MATH 3316: Project

Due Time/Date: Central Time (CT) 23:59 Aug 9, 2023

Please submit your project report online in .pdf format. Only one file is allowed to be submitted. Please do not send the report by email or hand in the report in person. Teamwork is strongly recommended, with the maximal size of each team as THREE. All project reports should be type-written. No hand-written version will be accepted.

For this project, we will use the North Carolina birth data set ncbirth1450.csv. In this assignment you will test hypotheses relating to mage, weeks, tounces, low, and smoke.

Variable Label	*Description*
plurality	Number of children born of the pregnancy
sex	Sex of child (1=Male, 2=Female)
mage	Age of mother (years)
weeks	Completed Weeks of Gestation (weeks)
marital	Marital status (1=married, 2=not married)
racemom	Race of Mother (0=Other Non-white, 1=White, 2=Black 3=American Indian, 4=Chinese, 5=Japanese, 6=Hawaiian, 7=Filipino, 8=Other Asian or Pacific Islander)
hispmom	Mother of Hispanic origin (C=Cuban, M=Mexican, N=Non-Hispanic, O=Other and Unknown Hispanic, P=Puerto Rican, S=Central/South American, U=Not Classifiable)
gained	Weight gained during pregnancy (pounds)
smoke	0=mother did not smoke during pregnancy 1=mother did smoke during pregnancy
drink	0=mother did not consume alcohol during pregnancy 1=mother did consume alcohol during pregnancy
tounces	Weight of child (ounces)
tgrams	Weight of child (grams)
low	0=infant was not low birth weight 1=infant was low birth weight
Premie	0=infant was not premature 1=infant was premature premature defined at 36 weeks or sooner

1. Please begin the assignment by deleting the items with missing values ("#NULL!") in it and do analysis only on the remained data set (hint: use the code provided in class). Please provide a frequency table for the percentage of low birth weights and a frequency table for the percentage of smokers. Create a summary table (mean, median, standard deviation, minimum, maximum) for the continuous variables of mage, weeks, and tounces.
2. With the information that you gather from this summary, test the following:
   (a) Determine if there is sufficient evidence to conclude the mean age of mothers giving birth in North Carolina is over 25 years of age at the 0.05 level of significance.
   (b) Determine if there is sufficient evidence to conclude the mean weeks of gestation of mothers giving birth in North Carolina is below 39 weeks.
   (c) Determine if there is sufficient evidence to conclude that the mean weight of babies born to mothers in North Carolina is above 7 lbs. (Note that there are 16 ounces in a pound.)

(d) Construct a side-by-side boxplot for tounces for smokers and non-smokers. Comment on whether you believe you will reject or fail to reject the null hypothesis. Determine if there is sufficient evidence to conclude the mean tounces of smoking mothers is lower than the mean birth weight for non-smoking mothers.

(e) Determine if there is sufficient evidence to conclude the percentage of low birth weight children in North Carolina is above 6%.

(f) Determine if there is sufficient evidence to conclude the percentage of mothers who smoke in North Carolina is above 10%.

(g) Determine if there is sufficient evidence to conclude the percentage of low birth weight for smoking mothers is **different** than the percentage of low birth weight for non-smoking mothers.

(h) Determine if there is sufficient evidence to conclude the percentage of low birth weight for smoking mothers is **lower** than the percentage of low birth weight for non-smoking mothers.

(i) Determine if there is sufficient evidence to conclude the percentage of low birth weight for mothers who did not consume alcohol during pregnancy is **lower** than the percentage of low birth weight for mothers who did consume alcohol during pregnancy.

(j) Determine if there is sufficient evidence to conclude the percentage of low birth weight has any relation with the races of mothers?

For each of the tests above, in your report, be sure to

1. Clearly state a null and alternative hypothesis
2. Check the necessary assumptions.
3. Give the value of the test statistic
4. Report the P-value
5. Clearly state your conclusion (i.e. Reject the Null is not sufficient)