**Econometrics Data Assignment 1**

Name: Utkarsh Arora

Roll No: 2020143

Q1)

First, we created datasets for gdp, beds and tap, using data from the internet from the links that were provided. Then we have added this data to the appended columns in the larger dataset.

We have performed data-cleaning as well. Our analysis in this assignment depends on the dependant variables v40, v42, v43, v44, v45, v46; and the explanatory variables index, gdp, beds, tap. For some states/districts, values for gdp and tap is not given (Eg Ladakh). For the years 2017-2019, we don’t have data for the dependant variables. In any of the rows where data for any of the 10 variables (dependant and independent) is missing, we have removed that row. This is to make sure that our vectors when conducting analysis do not have NULL values and are of equal sizes.

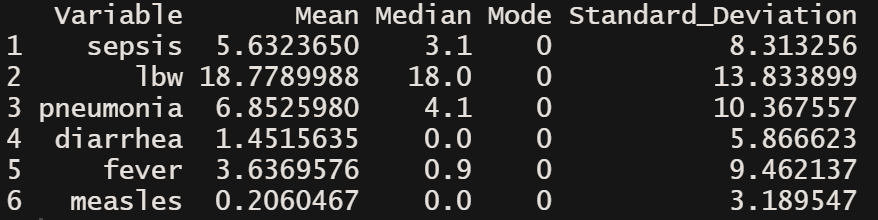
In other variables where we have missing data, we have put NA.

Input dataset: main.csv

Appended and cleaned dataset: main9.csv

Q2)

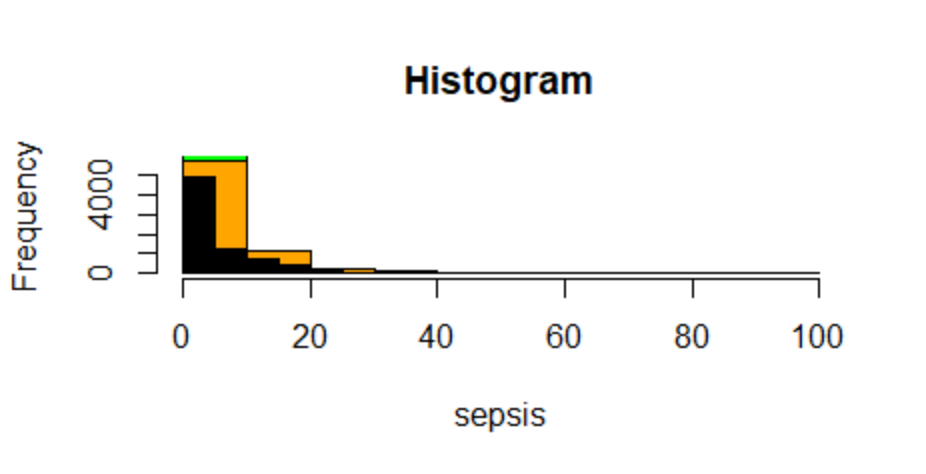
1. File: q2a.r



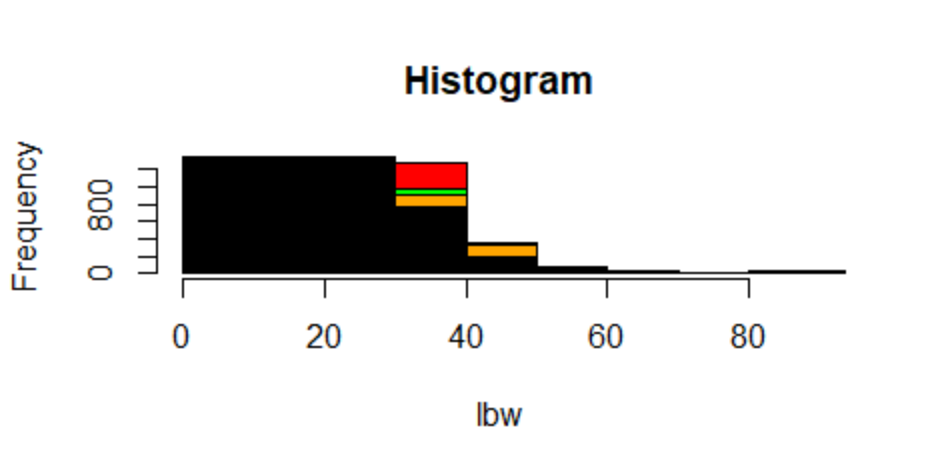
b)

**YEAR-**

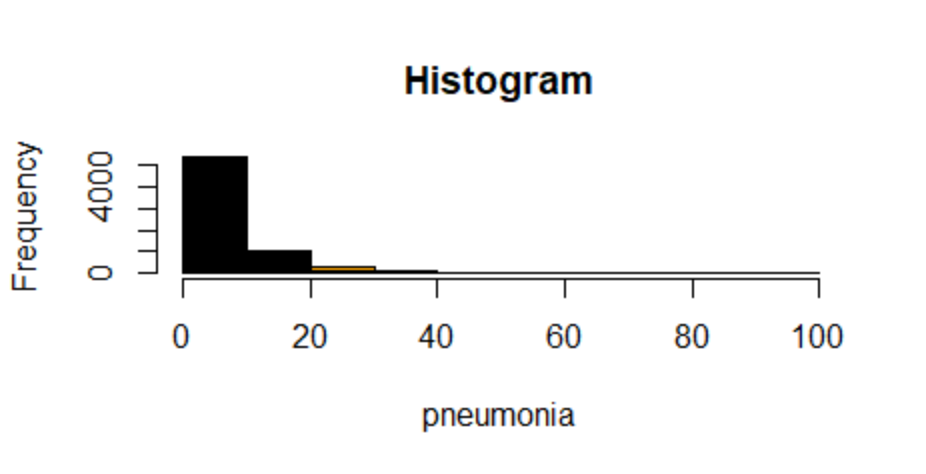
Sepsis



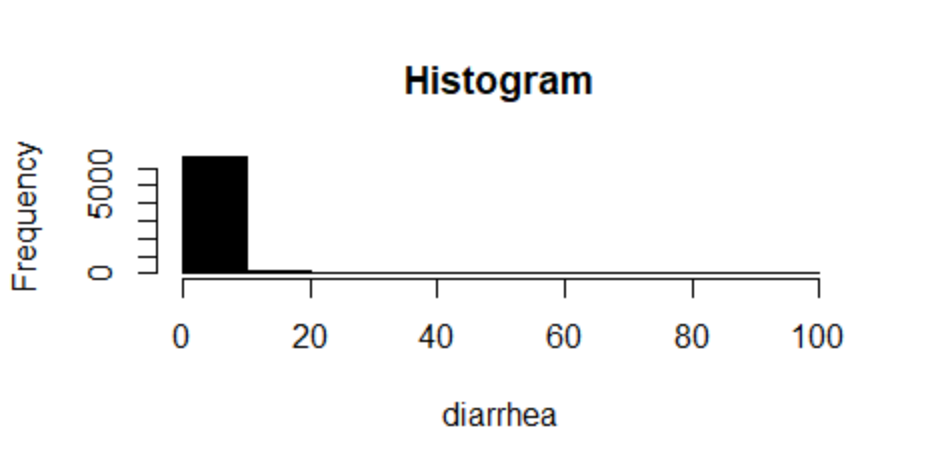
Lbw



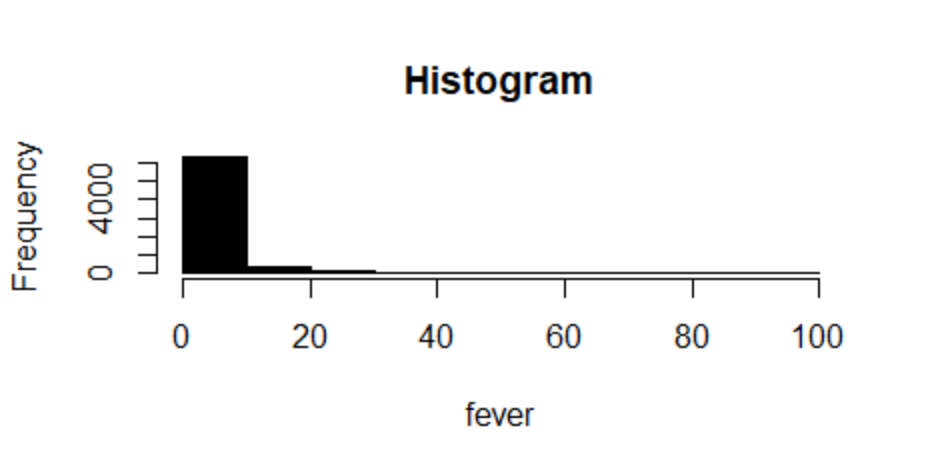
Pneumonia



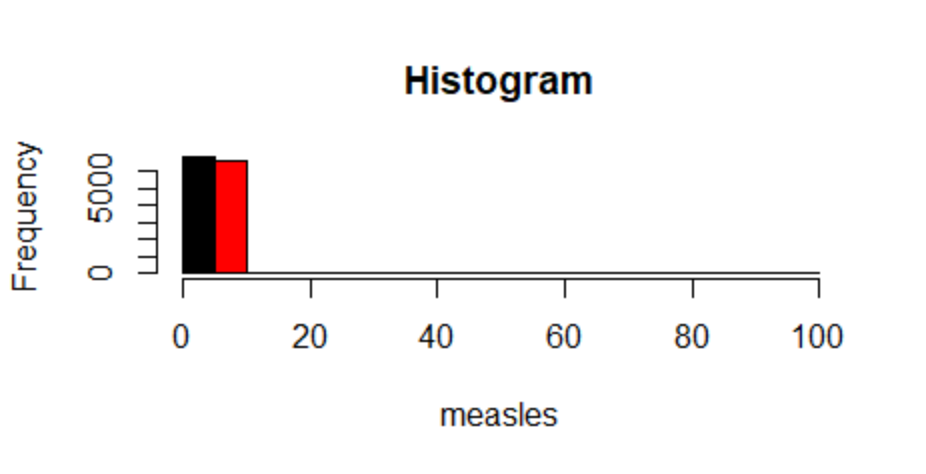
Diarrhoea



Fever

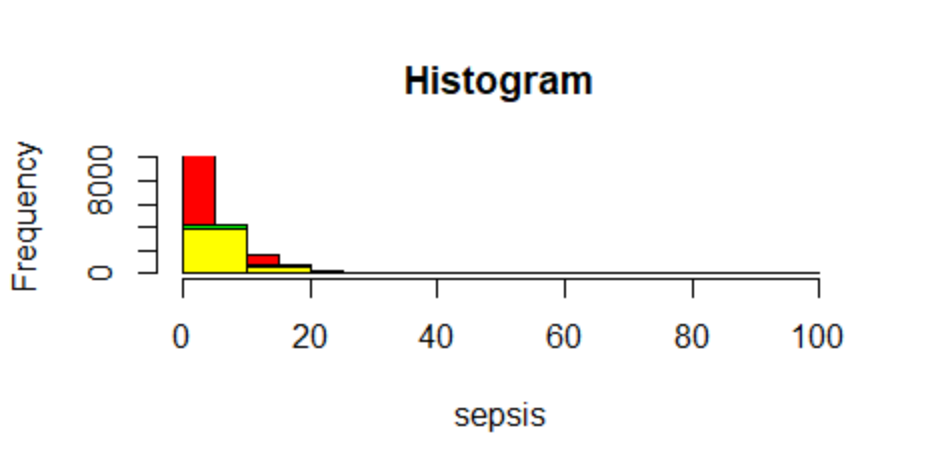


Measles

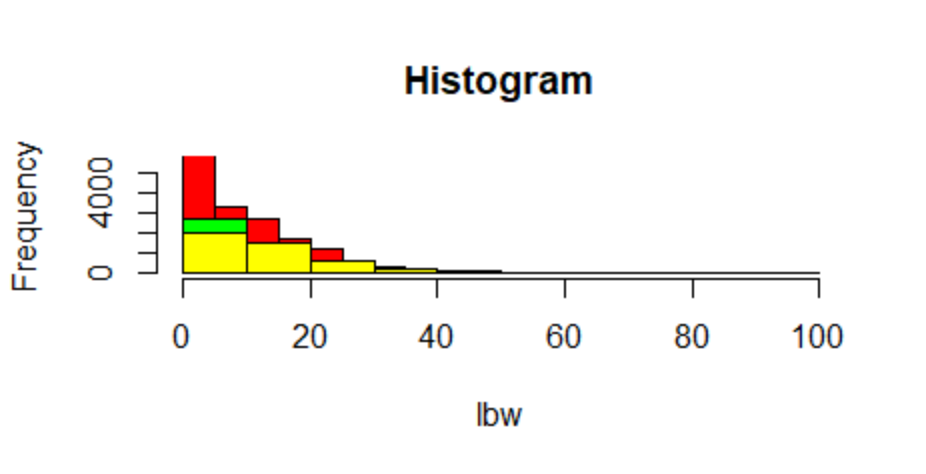


**Season--**

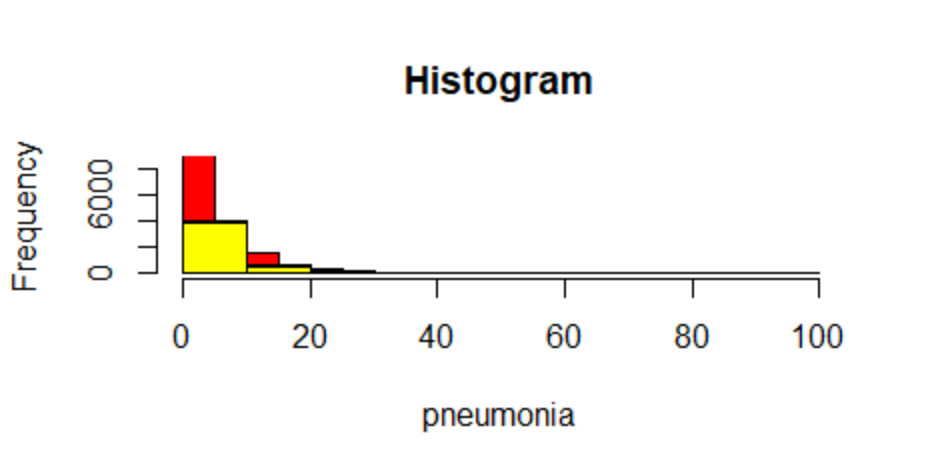
Sepsis



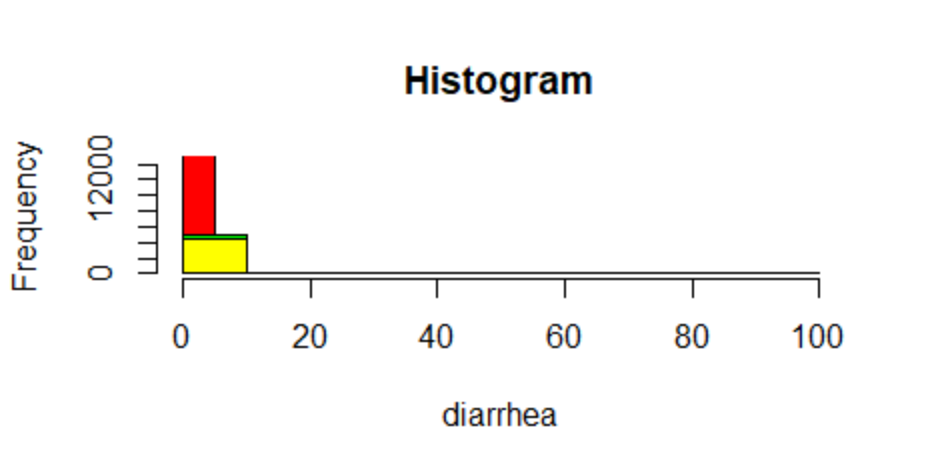
Lbw



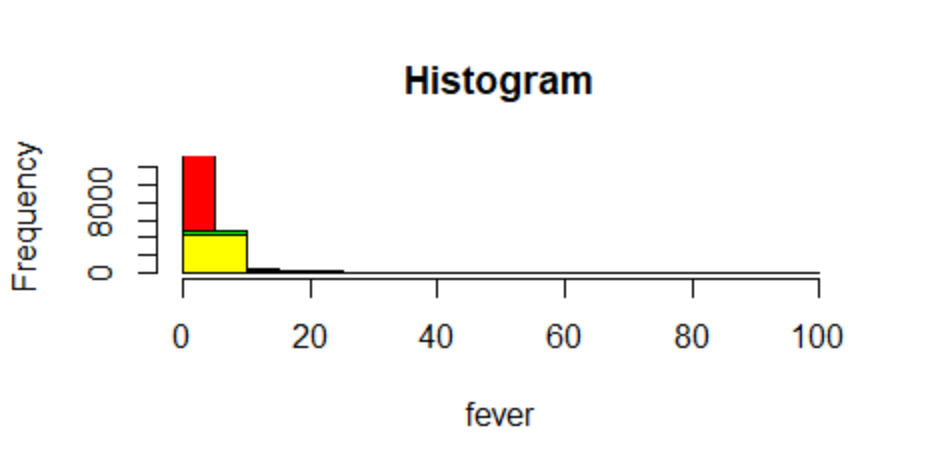
Pneumonia



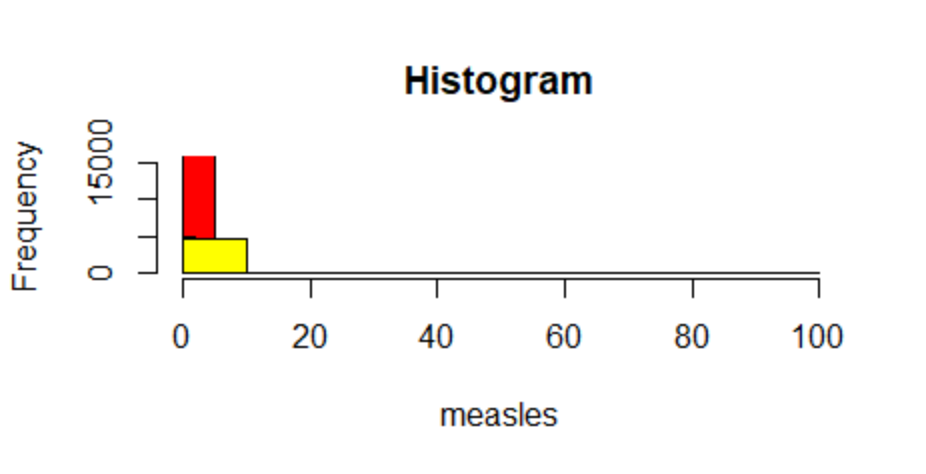
Diahhroea



Fever



Measles

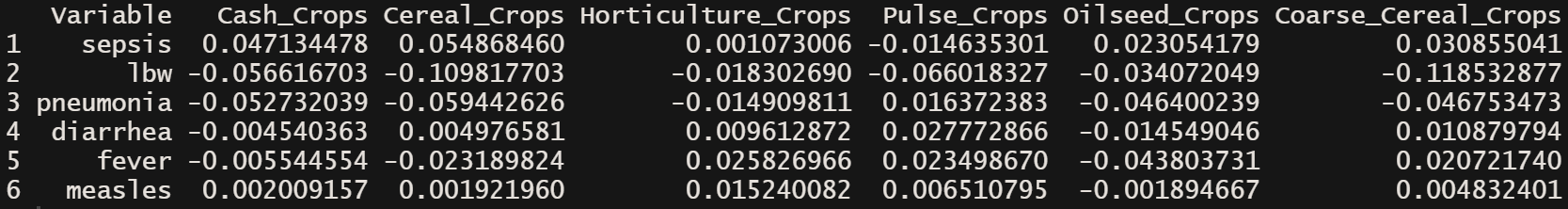


d)

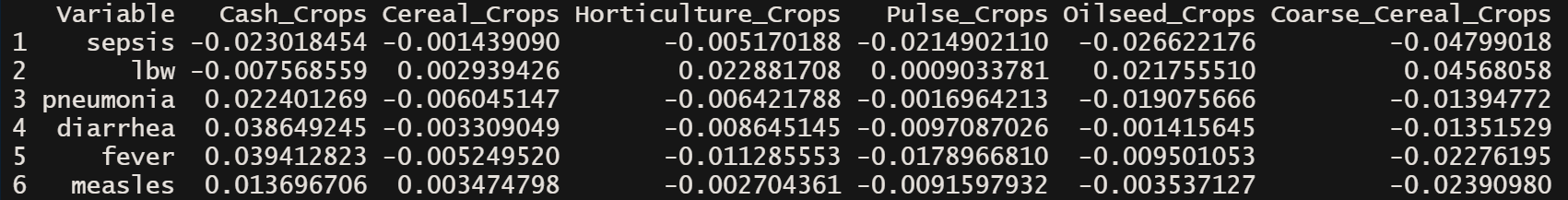
1) File: q2d1.r



2) File: q2d2.r

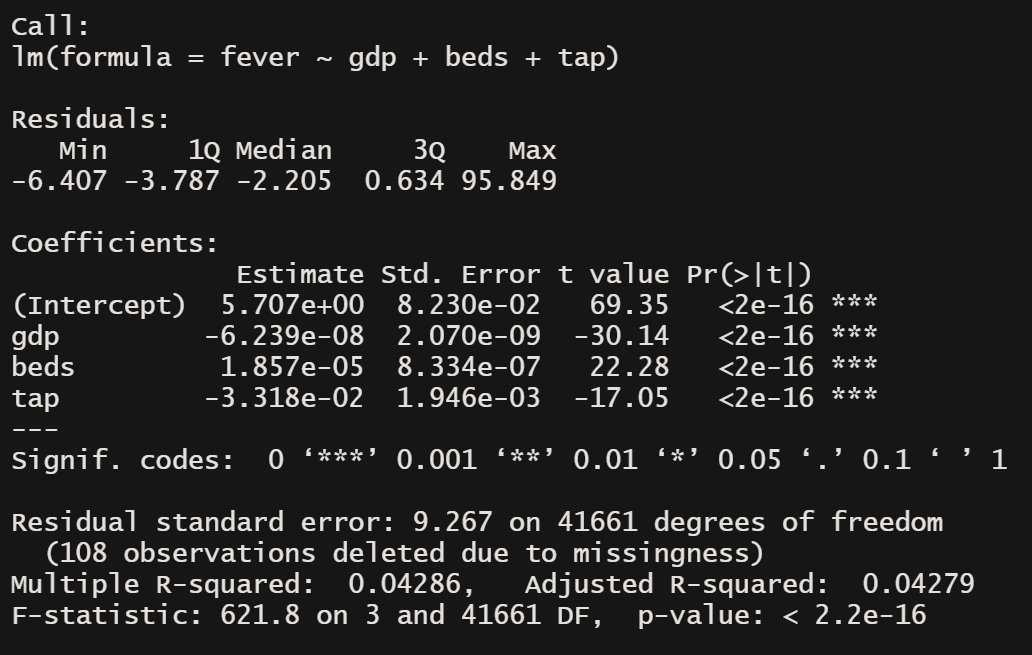


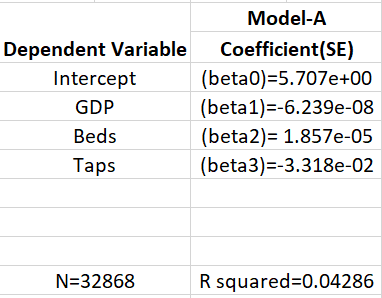
3) File: q2d3.r



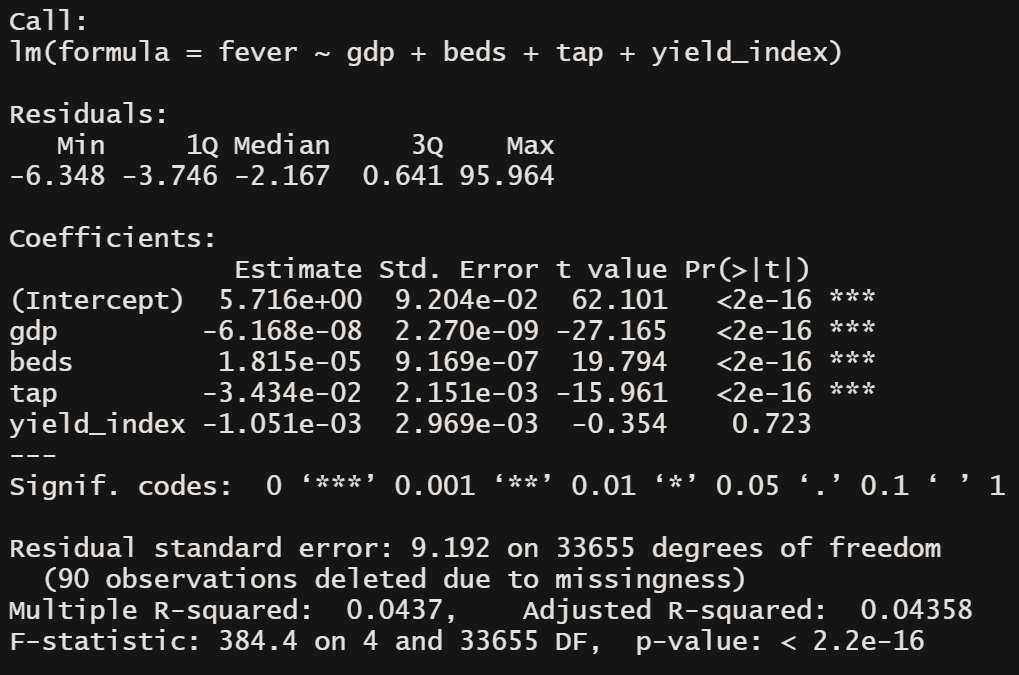
Q3)

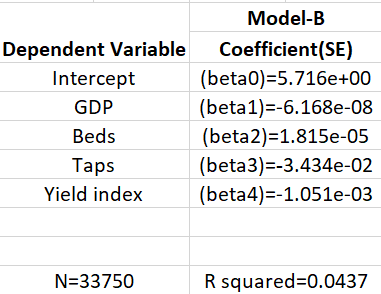
1. File: q3a.r



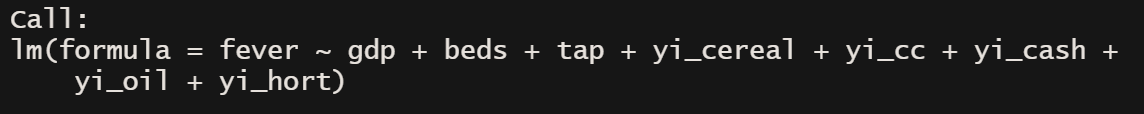


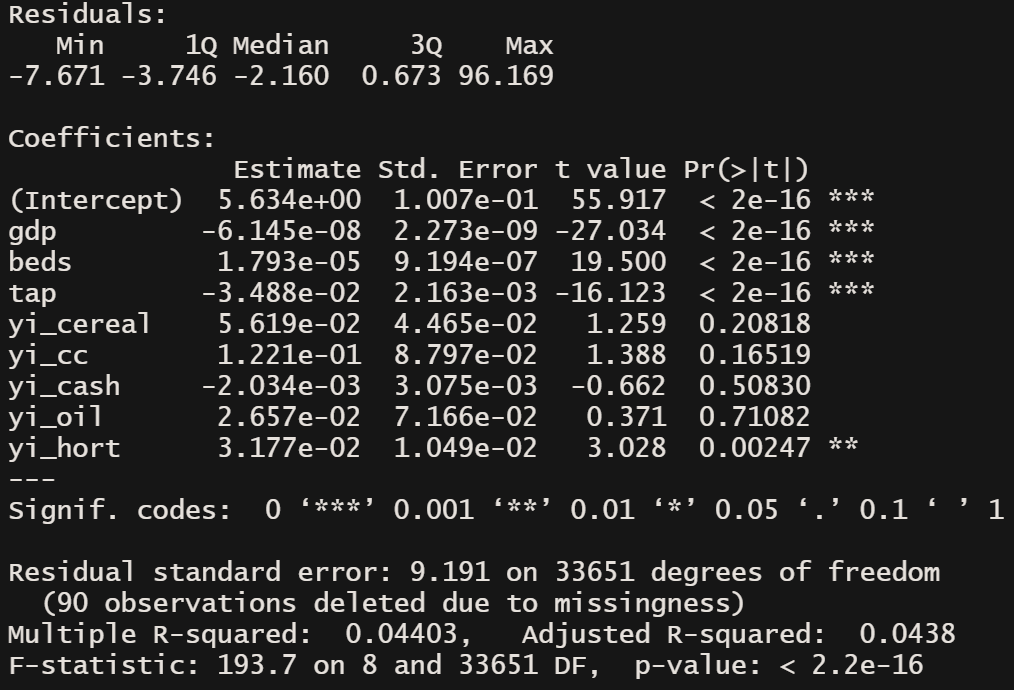
1. File: q3b.r

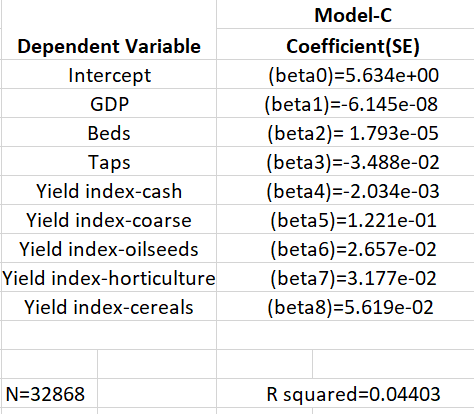




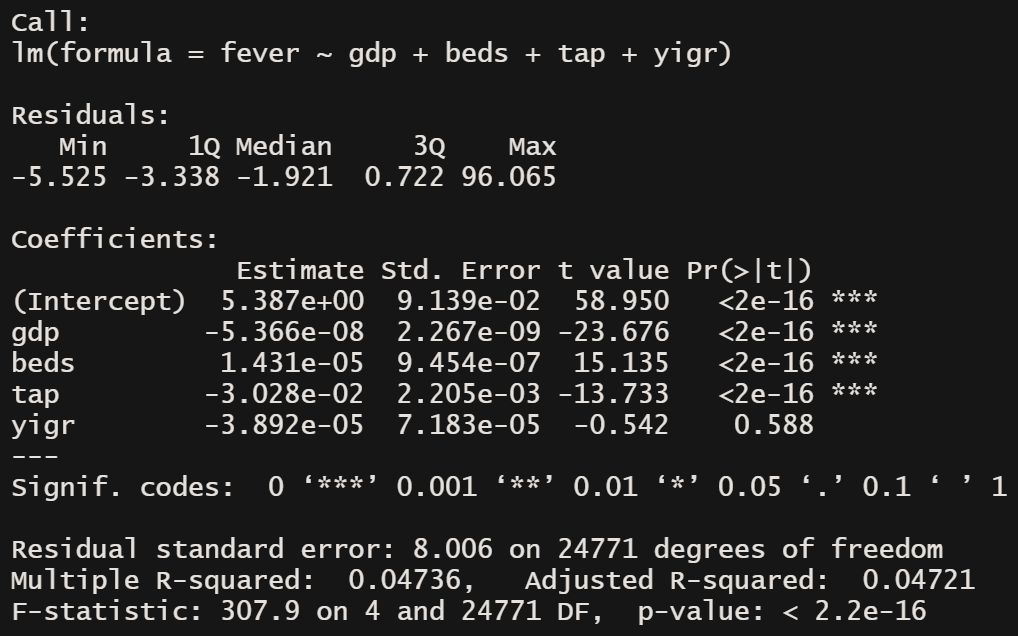
1. File: q3c.r

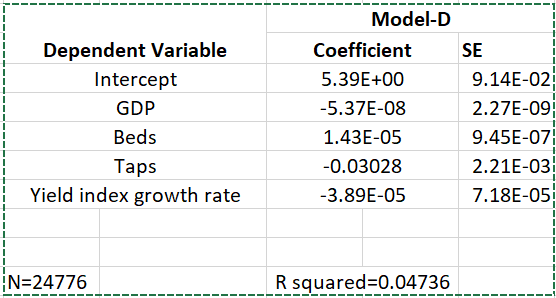




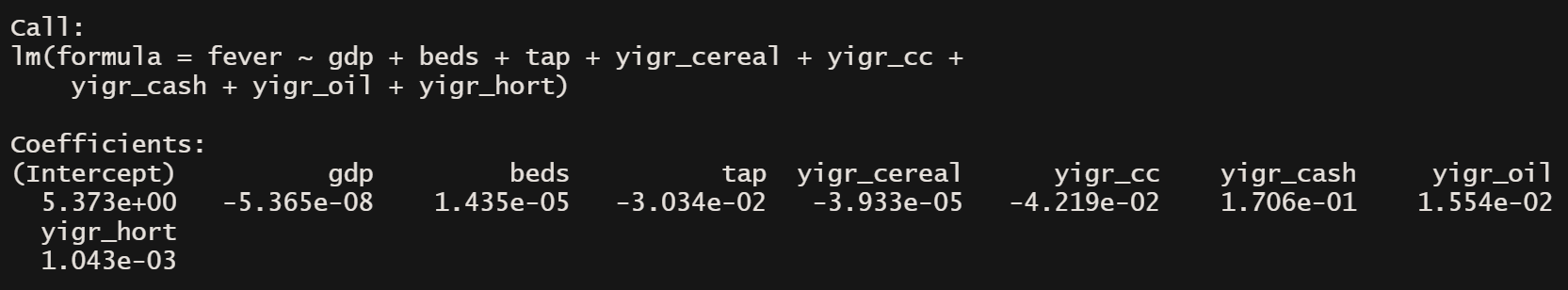


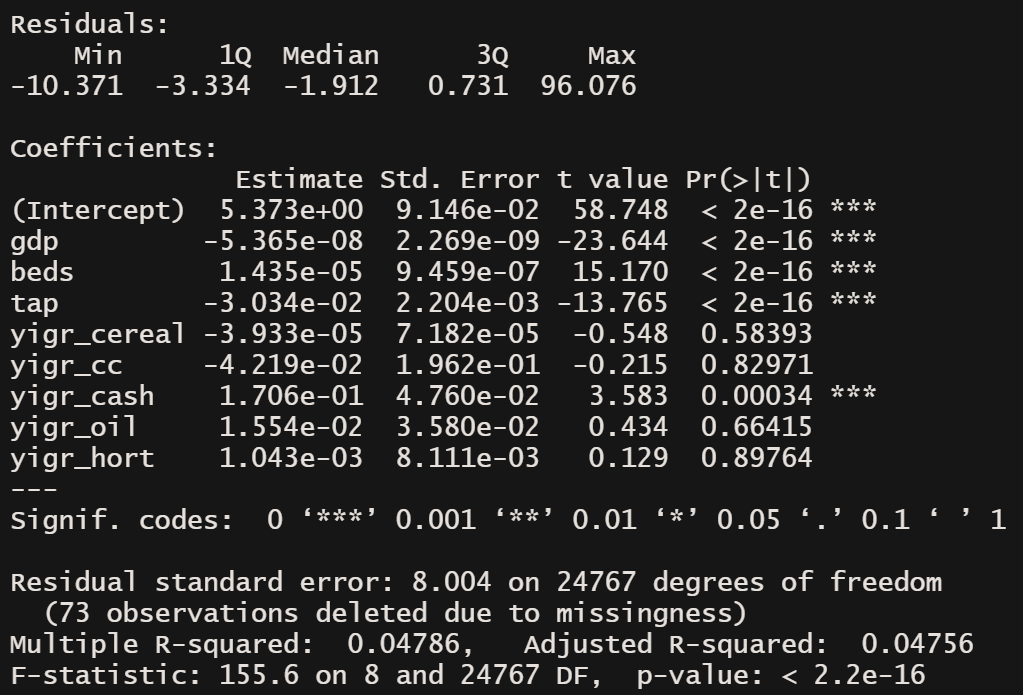
1. File: q3d.r

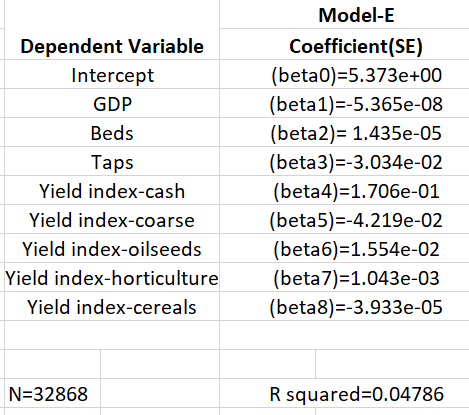




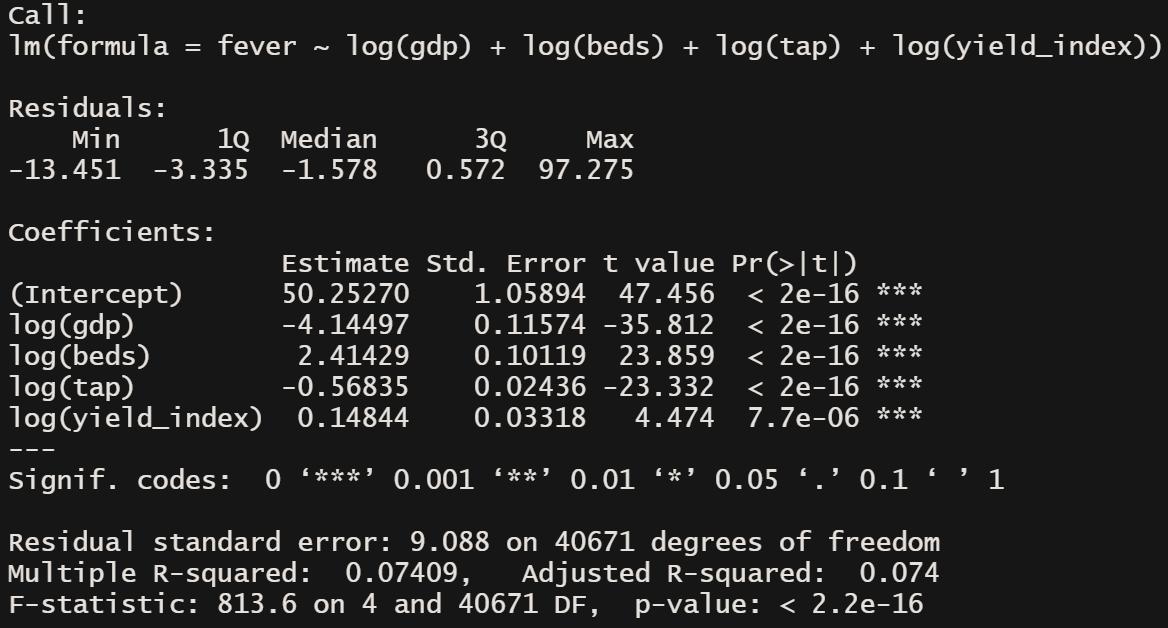
1. File: q3e.r

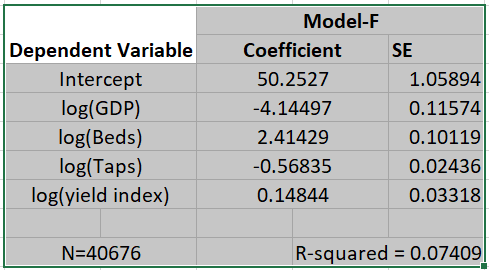




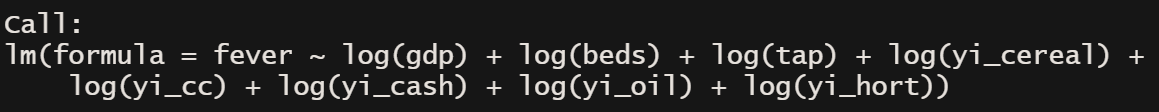


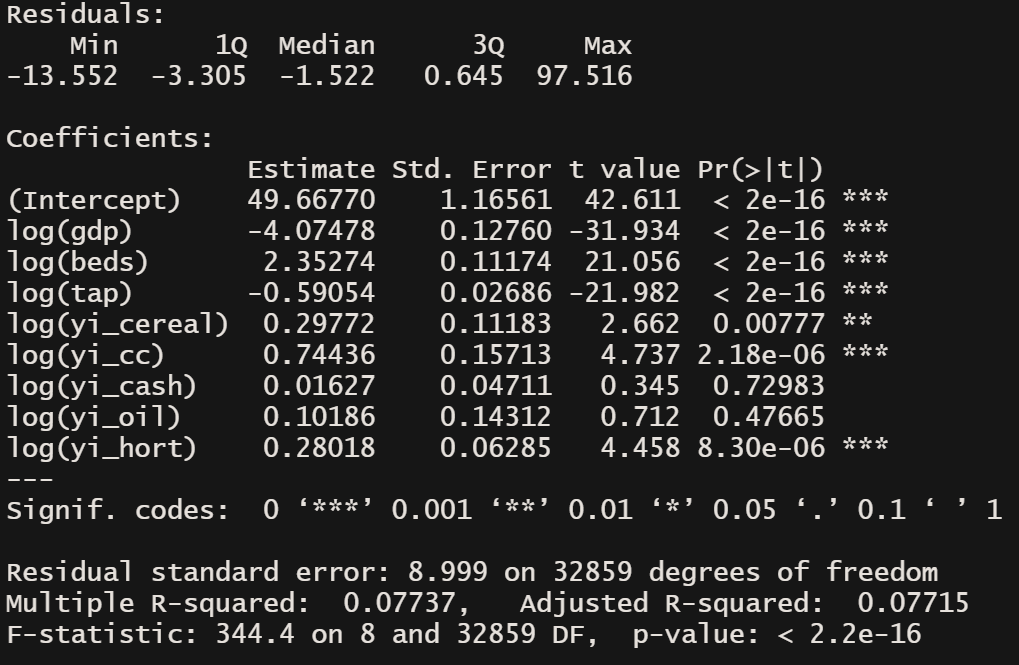
1. File: q3f.r

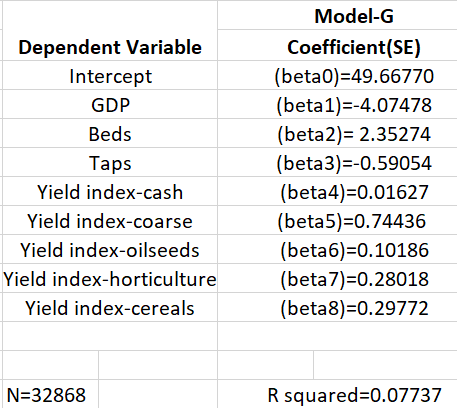




1. File: q3g.r







Q4) After running multiple analysis on the coefficient of correlation and R-Squared, we can conclude that in the given data, the theoretical relation between the coefficient of correlation and the goodness of fit does infact hold.

For a smaller example to display it better, correlation coefficient between fever and gdp is -0.1313496, and Goodness of Fit when fever is regressed on gdp is 0.01725; which is exactly the square of the correlation.

Q5)

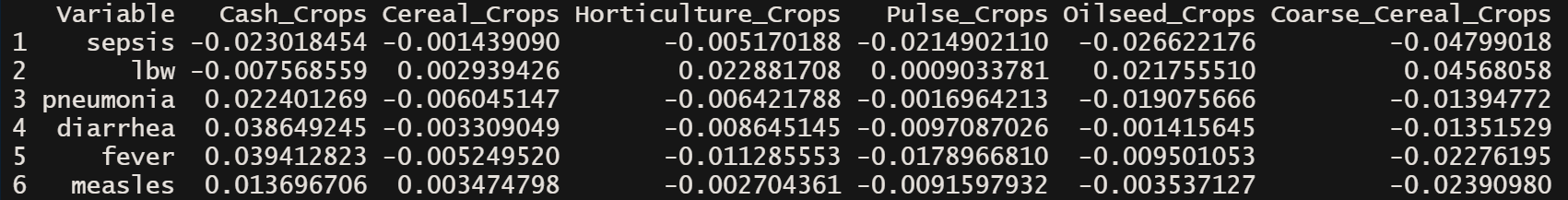
In the regression models we have analysed in which we take yield index of different crop categories separately, in many of them we can notice the fact:

If we keep all other explanatory variables constant, and just move yield index of cash crops, this will give an effect in the opposite direction, than if we kept all other explanatory variables constant and just moved yield index of coarse cereals.

Hence yield indexes of some of different crop categories have an opposing effect on the health indicator. This nuance is cancelled out and missing when we include the yield index for all six crop categories together. Thus we would be missing precision in exchange for generality.

Q6)

We obtained coefficients of correlation between yield growth and health indicators across crop categories in Q2D3



Sepsis has a negative correlation with yield growth across all crop categories.

lbw has a negative correlation with yield growth rate of cash crops, but a positive correlation with yield growth rates across other crop categories.

pneumonia has a positive correlation with yield growth rate of cash, but a negative correlation with yield growth rates across other crop categories.

Looking at this chart, we can conclude that the relation between yield growth and health indicators is not similar across crop categories.