**PH 717 Problem Set 5**

**Problem Set 5 has five (5) questions and is worth 100 points.** Adhere to the “Rules for Collaborating”.

* **Download** the data setbirthweight.csv
* **For hand calculations, show your work**. Do not round off during intermediate steps. Round to two decimal places for all final answers.
* **For R computations, insert your R code** (you can cut and paste into this document).
* For most questions, you can provide a full and thoughtful answer in 1-2 short sentences.

In this assignment, you will be using the dataset **birthweight.csv,** which contains data on a infants and their mothers.

Description of the variables:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Coding/Units** |
| ID | Study identifier |  |
| length | Length of baby (cm) | Continuous |
| Birthweight | Birth weight of baby (kg) | Continuous |
| gestation | Gestational age at birth (weeks) | Continuous |
| smoker | Mother smokes | 0=Non-smoker  1=Smoker |
| motherage | Maternal age | Continuous |
| mnocig | Number of cigarettes smoked per day by mother | Continuous |
| lowbwt | Low birth weight | 0=No  1=Yes |

1. First, compute descriptive statistics for the study sample. Fill in the values (as indicated by ‘**? (?)**’ – mean and standard deviation for continuous variables, frequency and percentages for non-continuous variables) for the following table and fill in the values using your output. Use the descriptive statistics to briefly describe the study sample (no more than one paragraph) **(2 points each; total=16 points)**

**Table 1: Patient Demographics**

|  |  |
| --- | --- |
|  | **Total number of infants**  **(N = ?)** |
|
| ***Continuous variables, mean (SD)*** |  |
| Length of baby (cm) | **? (?)** |
| Birth weight of baby (kg) | **? (?)** |
| Gestational age at birth (weeks) | **? (?)** |
| Maternal age | **? (?)** |
| Number of cigarettes smoked per day by mother | **? (?)** |
|  |  |
| ***Categorical variables, freq (%)*** |  |
| Mother smokes | **? (?)** |
| Low birth weight | **? (?)** |

1. Create a frequency histogram for the birthweights, and describe the distribution in a sentence. Then create a second histogram for the number of cigarettes smoked by the mother. Be sure to include labels for the x-axis of each figure (variable name) and main titles. Summarize each distribution in 1-2 sentences, and explain whether each roughly follows a normal distribution. **Include the histograms in your results**. **(4 points each graphs; 4 points for each summary = total 16 points)**
2. Consider the distribution of birthweight and use the normal probability model for the following computations. The following are to be done **by hand** (please show your work). You may use the mean and standard deviation of birth weight from Table 1 and you may also check your computations using R:
3. What is the probability that an infant has a birthweight above 4 kg? **(10 points)**
4. What is the probability that an infant has a birthweight between 3 and 3.5 kg? **(10 points)**

39%

1. What is the probability that an infant has a birthweight below 2.5 kg? **(10 points)**
2. Determine the 10th and 90th percentiles of birthweight among infants. **(10 points)**

When we 68.2%

1. Complete the following table by doing the following computations **by hand** (please show your work). You may check your results using R, but you must do the computations by hand. Be sure to include an interpretation of *both* of the 95% confidence intervals. **(12 points)**

**Table 2: 95% CIs for patient characteristics**

|  |  |
| --- | --- |
| **Patient Characteristic** | **95% Confidence Interval** |
| Maternal age |  |
| Low birthweight |  |

1. The BinaxNOW COVID-19 Antigen Self-Test is an at-home SARS-COV-2 rapid test that has received emergency use authorization from the FDA and is widely available for purchase. The manufacturer evaluated the performance of the BinaxNOW by comparing it to a gold standard “Comparator Method”. Overall, 117 subjects tested positive for SARS-COV-2 using the gold standard comparator method, and 99 of these had a positive BinaxNOW test. The other 343 subjects were negative based on the Comparator Method, and 5 of these had a positive BinaxNOW test.
2. Create a contingency table to summarize these results. **(4 points)**
3. Compute the sensitivity of the BinaxNOW screening test and describe the result in a sentence. Show your work. **(4 points)**
4. Compute the specificity of the BinaxNOW screening test and describe the result in a sentence. Show your work. **(4 points)**
5. Compute the positive predictive value of the BinaxNOW screening test and describe the result in a sentence. Show your work. **(4 points)**