Qns 1)

The bwght is not a normal random variable as the variable is not symmetric since the skewness is not 0 but rather equal to -0.14602351629957394 and the kurtosis level is not equal to 3 but rather 3.1633425253063665.

Qns 2)

There is a negative relationship between the 2 variables as the coefficient is -0.5137720928233965 which is less than zero, making it a negative relationship.

Qns 3)

H0: No significant difference between means mothers who don’t smoke is equal to mothers who smoke

H1: Got significant difference mothers who don’t smoke is not equal to mothers who smoke

Two-tailed t-test with α = 0.05

The mean birth weight for mothers who smoke is 111.14622641509433 while the mean birth weight for mothers who don’t smoke is 120.06122448979592. This means that non-smoking mothers have heavier babies by 8.914998074701586 meaning that both groups are statistically different. We can reject the null hypothesis as the p-value is 3.362457867694957e-09 which is < α = 0.05 meaning that the model is statistically significant.

Qns 4)

Smoking has a negative effect on infant health, which is to say the more a mother smokes, the less healthy the baby, as evidenced by the negative association between smoking and infant weight, where a higher number of cigarettes consumed leads to a lower infant weight, where 1 unit of cigarettes smoked leads to a 0.5137720928233965 decrease in infant weight. This is significant at the 5% level as the p-value is 1.661538170263396e-08 which is less than α= 0.05.

Qns 5)

The r-squared amount is 0.022729121106052963 which is less than the accepted 0.30, meaning to say that the model only explains 2.7% of the variation of infant weight over the number of cigarettes smoked. However, attention should be devoted to the issue of smoking and infant weight, as the model can be improved by adding more control variables, which are variables that affect the number of cigarettes smoked which are not included in the model, resulting in a more a higher r-square amount. Thus, the regression model is still useful.

Furthermore, even if the R-squared value is low, there are still statistically significant predictors as the p-value is significant at 5%, meaning that you can still conclude that there is a negative association between smoking and infant health.

Qns 6)

A variable qualifies as an omitted variable if it’s correlated with the number of cigarettes smoked and if it influences the dependent variable which is the weight of infants. Some potential omitted variables could be the father's education level and family income. Fathers with fewer years of education would smoke more as a lower education leads to less knowledge on the harmful effects of smoking and thus would affect the infant’s health as they would smoke more around the infant, impacting their health and weight. Furthermore, if family income is low, they might be more drawn to cigarettes as a vice resulting in more smoking and thus a lower infant weight.

Qns 7)

The new estimated effect of smoking is still negatively associated with infant health, which is to say that a higher number of cigarettes smoked leads to a lower infant weight and thus worse infant health, where 1 cigarette smoked leads to a decrease of 0.5908699310797849 pounds in infant weight. The coefficient of “male” is 3.716468348076219 while “white” is 4.519182849586225. This means that both coefficients are more than 0 and thus are positive, showing that there is a positive association between male and white child over infant weight, where white male infants are shown to be healthier as they are heavier.

Qns 8)

H0: faminc, fathedu, and motheduc probability of the f-statistic does not affect infant health

where probability of f-statistic is > α = 0.05

H1: faminc, fathedu, and motheduc probability of the f-statistic influences infant health where the probability of f-statistic is <= α = 0.05

By doing a regression of bwght = B0 + B1(faminc) + B2(fatheduc) + B3(motheduc) + ui

We find that we can reject the H0 null hypothesis at the significance level of α=0.05 as the probability of the f-statistic is 0.00936777471269788 which is < 0.05, meaning that the statistically significant

Qns 9)

There is a decreasing effect in the regression as the coefficient of (faminc \* cigs) is 0.00848563846260081 which is more than 0, indicating that given a constant amount of cigarette smoked, there will be an increase in infant weight and thus health given an increase in family income.

H0: the effect that the effects of smoking on infant health should be decreasing with the family income where the p-value is <= α = 0.05

H1: the effect that the effects of smoking on infant health are not decreasing with the family income where the p-value is > α = 0.05

By doing a regression of bwght = B0 + B1(cigs) + B2(faminc) + B3(fatheduc)+ B4(motheduc) + B5(faminc \* cigs) + ui

We find that we cannot reject H0 null hypothesis at the significance level of α=0.05 as the p-value of (faminc \* cigs) is 0.29345874159547103 which is > 0.05, meaning that this finding is not statistically significant.

Qns 10)

By running a regression of bwght = B0 + B1(cigs) + B2(faminc) + B3(fatheduc) + B4(motheduc) + B5(male \* cigs) + ui, we find that the smoking effect on infant weight for females

is -0.5674173101639224 while that of males is -0.5875196792947183, meaning that the effect for males is larger since -0.5875196792947183 < -0.5674173101639224.

The result is not statistically significant at the 5% level where α = 0.05 as the p-value of (male \* cigs) is equal to 0.9255457128676235 which is more than 0.05, thus meaning that the regression is not statistically significant, and we cannot reject the null hypothesis that the effect of smoking on birth weight is not stronger for infant boys.