**A) Use these sources, and/or any other to Estimate the Price Elasticity of Demand for Cigarette, Elasticity of Tax (all combined) rate on Cigarette. Test the Hypothesis that Tax rate has no bearing/impact on Cigarette consumption. What can you conclude about overall Cigarette consumption in the Country?**

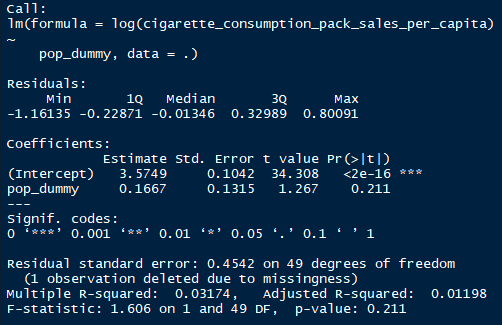
The price elasticity of demand for cigarettes is -1.31. The elasticity of combined taxes is -0.10. This suggests demand for cigarettes is elastic but that per-pack tax on cigarettes is not, meaning increasing the taxes on a pack of cigarettes will do little to deter smoking, if that is the intended outcome. This makes sense if we consider the features of addiction.

To test the hypothesis that the tax rate has no impact on cigarette consumption, we compared the t value of the coefficient of the total tax per pack to the corresponding t value for an alpha significance level of 0.05 with 48 degrees of freedom (51 observations minus 2 independent variables in the model minus 1 constant. Because the absolute value of the coefficient’s t value (-0.550) is less than the corresponding significant t value (1.684), we conclude that the effect of tax rates has no bearing no cigarette consumption.

**B) DIVIDE THE STATES INTO TWO CATEGORIES LESS THAN 6 MILLION IN POPULATION AND MORE THAN 10 MILLION. RUN A DUMMY VARIABLE MODEL TO TEST THE HYPOTHESIS THAT MORE POPULATED STATES HAVE A HIGHER PER-CAPITA CONSUMPTION OF CIGARETTES; AND**

It does not appear from the data that more populous states have higher per-capita cigarette consumption. We conducted a t-test to determine whether there were significant differences in the number of packs of cigarettes consumed between states with less than 6 million residents and those with more than 10 million. The average number of packs of cigarettes consumed per capita was 45.94 in states with fewer than 6 million residents, compared with 36.41 packs per person for states with more than 10 million residents. The t-test yielded a *t*-statistic of 1.6975 with 17.624 degrees of freedom. Because this *t*-statistic is less than the corresponding *t.95* value of 1.740, we fail to reject the null hypothesis that the two groups consume cigarettes at different rates.

Additionally, the linear model in which we regressed per-capita cigarette consumption on the population dummy variable we created returned the following results.



As you can see from the model output, the model is not significant because the F(1,49) value of 1.606 is less than the F(1,50) value of 4.0847 that corresponds to α = 0.05. The adjusted R2 of the model is very low; it explains roughly three percent of the variation in the data. Furthermore, the coefficient of the population dummy variable (0.1667) has a t-statistic of 1.267. Because this tcalculated is less than the ttable value of 1.684, we accept the null hypothesis that there is no significant difference in cigarette consumption between states with populations of at least 6 million and those with less than 6 million.

**C) Check to see if there is any problems of Multiple linear Regression that may be present here. If THERE IS CORRECT FOR IT**

We tested for multicollinearity by correlating the two independent variables of our initial model (listed below).

ŷ = 6.2084 + -1.3145 \* log(average\_cost\_per\_pack) + -0.1014 \* log(total\_state\_tax\_per\_pack).

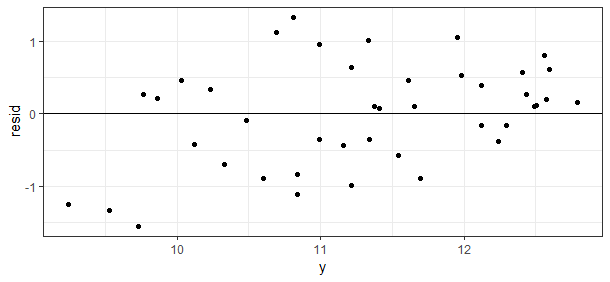
We found a correlation coefficient of 0.91 between the IVs of logged average cost per pack and the log of the total state tax per pack. This is not surprising when we consider that total taxes per pack usually represent a percentage of the total per-pack cost.

Because the estimated coefficient for the log of the total state tax per pack was not statistically significant (t = -0.550, p > 0.5), we removed this predictor from the model. Our revised model is below.

ŷ = 6.6892 - 1.6010 \* log(average\_cost\_per\_pack)

This model is statistically significant and explains about 50 percent of the variance in cigarette consumption (F1, 49 = 48.85, p < 0.001, R2 = 0.4992).

We further tested to see whether the residuals were normally distributed with a mean of 0 and variance of 1. The residual plot below indicates the residuals are evenly distributed around the mean of 0.



We also employed the RESET test to determine whether there were omitted variables in the model. We compared the F-statistic obtained from the squared and cubic values of the regressors (F2, 47 = 1.3839) to the corresponding F-statistic table value (F1, 60 = 4.0012). Because the calculated F-statistic is less than the corresponding table value, we conclude that variables were not omitted from the model (p > 0.25).