**PHC 6053 – Exam 2 (100 points)**

**Directions:**

* You will work in **groups of 3-5** to complete this exam. You can discuss your work with other groups.
* Each student can submit their own solution to any question for which agreement cannot be reached on a solution, however, I highly urge you to resolve conflicts in solutions as a group. Any dissenting solutions to questions must be clearly labeled and documented to be considered.
* Provide all group answers on this exam and SHOW ALL WORK.

**Names of Group Members:**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

|  |
| --- |
|  |

**All questions (or parts of questions) are worth 2 points unless otherwise stated.**

**Dataset Information:** Two datasets are used for this exam.

|  |
| --- |
| **PART 1 - BEARS:** The first contains body measurements and gender of black bears. The goal of the dataset it to accurately predict the weight of bears using easier to obtain quantities. The data contain information on black bears under the age of 96 months (8 years). The variables of interest are   * Outcome = Weight (pounds) * Predictors:   + Chest Girth (inches)   + Chest Group: S = less than 30, M = [30, 40), and L = 40 or larger   + Age (months)   + Gender (M=Male, F=Female)   Analyses for this dataset are available as a PDF file called **BEARSFULL.PDF** |
| **PART 2 - NHANES:** The second consists of data on 5104 individuals surveyed as part of the NHANES (National Health and Nutrition Examination Survey).  **The variable list from PROC CONTENTS is provided to the right. Review this list and the following information VERY carefully.**  **Goal: Investigate predictors of High Blood Pressure.**  **Categorical Variable Information:**  The needed information from the PROC FORMAT code for the three categorical variables is provided to the right.  **Other Important Comments:**  Only individuals 35 or older in the groups defined by our categorical variables are considered. Individuals who were classified as underweight or race = “Other” were not included in analyses.  You can use the phrase “among our population” or similar vague explanation of the population in your interpretations.    **Analyses for this data are included in this document.** |

**PART 1: Bears Data (50 points)**

1. **Model 1:** Estimate and interpret the effect of a 10-inch increase in chest girth based upon this model.

|  |
| --- |
|  |

1. **Model 2:** Estimate and interpret the effect of a 10-inch increase in chest girth based upon this model using the following comparisons:
   1. Increase from 20 inches to 30 inches

|  |
| --- |
|  |

* 1. Increase from 30 inches to 40 inches

|  |
| --- |
|  |

* 1. Increase from 40 inches to 50 inches

|  |
| --- |
|  |

1. **Model 3:** Estimate and interpret the effects below:
   1. Comparing Medium (M) to Small (S)

|  |
| --- |
|  |

* 1. Comparing Large (L) to Medium (M)

|  |
| --- |
|  |

1. **(6 points) Comparing Models 1, 2, and 3:** Which would you choose to provide the best prediction of the weights of bears under age 96 months? Be sure to address any regression assumptions which are of concern as well as any other issues you feel are important.

|  |
| --- |
|  |

1. **(6 points) Using Models 1, 2, and 3:** For the 5 NEW bears below (not used in our model), predict the weight of the bear and provide the residual error for each of these three models.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **weight** | **chest** | **Model 1** | **Model 1 Error** | **Model 2** | **Model 2 Error** | **Model 3** | **Model 3 Error** |
| 22 | 132 | 33 |  |  |  |  |  |  |
| 23 | 90 | 28 |  |  |  |  |  |  |
| 34 | 94 | 29 |  |  |  |  |  |  |
| 39 | 202 | 40 |  |  |  |  |  |  |
| 43 | 446 | 55 |  |  |  |  |  |  |

1. **Model 4:** Estimate and interpret the effects below:
   1. Effect of a 6 month increase in age for Gender = M

|  |
| --- |
|  |

* 1. Effect of a 6 month increase in age for Gender = F

|  |
| --- |
|  |

* 1. Comparing Males to Females at AGE = 24 months

|  |
| --- |
|  |

* 1. Comparing Males to Females at AGE = 60 months

|  |
| --- |
|  |

1. **Model 6:** Write the full estimated regression model for this model.

|  |
| --- |
|  |

1. **Model 6:** Estimate and interpret the following effects based upon this model using the following comparisons:
   1. Increase in chest girth from 20 inches to 30 inches

|  |
| --- |
|  |

* 1. Increase in chest girth from 40 inches to 50 inches

|  |
| --- |
|  |

* 1. Effect of a 6 month increase in age for Gender = M

|  |
| --- |
|  |

* 1. Effect of a 6 month increase in age for Gender = F

|  |
| --- |
|  |

* 1. Comparing Males to Females at AGE = 24 months

|  |
| --- |
|  |

* 1. Comparing Males to Females at AGE = 60 months

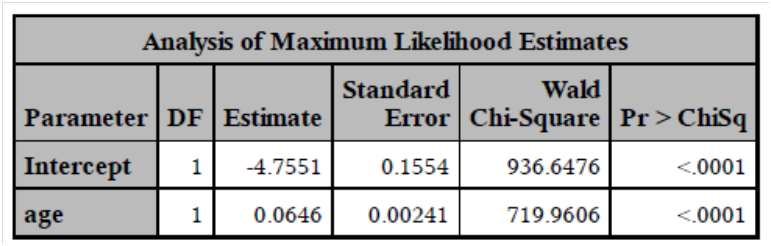
|  |
| --- |
|  |

1. **(4 points) Using Models 4 and 6:** For the 5 NEW bears below (not used in our model), predict the weight of the bear and provide the residual error using these two models.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **weight** | **chest** | **age** | **gender** | **Model 4** | **Model 4 Error** | **Model 6** | **Model 6 Error** |
| 22 | 132 | 33 | 81 | F |  |  |  |  |
| 23 | 90 | 28 | 21 | M |  |  |  |  |
| 34 | 94 | 29 | 10 | M |  |  |  |  |
| 39 | 202 | 40 | 58 | F |  |  |  |  |
| 43 | 446 | 55 | 70 | M |  |  |  |  |

**PART 2 – NHANES Data (50 points)**

1. The following is the SAS output for a simple logistic regression model evaluating the effect of age on the risk of high blood pressure (HBP). The probability of HBP = Yes is modeled.



* 1. Provide the estimated regression model.

|  |
| --- |
|  |

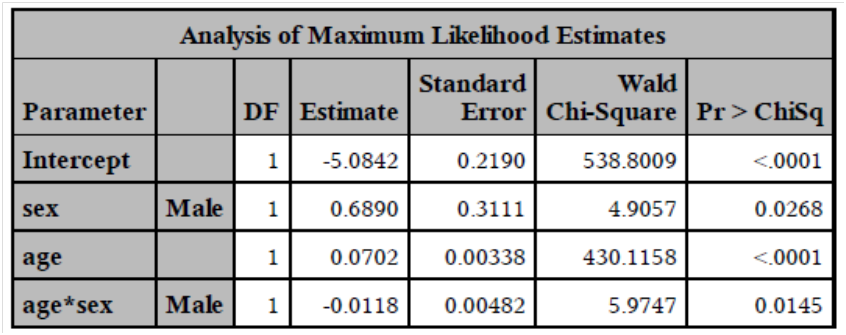
* 1. (3 points) Find the odds ratio of high blood pressure for a 1-year increase in age and the associated 95% confidence interval.

|  |
| --- |
|  |

* 1. Find the odds ratio of high blood pressure for a 10-year increase in age.

|  |
| --- |
|  |

1. The interaction effect between gender (SEX) and age was considered in a multiple logistic regression model. The output is provided below. The probability of HBP = Yes is modeled.



* 1. State the fitted (estimated) regression model.

|  |
| --- |
|  |

* 1. Calculate the odds ratio of high blood pressure comparing males to females for individuals age 40.

|  |
| --- |
|  |

* 1. Calculate the odds ratio of high blood pressure comparing males to females for individuals age 80.

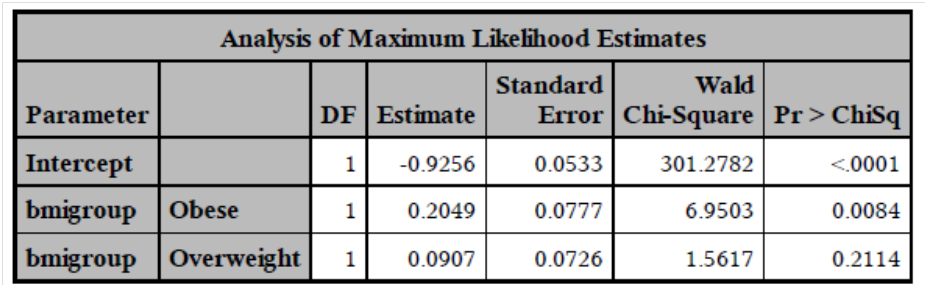
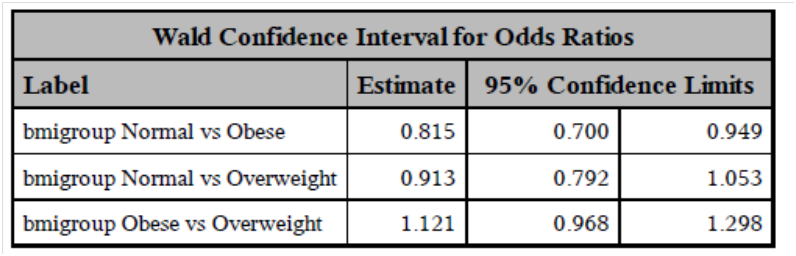
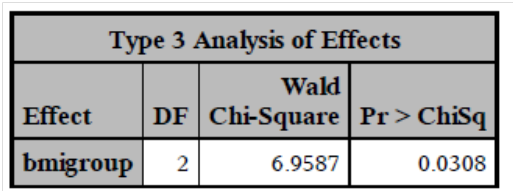
|  |
| --- |
|  |

* 1. Calculate the odds ratio of high blood pressure for a 10-year change in age among males.

|  |
| --- |
|  |

* 1. Calculate the predicted probability of high blood pressure for a male of age 40 and again for a female of age 40.

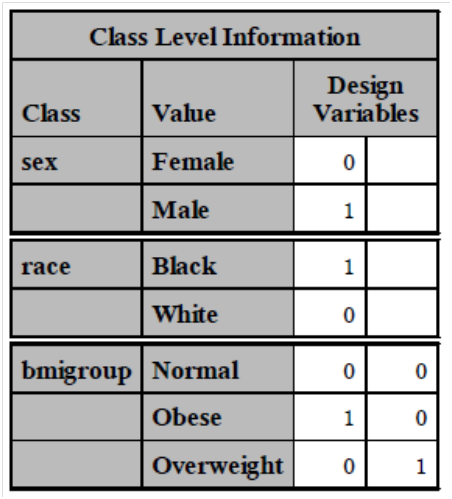
|  |
| --- |
|  |

1. The following is the SAS output for a simple logistic regression model evaluating the effect of BMI group on the risk of high blood pressure (HBP). The probability of HBP = Yes is modeled.   
     
      
     
     
     
     
     
     
     
     
      
     
     
     
   1. (6 points) Briefly discuss the conclusions for this predictor. Include interpretations of any significant odds ratios.

|  |
| --- |
|  |

* 1. (6 points) Calculate the predicted probability of high blood pressure in each group.

|  |
| --- |
|  |

1. A full multiple logistic regression model is performed. The output is provided as needed.
   1. (5 points) Consider the “Class Level Information” output from SAS. Explain clearly what this output indicates about the current regression model.

|  |
| --- |
|  |

* 1. (16 points) Using the output on the last page, correctly interpret all significant odds ratios presented.

|  |
| --- |
|  |

