**Epi 510, R assignment 4**

For this assignment, please submit two files: (1) **a file containing text and tables** (.doc, .docx, .pdf or .txt) answering questions posed in the assignment, and (2) **an R script** that performs the requested operation (.R).

**On this assignment, you’ll need to use the vipclsClean.rds from the Canvas site**. In homework 2 you should have created a clean version of the vipcls dataset with additional variables that you created, labels, and missing codes and outliers converted to missing. A few more cleaning steps have been performed on the vipcls dataset outside of these homework assignments, which is why you’ll need to download the version from Canvas. In this dataset, you should find binary variables coded as both 0/1 and 1/2 versions. You’ll also find three new variables:

* smokeFirst: indicates whether a mother smoked during the first trimester
* lbw: indicates whether birthweight was <2500 grams
* over25: whether maternal age is ≥25

1. Use commands such as str or glimpse and tabulate binary variables get a rough sense of the vipclsClean dataset and what new variables have been made for you. You do not need to submit output with this question. **(3 points)**
2. Explore the distribution of birth weight as follows and **submit the output**.
   1. Produce a histogram of the distribution of birth weight. **(4 points)**

Chart, histogram

Description automatically generated

* 1. Produce a boxplot for birth weight grouped by categories of smokeFirst (use the 1/2 version). **(4 points)**

Chart, box and whisker chart

Description automatically generated

* 1. Produce a boxplot for birth weight grouped by categories of over25 (use the 1/2 version). **(4 points)**

Chart, diagram, box and whisker chart

Description automatically generated

1. Let’s look at the relationship between smoking while pregnant and low birth weight. For the analyses in this question and question 5, we’ll define smoking as any smoking during the first trimester (i.e. the smokeFirst variable), and low birth weight as <2500 grams (i.e. the lbw variable).
   1. Produce a 2x2 table of smokeFirst and lbw. **Submit the table**. **(5 points)**

vipcls$lbw

vipcls$smokeFirst. Yes No

Yes 392 3362

No 606 8447

* 1. Use the epi.2by2 function to estimate the crude RR for the association between smoking and low birth weight.What is the RR? **(5 points)**

The crude RR is 1.56.

* 1. Use the epi.2by2 function to estimate the crude OR for the association between smoking and low birth weight. What is the OR?  **(5 points)**

The crude OR is 1.63.

1. We’ll continue our analysis of the relationship between smoking (smokeFirst) and low birth weight (lbw) by investigating the potential role of maternal age, using the variable over25. We’re going to use both stratified analysis and regression to determine if maternal age is either a confounder or effect modifier:
   1. Use the epi.2by2 function to estimate the RR for the association between any smoking and low birth weight, adjusted for maternal age older than 25. What are the adjusted, crude, and stratum-specific RRs, and their 95% CIs? **(5 points)**
   2. Use the epi.2by2 function to estimate the OR for the association between any smoking and low birth weight, adjusted for maternal age older than 25. What are the adjusted, crude, and stratum-specific ORs, and their 95% CIs? **(5 points)**
   3. Based on the above analyses, would you consider maternal age a confounder? Why or why not? **(5 points)**
   4. Based on the above analyses, would you consider maternal age an effect modifier? Why or why not? **(5 points)**