}{\pi_{j+1}(x) + \cdots + \pi_{J-1}(x)}, \quad j = 1, \ldots, J-1 \quad (3.1) \]

The CO or PO Form of cumulative logit model is:

\[ \text{logit}[P(Y \leq j/x)] = \alpha_j + \beta^T x \quad (3.2) \]

Where,

\( \alpha_j \) (Intercepts) can differ.

\( \beta \) (Slope) is constant.

Each cumulative logit has its intercept. The \( \{\alpha_j\} \) are increasing in \( j \) because \( P(Y \leq j/x) \) increases in \( j \) for fixed \( x \). As mentioned above, this model assumes that \( \beta \) have the same affect for each logit. For more illustration, in our study a three-category outcome will have two binary logit equations based on the following comparisons: 1 vs. 2&3, 1&2 vs. 3. The CO is used to predict the odds of a person being at or below any particular level of Psychological Distress (PD). PD categories are coded as: [1=low, 2=moderate, 3=high]. The following CO model was fitted to our data using the equation form (3.2):

**Low vs. (Moderate & High):**

\[ \text{logit}[P(Y \leq 1/x)] = \ln \frac{\pi_1}{\pi_2 + \pi_3} = \alpha_1 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \beta_7x_7 + \beta_8x_8 + \beta_9x_9 + \beta_{10}x_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{13}x_{13} + \beta_{14}x_{14} + \beta_{15}x_{15} + \beta_{16}x_{16} + \beta_{17}x_{17} \quad (3.4) \]
(Low & Moderate) vs. High:

\[
\text{logit}[P(Y \leq 2 / x)] = \ln \left( \frac{\pi_1 + \pi_2}{\pi_3} \right) = \alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \\
\beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14} + \beta_{15} x_{15} + \\
\beta_{16} x_{16} + \beta_{17} x_{17}
\] (3.5)

Where \(\pi_1, \pi_2, \pi_3\) are the probability of being in low level, moderate level and high level, respectively. \(\beta_1, ..., \beta_{17}\) are the slope parameters and \(x_1, ..., x_{17}\) are the covariate factors as defined in the next chapter with the results. More details of this model will be in the next chapter (Chapter 4: Data Analysis and Results).

The proportional odds assumption can be tested by a score test obtained by the Statistical software (e.g. SAS using PROC LOGISTIC provides a score test for the proportional odds assumption). This assumption implies that the independent variables have the same effect on the odds for each category of the model of interest. The proportional odds assumptions is valid when \((P - value > 0.05)\), meaning that the effect of the independent variables is not statistically different across the categories cumulative splits for the data (O’Connell, A. (2006)).

3.3.2 The Continuation Ratio Model (CR)

In this section, we show another alternative model for analyzing the ordinal responses. This called the Continuation Ratio Model (CR), which is to model the ratios of probabilities.

\[
\frac{\pi_1}{\pi_2}, \frac{\pi_1 + \pi_2}{\pi_3}, ..., \frac{\pi_1 + \cdots + \pi_{J-1}}{\pi_J}
\]

OR
These are called “continuation–ratio logits”. Thus, the model can be written as:

\[
\log \left( \frac{\pi_j}{\pi_{j+1} + \cdots + \pi_J} \right) = x_j^T \beta_j \quad (3.6)
\]

The continuation ratio model provides the log odds of the response being in the category \( j \).

For our study, \( J=3 \), we can estimate the odds of the respondents PD as “Low” vs. “Moderate” and the odds of these levels are in “Low” and “Moderate” versus “High” using:

\[
\log \left( \frac{\pi_1}{\pi_2} \right) \text{ and } \log \left( \frac{\pi_1 + \pi_2}{\pi_3} \right)
\]

The CR may be easier than CO in terms of interpretation if we are interested in finding the probability for individual categories \( \pi_j \).

### 3.3.3 The Adjacent Categories Model (AC)

The last alternative model we discuss in this study for analyzing ordinal responses is called the Adjacent Categories Model (AC). This model considers the ratios of probabilities for successive categories, for instance:

\[
\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_3}, \ldots, \frac{\pi_{J-1}}{\pi_J}
\]

The AC model can be written as:

\[
\log \left( \frac{\pi_j}{\pi_{j+1}} \right) = x_j^T \beta_j \quad (3.7)
\]
Which is equivalent to,

\[
\log(P(y_i = j)/P(y_i = j + 1)) = \log\left(\frac{\pi_j}{\pi_{j+1}}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1}
\]

This model assumes that the effect of each independent variable to be the same for all adjacent pairs of categories. The parameters \(\beta_k\) are interpreted as odds ratios using \(OR = \exp(\beta_k)\).

The question remains is that which of these models would be appropriate for our data (for more explanations please refer to Agresti, A. (2013), O’Connell, A. (2006), and Dobson, A. (2001)). For the purpose of this study, we fit all three models and assess the goodness of fit to the data with deviance measures.

### 3.4 Assessing Model Fit

In terms of the model diagnostics, there is a study that focused specifically on data diagnostics for ordinal outcomes. O’Connell and Liu (2011) reviewed the strategies for model diagnosis that may be helpful in examining model assumptions and also in identifying unusual cases for proportional odds models. This paper discussed the methods to assess the ordinal logistic regression model performance. In particular, they provided a similar example for these diagnostic methods to "the prediction of proficiency in early literacy for children drawn from the kindergarten cohort of the Early Childhood Longitudinal Study". After making some comparisons between the strategies, they concluded the paper with the following guidelines: First, residuals from ordinary least squares (OLS) and Binary Logistic Models give a good first look at the possible for unusual cases from the ordinal model. Another recommendation is that neither OLS nor the binary logistic analysis could catch all the unusual values. Hence, investigators should be cautious regarding the possibility of misleading cases by plotting as many different diagnostic methods as possible.
Thirdly, graphical strategies should help the investigator more about the data than a summary statistic. The last recommendation is that investigators should be aware to include residual diagnostics in all the presented or published papers. Thus, in this study we would follow the strategies to assess our model of interest.

3.5 Computational tools

All computations are conducted using SAS version 9.3 (SAS Institute, Cary, NC) and the R computing environment (Version 3.11, The R Project). SAS is used to manage the data and create analysis variables. The standard SAS procedure for ordinal logistic regression model considered here is PROC LOGISTIC. In R, we used the VIGAM package to fit all of the three models (Code is provided in Appendix C).

3.6 Concluding Remarks

In this chapter, we describe the methodology of the entire research work. The data used in the study are obtained from a large representative survey conducted to study the issue of the distress and cancer. Therefore, we covered a description of the variables used in the study. We also briefly discussed the statistical methods for analyzing the ordinal responses including three different models (Cumulative odds model, Continuation Ratio model, and Adjacent ratio model) followed by the assessing model fit section. Finally, we discussed the computational tools such as, SAS version 9.3 and the R computing environment (Version 3.11, The R Project) for data management and analysis.
4 Chapter Four: Data Analysis and Results

4.1 Introduction

In the analysis process, we start with simple summary statistics for the selected demographic and risk factor predictors. We also employ few graphical plots for the response variable and the selected predictors to have an idea about the distribution pattern of the study-variables. Bivariate analyses are performed to examine the association between the response variable and each of the selected predictors. Only those found to be significant are entered into logistic regression model to determine the degree of association. Based on literature, we have performed those types of logistic regression model; CO, AC, and CR. Using AIC, Deviance and so on, the best model selected to fit the data. Diagnostic tests are performed to assess the goodness-of-fit and the assumptions pertaining to ordinal logistic regression model. Further exploratory analysis is performed where it is thought to be necessary.

In addition, four self-reported cancer diagnoses examined separately. The four different types of cancer are Breast Cancer (BC), Colon Cancer (CC), Prostate Cancer (PC), and Lung Cancers (LC). Each cancer type has been analyzed in two parts: bivariate and multivariate analysis.

4.2 The Sample Characteristic

A total of 17765 NHIS participants were included in this study who completed the survey. Of the total 51% were female and 49% were male. Table 4.1 shows general demographic and socioeconomic characteristics of the sample based on the PD levels. The sample consists of 9131 (51.40%) female and they were at the group aged between 31 years and 64 years old by 68.03%. Approximately, the majority says that they are white American 13527(76.14%), followed by black/African American 2548(14.34%). About 55.46% of the participants were not married at the
time of survey. There is 36.31% of the respondents were from the south region followed by 26.93% from the west region. The education level of the respondents’ show that approximately 12042 (67.78%) have high school or higher. More than 62% says that they did not smoke. However, more than half of the respondents did some physical activities about 54.60%. Among the respondents, the majority has insurance about 80.15%. 34.57% of participates says that they are overweight. Among respondents 9550 (53.76%) of them were in the category of the income between 0$ and %34,999 annually.

Moreover, the proportions of being in the high level of PD for female was 1.60% higher comparing to the male 0.99%. We have noticed also unmarried people 1.86% have higher PD than married people 0.73%. Simple summary statistics (frequency and percentages) are calculated and are presented in table (4.1) for the all the variables.

At the beginning of analysis, bivariate analysis (based on Pearson Chi-square test) has been performed to examine the association between response variable and each of the selected predictors.

4.3 Bivariate Analysis

Figures 4.1 and 4.2 below show that the distribution of cancer patents in the red color among the PD levels and shows the proportion of having a cancer getting greater in high PD. We show also the distribution of the gender of the participants and it shows the proportion of PD levels in each. Moreover, we constructed bar graphs for all of the variables and show the distributions of PD among each of them (refer to the Appendix A)
Figure 4.1: The distribution of PD among Cancer patients

The Distribution Psychological Distress among Cancer Patients

Figure 4.2: The distribution of respondents’ gender and PD

The Distribution the respondents gender and Psychological Distress

Bivariate analysis explores the concept of association between two variables. Association is based on how two variables simultaneously change together. Bivariate descriptive statistics involves simultaneously analyzing (comparing) two variables to determine if there is a relationship
between the variables. The purpose of this chapter is to go beyond the univariate statistics, in which the analysis focuses on one variable at a time.

Table 4.1: Descriptive Statistics for All Variables (Cancer in general), N=17765

<table>
<thead>
<tr>
<th>The variables</th>
<th>Category</th>
<th>The Psychological Distress Level</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low (n=16018)</td>
<td>Moderate (n=1288)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>90.17%</td>
<td>7.25%</td>
</tr>
<tr>
<td>Region</td>
<td>Northeast</td>
<td>2556 (14.39%)</td>
<td>205(1.15%)</td>
</tr>
<tr>
<td></td>
<td>Midwest</td>
<td>3341(18.81%)</td>
<td>265(1.49%)</td>
</tr>
<tr>
<td></td>
<td>South</td>
<td>5830(32.82%)</td>
<td>449(2.53%)</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>4291(24.15%)</td>
<td>369(2.08%)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>7933(44.66%)</td>
<td>526(2.96%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8085(45.51%)</td>
<td>762(4.29%)</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>12212(68.74%)</td>
<td>976(5.49%)</td>
</tr>
<tr>
<td></td>
<td>Black/African</td>
<td>2284(12.86%)</td>
<td>199(1.12%)</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>1022(5.75%)</td>
<td>62(0.35%)</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>500(2.81%)</td>
<td>51(0.29%)</td>
</tr>
<tr>
<td>Age</td>
<td>(18-30)</td>
<td>4114(23.16%)</td>
<td>363(2.04%)</td>
</tr>
<tr>
<td></td>
<td>(31-64)</td>
<td>10870(61.19%)</td>
<td>876(4.93%)</td>
</tr>
<tr>
<td></td>
<td>65+</td>
<td>1034(5.82%)</td>
<td>49(0.28%)</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Married</td>
<td>7344(41.34%)</td>
<td>439(2.47%)</td>
</tr>
<tr>
<td></td>
<td>Unmarried</td>
<td>8674(48.83%)</td>
<td>849(4.78%)</td>
</tr>
<tr>
<td>Cancer</td>
<td>Yes</td>
<td>792(4.46%)</td>
<td>87(0.49%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>15226(85.71%)</td>
<td>1201(6.76%)</td>
</tr>
<tr>
<td>Smoke</td>
<td>Yes</td>
<td>5840(32.87%)</td>
<td>593(3.34%)</td>
</tr>
</tbody>
</table>
We found in the above Table 4.1 that 4% of the respondents having cancer have low distress level. We also found 44 of cancer patients have high distress level. Interestingly, we explored that people who are not married having high distress level (2%) compare to married people (0.73%). 73% of the people who have insurance have low distress level while 17% of people who do not have insurance are having low distress level. In terms of the income, people who have low income tend to have high distress compared to high class people.

We use cross tabulation technique for finding association among variables. Initially we test that two variables are associated or not. If two variables are associated then we find strength of this association by appropriate statistic. Cross tabulations can be produced by a range of statistical
packages, including some that are specialized for the task. The hypothesis to assess the association between response variable and each of the predictors as follows:

\[ H_0: \text{There is no association between PD and } i^{th} \text{ predictors} \]

Vs.

\[ H_1: \text{There is an association between PD and } i^{th} \text{ predictors} \]

Pearson Chi-square has been performed at 5% level (shown in Table 4.2). It shows most of the predictors (except mother's marital status, smoking, pre-pregnancy diabetes, gestational diabetes, pre-pregnancy hypertension, and previous cesarean deliveries) are significantly associated with the response variable.

Table 4.2: The association between PD of the respondents and predictors for (the cancer sample)

<table>
<thead>
<tr>
<th>The variables</th>
<th>d.f</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>6</td>
<td>3.4235</td>
<td>0.7541</td>
</tr>
<tr>
<td>SEX</td>
<td>2</td>
<td>56.7093</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Race</td>
<td>6</td>
<td>29.4704</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age (Group in years)</td>
<td>4</td>
<td>26.5571</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Marital Status</td>
<td>2</td>
<td>118.3031</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Cancer</td>
<td>2</td>
<td>26.3698</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Smoking status</td>
<td>2</td>
<td>105.0631</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>2</td>
<td>68.8511</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>2</td>
<td>0.1908</td>
<td>0.9090</td>
</tr>
<tr>
<td>BMI</td>
<td>6</td>
<td>43.7622</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Insurance</td>
<td>2</td>
<td>176.8340</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>68.8976</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Income</td>
<td>4</td>
<td>306.9797</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
All the significant predictors are then included (except the region and Alcohol consumption) in the CO, CR and AC model. We applied these three models for estimating regression parameter ($\beta$) including p-values based on Wald statistics. The statistical software package R (Studio) is used for fitting all of the three models and extracting the information from NHIS 2013, recoding and parameter estimation of the study models. We also made a comparison between CO, CR and AC model based on Akike Information Criteria (AIC).

4.4 Multivariate Analysis (Ordinal Logistic Model)

A multivariate analysis is conducted to determine the effect of predictor variables (social characteristics, other risk factors e.g. activity, smoke etc.) on the dependent variable (PD). Ordinal logistic regression is suitable here because it predicts an ordinal outcome (low, moderate, high). Three common ordinal logit models are fitted and compared in terms of goodness of fit to the data. And based on the diagnostics tools, we select the best model that fit the data. The variables in this analysis are all categorical, allowing a convenient interpretation of the logistic regression coefficients as odds ratios. Odds ratios are valuable because they demonstrate how much higher or lower the odds are of a positive outcome for a comparison group relative to the reference group. The logit model is preferred in epidemiology, demography, and public health research because of the close similarity between odds ratio and relative risk.

4.4.1 Model Selection

First, a CO model was fit with eleven explanatory variables, which is referred to as the full model. Table 4.4 displays the results of fitting from the full model. Before interpreting the results of the full model, the proportional odds assumption was first examined. From the table, labeled Score test (proportional assumption), we found that the score test$= 58.5273$, $p – value = <$
.0001, indicating that the proportional odds assumption for the full model was not upheld. This suggests that the pattern of effects for one or more of the independent variables is likely to be different. O’Connell (2006) mentioned that the violation of the assumption may be caused by the large sample size, which the score test will nearly always indicate rejection of the assumption of proportional odds, and then it should be interpreted with caution. The log likelihood ratio Chi-Square test, \( LR = 669.57 \) with 17 d.f, \( p = .0001 \), indicating that the full model with eleven predictor provided better fit than the null model with no independent variables in predicting cumulative probability for PD. In the same way, we fit other two candidate modes AC and CR with the same eleven explanatory variables. The results of the adequacy of the three models are given in Table (4.3).

Table 4.3: The adequacy tool for model selection

<table>
<thead>
<tr>
<th>Ordinal logistic regression model</th>
<th>Cumulative Odds (CO)</th>
<th>Adjacent category logit model (AC)</th>
<th>Continuation ratio model (CR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>12800.63</td>
<td>12765.06</td>
<td>12806.01</td>
</tr>
</tbody>
</table>

From the above table, we see the AIC for AC model is the smallest. So, we should consider AC as the best model. Now, we fit the data by AC model with the predictors that we found significant in the bivariate analysis.

**4.4.2 Model Fitting and Interpretation**

Here, we run all the three candidates models and the coefficients are displayed in the tables below.
Table 4.4: The results of the ordinal logistic regression models.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE) OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE) OR (95% CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>Intercept (\alpha_1)</td>
<td>2.75 (0.23) 2.87 (0.17)</td>
<td></td>
<td></td>
<td>3.06 (0.22)</td>
</tr>
<tr>
<td>Intercept (\alpha_2)</td>
<td>4.2 (0.234) 1.64 (0.185)</td>
<td></td>
<td></td>
<td>4.21 (0.23)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td>0.29 (0.054) 1.34 (0.70-0.85)</td>
<td>&lt;.0001 0.21 (0.04) (1.14-1.34)</td>
<td>&lt;.0001 0.29 (0.05) (1.12-1.48)</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td>0.28 (0.12) 1.33 (0.57-0.87)</td>
<td>0.002 0.24 (0.08) (1.07-1.52)</td>
<td>0.0004 0.26 (0.11) (1.03-1.65)</td>
</tr>
<tr>
<td>White</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.41 (0.13) 1.51 (0.49-0.79)</td>
<td>0.0002 0.35 (0.09) (1.17-1.72)</td>
<td>&lt;.0001 0.001 0.39 (0.13) (1.14-1.92)</td>
<td>0.0003</td>
</tr>
<tr>
<td>Asian</td>
<td>0.29 (0.16) 1.34 (0.54-0.97)</td>
<td>0.007 0.23 (0.12) (0.99-1.60)</td>
<td>0.005 0.27 (0.16) (0.95-1.81)</td>
<td>0.009</td>
</tr>
<tr>
<td>Other</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td>-0.61 (0.14) 0.53 (1.2-2.1)</td>
<td>&lt;.0001 -0.44 (0.11) (0.51-0.8)</td>
<td>&lt;.0001 -0.6 (0.14) (0.41-0.71)</td>
</tr>
<tr>
<td>(18-30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(31-64)</td>
<td>-0.80 (0.13) 0.44 (1.5-2.5)</td>
<td>&lt;.0001 -0.60 (0.10) (0.44-0.67)</td>
<td>&lt;.0001 -0.78 (0.13) (0.35-0.59)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>65+</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td>0.46 (0.056) 1.58 (0.58-0.72)</td>
<td>&lt;.0001 0.35 (0.04) (1.30-1.54)</td>
<td>&lt;.0001 0.44 (0.05) (1.40-1.74)</td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
<td>-0.569 (0.10) 0.56 (1.39-2.0)</td>
<td>&lt;.0001 -0.42 (0.07) (0.56-0.75)</td>
<td>&lt;.0001 -0.55 (0.09) (0.47-0.69)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoke</td>
<td></td>
<td>-0.42 (0.052) 0.65 (1.34-0.62)</td>
<td>&lt;.0001 -0.32 (0.03) (0.66-0.78)</td>
<td>&lt;.0001 -0.41 (0.05) (0.59-0.72)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td></td>
<td>0.20 (0.053) 1.22 (0.73-0.89)</td>
<td>0.00001 0.16 (0.04) (1.08-1.27)</td>
<td>&lt;.0001 0.19 (0.05) (1.09-1.34)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE=Standard Error of the estimate, OR=odds ratio and CI=Confidence interval
Table 4.5: The results of the ordinal logistic regression models. (Con’t)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>OR (95% CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>0.321 (0.22)</td>
<td>1.37 (0.52,1.13)</td>
<td>0.01</td>
</tr>
<tr>
<td>Normal</td>
<td>0.27 (0.06)</td>
<td>1.31 (0.68,0.86)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Overweight</td>
<td>0.19 (0.06)</td>
<td>1.21 (0.72,0.90)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Obese</td>
<td>Ref.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Covered</td>
<td>0.32 (0.06)</td>
<td>1.38 (0.65,0.81)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Uncovered</td>
<td>Ref.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>Ref.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt; high school</td>
<td>0.09 (0.055)</td>
<td>1.10 (0.83,1.101)</td>
<td>0.008</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($0-$34,999)</td>
<td>-1.063 (0.11)</td>
<td>0.34 (2.16,3.28)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>($35,000-$74,999)</td>
<td>-0.42 (0.11)</td>
<td>0.65 (1.15,1.7)</td>
<td>0.00004</td>
</tr>
<tr>
<td>$75,000+</td>
<td>Ref.</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Score test (proportional assumption) 58.5273
p-value <.0001

SE=Standard Error of the estimate, OR=odds ratio and CI=Confidence interval

All the predictors are found significant at 5% level of significant under AC model. Male are 1.23 times more likely to have low or moderate distress level than female. In other sense, female are usually with high distress. White American are more likely to have low or moderate distress level compare to other groups. In exact, they are 28% more relaxed than other group whereas Black/African American and Asian are 42% and 26% more likely to have low or moderate distress level, respectively. Respondents of aged 18-30 years are 36% less likely to have low or moderate distress level compare to the respondents of aged 65+ years. People who are at the age of 31-64 years are 46% less likely to have low or moderate distress level compare to the aged 65+. This is interesting to note that older people are more relaxed than the younger people. If a person
had a cancer, then he/she is more likely to have high distress level which we found in our study also. In fact, we have found a person having cancer is 35% less likely to have low or moderate distress level. In our study, we have also discovered that married people are 42% more likely to have low or moderate distress level than unmarried people. Smokers are less likely to have lower distress level than nonsmokers. If a person does some physical activity at least once a week, then he/she is 1.17 times higher chance to have low or moderate distress level. Obese people are more likely to have high distress level. We have noticed that people who insurance are more likely to have low or moderate distress than people who do not have insurance. In other words, covered respondents tend to be 28% times higher to have low or moderate distress level. Persons having at least high school diploma are more likely to have low or moderate distress level. Low income group people are less likely to have lower distress level.

4.5 Psychological Distress and Types of Cancer

The following figure 4.3 shows that the psychological distress levels are nominal over all the types of cancer. We found the same result when we use the cross tabulation technique for testing the homogeneity.

Figure 4.3: The Distribution of Cancer Types and PD.
After fitting the model, now we are interested to see if there is any heterogeneity in the distress level among the different types of cancer. In our study we have found 923 people have at least one type of cancer. Out of these 923, 148 have breast cancer, 24 have colon cancer, 11 have lung cancer, 82 have prostate cancer and rest of them have other types of cancer or they won’t disclose it. Now the question arises into mind, is the distress level same for all types of cancer? To answer this question, we employ a contingency analysis and use chi-square statistic to investigate the homogeneity of the distress level.

**Hypothesis:**

\[ H_0: \text{Distress level are homogenous over all types of cancer} \]

vs.

\[ H_1: \text{Distress level are NOT homogenous over all types of cancer} \]

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Psychological Distress Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Freq. (%)</td>
</tr>
<tr>
<td>Breast</td>
<td>129 (13.98%)</td>
</tr>
<tr>
<td>Colon</td>
<td>19 (2.06%)</td>
</tr>
<tr>
<td>Lung</td>
<td>8 (0.87%)</td>
</tr>
<tr>
<td>Prostate</td>
<td>75 (8.13%)</td>
</tr>
<tr>
<td>Other</td>
<td>561 (60.78%)</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>8.0121</td>
</tr>
<tr>
<td>d.f.</td>
<td>8</td>
</tr>
<tr>
<td>p-value</td>
<td>0.4323</td>
</tr>
</tbody>
</table>

Based on the diagnostic tools given in the above table 4.6, we fail to reject the hypothesis of homogeneity of the distress level among the different types of cancer, as p-value for all diagnostic tools are higher than classical 5% level. So we can conclude that there is no difference in distress level among different types of cancer.
4.6 Psychological Distress and Race

We are now interested to see if there any heterogeneity in the distress level in terms of race. In our study we have found that most of the respondents were white American 76% followed by African American by 14%. Now the question arises into mind, is the distress level same for all people who have different races? To answer this question, we employ a contingency analysis and use chi-square statistic to investigate the homogeneity of the distress level.

**Hypothesis:**

\[ H_0: \text{Distress level is homogenous over all people in terms of races} \]

vs.

\[ H_1: \text{Distress level is NOT homogenous over all people in terms of races} \]

<table>
<thead>
<tr>
<th>Race</th>
<th>Low Freq. (%)</th>
<th>Moderate Freq. (%)</th>
<th>High Freq. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White American</td>
<td>12212 (68.74%)</td>
<td>976 (5.49%)</td>
<td>339 (1.91%)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>2284 (12.86%)</td>
<td>199 (1.12%)</td>
<td>65 (0.37%)</td>
</tr>
<tr>
<td>Asian</td>
<td>1022 (5.75%)</td>
<td>62 (0.35%)</td>
<td>23 (0.37%)</td>
</tr>
<tr>
<td>Others</td>
<td>500 (2.81%)</td>
<td>51 (0.29%)</td>
<td>32 (0.18%)</td>
</tr>
</tbody>
</table>

Chi-Square d.f. p-value

| 29.47 | 6 | <.0001 |

Based on the diagnostic tools given in the above table 4.7, we reject the hypothesis of homogeneity of the distress level over all people who have different races, as p-value for all diagnostic tools are much smaller than classical 5% level. So we can conclude that there is difference in distress level for different race, which support what we found in the literature (Cukier, Y. (2013) & Krieger, N. (2011)).
4.7 Residual Analysis

In order to assess the best model of fit, O'Connell and Liu (2011) suggested some strategies to do so. The graphical method is considered to see how the residual points behave. Figure 4.4 shows that plot of the two different equations of fitting the adjacent model. It is clearly difficult to see any patterns in the residual plots, which suggests that there is no potential problem with the assumptions of the adjacent category model.

Figure 4.4: The residual plots of the adjacent model

4.8 Discussion

To examine the relationship between the psychological distress in different types of cancer, we use cross tabulation technique for finding association among variables in each sub-studies for different kind of cancer. Interestingly, we do not find any association between PD and selected types of cancer while we find that there is association between PD and cancer in general. The results are displayed in the tables in Appendix B (Table 6.1-6.4). One explanation for non-significance of PD and subtypes of cancer could be smaller sample size and presence of missing values in this sample. For the breast, colon, and lung sub-samples, we found that region, sex, alcohol consumption and BMI are insignificant, indicating that these covariates are not associated
with PD if we fix the cancer type. For the prostate sub-sample, in addition to the aforementioned covariates, smoking status, physical activities, and income are also found to be non-significant. However, we fit the three commonly used models for ordinal response to these sub-samples including the different types of cancer into the models. The results are shown in Appendix B (Table 6.5-6.12).

In general, the results suggest that psychological distress is a burden among participants with the cancer, and it may be linked with particular types of cancer that are not included in our study. The results from the logistic regression models suggest that individuals with self-reported cancer are more likely to experience PD when compared with those non-diseases.

Finally, the psychological distress of cancer patients is not a straightforward subject to study. Numerous factors must be taken into account. However, the main advantage of this study is the availability of a large representative population-based sample. The NHIS has been conducted every year since 1957.

There are some limitations of this study. One limitation is that the self-reported nature of the data. Another limitation is that a bulk of information is missing. There is not enough data to run the PD analysis across different subtypes of the cancer.
5 Chapter Five: Conclusion and Discussion

Clearly, in terms of mental health, this study has found that cancer patients suffer from higher level of psychological distress than the general population. That is individuals with cancer diagnostics are more likely to experience higher psychological distress than non-diseased people. This study is conducted to determine the prevalence of psychological distress among cancer patients and to investigate the association of psychological distress with various socio-demographic factors. We found a number of socio-demographic factors that contributed PD other than cancer.

In the second chapter, we reviewed the literature on the research that were conducted on psychological distress and cancer. We discussed the prevalence of psychological distress and the associated factors among individuals with different types of cancer in a large representative sample of adults in the US. We present discussions of several prior research works on PD, including definition, and the measure to assess the PD.

In the third chapter, we discussed the research methodology including the research process and data management steps. We also conducted an exploratory data analysis including univariate and bivariate analysis to explore our data and to address our preliminary research questions.

In the fourth chapter, we perform the major analysis for this study. The descriptive statistics of the data is obtained according to psychological distress level and the types of cancer. In our study, of 17765 NHIS participants 51% were female and 49% were male. Most of the individuals were in the aged group 31-64 by 68.03%. The majority were white American 13527(76.14%), followed by black/African American 2548(14.34%). About 55.46% of the participants were not married at the time of survey. There is 36.31% of the respondents were from the south region followed by 26.93% from the west region. The education level of the respondents’ show that
approximately 12042 (67.78%) have high school or higher. More than 62% says that they did not smoke. However, more than half of the respondents did some physical activities about 54.60%. Among the respondents, the majority has insurance about 80.15%. 34.57% of particpates says that they are overweight. Among respondents 9550 (53.76%) of them were in the category of the income between 0$ and 34,999 annually. We then employed few graphical plots for the response variable and the selected predictors to have an idea about the distribution pattern of the study-variables. Bivariate analyses are performed to examine the association between the response variable and each of the selected predictors. Three ordinal logistic regression models are fitted to the data and compared to pick the better model that fits the data well. Moreover, we test the homogeneity between PD levels among the different types of cancer as well as the race variable. Finally, in terms of the model diagnostics, the residual analysis is performed. According to the goodness of fit criteria, AIC and deviance, we select the adjacent category model as best model for PD. All the predictors along with afflicted by cancer that were found to be significant in bivariate analysis, are also found to be significant determinants of PD in the multivariate analysis. Subgroup analysis of PD among the subtypes of cancer (breast, colon, lung, and prostate) do not demonstrate any significant determinants of PD.

Psychological distress is found to be significantly prevalent among cancer patients that adds extra burden on them. An important finding is that differential psychological distress level exists across different race of overall cancer patients. However, among the different sub types of cancer (breast, colon, lung, and prostate) PD is not found to be different across race.
Bibliography


6 The Appendixes

Appendix A: Figures

Figure 6.1: The psychological distress distribution

The Psychological Distress Distribution

PERCENT

0 10 20 30 40 50 60 70 80 90 100

Psychological Distress Levels

Low Moderate High
Figure 6.2: the distribution of the respondents’ region and psychological distress

The Distribution the respondents Region and Psychological Distress

Psychological Distress Levels
- Low
- Moderate
- High

Region
- Northeast
- Midwest
- South
- West

Figure 6.3: the distribution of the respondents’ race and psychological distress

The Distribution the respondents Race and Psychological Distress

Psychological Distress Levels
- Low
- Moderate
- High

Race
- White
- Black/African American
- Asian
- Others
Figure 6.4: the distribution of the respondents’ age and psychological distress

The Distribution the respondents Age and Psychological Distress

Figure 6.5: the distribution of the respondents’ Marital Status and psychological distress

The Distribution the respondents Marital Status and Psychological Distress
Figure 6.6: the distribution of the respondents’ smoking status and psychological distress

The Distribution the respondents Smoking Status and Psychological Distress

Figure 6.7: the distribution of the respondents’ physical activity and psychological distress

The Distribution the respondents Activity and Psychological Distress
Figure 6.8: the distribution of the respondents’ BMI and psychological distress

The Distribution the respondents BMI and Psychological Distress

Figure 6.9: the distribution of the respondents’ insurance and psychological distress

The Distribution the respondents Insurance and Psychological Distress
Figure 6.10: the distribution of the respondents’ Education level and psychological distress

The Distribution the respondents Education level and Psychological Distress

Figure 6.11: the distribution of the respondents’ income and psychological distress

The Distribution the respondents Income and Psychological Distress
## Appendix B: Tables

### Table 6.1: The association between PD of the respondents and predictors for (the breast cancer sub-sample)

<table>
<thead>
<tr>
<th>The variables</th>
<th>d.f</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>6</td>
<td>2.5726</td>
<td>0.8603</td>
</tr>
<tr>
<td>SEX</td>
<td>2</td>
<td>2.7711</td>
<td>0.2502</td>
</tr>
<tr>
<td>Race</td>
<td>6</td>
<td>13.8781</td>
<td>0.0310</td>
</tr>
<tr>
<td>Age (Group in years)</td>
<td>4</td>
<td>20.3982</td>
<td>0.0004</td>
</tr>
<tr>
<td>Marital Status</td>
<td>2</td>
<td>11.7366</td>
<td>0.0028</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>2</td>
<td>0.3373</td>
<td>0.8448</td>
</tr>
<tr>
<td>Smoking status</td>
<td>2</td>
<td>12.6268</td>
<td>0.0018</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>2</td>
<td>8.9291</td>
<td>0.0115</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>2</td>
<td>0.5382</td>
<td>0.7641</td>
</tr>
<tr>
<td>BMI</td>
<td>6</td>
<td>1.9340</td>
<td>0.9257</td>
</tr>
<tr>
<td>Insurance</td>
<td>2</td>
<td>40.4833</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>12.2642</td>
<td>0.0022</td>
</tr>
<tr>
<td>Income</td>
<td>4</td>
<td>27.3373</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### Table 6.2: The association between PD of the respondents and predictors for (the colon cancer sub-sample)

<table>
<thead>
<tr>
<th>The variables</th>
<th>d.f</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>6</td>
<td>2.5726</td>
<td>0.8603</td>
</tr>
<tr>
<td>SEX</td>
<td>2</td>
<td>2.7711</td>
<td>0.2502</td>
</tr>
<tr>
<td>Race</td>
<td>6</td>
<td>13.8781</td>
<td>0.0310</td>
</tr>
<tr>
<td>Age (Group in years)</td>
<td>4</td>
<td>20.3982</td>
<td>0.0004</td>
</tr>
<tr>
<td>Marital Status</td>
<td>2</td>
<td>11.7366</td>
<td>0.0028</td>
</tr>
<tr>
<td>Colon Cancer</td>
<td>2</td>
<td>0.8704</td>
<td>0.6471</td>
</tr>
<tr>
<td>Smoking status</td>
<td>2</td>
<td>12.6268</td>
<td>0.0018</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>2</td>
<td>8.9291</td>
<td>0.0115</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>2</td>
<td>0.5382</td>
<td>0.7641</td>
</tr>
<tr>
<td>BMI</td>
<td>6</td>
<td>1.9340</td>
<td>0.9257</td>
</tr>
<tr>
<td>Insurance</td>
<td>2</td>
<td>40.4833</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>12.2642</td>
<td>0.0022</td>
</tr>
<tr>
<td>Income</td>
<td>4</td>
<td>27.3373</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Table 6.3: The association between PD of the respondents and predictors for (the lung cancer sub-sample)

<table>
<thead>
<tr>
<th>The variables</th>
<th>d.f</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>6</td>
<td>2.5726</td>
<td>0.8603</td>
</tr>
<tr>
<td>SEX</td>
<td>2</td>
<td>2.7711</td>
<td>0.2502</td>
</tr>
<tr>
<td>Race</td>
<td>6</td>
<td>13.8781</td>
<td>0.0310</td>
</tr>
<tr>
<td>Age (Group in years)</td>
<td>4</td>
<td>20.3982</td>
<td>0.0004</td>
</tr>
<tr>
<td>Marital Status</td>
<td>2</td>
<td>11.7366</td>
<td>0.0028</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>2</td>
<td>3.7645</td>
<td>0.1522</td>
</tr>
<tr>
<td>Smoking status</td>
<td>2</td>
<td>12.6268</td>
<td>0.0018</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>2</td>
<td>8.9291</td>
<td>0.0115</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>2</td>
<td>0.5382</td>
<td>0.7641</td>
</tr>
<tr>
<td>BMI</td>
<td>6</td>
<td>1.9340</td>
<td>0.9257</td>
</tr>
<tr>
<td>Insurance</td>
<td>2</td>
<td>40.4833</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>12.2642</td>
<td>0.0022</td>
</tr>
<tr>
<td>Income</td>
<td>4</td>
<td>27.3373</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Table 6.4: The association between PD of the respondents and predictors for (the prostate cancer sub-sample)

<table>
<thead>
<tr>
<th>The variables</th>
<th>d.f</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>6</td>
<td>1.9964</td>
<td>0.9200</td>
</tr>
<tr>
<td>SEX*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Race</td>
<td>6</td>
<td>15.9370</td>
<td>0.0141</td>
</tr>
<tr>
<td>Age (Group in years)</td>
<td>4</td>
<td>15.2559</td>
<td>0.0042</td>
</tr>
<tr>
<td>Marital Status</td>
<td>2</td>
<td>11.6702</td>
<td>0.0029</td>
</tr>
<tr>
<td>Prostate Cancer</td>
<td>2</td>
<td>1.0772</td>
<td>0.5836</td>
</tr>
<tr>
<td>Smoking status</td>
<td>2</td>
<td>2.7617</td>
<td>0.2514</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>2</td>
<td>2.1229</td>
<td>0.3459</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>2</td>
<td>0.4335</td>
<td>0.8051</td>
</tr>
<tr>
<td>BMI</td>
<td>6</td>
<td>2.6612</td>
<td>0.8500</td>
</tr>
<tr>
<td>Insurance</td>
<td>2</td>
<td>18.5822</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>7.6197</td>
<td>0.0222</td>
</tr>
<tr>
<td>Income</td>
<td>4</td>
<td>8.0091</td>
<td>0.0912</td>
</tr>
</tbody>
</table>
Table 6.5: Results of ordinal logistic models for the breast cancer sub-sample

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE)</td>
</tr>
<tr>
<td>Intercept_1((\alpha_1))</td>
<td>2.52 (0.82)</td>
<td>2.83 (2.83)</td>
<td></td>
<td>2.95 (0.79)</td>
</tr>
<tr>
<td>Intercept_2((\alpha_2))</td>
<td>3.80 (0.83)</td>
<td>1.68 (0.65)</td>
<td></td>
<td>3.81 (0.80)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Female</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.25 (0.52)</td>
<td>1.29 (0.46,3.63)</td>
<td>-0.21 (0.36)</td>
<td>1.23 (0.60,2.51)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.44 (0.59)</td>
<td>0.63 (0.20,2.03)</td>
<td>-0.27 (0.40)</td>
<td>0.75 (0.34,1.67)</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.26 (0.774)</td>
<td>0.76 (0.16,3.49)</td>
<td>-0.23 (0.52)</td>
<td>0.79 (0.28,2.21)</td>
</tr>
<tr>
<td>Other</td>
<td>Ref.</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18-30)</td>
<td>-1.37 (0.47)</td>
<td>0.25 (0.09,0.63)</td>
<td>-0.84 (0.33)</td>
<td>0.43 (0.22,0.82)</td>
</tr>
<tr>
<td>(31-64)</td>
<td>-1.23 (0.33)</td>
<td>0.29 (0.14,0.56)</td>
<td>-0.77 (0.24)</td>
<td>0.45 (0.28,0.73)</td>
</tr>
<tr>
<td>65+</td>
<td>Ref.</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.47 (0.20)</td>
<td>1.60 (1.07,2.40)</td>
<td>0.33 (0.14)</td>
<td>1.39 (1.04,1.86)</td>
</tr>
<tr>
<td>Unmarried</td>
<td>Ref.</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.09 (0.27)</td>
<td>1.0 (0.63,1.89)</td>
<td>0.09 (0.20)</td>
<td>1.09 (0.73,1.62)</td>
</tr>
<tr>
<td>No</td>
<td>Ref.</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Smoke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.43 (0.20)</td>
<td>0.64 (0.43,0.96)</td>
<td>-0.37 (0.14)</td>
<td>0.68 (0.51,0.92)</td>
</tr>
<tr>
<td>No</td>
<td>Ref.</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.38 (0.20)</td>
<td>1.47 (0.97,2.20)</td>
<td>0.28 (0.14)</td>
<td>1.33 (0.99,1.78)</td>
</tr>
<tr>
<td>No</td>
<td>Ref.</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Table 6.6: Con't

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overweight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Obese</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covered</td>
<td>0.86</td>
<td>2.38</td>
<td>9.984503e-04</td>
<td>0.53</td>
</tr>
<tr>
<td>Uncovered</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>Ref.</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>0.20</td>
<td>1.22</td>
<td>3.492382e-01</td>
<td>0.17</td>
</tr>
<tr>
<td>&gt; high school</td>
<td>(0.215)</td>
<td>(0.80,1.86)</td>
<td>01</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($0-$34,999)</td>
<td>-1.13</td>
<td>0.31</td>
<td>7.897483e-03</td>
<td>-1.05</td>
</tr>
<tr>
<td>($35,000-$74,999)</td>
<td>-0.75</td>
<td>0.46</td>
<td>8.236507e-03</td>
<td>-0.74</td>
</tr>
<tr>
<td>$75,000+</td>
<td>(0.43)</td>
<td>(0.20,1.10)</td>
<td>02</td>
<td>(0.37)</td>
</tr>
<tr>
<td></td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>Ref.</td>
</tr>
<tr>
<td>Score test</td>
<td>10.6518</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(proportional assumption)</td>
<td>p-value</td>
<td>0.5590</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance</td>
<td>87.10096</td>
<td></td>
<td>88.79918</td>
<td></td>
</tr>
<tr>
<td>D.f.</td>
<td>13</td>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;.0001</td>
<td></td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>854.0275</td>
<td></td>
<td>852.3293</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.7: Results of ordinal logistic models for the colon cancer sub-sample

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE) OR (95% CI) P-value</td>
<td>Estimate (SE) OR (95% CI) P-value</td>
<td>Estimate (SE) OR (95% CI) P-value</td>
<td>Estimate (SE) OR (95% CI) P-value</td>
</tr>
<tr>
<td>Intercept(_1(\alpha_1))</td>
<td>2.51 (0.82) 2.84 (0.61)</td>
<td>2.95 (0.79)</td>
<td>2.51 (0.82) 2.84 (0.61)</td>
<td>2.95 (0.79)</td>
</tr>
<tr>
<td>Intercept(_2(\alpha_2))</td>
<td>3.79 (0.83) 1.69 (0.65)</td>
<td>3.81 (0.80)</td>
<td>3.79 (0.83) 1.69 (0.65)</td>
<td>3.81 (0.80)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.26 (0.52) 1.30 (0.46,3.65) 0.21 (0.36) 1.23 (0.60,2.51)</td>
<td>0.26 (0.52) 1.30 (0.46,3.65) 0.21 (0.36) 1.23 (0.60,2.51)</td>
<td>0.26 (0.52) 1.30 (0.46,3.65) 0.21 (0.36) 1.23 (0.60,2.51)</td>
<td>0.26 (0.52) 1.30 (0.46,3.65) 0.21 (0.36) 1.23 (0.60,2.51)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.41 (0.59) 0.65 (0.20,2.10) -0.25 (0.40) 0.77 (0.35,1.71)</td>
<td>-0.41 (0.59) 0.65 (0.20,2.10) -0.25 (0.40) 0.77 (0.35,1.71)</td>
<td>-0.41 (0.59) 0.65 (0.20,2.10) -0.25 (0.40) 0.77 (0.35,1.71)</td>
<td>-0.41 (0.59) 0.65 (0.20,2.10) -0.25 (0.40) 0.77 (0.35,1.71)</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.28 (0.77) 0.749 (0.16,3.41) -0.25 (0.52) 0.77 (0.27,2.17)</td>
<td>-0.28 (0.77) 0.749 (0.16,3.41) -0.25 (0.52) 0.77 (0.27,2.17)</td>
<td>-0.28 (0.77) 0.749 (0.16,3.41) -0.25 (0.52) 0.77 (0.27,2.17)</td>
<td>-0.28 (0.77) 0.749 (0.16,3.41) -0.25 (0.52) 0.77 (0.27,2.17)</td>
</tr>
<tr>
<td>Other</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18-30)</td>
<td>-1.38 (0.47) 0.24 (0.09,0.63) -0.85 (0.33) 0.42 (0.22,0.81)</td>
<td>-1.38 (0.47) 0.24 (0.09,0.63) -0.85 (0.33) 0.42 (0.22,0.81)</td>
<td>-1.38 (0.47) 0.24 (0.09,0.63) -0.85 (0.33) 0.42 (0.22,0.81)</td>
<td>-1.38 (0.47) 0.24 (0.09,0.63) -0.85 (0.33) 0.42 (0.22,0.81)</td>
</tr>
<tr>
<td>(31-64)</td>
<td>-1.21 (0.33) 0.29 (0.15,0.57) -0.76 (0.24) 0.46 (0.28,0.74)</td>
<td>-1.21 (0.33) 0.29 (0.15,0.57) -0.76 (0.24) 0.46 (0.28,0.74)</td>
<td>-1.21 (0.33) 0.29 (0.15,0.57) -0.76 (0.24) 0.46 (0.28,0.74)</td>
<td>-1.21 (0.33) 0.29 (0.15,0.57) -0.76 (0.24) 0.46 (0.28,0.74)</td>
</tr>
<tr>
<td>65+</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.47 (0.20) 1.60 (1.07,2.40) 0.33 (0.14) 1.39 (1.04,1.86)</td>
<td>0.47 (0.20) 1.60 (1.07,2.40) 0.33 (0.14) 1.39 (1.04,1.86)</td>
<td>0.47 (0.20) 1.60 (1.07,2.40) 0.33 (0.14) 1.39 (1.04,1.86)</td>
<td>0.47 (0.20) 1.60 (1.07,2.40) 0.33 (0.14) 1.39 (1.04,1.86)</td>
</tr>
<tr>
<td>Unmarried</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.51 (0.53) 0.59 (0.20,1.70) -0.38 (0.36) 0.67 (0.32,1.39)</td>
<td>-0.51 (0.53) 0.59 (0.20,1.70) -0.38 (0.36) 0.67 (0.32,1.39)</td>
<td>-0.51 (0.53) 0.59 (0.20,1.70) -0.38 (0.36) 0.67 (0.32,1.39)</td>
<td>-0.51 (0.53) 0.59 (0.20,1.70) -0.38 (0.36) 0.67 (0.32,1.39)</td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.43 (0.20) 0.64 (0.43,0.97) -0.37 (0.148) 0.68 (0.51,0.92)</td>
<td>-0.43 (0.20) 0.64 (0.43,0.97) -0.37 (0.148) 0.68 (0.51,0.92)</td>
<td>-0.43 (0.20) 0.64 (0.43,0.97) -0.37 (0.148) 0.68 (0.51,0.92)</td>
<td>-0.43 (0.20) 0.64 (0.43,0.97) -0.37 (0.148) 0.68 (0.51,0.92)</td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.38 (0.20) 1.46 (0.97,2.20) 0.28 (0.14) 1.33 (0.99,1.78)</td>
<td>0.38 (0.20) 1.46 (0.97,2.20) 0.28 (0.14) 1.33 (0.99,1.78)</td>
<td>0.38 (0.20) 1.46 (0.97,2.20) 0.28 (0.14) 1.33 (0.99,1.78)</td>
<td>0.38 (0.20) 1.46 (0.97,2.20) 0.28 (0.14) 1.33 (0.99,1.78)</td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 6.8: Con’t

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE) OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE) OR (95% CI)</td>
<td>P-value</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overweight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Obese</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Insurance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covered</td>
<td>0.89 (0.26) 2.44 (1.45, 4.09)</td>
<td>0.55 (0.17) 0.54 (1.22, 2.46)</td>
<td>0.82 (0.24) 2.82 (1.40, 3.70)</td>
<td>0.4844 8.106299e-04</td>
</tr>
<tr>
<td>Uncovered</td>
<td>Ref. 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>0.19 (0.21) 1.22 (0.79, 1.86)</td>
<td>0.17 (0.15) 1.18 (0.88, 1.59)</td>
<td>0.17 (0.20) 1.19 (0.79, 1.78)</td>
<td>3.906917e-01</td>
</tr>
<tr>
<td>&gt; high school</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($0-$34,999)</td>
<td>-1.14 (0.42) 0.31 (0.13, 0.73)</td>
<td>-1.06 (0.37) 0.34 (0.16, 0.71)</td>
<td>-1.11 (0.42) 0.32 (0.14, 0.75)</td>
<td>8.243150e-03</td>
</tr>
<tr>
<td>($35,000-$74,999)</td>
<td>-0.75 (0.43) 0.47 (0.20, 1.10)</td>
<td>-0.74 (0.37) 0.47 (0.22, 0.99)</td>
<td>-0.73 (0.42) 0.47 (0.20, 1.10)</td>
<td>8.355503e-02</td>
</tr>
<tr>
<td>$75,000+</td>
<td>Ref. 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Score test (proportional assumption)</td>
<td>12.5345</td>
<td></td>
<td>0.4844</td>
<td>87.3069</td>
</tr>
</tbody>
</table>
Table 6.9: Results of ordinal logistic models for the lung cancer sub-sample

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE)</td>
</tr>
<tr>
<td>Intercept(_1(\alpha_1))</td>
<td>2.64 (0.82)</td>
<td>2.93 (0.61)</td>
<td>2.97 (0.79)</td>
<td>3.07 (0.80)</td>
</tr>
<tr>
<td>Intercept(_2(\alpha_2))</td>
<td>3.92 (0.83)</td>
<td>1.78 (0.65)</td>
<td>3.94 (0.80)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.23 (0.52)</td>
<td>1.25 (0.44,3.53)</td>
<td>0.18 (0.36)</td>
<td>1.20 (0.59,2.4)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.49 (0.59)</td>
<td>0.61 (0.19,1.94)</td>
<td>-0.31 (0.40)</td>
<td>0.73 (0.33,1.6)</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.31 (0.77)</td>
<td>0.72 (0.15,3.3)</td>
<td>-0.28 (0.52)</td>
<td>0.75 (0.26,2.1)</td>
</tr>
<tr>
<td>Other</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18-30)</td>
<td>-1.43 (0.47)</td>
<td>0.23 (0.09,0.60)</td>
<td>-0.89 (0.33)</td>
<td>0.40 (0.21,0.78)</td>
</tr>
<tr>
<td>(31-64)</td>
<td>-1.26 (0.34)</td>
<td>0.28 (0.14,0.54)</td>
<td>-0.8 (0.24)</td>
<td>0.44 (0.27,0.7)</td>
</tr>
<tr>
<td>65+</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.45 (0.20)</td>
<td>1.58 (1.05,2.36)</td>
<td>0.32 (0.14)</td>
<td>1.38 (1.03,1.8)</td>
</tr>
<tr>
<td>Unmarried</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.98 (0.68)</td>
<td>0.37 (0.09,1.43)</td>
<td>-0.73 (0.44)</td>
<td>0.47 (0.1,1.14)</td>
</tr>
<tr>
<td>No</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Smoke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.42 (0.20)</td>
<td>0.65 (0.43,0.97)</td>
<td>-0.36 (0.14)</td>
<td>0.69 (0.5,0.9)</td>
</tr>
<tr>
<td>No</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.37 (0.20)</td>
<td>1.45 (0.96,2.18)</td>
<td>0.27 (0.14)</td>
<td>1.32 (0.98,1.7)</td>
</tr>
<tr>
<td>No</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Predictors</td>
<td>Proportional odds model</td>
<td>Adjacent category logit model</td>
<td>Continuation ratio model</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurance</td>
<td>Covered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covered</td>
<td>0.87 (0.26)</td>
<td>2.40 (1.43, 4.02)</td>
<td>0.53 (0.17)</td>
<td>1.71 (1.21, 2.4)</td>
</tr>
<tr>
<td>Uncovered</td>
<td>Ref. 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>High school or less</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt; high school</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.19 (0.21)</td>
<td>1.21 (0.79, 1.85)</td>
<td>0.17 (0.151)</td>
<td>1.18 (0.88, 1.5)</td>
</tr>
<tr>
<td>Income</td>
<td>($0-$34,999)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.16 (0.42)</td>
<td>0.31 (0.13, 0.72)</td>
<td>-1.07 (0.37)</td>
<td>0.33 (0.16, 0.7)</td>
</tr>
<tr>
<td>$35,000-$74,999</td>
<td>-0.78 (0.43)</td>
<td>0.45 (0.19, 1.07)</td>
<td>-0.77 (0.38)</td>
<td>0.46 (0.21, 0.96)</td>
</tr>
<tr>
<td>$75,000+</td>
<td>Ref. 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score test</td>
<td></td>
<td>14.3743</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>0.348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance</td>
<td></td>
<td>88.60655</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.f.</td>
<td></td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>852.5219</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.11: Results of ordinal logistic models for the prostate cancer sub-sample

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Proportional odds model</th>
<th>Ordinal logistic regression model</th>
<th>Adjacent category logit model</th>
<th>Continuation ratio model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE) OR (95% CI) P-value</td>
<td>Estimate (SE) OR (95% CI) P-value</td>
<td>Estimate (SE) OR (95% CI) P-value</td>
<td>Estimate (SE) OR (95% CI) P-value</td>
</tr>
<tr>
<td>Intercept1(α1)</td>
<td>1.68 (1.37) 1.95 (1.06)</td>
<td>2.11 (1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept2(α2)</td>
<td>3.00 (1.38) 0.72 (1.10)</td>
<td>3.04 (1.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
</tr>
<tr>
<td>Female</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.18 (1.12) 1.20 (0.13, 10.7) 0.87 (0.91) 0.071 (0.02, 1.9) 0.93 (1.08) 1.19 (0.14, 10.6)</td>
<td>0.34 (0.05, 2.2) 0.77 (0.42, 1.9) 0.82 (1.52) 0.14 (0.04, 15.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>-1.33 (1.17) 0.26 (0.02, 2.6) 0.25 (0.94) 0.25 (0.03, 2.59)</td>
<td>-1.27 (1.33) 0.27 (0.03, 2.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>-0.23 (1.56) 0.78 (0.03, 16.9) 0.87 (1.28) 0.53 (0.06, 9.3)</td>
<td>0.22 (1.27) 0.79 (1.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18-30)</td>
<td>-1.44 (1.07) 0.23 (0.02, 1.9) 0.18 (0.77) 0.53 (0.12, 2.6) 0.48 (1.02) 0.22 (0.02, 1.67)</td>
<td>-1.49 (1.02) 0.22 (0.02, 1.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(31-64)</td>
<td>-1.60 (0.63) 0.20 (0.05, 0.69) 0.01 (0.42) 0.80 (0.19, 1.02) 0.05 (0.62) 0.20 (0.06, 0.70)</td>
<td>-1.57 (0.62) 0.20 (0.06, 0.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>0.59 (0.37) 1.81 (0.83, 3.77) 0.10 (0.27) 0.53 (0.99, 2.9) 0.05 (0.36) 1.75 (0.86, 3.55)</td>
<td>0.56 (0.36) 1.75 (0.86, 3.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.15 (0.49) 1.16 (0.44, 3.07) 0.751 (0.358) 0.16 (0.58, 2.3) 0.64 (0.47) 1.17 (0.46, 2.99)</td>
<td>0.16 (0.47) 1.17 (0.46, 2.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Ref. 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoke</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
</tr>
<tr>
<td>No</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
</tr>
<tr>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
</tr>
<tr>
<td>No</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
<td>- - - - - - - -</td>
</tr>
<tr>
<td>Predictors</td>
<td>Proportional odds model</td>
<td>Ordinal logistic regression model</td>
<td>Adjacent category logit model</td>
<td>Continuation ratio model</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>OR (95% CI)</td>
<td>P-value</td>
<td>Estimate (SE)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overweight</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Obese</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Insurance</td>
<td>Covered</td>
<td>1.28 (0.51)</td>
<td>3.61 (1.32, 9.8)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Uncovered</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Education</td>
<td>High school or less</td>
<td>0.32 (0.37)</td>
<td>1.37 (0.6, 2.8)</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>&gt; high school</td>
<td>Ref.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Income</td>
<td>($0-$34,999)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>($35,000-$74,999)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$75,000+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Score test (proportional assumption)</td>
<td>154.0184</td>
<td>&lt;.0001</td>
<td>38.05671</td>
<td>35.74752</td>
</tr>
<tr>
<td>Deviance</td>
<td>38.05671</td>
<td>35.74752</td>
<td>37.83033</td>
<td>9</td>
</tr>
<tr>
<td>D.f.</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>AIC</td>
<td>296.7765</td>
<td>299.0856</td>
<td>297.0028</td>
<td>297.0028</td>
</tr>
</tbody>
</table>
Appendix C : SAS & R Code

SAS code:
libname thesis "D:\BSU\Thesis\cancer";
DATA thesis.adult;
  INFILE "D:\BSU\Thesis\cancer\dataset\samadult.dat"
  LRECL=2000 DLM=' ';
  input HHX 7-12 FMX 16-17 FPX 18-19
            REGION 33
            SEX 39 RACE 42-43 AGE 48-49 MARITL 50
            CANCER 120 C_BREAST 125 C_COLON 127 C_LUNG 134 C_PROSTATE 140
            SMOKE 619 ACTIVITY 643 ALCOHOL 672 BMİ 699-702
            SAD 1002 NERVOUS 1003 RESTLESS 1004 HOPELESS 1005 EFFORT 1006
  WORTHLESS 1007;

  /* set up missing values codes */
  if REGION not=. then do;
    if REGION=1 then REGION_CAT=1; /*Northeast*/
    if REGION=2 then REGION_CAT=2; /*Midwest*/
    if REGION=3 then REGION_CAT=3; /*South*/
    if REGION=4 then REGION_CAT=4; /*West*/
  end;

  if SEX not=. then do;
    if SEX=1 then SEX_CAT=1; /*Male*/
    if SEX=2 then SEX_CAT=0; /*Female*/
  end;

  if RACE not=. then do;
    if RACE=1 then RACE_CAT=1; /*White*/
    if RACE=2 then RACE_CAT=2; /*Black/African american*/
    if RACE=4 then RACE_CAT=3; /*Asian*/
    if RACE in (3,5,6) then RACE_CAT=4; /*Others= AIAN,Race group not releasable ,Multiple race */
  end;

  if AGE not=. then do;
    if (18 <= AGE <=30) then AGE_CAT=1; /* 18-30*/
    if (31 <= AGE <=64) then AGE_CAT=2; /* 31-64*/
    if AGE > 65 then AGE_CAT=3; /* 65+*/
  end;

  if MARITL=9 then MAR_CAT=.;
  if MARITL not=. then do;
    if MARITL in (1,2,3) then MAR_CAT=1; /*Married*/
    if MARITL in (4,5,6,7,8) then MAR_CAT=0; /*Unmarried*/
  end;

  if CANCER in (7,8,9) then CANCER_CAT=.;
  if CANCER not=. then do;
    if CANCER=1 then CANCER_CAT=1; /*Yes*/
    if CANCER=2 then CANCER_CAT=0; /*No*/
  end;

  if C_BREAST in (7,8,9) then BREAST_CAT=.;
  if C_BREAST not=. then do;
    if C_BREAST=1 then BREAST_CAT=1; /*Yes*/
    if C_BREAST=2 then BREAST_CAT=0; /*No*/
  end;

  if C_COLON in (7,8,9) then COLON_CAT=.;
  if C_COLON not=. then do;
if C_COLON=1 then COLON_CAT=1; /*Yes*/
if C_COLON=2 then COLON_CAT=0; /*NO*/
end;
if C_LUNG in (7,8,9) then LUNG_CAT=.
if C_LUNG not=. then do;
    if C_LUNG=1 then LUNG_CAT=1; /*Yes*/
    if C_LUNG=2 then LUNG_CAT=0; /*NO*/
end;
if C_PROSTATE in (7,8,9) then PROSTATE_CAT=.
if C_PROSTATE not=. then do;
    if C_PROSTATE=1 then PROSTATE_CAT=1; /*Yes*/
    if C_PROSTATE=2 then PROSTATE_CAT=0; /*NO*/
end;
if SMOKE=9 then SMOKE_CAT=.
if SMOKE not=. then do;
    if SMOKE in (1,2,3,5) then SMOKE_CAT=1; /*Yes*/
    if SMOKE=4 then SMOKE_CAT=0; /*NO*/
end;
if ACTIVITY in (7,8,9) then ACTIV_CAT=.
if ACTIVITY not=. then do;
    if ACTIVITY in (0,6) then ACTIV_CAT=0; /*NO*/
    if ACTIVITY in (1,2,3,4) then ACTIV_CAT=1; /*YES*/
end;
if ALCOHOL in (7,8,9) then ALCOHOL_CAT=.
if ALCOHOL not=. then do;
    if ALCOHOL=1 then ALCOHOL_CAT=1; /*Yes*/
    if ALCOHOL=2 then ALCOHOL_CAT=0; /*NO*/
end;
if BMI=9999 then BMI_CAT=.
if BMI not=. then do;
    if BMI <1850 then BMI_CAT=1; /*Underwieght*/
    if (1850<= BMI <2499) then BMI_CAT=2; /*Normal*/
    if (2500<= BMI <3000) then BMI_CAT=3; /*Overwieght*/
    if BMI >=3000 then BMI_CAT=4; /*Obese*/
end;
if SAD in (7,8,9) then SAD_CAT=.
if SAD not=. then do;
    if SAD=5 then SAD_CAT=0; /*None of the time*/
    if SAD=4 then SAD_CAT=1; /*A little of the time*/
    if SAD=3 then SAD_CAT=2; /*some of the time*/
    if SAD=2 then SAD_CAT=3; /*Most of the time*/
    if SAD=1 then SAD_CAT=4; /*All of the time*/
end;
if NERVOUS in (7,8,9) then NERVOUS_CAT=.
if NERVOUS not=. then do;
    if NERVOUS=5 then NERVOUS_CAT=0; /*None of the time*/
    if NERVOUS=4 then NERVOUS_CAT=1; /*A little of the time*/
    if NERVOUS=3 then NERVOUS_CAT=2; /*some of the time*/
    if NERVOUS=2 then NERVOUS_CAT=3; /*Most of the time*/
    if NERVOUS=1 then NERVOUS_CAT=4; /*All of the time*/
end;
if RESTLESS in (7,8,9) then RESTLESS_CAT=.
if RESTLESS not=. then do;
    if RESTLESS=5 then RESTLESS_CAT=0; /*None of the time*/
    if RESTLESS=4 then RESTLESS_CAT=1; /*A little of the time*/
    if RESTLESS=3 then RESTLESS_CAT=2; /*some of the time*/
    if RESTLESS=2 then RESTLESS_CAT=3; /*Most of the time*/
if RESTLESS=1 then RESTLESS_CAT=4; /*All of the time*/
end;
if HOPELESS in (7, 8, 9) then HOPELESS_CAT=.;
if HOPELESS not=. then do;
   if HOPELESS=5 then HOPELESS_CAT=0; /*None of the time*/
   if HOPELESS=4 then HOPELESS_CAT=1; /*A little of the time*/
   if HOPELESS=3 then HOPELESS_CAT=2; /*some of the time*/
   if HOPELESS=2 then HOPELESS_CAT=3; /*Most of the time*/
   if HOPELESS=1 then HOPELESS_CAT=4; /*All of the time*/
end;
if EFFORT in (7, 8, 9) then EFFORT_CAT=.;
if EFFORT not=. then do;
   if EFFORT=5 then EFFORT_CAT=0; /*None of the time*/
   if EFFORT=4 then EFFORT_CAT=1; /*A little of the time*/
   if EFFORT=3 then EFFORT_CAT=2; /*some of the time*/
   if EFFORT=2 then EFFORT_CAT=3; /*Most of the time*/
   if EFFORT=1 then EFFORT_CAT=4; /*All of the time*/
end;
if WORTHLESS in (7, 8, 9) then WORTHLESS_CAT=.;
if WORTHLESS not=. then do;
   if WORTHLESS=5 then WORTHLESS_CAT=0; /*None of the time*/
   if WORTHLESS=4 then WORTHLESS_CAT=1; /*A little of the time*/
   if WORTHLESS=3 then WORTHLESS_CAT=2; /*some of the time*/
   if WORTHLESS=2 then WORTHLESS_CAT=3; /*Most of the time*/
   if WORTHLESS=1 then WORTHLESS_CAT=4; /*All of the time*/
end;
if SAD_CAT | NERVOUS_CAT | RESTLESS_CAT | HOPELESS_CAT | EFFORT_CAT | WORTHLESS_CAT not=.
   then do;
      PD=SAD_CAT+NERVOUS_CAT+RESTLESS_CAT+HOPELESS_CAT+EFFORT_CAT+WORTHLESS_CAT;
   end;
if PD not=.
   then do;
      if (0<= PD <=7) then PD_CAT=1; /*Low*/
      if (8 <= PD <=12) then PD_CAT=2; /*Moderate*/
      if (13 <= PD <=24) then PD_CAT=3; /*High*/
   end;
run;
DATA thesis.person;
   INFILE "D:\BSU\Thesis\cancer\dataset\PERSONSX.dat"
   LRECL=2000 DLM=' ';
   input HHX 7-12 FMX 16-17 FPX 18-19 INSURANCE 544 EDUC 754-755 WORK 773
   INCOME 776-777;
   if INSURANCE=9 then INSURANCE_CAT=.;
   if INSURANCE not=. then do;
      if INSURANCE=1 then INSURANCE_CAT=0; /*Not covered*/
      if INSURANCE=2 then INSURANCE_CAT=1; /*covered*/
   end;
   if EDUC >= 96 then EDUC_CAT=.;
   if EDUC not=. then do;
      if (0 <= EDUC <=14) then EDUC_CAT=0; /*high school or below*/
      if (15 <= EDUC <=21) then EDUC_CAT=1; /*more than high school*/
   end;
   if WORK >= 7 then WORK_CAT=.;
   if WORK not=. then do;
      if WORK=1 then WORK_CAT=1; /*employed in 2012*/
      if WORK=2 then WORK_CAT=0; /*unemployed in 2012*/
   end;
   if INCOME >= 97 then INCOME_CAT=.;
if INCOME not= then do;
   if INCOME in (1, 2, 3, 4, 5, 6) then INCOME_CAT=1; /* $0-$34,999 */
   if INCOME in (7, 8, 9, 10) then INCOME_CAT=2; /* $35,000-$74,999 */
   if INCOME=11 then INCOME_CAT=3;
/* >= $75,000 */
end;
run;
/*** start merging the files ***/
proc sort data=thesis.person;
   by HHX FMX FPX;
run;
proc sort data=thesis.adult;
   by HHX FMX FPX;
run;
data thesis.data;
   merge thesis.person thesis.adult;
   by HHX FMX FPX;
run;
data data1;
   set thesis.data; where AGE>=18;
run;
/*** before missing values have been dropped*******/
data data2;
   set data1;
run;
proc freq data=data2;run;
/********* drop missing values***********/
data NOMISS;
   set data2;
   array miss{14} REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT CANCER_CAT SMOKE_CAT ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT PD_CAT;
   Do i=1 to 14;
      if miss[i]='' then delete;
      if miss[i] ^='';
   end;
run;
proc freq data=NOMISS;run;
libname thesis "D:\BSU\Thesis\cancer";
data thesis.analysis;
   set thesis.nomiss;
   LABEL
REGION_CAT = 'Region'
SEX_CAT = 'SEX'
RACE_CAT = 'Race'
AGE_CAT = 'Age group'
MAR_CAT = 'Marital Status'
CANCER_CAT = 'Cancer'
BREAST_CAT = 'Breast Cancer'
COLON_CAT = 'Colon Cancer'
LUNG_CAT = 'Lung Cancer'
PROSTATE_CAT = 'Prostate Cancer'
SMOKE_CAT = 'Smoking status'
ACTIV Cat='Physical Activity'
ALCOHOL Cat='Alcohol consumption'
BMI Cat='BMI'
INSURANCE Cat='Insurance'
EDUC Cat='Education'
INCOME Cat='Income'
PD Cat='Psychological Distress Levels';

run;

****** labeling the categories ******

proc format;
value REGION_CATFmt 1='Northeast'
2='Midwest'
3='South'
4='West';
value SEX_CATFmt 1='Male'
0='Female';
value RACE_CATFmt 1='White'
2='Black/African american'
3='Asian'
4='Others';
value AGE_CATFmt 1='18-30'
2='31-64'
3='65+';
value MAR_CATFmt 1='Married'
0='Unmarried';
value CANCER_CATFmt 1='Yes'
0='No';
value BREAST_CATFmt 1='Yes'
0='No';
value COLON_CATFmt 1='Yes'
0='No';
value LUNG_CATFmt 1='Yes'
0='No';
value PROSTATE_CATFmt 1='Yes'
0='No';
value SMOKE_CATFmt 1='Yes'
0='No';
value ACTIV_CATFmt 0='No'
1='Yes';
value ALCOHOL_CATFmt 1='Yes'
0='No';
value BMI_CATFmt 1='Underweight'
2='Normal'
3='Overweight'
4='Obese';
value INSURANCE_CATFmt 0='Not covered'
1='covered';
value EDUC_CATFmt 0='high school or below'
1='more than high school';
value INCOME_CATFmt 1='<= $34,999'
2='$35,000-$74,999'
3='> $75,000';
value PD_CATFmt 1='Low'
2='Moderate'
3='High';

run;

************* test the association *************/
PROC freq Data=thesis.analysis;
format REGION_CAT REGION_CATFmt. SEX_CAT SEX_CATFmt. RACE_CAT
RACE_CATFmt. AGE_CAT AGE_CATFmt.
MAR_CAT MAR_CATFmt. CANCER_CAT CANCER_CATFmt. BREAST_CAT
BREAST_CATFmt.
PROSTATE_CAT
PROSTATE_CATFmt.
SMOKE_CAT SMOKE_CATFmt. ACTIV_CAT ACTIV_CATFmt. ALCOHOL_CAT
ALCOHOL_CATFmt. BMI_CAT BMI_CATFmt.
INSURANCE_CAT INSURANCE_CATFmt. EDUC_CAT EDUC_CATFmt. INCOME_CAT
INCOME_CATFmt.
PD_CAT PD_CATFmt.;
tables PD_CAT*(REGION_CAT SEX_CAT RACE_CAT AGE_CAT  MAR_CAT
CANCER_CAT  SMOKE_CAT
ACTIV_CAT  ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT
INCOME_CAT) / chisq;
title "Relationships Between Psychological Distress";
title2 "And Each of the Categorical";
title3 "Explanatory Variables";
RUN;
/*************** Sub-data analysis *************************************/
/*************************** Cancer in general *****************************/
data thesis.analysis_Cancer;
set thesis.analysis;
keep PD_CAT REGION_CAT SEX_CAT  RACE_CAT AGE_CAT  MAR_CAT CANCER_CAT
SMOKE_CAT  ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT
INCOME_CAT;
array miss{1} CANCER_CAT;
Do i=1 to 1;
  if miss[i]='' then delete;
  if miss[i]='';
end;
run;
Proc freq data=thesis.analysis_Cancer;
format REGION_CAT REGION_CATFmt. SEX_CAT SEX_CATFmt. RACE_CAT
RACE_CATFmt. AGE_CAT AGE_CATFmt.
MAR_CAT MAR_CATFmt. CANCER_CAT CANCER_CATFmt. SMOKE_CAT
SMOKE_CATFmt. ACTIV_CAT ACTIV_CATFmt. ALCOHOL_CAT ALCOHOL_CATFmt. BMI_CAT
BMI_CATFmt. INSURANCE_CAT INSURANCE_CATFmt. EDUC_CAT EDUC_CATFmt. INCOME_CAT
INCOME_CATFmt.
PD_CAT PD_CATFmt.;
table PD_CAT*(REGION_CAT SEX_CAT RACE_CAT AGE_CAT  MAR_CAT
CANCER_CAT  SMOKE_CAT
ACTIV_CAT  ALCOHOL_CAT BMI_CAT INSURANCE_CAT
EDUC_CAT INCOME_CAT);
  / chisq;
title "Relationships Between Psychological Distress";
title2 "And Each of the Categorical";
title3 "Explanatory Variables For Cancer Patients";
run;
/*************** start fitting the ordinal model - Cancer - **************/
/******************** Cumulative Odds Model
**********************************************************************/
proc logistic data=thesis.analysis_Cancer; * proportional odds cumulative
logit model;
class SEX_CAT RACE_CAT AGE_CAT  MAR_CAT CANCER_CAT SMOKE_CAT

ACTIV_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT / param=ref;
model PD_CAT = SEX_CAT RACE_CAT AGE_CAT MAR_CAT CANCER_CAT SMOKE_CAT ACTIV_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT /
link=cloglog aggregate scale=none expb;
titl1 'proportional odds cumulative logit model - Cancer -';
run;
/* Types of cancer */
data type;
where CANCER_CAT=1;
if BREAST_CAT=1 then CT=1;
if COLON_CAT=1 then CT=2;
if LUNG_CAT=1 then CT=3;
if PROSTATE_CAT=1 then CT=4;
if (BREAST_CAT^=1) and(COLON_CAT^=1) and (LUNG_CAT^=1) and (PROSTATE_CAT^=1) then CT=5;
keep PD_CAT SEX_CAT CANCER_CAT BREAST_CAT COLON_CAT LUNG_CAT PROSTATE_CAT CT;
run;
proc print data=type; run;
proc freq data=type;
table CT*PD_CAT/norow chisq;
run;
proc format;
value CTFmt 1='Breast'
2='Colon'
3='Lung'
4='Prostate'
5='Other';
run;
/* graphs*/
proc gchart data=type;
format PD_CAT PD_CATFmt. CT CTFmt.;
vbar CT/discrete type=percent group=PD_CAT;
titl1 'The Psychological Distress Distribution';
run;
proc gchart data=type;
format PD_CAT PD_CATFmt. CT CTFmt.;
vbar CT/discrete type=percent subgroup=PD_CAT;
titl1 'The Distribution cancer types and Psychological Distress';
run;
/** Race ***/
Data race;
set thesis.analysis;
keep PD_CAT RACE_CAT;
run;
proc freq data=race;
table RACE_CAT*PD_CAT/norow chisq;
run;

/******************** Breast Cancer ********************/
Data thesis.analysis_B;
set thesis.analysis;
keep PD_CAT REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT CANCER_CAT
BREAST_CAT  SMOKE_CAT  ACTIV_CAT  ALCOHOL_CAT  BMI_CAT  INSURANCE_CAT
EDUC_CAT  INCOME_CAT;
array miss{2}  CANCER_CAT  BREAST_CAT;
    Do i=1 to 2;
    if miss[i]='' then delete;
    if miss[i]^='';
end;
run;

Proc freq data=thesis.analysis_B;
format  REGION_CAT  REGION_CATFmt.  SEX_CAT  SEX_CATFmt.  RACE_CAT
        AGE_CAT  AGE_CATFmt.  MAR_CAT  MAR_CATFmt.  BREAST_CAT  BREAST_CATFmt.  SMOKE_CAT
        SMOKE_CATFmt.  ACTIV_CAT  ACTIV_CATFmt.  ALCOHOL_CAT  ALCOHOL_CATFmt.  BMI_CAT
        BMI_CATFmt.  INSURANCE_CAT  INSURANCE_CATFmt.  EDUC_CAT  EDUC_CATFmt.  INCOME_CAT
        PD_CAT  PD_CATFmt.;
table  PD_CAT*(REGION_CAT  SEX_CAT  RACE_CAT  AGE_CAT  MAR_CAT
        BREAST_CAT  SMOKE_CAT  ACTIV_CAT  ALCOHOL_CAT  BMI_CAT  INSURANCE_CAT
        EDUC_CAT  INCOME_CAT);
/ chisq;
title "Relationships Between Psychological Distress"
    "And Each of the Categorical"
    "Explanatory Variables For Breast Cancer Patients";
run;

/************* Colon Cancer ***********************************/

Data  thesis.analysis_C;
set  thesis.analysis;
keep  PD_CAT  REGION_CAT  SEX_CAT  RACE_CAT  AGE_CAT  MAR_CAT  CANCER_CAT
        COLON_CAT  SMOKE_CAT  ACTIV_CAT  ALCOHOL_CAT  BMI_CAT  INSURANCE_CAT
        EDUC_CAT  INCOME_CAT;
array miss{2}  COLON_CAT  CANCER_CAT;
    Do i=1 to 2;
    if miss[i]='' then delete;
    if miss[i]^='';
end;
run;

Proc freq data=thesis.analysis_C;
format  REGION_CAT  REGION_CATFmt.  SEX_CAT  SEX_CATFmt.  RACE_CAT
        AGE_CAT  AGE_CATFmt.  MAR_CAT  MAR_CATFmt.  CANCER_CAT  CANCER_CATFmt.  COLON_CAT
        COLON_CATFmt.  SMOKE_CAT  SMOKE_CATFmt.  ACTIV_CAT  ACTIV_CATFmt.  ALCOHOL_CAT
        ALCOHOL_CATFmt.  BMI_CAT  BMI_CATFmt.  INSURANCE_CAT  INSURANCE_CATFmt.  EDUC_CAT
        EDUC_CATFmt.  INCOME_CATFmt.  PD_CAT  PD_CATFmt.;
table  PD_CAT*(REGION_CAT  SEX_CAT  RACE_CAT  AGE_CAT  MAR_CAT
        COLON_CAT  SMOKE_CAT  ACTIV_CAT  ALCOHOL_CAT  BMI_CAT  INSURANCE_CAT
        EDUC_CAT  INCOME_CAT);
/ chisq;
title "Relationships Between Psychological Distress"
    "And Each of the Categorical"
    "Explanatory Variables For Colon Cancer Patients";
run;

/************* LUNG Cancer *************************************/
Data thesis.analysis_L;
set thesis.analysis;
keep PD_CAT REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT CANCER_CAT LUNG_CAT SMOKE_CAT ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT;
array miss{2} LUNG_CAT CANCER_CAT;
Do i=1 to 2;
  if miss[i]='' then delete;
  if miss[i]='#';
end;
run;
Proc freq data=thesis.analysis_L;
  PD_CAT PD_CATFmt.;
table PD_CAT*(REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT LUNG_CAT SMOKE_CAT ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT) / chisq;
title "Relationships Between Psychological Distress";
title2 "And Each of the Categorical"
title3 "Explanatory Variables For Lung Cancer Patients";
run;
/************* PROSTATE Cancer *******************************************/
Data thesis.analysis_P;
set thesis.analysis;
keep PD_CAT REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT CANCER_CAT PROSTATE_CAT SMOKE_CAT ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT;
array miss{2} PROSTATE_CAT CANCER_CAT;
Do i=1 to 2;
  if miss[i]='' then delete;
  if miss[i]='#';
end;
run;
Proc freq data=thesis.analysis_P;
  SMOKE_CAT SMOKE_CATFmt. ACTIV_CAT ACTIV_CATFmt. ALCOHOL_CAT ALCOHOL_CATFmt. BMI_CAT BMI_CATFmt. INSURANCE_CAT INSURANCE_CATFmt. EDUC_CAT EDUC_CATFmt. INCOME_CAT INCOME_CATFmt.
  PD_CAT PD_CATFmt.;
table PD_CAT*(REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT PROSTATE_CAT SMOKE_CAT ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT) / chisq;
title "Relationships Between Psychological Distress";
title2 "And Each of the Categorical"
title3 "Explanatory Variables For Prostate Cancer Patients";
run;
Proc freq data=thesis.analysis_CANCER;
format
REGION_CAT REGION_CATFmt. SEX_CAT SEX_CATFmt. RACE_CAT
RACE_CATFmt. AGE_CAT AGE_CATFmt.
MAR_CAT MAR_CATFmt. CANCER_CAT CANCER_CATFmt.
SMOKE_CAT SMOKE_CATFmt. ACTIV_CAT ACTIV_CATFmt. ALCOHOL_CAT
ALCOHOL_CATFmt. BMI_CAT BMI_CATFmt.
INSURANCE_CAT INSURANCE_CATFmt. EDUC_CAT EDUC_CATFmt. INCOME_CAT
INCOME_CATFmt.
table CANCER_CAT*(REGION_CAT SEX_CAT RACE_CAT AGE_CAT MAR_CAT
SMOKE_CAT ACTIV_CAT ALCOHOL_CAT BMI_CAT INSURANCE_CAT EDUC_CAT
INCOME_CAT)
/ chisq;
title "Relationships Between Psychological Distress"
title2 "And Each of the Categorical"
title3 "Explanatory Variables For Prostate Cancer Patients";
run;
/***************************************************************************
*********** start fitting the ordinal model - Breast Cancer -
************
************ Cumulative Odds Model
***************************************************************************/
proc logistic data=thesis.analysis_B; * proportional odds cumulative logit
model;
class RACE_CAT AGE_CAT MAR_CAT BREAST_CAT SMOKE_CAT
ACTIV_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT / param=ref;
model PD_CAT = RACE_CAT AGE_CAT MAR_CAT BREAST_CAT SMOKE_CAT
ACTIV_CAT INSURANCE_CAT EDUC_CAT INCOME_CAT / aggregate scale=none
expb;
title1 'proportional odds cumulative logit model - Breast Cancer -';
run;
end;
*proc genmod data=analysis1;
*class AGE_CAT MAR_CAT BREAST_CAT SMOKE_CAT
ACTIV_CAT INSURANCE_CAT INCOME_CAT;
*model PD_CAT = AGE_CAT MAR_CAT BREAST_CAT SMOKE_CAT
ACTIV_CAT INSURANCE_CAT INCOME_CAT / dist=multinomial
link=clogit;
*run;
*proc freq data=analysis;
*table Breast_CAT*PD_CAT CANCER_CAT*PD_CAT COLON_CAT*PD_CAT LUNG_CAT*PD_CAT
PROSTATE_CAT*PD_CAT / chisq; *run;
/* trying to treat missing values */
proc logistic data=analysis;
class SMOKE_CAT ;
model LUNG_CAT= SMOKE_CAT;
output pred=breast_p out=miss1;
run;
proc print data=miss1;
run;
data miss2;
set miss1;
if breast_p>.5 then breast_c =1;
else breast_c=0;
run;
proc print data=miss2;run;
data="miss1";

var breast_p ;
histogram;
run;

proc freq data=analysis;run;

/******************** start fitting the ordinal model - Colon Cancer -
***************/
/******************** Cumulative Odds Model
***************************************************************************
proc logistic data=thesis.analysis_C; * proportional odds cumulative logit
model;
class RACE_CAT AGE_CAT MAR_CAT COLON_CAT SMOKE_CAT
   ACTIV_CAT _INSURANCE_CAT EDUC_CAT INCOME_CAT / param=ref;
model PD_CAT = RACE_CAT AGE_CAT MAR_CAT COLON_CAT SMOKE_CAT
   ACTIV_CAT _INSURANCE_CAT EDUC_CAT INCOME_CAT / aggregate
scale=none expb;
title1 'proportional odds cumulative logit model - Colon Cancer -';
run;
end;
/******************** start fitting the ordinal model - LUNG Cancer -
***************/
/******************** Cumulative Odds Model
***************************************************************************
proc logistic data=thesis.analysis_L; * proportional odds cumulative logit
model;
class RACE_CAT AGE_CAT MAR_CAT LUNG_CAT SMOKE_CAT
   ACTIV_CAT _INSURANCE_CAT EDUC_CAT INCOME_CAT / param=ref;
model PD_CAT = RACE_CAT AGE_CAT MAR_CAT LUNG_CAT SMOKE_CAT
   ACTIV_CAT _INSURANCE_CAT EDUC_CAT INCOME_CAT / aggregate
scale=none expb;
title1 'proportional odds cumulative logit model - Lung Cancer -';
run;
end;
/******************** start fitting the ordinal model - PROSTATE Cancer -
***************/
/******************** Cumulative Odds Model
***************************************************************************
proc logistic data=thesis.analysis_P;
class RACE_CAT AGE_CAT MAR_CAT PROSTATE_CAT INSURANCE_CAT EDUC_CAT /
   param=ref;
model PD_CAT = RACE_CAT AGE_CAT MAR_CAT PROSTATE_CAT INSURANCE_CAT EDUC_CAT
   / aggregate scale=none expb;
title1 'Proportional Odds Model- PROSTATE Cancer - Prostate Cancer -';
run;

proc logistic data=thesis.analysis_P;
class RACE_CAT AGE_CAT MAR_CAT PROSTATE_CAT INSURANCE_CAT EDUC_CAT /
   param=ref;
model PD_CAT = RACE_CAT AGE_CAT MAR_CAT PROSTATE_CAT INSURANCE_CAT EDUC_CAT
   / link=cloglog aggregate scale=none expb;
title1 'Proportional Odds Model- PROSTATE Cancer - Prostate Cancer -';
run;
end;
proc freq data=analysis4;run;
/******************** Continuation Ratio Model in R
***************************************************************************/
/******************** Adjacent Categories Model in R 
***************************************************/

/******************************* graphing **********************/
proc gchart data=thesis.analysis;
   format PD_CAT PD_CATFmt. PD_CAT PD_CATFmt.;
   vbar PD_CAT discrete type=percent ;
   title1 'The Psychological Distress Distribution';
run;

proc gchart data=thesis.analysis;
   format REGION_CAT REGION_CATFmt. PD_CAT PD_CATFmt.;
   vbar REGION_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Region and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format SEX_CAT SEX_CATFmt. PD_CAT PD_CATFmt.;
   vbar SEX_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents gender and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format RACE_CAT RACE_CATFmt. PD_CAT PD_CATFmt.;
   vbar RACE_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Race and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format AGE_CAT AGE_CATFmt. PD_CAT PD_CATFmt.;
   vbar AGE_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Age and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format MAR_CAT MAR_CATFmt. PD_CAT PD_CATFmt.;
   vbar MAR_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Marital Status and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format SMOKE_CAT SMOKE_CATFmt. PD_CAT PD_CATFmt.;
   vbar SMOKE_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Smoking Status and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format ACTIV_CAT ACTIV_CATFmt. PD_CAT PD_CATFmt.;
   vbar ACTIV_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Activity and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format BMI_CAT BMI_CATFmt. PD_CAT PD_CATFmt.;
   vbar BMI_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents BMI and Psychological Distress';
run;

proc gchart data=thesis.analysis;
   format INSURANCE_CAT INSURANCE_CATFmt. PD_CAT PD_CATFmt.;
   vbar INSURANCE_CAT discrete type=percent subgroup=PD_CAT;
   title1 'The Distribution the respondents Insurance and Psychological Distress';
run;
```
proc gchart data=thesis.analysis;
format EDUC_CAT EDUC_CATFmt.  PD_CAT PD_CATFmt.;
vbar EDUC_CAT/ discrete type=percent subgroup=PD_CAT;
title='The Distribution the respondents Education level and Psychological Distress';
run;
proc gchart data=thesis.analysis;
format INCOME_CAT INCOME_CATFmt.  PD_CAT PD_CATFmt.;
vbar INCOME_CAT/ discrete type=percent subgroup=PD_CAT;
title='The Distribution the respondents Income and Psychological Distress';
run;
proc gchart data=thesis.analysis;
format CANCER_CAT CANCER_CATFmt.  PD_CAT PD_CATFmt.;
vbar PD_CAT/ discrete type=percent subgroup=CANCER_CAT;
title='The Distribution Psychological Distress among Cancer Patients';
run;
proc gchart data=thesis.analysis;
format BREAST_CAT BREAST_CATFmt.  PD_CAT PD_CATFmt.;
vbar PD_CAT/ discrete type=percent subgroup=BREAST_CAT;
run;
proc gchart data=thesis.analysis;
format COLON_CAT COLON_CATFmt.  PD_CAT PD_CATFmt.;
vbar PD_CAT/ discrete type=percent subgroup=COLON_CAT;
run;
proc gchart data=thesis.analysis;
format LUNG_CAT LUNG_CATFmt.  PD_CAT PD_CATFmt.;
vbar PD_CAT/ discrete type=percent subgroup=LUNG_CAT;
run;
proc gchart data=thesis.analysis;
format PROSTATE_CAT PROSTATE_CATFmt.  PD_CAT PD_CATFmt.;
vbar PD_CAT/ discrete type=percent subgroup=PROSTATE_CAT;
run;
```

**R code:**

```r
# For Cancer sample:
library(sas7bdat)
library(VGAM)
library(rms)
library(ordinal)
##1 Reading the SAS file --- Cancer sample
data.cancer <- read.sas7bdat("D:/BSU/Thesis/cancer/analysis_Cancer.sas7bdat")
head(data.cancer)
tail(data.cancer)
class(data.cancer)
attach(data.cancer)

# Creating dummy variables
Race_1 = as.numeric(data.cancer$RACE_CAT==1)
Race_2 = as.numeric(data.cancer$RACE_CAT==2)
Race_3 = as.numeric(data.cancer$RACE_CAT==3)
Age_1 = as.numeric(data.cancer$AGE_CAT==1)
```
Age_2 = as.numeric(data.cancer$AGE_CAT==2)
BMI_1 = as.numeric(data.cancer$BMI_CAT==1)
BMI_2 = as.numeric(data.cancer$BMI_CAT==2)
BMI_3 = as.numeric(data.cancer$BMI_CAT==3)
Income_1 = as.numeric(data.cancer$INCOME_CAT==1)
Income_2 = as.numeric(data.cancer$INCOME_CAT==2)

# proportional odds model
fit.prop <- vglm(PD_CAT ~ SEX_CAT+Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+CANCER_CAT+SMOKE_CAT+ACTIV_CAT +BMI_1+BMI_2+BMI_3+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2, 
      family=cumulative(parallel=TRUE), data=data.cancer)
fit.prop1 <- vglm(PD_CAT ~ 1, 
      family=cumulative(parallel=TRUE), data=data.cancer)
summary(fit.prop)
summary(fit.prop1)
ctable1 = coef(summary(fit.prop))
p1 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR = exp(coef(fit.prop))
(ctable1 <- cbind(ctable1, `p value` = p1, OR))

coeffs = coef(fit.prop)
se = diag(vcov(fit.prop))^0.5
L = coeffs-1.96*se
U = coeffs+1.96*se
ci.or.prop = exp(cbind(L,U))
ci.or.prop

def.prop=fitted(fit.prop)
res=resid(fit.prop)
r1=res[,1]
r2=res[,2]
plot(r1)
plot(r2)
AIC(fit.prop)

# the deviance
pchisq(deviance(fit.prop1)-deviance(fit.prop),df=df.residual(fit.prop1)-df.residual(fit.prop),lower.tail=FALSE)

#Adjacent-Categories Logit Model:
fit.adj <- vglm(PD_CAT ~ SEX_CAT+Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+CANCER_CAT+SMOKE_CAT+ACTIV_CAT +BMI_1+BMI_2+BMI_3+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2, 
      family=acat(reverse=TRUE, parallel=TRUE), data=data.cancer)
fit.adj1 <- vglm(PD_CAT ~ 1, 
      family=acat(reverse=TRUE, parallel=TRUE), data=data.cancer)
summary(fit.adj)
summary(fit.adj1)
ctable1 = coef(summary(fit.adj))
p2 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR = exp(coef(fit.adj))
(ctable1 <- cbind(ctable1, `p value` = p2, OR))

coeffs2 = coef(fit.adj)
se2 = diag(vcov(fit.adj))^0.5

L = coeffs2 - 1.96*se2
U = coeffs2 + 1.96*se2

ci.or.adj = exp(cbind(L, U))
ci.or.adj

fitted(fit.adj)
AIC(fit.adj)
pchisq(deviance(fit.adj1) - deviance(fit.adj), df = df.residual(fit.adj1) - df.residual(fit.adj), lower.tail = FALSE)

# Continuation-Ratio Logit Model:
fit.cratio <- vglm(PD_CAT ~ SEX_CAT + Race_1 + Race_2 + Race_3 + Age_1 + Age_2 + MAR_CAT + CANCER_CAT + SMOKE_CAT + ACTIV_CAT + BMI_1 + BMI_2 + BMI_3 + INSURANCE_CAT + EDUC_CAT + Income_1 + Income_2,
family = cratio(reverse = FALSE, parallel = TRUE), data = data.cancer)
fit.cratio1 <- vglm(PD_CAT ~ 1,
family = cratio(reverse = FALSE, parallel = TRUE), data = data.cancer)

summary(fit.cratio)
summary(fit.cratio1)
ctable1 = coef(summary(fit.cratio))
p3 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR = exp(coef(fit.cratio))
(ctable1 <- cbind(ctable1, `p value` = p3, OR))

coeffs3 = coef(fit.cratio)
se3 = diag(vcov(fit.cratio))^0.5

L = coeffs3 - 1.96*se3
U = coeffs3 + 1.96*se3

ci.or.cratio = exp(cbind(L, U))
ci.or.cratio

pred.cont = fitted(fit.cratio)
AIC(fit.cratio)
pchisq(deviance(fit.cratio1) - deviance(fit.cratio), df = df.residual(fit.cratio1) - df.residual(fit.cratio), lower.tail = FALSE)

## barplots
# Psychological Distress vs. Gender
bartable_1 = table(data.cancer$PD_CAT, data.cancer$SEX)  ## get the cross tab
barplot(bartable_1,
     main="The distribution of
the Psychological Distress
Based on Gender",
     ylab= "Frequency", names.arg=c("Female","Male"),
     beside=TRUE, col=heat.colors(3))
legend("topright", c("Low","Moderate","High"), cex=0.6,
     bty="n", fill=heat.colors(3))

# Breast Cancer vs. Gender
bartable_2 = table(data.cancer$CANCER_CAT, data.cancer$SEX_CAT)  ## get the cross tab
barplot(bartable_2,
     main="The distribution of
The Breast Cancer
Based on the Gender", xlab= "Gender",
     ylab= "Frequency", names.arg=c("Female","Male"),beside=TRUE,
     col=c("darkblue","red"))
legend("topright", c("No","Yes"), cex=0.6,
     bty="n",
     fill=c("darkblue","red"))

# Psychological Distress vs. Race
bartable_3 = table(data_b$PD_CAT, data_b$RACE_CAT)  ## get the cross tab
barplot(bartable_3,
     main="The distribution of
the Psychological Distress
Based on the Race",
     xlab= "Race",
     ylab= "Frequency", names.arg=c("White","Black","Asian","Others"),beside=TRUE, col=heat.colors(3))
legend("topright", c("Low","Moderate","High"), cex=0.6,
     bty="n",
     fill=heat.colors(3))

# Breast Cancer vs. Race
bartable_4 = table(data_b$BREAST_CAT, data_b$RACE_CAT)  ## get the cross tab
barplot(bartable_4,
     main="The distribution of
The Breast Cancer
Based on the Race", xlab= "Race",
     ylab= "Frequency", names.arg=c("White","Black","Asian","Others"),beside=TRUE, col=heat.colors(3))
legend("topright", c("No","Yes"), cex=0.6,
     bty="n",
     fill=c("darkblue","red"))

# Psychological Distress vs. Age
bartable_2 = table(data_b$PD_CAT, data_b$RACE_CAT)  ## get the cross tab
barplot(bartable_2,
     main="The distribution of
the Psychological Distress
Based on the Race",
     ylab= "Frequency", names.arg=c("White","Black/African","Asian","Others"),beside=TRUE, col=heat.colors(3))
legend("topright", c("Low","Moderate","High"), cex=0.6,
     bty="n",
     fill=heat.colors(3))

>> For Breast Cancer sub-sample:
## Reading the SAS file --- Breast Cancer sub-sample

data_b <- read.sas7bdat("D:/BSU/Thesis/cancer/analysis_b.sas7bdat")
head(data_b)
tail(data_b)
class(data_b)
attach(data_b)

# Creating dummy variables
Race_1 = as.numeric(data_b$RACE_CAT==1)
Race_2 = as.numeric(data_b$RACE_CAT==2)
Race_3 = as.numeric(data_b$RACE_CAT==3)

Age_1 = as.numeric(data_b$AGE_CAT==1)
Age_2 = as.numeric(data_b$AGE_CAT==2)

BMI_1 = as.numeric(data_b$BMI_CAT==1)
BMI_2 = as.numeric(data_b$BMI_CAT==2)
BMI_3 = as.numeric(data_b$BMI_CAT==3)

Income_1 = as.numeric(data_b$INCOME_CAT==1)
Income_2 = as.numeric(data_b$INCOME_CAT==2)

# proportional odds model
fit.prop <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+BREAST_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2,
family=cumulative(parallel=TRUE), data=data_b)
fit.prop1 <- vglm(PD_CAT ~ 1, family=cumulative(parallel=TRUE), data=data_b)
summary(fit.prop)
summary(fit.prop1)
ctable1=coef(summary(fit.prop))
p1 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.prop))
(ctable1 <- cbind(ctable1, `p value` = p1, OR))

coeffs = coef(fit.prop)
se = diag(vcov(fit.prop))^0.5
L = coeffs-1.96*se
U = coeffs+1.96*se
ci.or.prop = exp(cbind(L,U))
ci.or.prop

pred.prop=fitted(fit.prop)
res=resid(fit.prop)
r1=res[,1]
r2=res[,2]
plot(r1)
plot(r2)
AIC(fit.prop)

# the deviance
pchisq(deviance(fit.prop1)-deviance(fit.prop),df=df.residual(fit.prop1)-df.residual(fit.prop),lower.tail=FALSE)
# Adjacent-Categories Logit Model:

```r
fit.adj <- vglm(PD_CAT ~ 
Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+BREAST_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+
EDUC_CAT+Income_1+Income_2,
    family=acat(reverse=TRUE, parallel=TRUE), data=data_b)
```

```r
fit.adj1 <- vglm(PD_CAT ~ 1, family=acat(reverse=TRUE, parallel=TRUE), data=data_b)
```

```r
summary(fit.adj)
summary(fit.adj1)
```

```r
ctable1=coef(summary(fit.adj))
p2 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.adj))
(ctable1 <- cbind(ctable1, `p value` = p2, OR))
```

```r
coeffs2 = coef(fit.adj)
se2 = diag(vcov(fit.adj))^.5
```

```r
L = coeffs2-1.96*se2
U = coeffs2+1.96*se2
```

```r
ci.or.adj = exp(cbind(L,U))
ci.or.adj
fitted(fit.adj)
AIC(fit.adj)
pchisq(deviance(fit.adj1)-deviance(fit.adj),df=df.residual(fit.adj1)-df.residual(fit.adj),lower.tail=FALSE)
```

# Continuation-Ratio Logit Model:

```r
fit.cratio <- vglm(PD_CAT ~ 
Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+BREAST_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+
EDUC_CAT+Income_1+Income_2,
    family=cratio(reverse=FALSE, parallel=TRUE), data=data_b)
```

```r
fit.cratio1 <- vglm(PD_CAT ~ 1, family=cratio(reverse=FALSE, parallel=TRUE), data=data_b)
```

```r
summary(fit.cratio)
summary(fit.cratio1)
```

```r
ctable1=coef(summary(fit.cratio))
p3 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.cratio))
(ctable1 <- cbind(ctable1, `p value` = p3, OR))
```

```r
coeffs3 = coef(fit.cratio)
se3 = diag(vcov(fit.cratio))^.5
```

```r
L = coeffs3-1.96*se3
U = coeffs3+1.96*se3
```

```r
ci.or.cratio = exp(cbind(L,U))
ci.or.cratio
pred.cont=fitted(fit.cratio)
AIC(fit.cratio)
```
pchisq(deviance(fit.cratio1) - deviance(fit.cratio), df = df.residual(fit.cratio1) - df.residual(fit.cratio), lower.tail = FALSE)

>> For Colon Cancer sub-sample:

##1 Reading the SAS file --- Colon Cancer sub-sample
data_c <- read.sas7bdat("D:/BSU/Thesis/cancer/analysis_c.sas7bdat")
head(data_c)
tail(data_c)
class(data_c)
attach(data_c)

# Creating dummy variables
Race_1 = as.numeric(data_c$RACE_CAT==1)
Race_2 = as.numeric(data_c$RACE_CAT==2)
Race_3 = as.numeric(data_c$RACE_CAT==3)

Age_1 = as.numeric(data_c$AGE_CAT==1)
Age_2 = as.numeric(data_c$AGE_CAT==2)

BMI_1 = as.numeric(data_c$BMI_CAT==1)
BMI_2 = as.numeric(data_c$BMI_CAT==2)
BMI_3 = as.numeric(data_c$BMI_CAT==3)

Income_1 = as.numeric(data_c$INCOME_CAT==1)
Income_2 = as.numeric(data_c$INCOME_CAT==2)

# proportional odds model
fit.prop <- vglm(PD_CAT ~ Race_1 + Race_2 + Race_3 + Age_1 + Age_2 + MAR_CAT + COLON_CAT + SMOKE_CAT + ACTIV_CAT + INSURANCE_CAT + EDUC_CAT + Income_1 + Income_2,
             family = cumulative(parallel = TRUE), data = data_c)
fit.prop1 <- vglm(PD_CAT ~ 1,
             family = cumulative(parallel = TRUE), data = data_c)
summary(fit.prop)
summary(fit.prop1)
ctable1 = coef(summary(fit.prop))
p1 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR = exp(coef(fit.prop))
(ctable1 <- cbind(ctable1, `p value` = p1, OR))

coeffs = coef(fit.prop)
se = diag(vcov(fit.prop))^.5

L = coeffs - 1.96*se
U = coeffs + 1.96*se
ci.or.prop = exp(cbind(L, U))
ci.or.prop

pred.prop = fitted(fit.prop)
res = resid(fit.prop)
r1 = res[, 1]
r2 = res[, 2]
plot(r1)
plot(r2)
AIC(fit.prop)

# the deviance
pchisq(deviance(fit.prop1)-deviance(fit.prop),df=df.residual(fit.prop1)-df.residual(fit.prop),lower.tail=FALSE)

#Adjacent-Categories Logit Model:
fit.adj <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+COLON_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2, family=acat(reverse=TRUE, parallel=TRUE), data=data_c)
fitted(fit.adj)
AIC(fit.adj)

ctable1=coef(summary(fit.adj))
p2 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.adj))
(ctable1 <- cbind(ctable1, `p value` = p2,OR))

coeffs2 = coef(fit.adj)
se2 = diag(vcov(fit.adj))^.5
L = coeffs2-1.96*se2
U = coeffs2+1.96*se2
ci.or.adj = exp(cbind(L,U))
ci.or.adj

fitted(fit.adj)
AIC(fit.adj)

pchisq(deviance(fit.adj1)-deviance(fit.adj),df=df.residual(fit.adj1)-df.residual(fit.adj),lower.tail=FALSE)

# Continuation-Ratio Logit Model:
fit.cratio <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+COLON_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2, family=cratio(reverse=FALSE, parallel=TRUE), data=data_c)
fitted(fit.cratio)
AIC(fit.cratio)

ctable1=coef(summary(fit.cratio))
p3 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.cratio))
(ctable1 <- cbind(ctable1, `p value` = p3,OR))

coeffs3 = coef(fit.cratio)
se3 = diag(vcov(fit.cratio))^.5
L = coeffs3-1.96*se3
U = coeffs3+1.96*se3

ci.or.cratio = exp(cbind(L,U))
ci.or.cratio

pred.cont=fitted(fit.cratio)
AIC(fit.cratio)

pchisq(deviance(fit.cratio1)-deviance(fit.cratio),df=df.residual(fit.cratio1)-df.residual(fit.cratio),lower.tail=FALSE)

>> For Lung Cancer sub-sample:

# Reading the SAS file --- Lung Cancer sub-sample
data_l <- read.sas7bdat("D:/BSU/Thesis/cancer/analysis_l.sas7bdat")
head(data_l)
tail(data_l)
class(data_l)
attach(data_l)

# Creating dummy variables
Race_1 = as.numeric(data_l$RACE_CAT==1)
Race_2 = as.numeric(data_l$RACE_CAT==2)
Race_3 = as.numeric(data_l$RACE_CAT==3)
Age_1 = as.numeric(data_l$AGE_CAT==1)
Age_2 = as.numeric(data_l$AGE_CAT==2)
BMI_1 = as.numeric(data_l$BMI_CAT==1)
BMI_2 = as.numeric(data_l$BMI_CAT==2)
BMI_3 = as.numeric(data_l$BMI_CAT==3)
Income_1 = as.numeric(data_l$INCOME_CAT==1)
Income_2 = as.numeric(data_l$INCOME_CAT==2)

# proportional odds model
fit.prop <- vglm(PD_CAT ~
Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+LUNG_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE
_CAT+EDUC_CAT+Income_1+Income_2,
family=cumulative(parallel=TRUE), data=data_l)
fit.prop1 <- vglm(PD_CAT ~ 1,
family=cumulative(parallel=TRUE), data=data_l)
summary(fit.prop)
summary(fit.prop1)
ctable1=coef(summary(fit.prop))
p1 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.prop))
(ctable1 <- cbind(ctable1, `p value` = p1,OR))

coeffs = coef(fit.prop)
se = diag(vcov(fit.prop))^=.5

L = coeffs-1.96*se
U = coeffs+1.96*se

ci.or.prop = exp(cbind(L,U))
ci.or.prop

pred.prop=fitted(fit.prop)
res=resid(fit.prop)
r1=res[,1]
r2=res[,2]
plot(r1)
plot(r2)
AIC(fit.prop)

# the deviance
pchisq(deviance(fit.prop1)-deviance(fit.prop),df=df.residual(fit.prop1)-df.residual(fit.prop),lower.tail=FALSE)

#Adjacent-Categories Logit Model:
fit.adj <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+LUNG_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2,
               family=acat(reverse=TRUE, parallel=TRUE), data=data_l)
fit.adj1 <- vglm(PD_CAT ~ 1,
                family=acat(reverse=TRUE, parallel=TRUE), data=data_l)
summary(fit.adj)
summary(fit.adj1)
ctable1=coef(summary(fit.adj))
p2 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.adj))
(ctable1 <- cbind(ctable1, `p value` = p2,OR))

coeffs2 = coef(fit.adj)
se2 = diag(vcov(fit.adj))^.5

L = coeffs2-1.96*se2
U = coeffs2+1.96*se2

ci.or.adj = exp(cbind(L,U))
ci.or.adj

fitted(fit.adj)
AIC(fit.adj)

pchisq(deviance(fit.adj1)-deviance(fit.adj),df=df.residual(fit.adj1)-df.residual(fit.adj),lower.tail=FALSE)

# Continuation-Ratio Logit Model:
fit.cratio <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+LUNG_CAT+SMOKE_CAT+ACTIV_CAT+INSURANCE_CAT+EDUC_CAT+Income_1+Income_2,
                  family=cratio(reverse=FALSE, parallel=TRUE), data=data_l)
fit.cratio1 <- vglm(PD_CAT ~ 1,
                    family=cratio(reverse=FALSE, parallel=TRUE), data=data_l)
summary(fit.cratio)
summary(fit.cratio1)
table1=coef(summary(fit.cratio))
c

c <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.cratio))
(ctype1 <- cbind(ctable1, `p value` = p3,OR))

coeffs3 = coef(fit.cratio)
se3 = diag(vcov(fit.cratio))^0.5

L = coeffs3-1.96*se3
U = coeffs3+1.96*se3

ci.or.cratio = exp(cbind(L,U))
ci.or.cratio

pred.cont=fitted(fit.cratio)
AIC(fit.cratio)

pchisq(deviance(fit.cratio1)-deviance(fit.cratio),df=df.residual(fit.cratio1)-df.residual(fit.cratio),lower.tail=FALSE)

>> For Prostate Cancer sub-sample:

#1 Reading the SAS file -- prostate Cancer sub-sample
data_p = read.sas7bdat("D:/BSU/Thesis/cancer/analysis_p.sas7bdat")
head(data_p)
tail(data_p)
class(data_p)
attach(data_p)

# Creating dummy variables
Race_1 = as.numeric(data_p$RACE_CAT==1)
Race_2 = as.numeric(data_p$RACE_CAT==2)
Race_3 = as.numeric(data_p$RACE_CAT==3)
Age_1 = as.numeric(data_p$AGE_CAT==1)
Age_2 = as.numeric(data_p$AGE_CAT==2)
BMI_1 = as.numeric(data_p$BMI_CAT==1)
BMI_2 = as.numeric(data_p$BMI_CAT==2)
BMI_3 = as.numeric(data_p$BMI_CAT==3)
Income_1 = as.numeric(data_p$INCOME_CAT==1)
Income_2 = as.numeric(data_p$INCOME_CAT==2)

# proportional odds model
fit.prop <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+PROSTATE_CAT+INSURANCE_CAT+EDUC_CAT,
family=cumulative(parallel=TRUE), data=data_p)
fit.prop1 <- vglm(PD_CAT ~ 1,
family=cumulative(parallel=TRUE), data=data_p)
summary(fit.prop)
s
summary(fit.prop1)
c
ctable1=coef(summary(fit.prop))
p1 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.prop))
(ctable1 <- cbind(ctable1, `p value` = p1,OR))

coeffs = coef(fit.prop)
se = diag(vcov(fit.prop))^0.5

L = coeffs-1.96*se
U = coeffs+1.96*se

ci.or.prop = exp(cbind(L,U))
ci.or.prop

pred.prop=fitted(fit.prop)
res=resid(fit.prop)
r1=res[,1]
r2=res[,2]
plot(r1)
plot(r2)
AIC(fit.prop)

# the deviance
pchisq(deviance(fit.prop1)-deviance(fit.prop),df=df.residual(fit.prop1)-df.residual(fit.prop),lower.tail=FALSE)

#Adjacent-Categories Logit Model:
fit.adj <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+PROSTATE_CAT+INSURANCE_CAT+EDUC_CAT,
family=acat(reverse=TRUE, parallel=TRUE), data=data_p)
fit.adj1 <- vglm(PD_CAT ~ 1,
family=acat(reverse=TRUE, parallel=TRUE), data=data_p)
summary(fit.adj)
summary(fit.adj1)
ctable1=coef(summary(fit.adj))
p2 <- pnorm(abs(ctable1[, "z value"])), lower.tail = FALSE) * 2
OR=exp(coef(fit.adj))
(ctable1 <- cbind(ctable1, `p value` = p2,OR))

coeffs2 = coef(fit.adj)
se2 = diag(vcov(fit.adj))^0.5

L = coeffs2-1.96*se2
U = coeffs2+1.96*se2

ci.or.adj = exp(cbind(L,U))
ci.or.adj

fitted(fit.adj)
AIC(fit.adj)

pchisq(deviance(fit.adj1)-deviance(fit.adj),df=df.residual(fit.adj1)-df.residual(fit.adj),lower.tail=FALSE)

# Continuation-Ratio Logit Model:
fit.cratio <- vglm(PD_CAT ~ Race_1+Race_2+Race_3+Age_1+Age_2+MAR_CAT+PROSTATE_CAT+INSURANCE_CAT+EDUC_CAT, 
family=cratio(reverse=FALSE, parallel=TRUE), data=data_p)

fit.cratio1 <- vglm(PD_CAT ~ 1, 
family=cratio(reverse=FALSE, parallel=TRUE), data=data_p)

summary(fit.cratio)
summary(fit.cratio1)

ctable1=coef(summary(fit.cratio))
p3 <- pnorm(abs(ctable1[, "z value"]), lower.tail = FALSE) * 2
OR=exp(coef(fit.cratio))
(ctable1 <- cbind(ctable1, `p value` = p3,OR))

coeffs3 = coef(fit.cratio)
se3 = diag(vcov(fit.cratio))^0.5

L = coeffs3-1.96*se3
U = coeffs3+1.96*se3

ci.or.cratio = exp(cbind(L,U))
ci.or.cratio

pred.cont=fitted(fit.cratio)
AIC(fit.cratio)

pchisq(deviance(fit.cratio1)-deviance(fit.cratio),df=df.residual(fit.cratio1)-df.residual(fit.cratio),lower.tail=FALSE)