The study of Effect of Higher Excise Tax on Smoking Behavior of Various Age Groups

GAURAV SHAHANE
KUSHAL THAKKAR
INST627, MIM, FALL 2015

Contents

1. Introduction

As per the Centers of Disease Control and Prevention, Smoking remains the primary cause of preventable deaths in the US, Killing an estimate of over 480,000 citizens each year. Government has taken up several initiatives to curb death and illnesses caused due chronic use of cigarettes. To name a few: There is a free phone based service – 1-800-QUIT-NOW consisting of mentorship, help with local resources to quit the use of tobacco. Also there are dedicated web pages offering cigarette quitting tips on www.cdc.gov which is an official site of Centers of Disease Control and Prevention and also on www.cancer.org , official site American Cancer Society. These measures cater to smokers who are proactive about the fact that they want to quit smoking. However, to target a bigger population of cigarette smokers' state governments usually use means of increasing excise tax on per pack of cigarette. Basis of their argument is making cigarette expensive will reduce its consumption and this is how they can address the entire issue of illnesses and death caused due to cigarette smoking.

Our rationale behind picking this problem was this: Let's revisit the CDC data of more than 480,000 deaths caused due to smoking. By any standards that is a very big number and this is true for each year. Our study focuses on whether governments' changes in state excise tax on cigarette are really effective on the ground level? Are the right things done and are they done to right extent to save some or more of these 'preventable' deaths. Through this project we look to study whether increasing taxes will have any effect on the smoking population? And will it affect each age group equally? To uncover the answers to these questions, we decided to take verify the relationship between taxation and behavior of smoking population belonging to various age groups in all the states of USA.

2. Research Design

2.1 Data collection

Our primary source of information is the website of Centers for Disease Control and Prevention (www.cdc.gov). CDC has a huge repository of data repository pertaining to various health related conditions. They have a separate section for Smoking and Tobacco. We will use the behavioral risk factor data related to tobacco use from the year 1996 to 2014 for our analysis. We will also use data sets provided by CDC on State Excise Tax on cigarettes over the years.

2.2 Datasets

2.2.1 Behavioral Risk Factor Data on Tobacco Use Survey

This dataset contains record of surveys conducted every year for determining the tobacco usage including cigarette smoking in United States. The data is collected from every state. Screenshot of sample data used for our project below:

YEAR	LocationAl	LocationD	TopicDesc	Data_	Data_Value	Sample_Size	Gender	Race	Age
2013	AL	Alabama	Cigarette Use (Adults)	%	21.5	6383	Overall	All Races	All Ages
2013	AL	Alabama	Cigarette Use (Adults)	%	18.2	4222	Female	All Races	All Ages
2013	AL	Alabama	Cigarette Use (Adults)	%	24	2567	Overall	All Races	45 to 64 Years
2013	AL	Alabama	Cigarette Use (Adults)	%	25.1	2161	Male	All Races	All Ages
2013	AL	Alabama	Cigarette Use (Adults)	%	28.2	1168	Overall	All Races	25 to 44 Years
2013	AL	Alabama	Cigarette Use (Adults)	%	8.9	2370	Overall	All Races	65 Years and Older
2013	AL	Alabama	Cigarette Use (Adults)	%	17.1	278	Overall	All Races	18 to 24 Years
2013	AK	Alaska	Cigarette Use (Adults)	%	11.7	844	Overall	All Races	65 Years and Older
2013	AK	Alaska	Cigarette Use (Adults)	%	20.9	345	Overall	All Races	18 to 24 Years
2013	AK	Alaska	Cigarette Use (Adults)	%	22.2	2395	Female	All Races	All Ages
2013	AK	Alaska	Cigarette Use (Adults)	%	22.6	4476	Overall	All Races	All Ages
2013	AK	Alaska	Cigarette Use (Adults)	%	22.9	2081	Male	All Races	All Ages
2013	AK	Alaska	Cigarette Use (Adults)	%	23.9	1938	Overall	All Races	45 to 64 Years
2013	AK	Alaska	Cigarette Use (Adults)	%	25.6	1349	Overall	All Races	25 to 44 Years

The dataset includes variables such as year, location, percentage of population smoking cigarette, age groups. These can be used for analysis of the data.

2.2.2 Excise tax on Combustible Tobacco Products

The below dataset records excise tax for cigarette per pack in dollars for all the states from 1996. We combined this data set with Behavioral Risk Factor Data on Tobacco Use dataset and are using it to measure effect of these taxes on % of smoking population in every state in United States.

The data is published by Centre for Disease Control & Prevention and the same is available on their web site: http://www.cdc.gov/

Year	Location/	LocationD(	TopicDesc	MeasureDesc	ProvisionDesc	ProvisionValue
2018	OR	Oregon	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	1.33
2016	OR	Oregon	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	1.32
2015	KS	Kansas	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	1.29
2015	LA	Louisiana	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	0.86
2015	NV	Nevada	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	1.8
2015	OH	Ohio	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	1.6
2015	RI	Rhode Isla(	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	3.75
2015	VT	Vermont	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	3.08
2015	AL	Alabama	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	0.425
2015	AK	Alaska	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	2
2015	AZ	Arizona	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	2
2015	AR	Arkansas	Legislation - Tax Combustible Tobacco	Cigarette	Cigarette Tax ($ per pack)	1.15

2.3 Population

The population for our study will be state wise yearly records on smoking population and state excise tax. Thus our level of analysis will be 'State'.

2.4 Type of Study

The study is an observational study as the population and the sample were not controlled for the research purpose. The independent variables are result of nationwide state level surveys conducted by the governmental agencies.

2.5 Research Questions

a. **Is there a relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state?**

We do the above analysis based on the aggregated data for all the states (includes yearly state excise tax, state wise yearly smoking population percentage and so on.) from the years 1996 to 2014.Here, we want to study the relationship between state excise taxes on cigarette and its effect on percentage of cigarette smokers over the last 15 years for each state.

Government justifies taxes on cigarettes also as a means of making cigarettes slightly (or more) unaffordable to the people and hopes that this will cause reduced number of smoking population. We will try to find out whether there is any positive or negative correlation between the excise taxes per pack of cigarette and percent of smoking population. Since we have the data for state excise rates on cigarettes over the years and nationwide survey on behavior risk factor data of smoking, we can identify the appropriate variables and use them to determine the relation between the two. The research outcome can help policy makers understand and act in accordance with the impact of taxes on cigarette consumption.

b. **If there is a relationship between the state excise tax and percent of smoking population, is the relation same for all age groups?**

We also plan to measure the degree of relationship certain increase in state excise tax to smoking population belonging to a certain age groups. Are some age groups more responsive to the change in tax over other? Age groups under consideration would be: 18-24, 25-44, 45-64 and 65-older. This will allow us to deep dive into what level of reach these tax regulations have on each age group.

The motivation for this research question is that, we believe the median income of the age groups 18-24, 25-44, 45-64 and 65-older will be different and burdening cigarettes with more tax is likely to curb smokers belonging to rather lower income group more than smokers belonging to a relatively higher income group.

2.6 Sampling Methods

The samples for the behavioral Risk Factor Data on Tobacco use survey have been gathered by private companies in all the states of US. The surveys are conducted as per the BRFSS guidelines, on phone calls, ensuring respondents of their confidentiality so that they can respond correctly. The samples are collected within the same month's timeframe across all states.

They have followed a random sampling methodology where telephone and cellular mobile numbers have been picked up for survey purpose. The samples are representative of the population as every person in the population had the equal chance of getting chosen. The sampling of telephone or cellular phones was done based on geography and high-medium density blocks of that area within each state. There was no bias as such when it came to sample from such a huge population.

2.7 Level of Analysis

Since the observations from the surveys of smoking behavior are aggregated over state level and we are comparing those to the state level excise tax. Our level of analysis will be states. Any conclusions we will draw will be directed from and towards overall 'State'.

2.8 Variables

a. **Research Question 1: Is there a relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state?**

- Independent variables:

Excise Tax on a Pack of Cigarette ($ per pack)

The excise tax on a pack of cigarette is a value in dollars per pack. This variable is a ratio type of variable. Since this is a monetary value it can be plotted from continuously on an axis on equal intervals and also has a clear definition of zero, which here means no tax on cigarette.

- Dependent Variable:

% Percentage of smoking population

The variable is percentage of people from sample population who smoke currently. This is a ratio type of variable.

Fig. Independent and Dependent Variables

b. Research Question 2: Does state excise tax value on a pack of cigarette impact % of smoking population belonging to all age groups equally?

- Independent variables:

1. Excise Tax on Pack of Cigarette

The excise tax on a pack of cigarette is a value in dollars per pack. For the nature of the research question we convert this ratio variable into ordinal variable. The variable will hold values “Very Low”, ”Low”, ”Medium”, ”High”. The categorization is done by ordering and plotting percentile values from the data available. The bottom 25 percentiles goes into “Very Low”, next 25 percentiles “Low”, similarly followed by “Medium” and “High”

2. Age Groups

Age groups under consideration would be: 18-24, 25-44, 45-64 and 65-older. The measure of the variable is ordinal.

- Dependent Variable:

% Percentage of smoking population

The variable is a percentage of people from sample population who smoke currently. Hence, this is a ratio type of variable. As per the definition of the ratio type of variable, this variable percentage could be zero also if no respondent in the sample smokes.

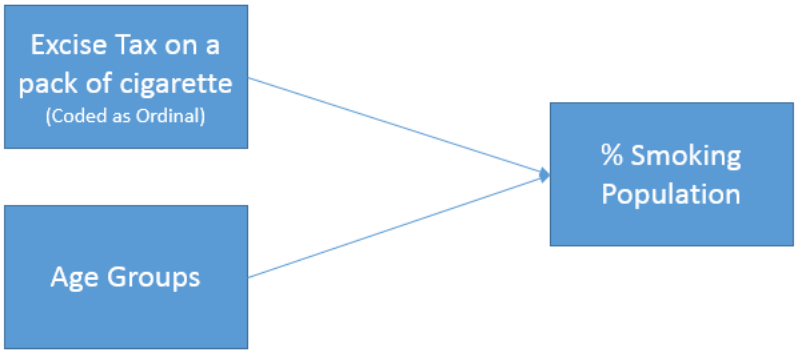


Fig. Independent and Dependent Variables

3. Statistical Tests

Is there a relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state? –For this question we have predictor variable as Excise Tax on a Pack of Cigarette ($ per pack) which is of the measure Ratio. % Percentage of smoking population for a given state which is also of Ratio measure. We conclude that we will need to perform a correlation and simple linear regression tests for our analysis.

If there is a relationship between the two, does state excise tax value on a pack of cigarette impact % of smoking population belonging to all age groups equally?–For this question we have predictor variables as Excise Tax on a Pack of Cigarette coded into an ordinal variable and Age groups. And, outcome variable as % Percentage of smoking population for a given state which is ratio measure. We conclude that we will need to perform factorial ANOVA with interactions for our analysis. The test will help us identify whether effect of State Excise Tax on Cigarette on % Smoking populations depends on the age group or not. If we find significant difference in mean of outcome variable with respect our independent variable for different levels of the other independent variable. We will also perform below correlation and linear regression tests to have detailed findings of the same: o

1. Excise Tax vs % Smoking Population between ages 18 to 24
2. State Excise Tax vs % Smoking Population between ages 25 to 44
3. State Excise Tax vs % Smoking Population between ages 45 to 64
4. State Excise Tax vs % Smoking Population ages 65 and over

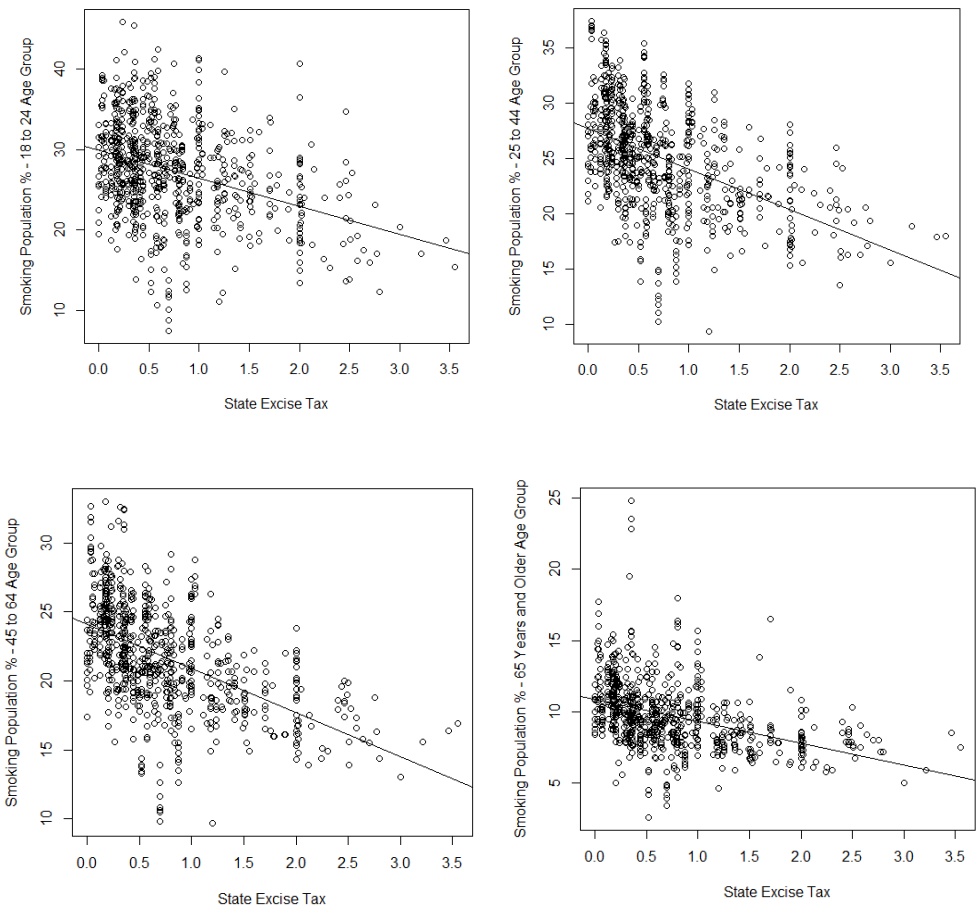
3.1 Descriptive Analysis

Exploratory Data Analysis:
To begin with, we studied the distribution of the variables relevant to our studies. Based on our intuition and our research on the problem domain, we also plotted potential predictor and outcome variables to understand if we were in the right direction.

Below are our initial findings of some key facts based on the data available:

Mean Tax on Cigarette	$ 0.74 per pack	All states, 1996-2014
Minimum of Tax on Cigarette, all states, 1996 - 2014	$ 0.00 per pack	In New York, 2007
Maximum of Tax on Cigarette, all states, 1996 - 2014	$ 3.55 per pack	In New York, 2010
Minimum Smoking Population at any time	9.1 %	In Utah, 2010
Maximum Smoking Population at any time	32.6 %	In Kentucky, 2002

To understand relation between potential predictor and outcome variables we plotted the below graphs:



We found that there was some relation between the variables of our interest. So we went ahead with our hypotheses testing.

3.2 Hypothesis Testing

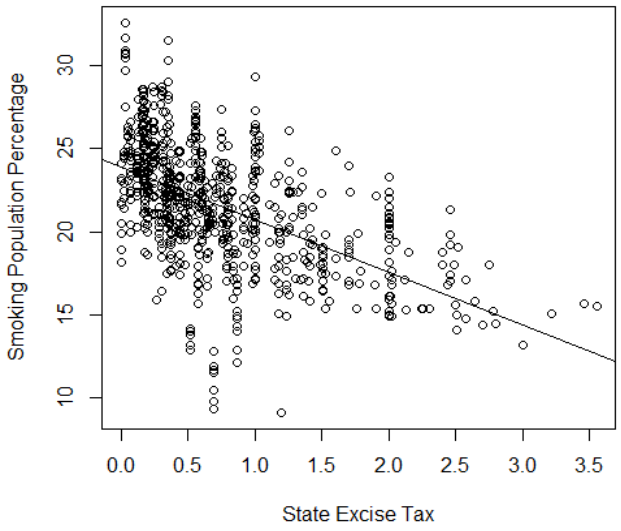
a. Is there a relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state?

Hypothesis:

H_o: ρ = 0: There is no relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state.

H_a: ρ ≠ 0: There is relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state.

Relevant test statistics:



Correlation test: r(762) = -0.5422938, p-value<2.2e-16

Linear Regression model: %SmokingPopulation = b_0 + b_1StateExciseTax = 23.8826 - 3.1696StateExciseTax

Summary of test results:

Based on the correlation test conducted we get that t-statistic is -17.817. Also, p-value < 2.2e-16 which is less than our significant value on 0.05. We reject the null hypothesis. This implies that the alternate hypothesis is true. Hence we can conclude that there is a significant relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state. Based on the r-value of the test, which is -0.5422938 we can further conclude that there is a negative relation between the two variables. That is: An increase in state excise tax is associated with decrease in % smoking population in a state.

Also, $R^2 = (-0.5422938)^2 = 0.29$, which means 29 percentage of variance in % smoking population is explained by State Excise Tax on per pack of cigarette.

b. If there is a relationship between the two, does state excise tax value on a pack of cigarette impact % of smoking population belonging to all age groups equally?

Hypothesis:

H_0: Difference between means of % smoking population for different categories of State Excise tax are equal for all age groups.

H_a: Difference between means of % smoking population for different categories of State Excise tax is different for at least one pair of age groups.

Here we conduct a two way ANOVA interaction test. A significant interaction would here mean that we should ignore the main effects for our independent variables and we should go about dividing our dataset for each level of age groups and calculate simple main effects for our independent variable State Excise Tax Category.

Relevant Test Statistics:

```
> m = aov(Data_Value~State_Tax_Category*Age,data=mydata)
> summary(m)
                          Df Sum Sq Mean Sq F value   Pr(>F)
State_Tax_Category          3  10388    3463  211.60  < 2e-16 ***
Age                         3 140258   46753 2857.15  < 2e-16 ***
State_Tax_Category:Age      9    960     107    6.52 2.99e-09 ***
Residuals                3040  49745      16
---
```

Calculating effect size, we get:

Overall R^2 = SSbetween/SStotal = 151606/201351= 0.7529

State_Tax_Category eta^2 = SSbetween/SStotal = 10388/201351= 0.0516

Age eta^2 = Between/SStotal = 140258/201351= 0.6965

State_Tax_CategoryXAge eta^2 = SSbetween/SStotal = 960/ 201351 = 0.005

State_Tax_Category*Age: The test statistic for the interaction is 6.52. P-value = 2.99e-09.

Summary of test results:

We reject the null hypothesis. There is at least one pair of age groups for which the difference in means of % smoking population is different for different State excise tax categories. Also, the Percent of variance explained by the model is 75.29%. Since interaction is significant we perform one-way ANOVAs for each age group.

Performing test to determine simple effects. This will help us determine effect of one independent variable on the dependent variable at a given level of other independent variable. Since we perform 4 individual test. We need to re-adjust the alpha level to minimize family wise error. 0.05/4 = 0.0125. Therefore, our new **alpha level is 0.0125**.

Hypotheses test:

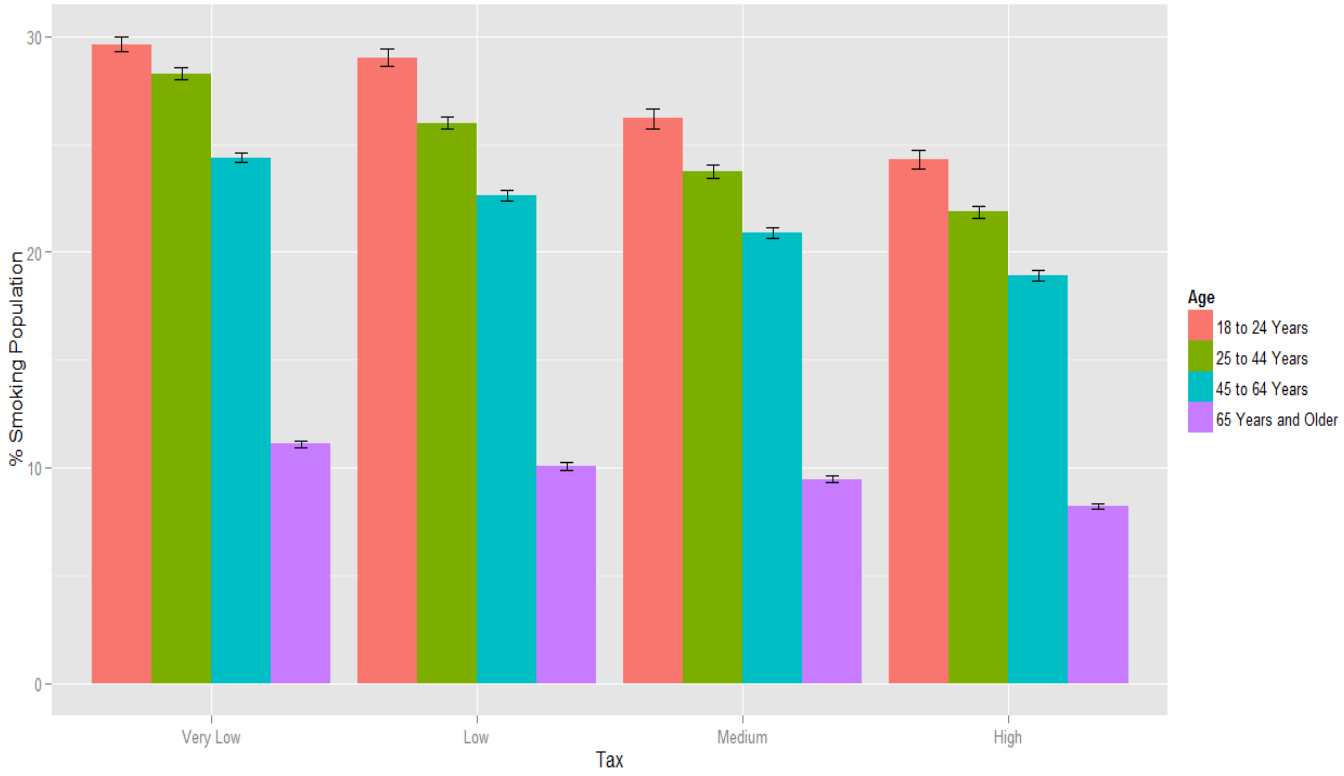
i. H_o: No difference in % smoking population means for different state excise tax levels for age group 18-24 years.
   H_a: At least one difference in % smoking population means for different state excise levels for age groups 18-24 years.
ii. H_o: No difference in % smoking population means for different state excise tax levels for age group 25-44 years.
   H_a: At least one difference in % smoking population means for different state excise levels for age groups 25-44 years.
iii. H_o: No difference in % smoking population means for different state excise tax levels for age group 45-64 years.
   H_a: At least one difference in % smoking population means for different state excise levels for age groups 45-64 years.
iv. H_o: No difference in % smoking population means for different state excise tax levels for age group 65 and older.
   H_a: At least one difference in % smoking population means for different state excise levels for age groups 65 and older.

Relevant test statistics for the above tests:

i. $F(3,760) = 34.1$, p-value< 2e-16
ii. $F(3,760) = 84.86$, p-value <2e-16
iii. $F(3,760) = 90.87$, p-value< 2e-16
iv. $F(3,760) = 52.68$, p-value< 2e-16

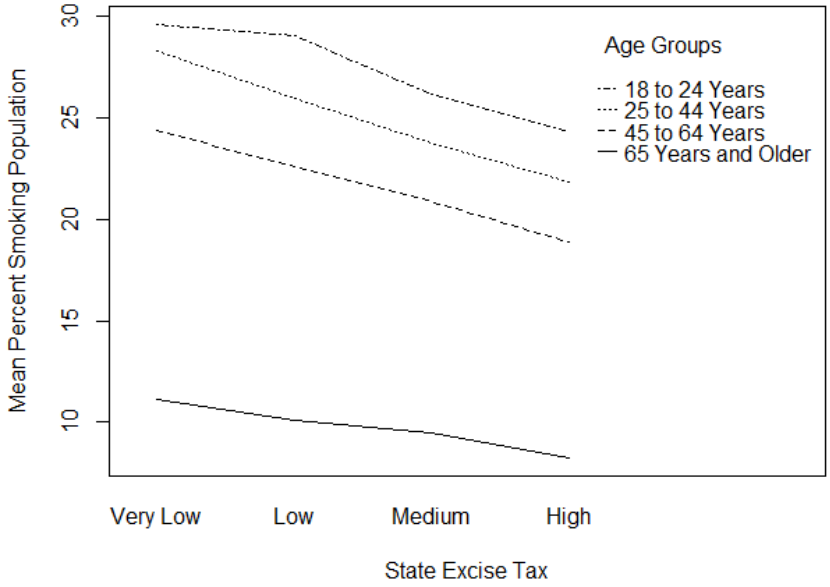
All the above test results are significant at alpha level is 0.0125. Rejecting all null hypotheses.

Since ANOVA is an omnibus test it doesn't tell us population means for which pair is different. So, we plot the below graphs highlighting interaction.



Based on the above graph we can infer the following things:

a. There is a general trend across all groups that with the increase in the state excise tax there is a decrease in smoking population.
b. For various tax groups the difference in percent smoking population is least affected for the age group 65 years and older
c. This brings us to two intuitions:
    i. This is the older population, most likely they have been smoking since a long time and hence the degree of addiction is such that tax raise doesn't help reduce the smoking population much.
    ii. Also if we see there is a steep fall of percent of smoking population after 45-64 years. Giving this observation we feel that this is a potential indicator that many smokers belonging to this age group might not make it 65 years or more



The above interaction graph suggests that the difference in % smoking population with an increase in tax is different for different age groups.

For the policy makers to have a better picture of how different age groups respond to changes in tax. We also performed individual linear regression tests for different age groups to arrive at a model that policy makers can use for predicting the change in smoking population. Since we perform 4 individual test. We need re-adjusted the alpha level to minimize family wise error. 0.05/4 = 0.0125. Therefore, our new **alpha level is 0.0125**. Here we subset data sets for each age group and ran individual correlation and linear regression tests to determine which age groups are related and to what extent. The scope of each test is restricted to a given age group.

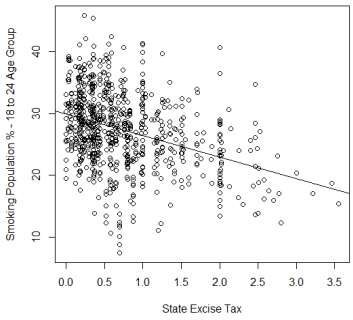
Note: Here the State Excise Tax is of the measure ratio.

i. State Excise Tax vs % Smoking Population between ages 18 to 24

Hypothesis:

H_o: ρ = 0: There is no relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state.

H_a: ρ ≠ 0: There is relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state.



Relevant test statistics:

Correlation test: r(762) = -0.3537829, p-value<2.2e-16
Linear Regression: %SmokingPopulation_Age18to24 = b_0 + b_1StateExciseTax = 29.8720 - 3.4638*StateExciseTax

Summary of test results:

Based on the correlation test conducted we get that t-statistic is -10.441. Also, p-value < 2.2e-16 which is less than our significant value on 0.05. We reject the null hypothesis. This implies that the alternate hypothesis is true**. Hence we can conclude that there is a significant relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population of age group 18 to 24 in that state**. Based on the r-value of the test, which is -0.3537829 we can further conclude that there is a negative relation between the two variables. That is: An increase in state excise tax is associated with decrease in % smoking population in a state.

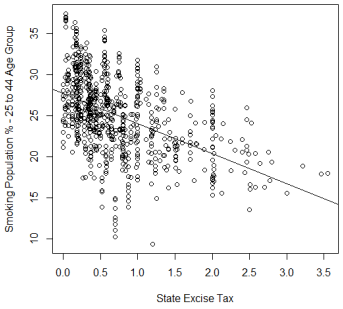
Here $R^2 = (-0.3537829)^2 = 0.124$, which means 12.4 % percentage of variance in % smoking population is explained by State Excise Tax on per pack of cigarette.

ii. State Excise Tax vs % Smoking Population between ages 25 to 44

Hypothesis:

H_o: ρ = 0: There is no relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population between ages 25 to 44 in that state.

H_a: ρ ≠ 0: There is relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population between ages 25 to 44 in that state.



Relevant test statistics

Correlation test: r(762) = -0.4862674, p-value<2.2e-16
Linear Regression: %SmokingPopulation_Age25to44 = b_0 + b_1StateExciseTax =27.6920 - 3.6560*StateExciseTax

Summary of test results:

Based on the correlation test conducted we get that t-statistic is -15.362. Also, p-value < 2.2e-16 which is less than our significant value on 0.05. We reject the null hypothesis. This implies that the alternate hypothesis is true**. Hence we can conclude that there is a significant relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population of age group 25 to 44 in that state**. Based on the r-value of the test, which is -0.4862674. We can further conclude that there is a negative relation between the two variables. That is: An increase in state excise tax is associated with decrease in % smoking population in a state.

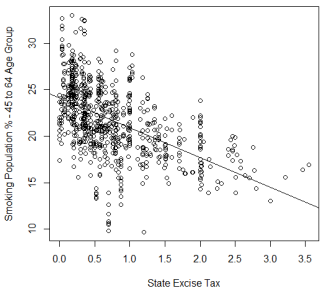
Here R^2 =(-0.4862674)^2 = 0.2355, which means 23.5 % percentage of variance in % smoking population is explained by State Excise Tax on per pack of cigarette

iii. State Excise Tax vs % Smoking Population between ages 45 to 64

Hypothesis

H_o: ρ = 0: There is no relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population between ages 45 to 64in that state.

H_a: ρ ≠ 0: There is relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population between ages 45 to 64 in that state.



Relevant test statistics:

Correlation test: r(762) = -0.5213718, p-value<2.2e-16

Linear Regression: %SmokingPopulation_Age45to65 = b_0 + b_1StateExciseTax =24.0853-3.2025*StateExciseTax

Summary of test results:

Based on the correlation test conducted we get that t-statistic is -16.866. Also, p-value < 2.2e-16 which is less than our significant value on 0.05. We reject the null hypothesis. This implies that the alternate hypothesis is true**. Hence we can conclude that there is a significant relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population of age group 45 to 65 in that state**. Based on the r-value of the test, which is -0.5213718. we can further conclude that there is a negative relation between the two variables. That is: An increase in state excise tax is associated with decrease in % smoking population in a state.

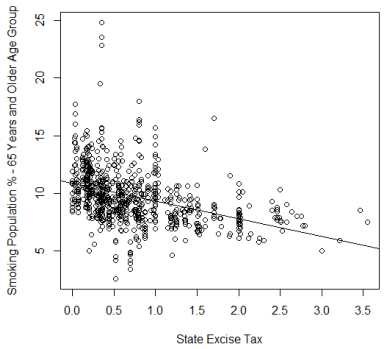
Here R^2 =(-0.5213718)^2 = 0.2709, which means 27.09 % percentage of variance in % smoking population is explained by State Excise Tax on per pack of cigarette

iv. State Excise Tax vs % Smoking Population ages 65 and over.

Hypothesis

H_o: ρ = 0: There is no relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population between ages 65 and over in that state.

H_a: ρ ≠ 0: There is relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population between ages 65 and over in that state.



Relevant test statistics:

Correlation test: r(762) = -0.394692, p-value<2.2e-16

Linear Regression: %SmokingPopulation_Age65plus = b_0 + b_1StateExciseTax=10.8653-1.5286*StateExciseTax

Summary of test results:

Based on the correlation test conducted we get that t-statistic is -11.858. Also, p-value < 2.2e-16 which is less than our significant value on 0.05. We reject the null hypothesis. This implies that the alternate hypothesis is true**. Hence we can conclude that there is a significant relation between state level excise tax on a pack of cigarettes and % of cigarette smoking population of age group of 65 and older in that state**. Based on the R-value of the test, which is -0.394692. We can further conclude that there is a negative relation between the two variables. That is: An increase in state excise tax is associated with decrease in % smoking population in a state.

Here R^2 =(-0.394692)^2 = 0.1547, which means 15.47 % percentage of variance in % smoking population is explained by State Excise Tax on per pack of cigarette

4. Assumptions and Limitations:

Research Question 1: Is there a relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state?

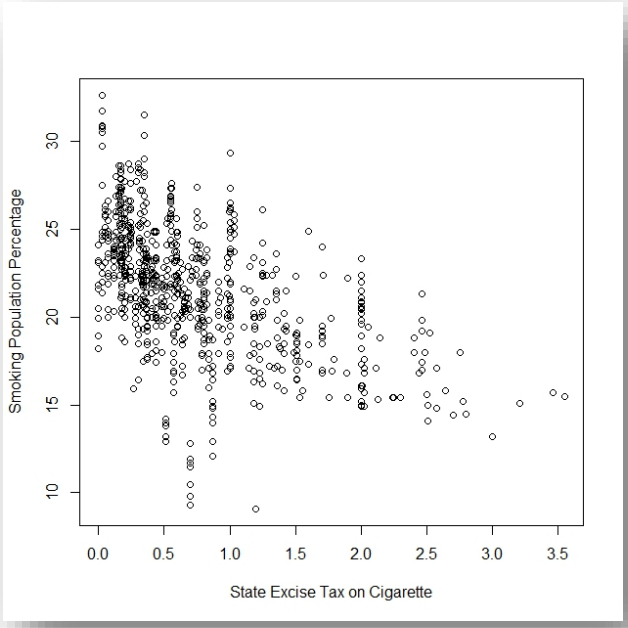


Fig. Scatterplot between State Excise Tax on Cigarette and related Smoking Population Percentage

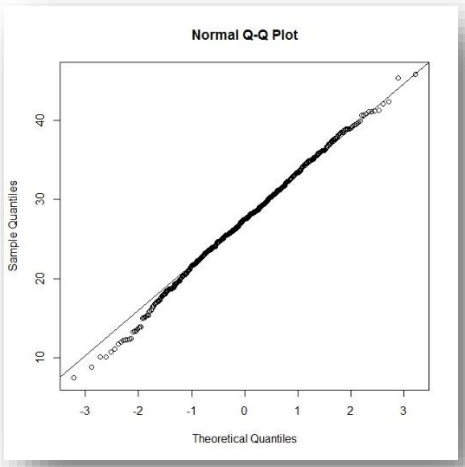
1. **Restricted Range:** The data was available for last 19 years, 1996 – 2014 for the analysis. And hence, the best fit line is applicable only for the limited range of period it has been analyzed for.

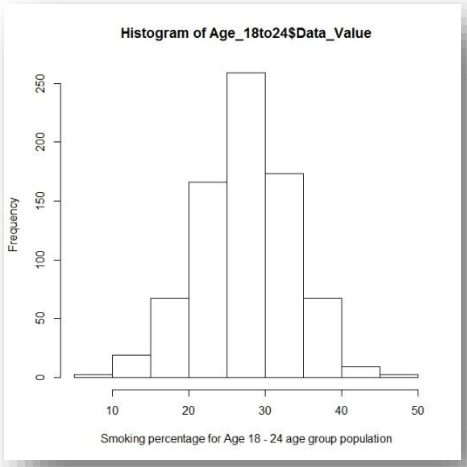
2. **Linearity of data:** As seen from the above scatterplot, the data is slightly nonlinear.

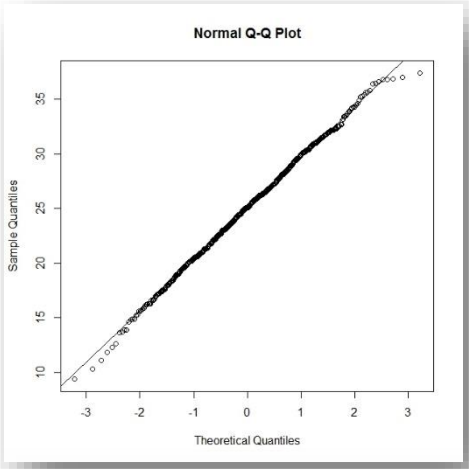
3. **Outliers:** As seen from the scatterplot, there are some outliers present in every set of data. Since our co-efficient prediction is based on minimization of squared error, few potential big outliers can influence our estimations causing our co-efficient and in turn confidence intervals to not be accurate.

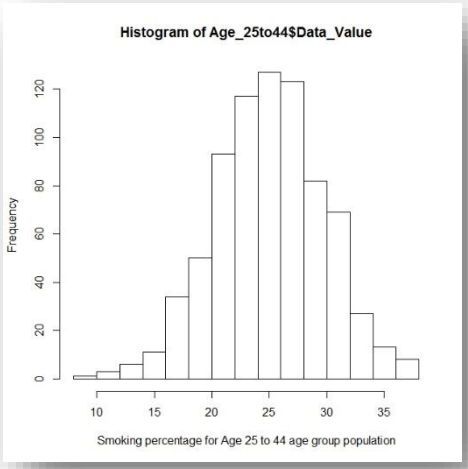
4. **Causation:** Correlation doesn't imply causation. There could be other variables such as income, rehabilitation, or anything else that may have reduced smoking population.

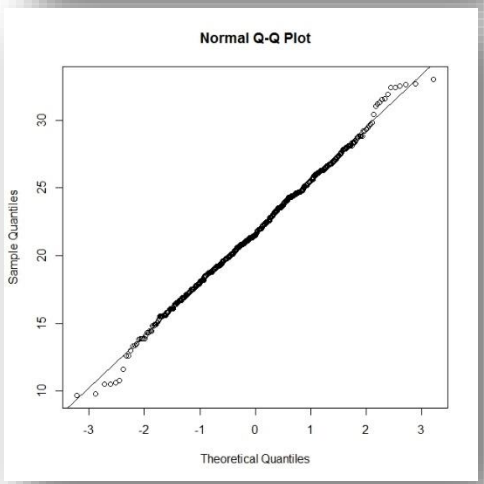
Research Question 2. If there a relationship between state level excise tax on a pack of cigarettes and % of cigarette smoking population in that state, is the relation same to all age groups?

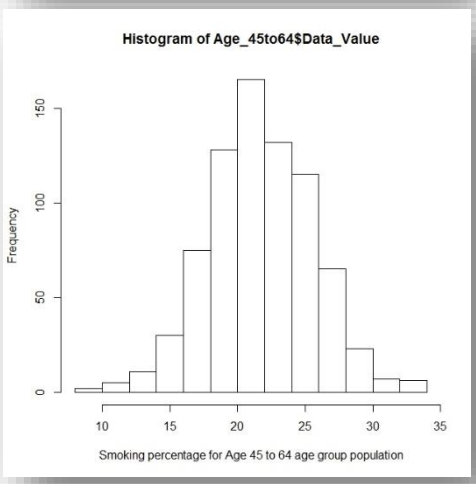


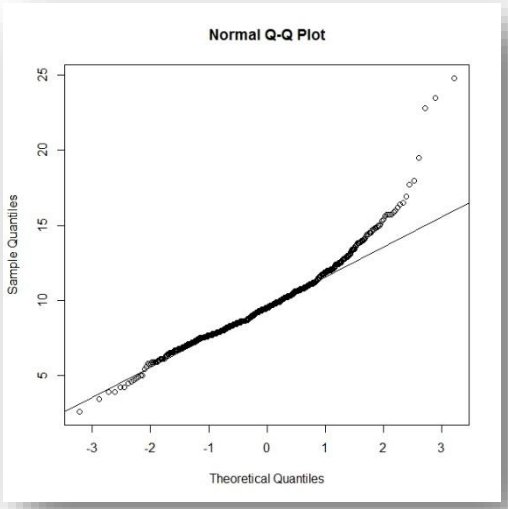












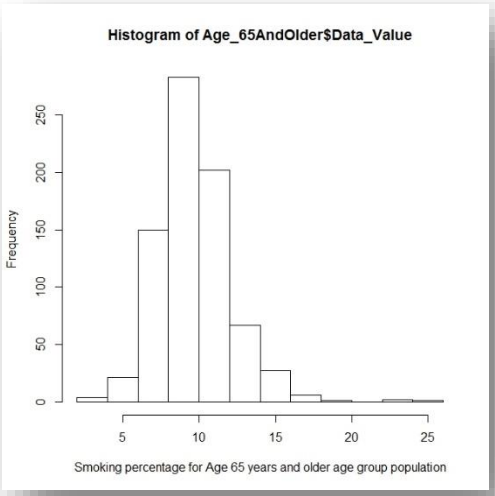


Fig. QQ plots to test Normality assumption

1. **Normality:**
   As seen from above QQ plots, the data is almost normal. Hence, we can say that the normality assumption is not violated.

2. **Homogeneity of Variance:**
   We conducted **Levene's test** to see that Null hypothesis rejected. It means at least one variance is not equal
   Assumption violated, but it is not so strict in this case.

3. **Independence of Observations:**
   The observations are independent of each other. The state excise tax for a given year in a given state is independent of other observation.

5. Conclusion

Conclusion of RQ1:

There is a significant and negative relationship between State Excise Tax on per pack of cigarette and % Percent of smoking population. We found that 1 $ rise in Excise Tax on Cigarette packet could result in decrease in smoking population by approximately 3 %. This led us to deep dive into how is the effect distributed among various age groups. So we designed our second research question to address the same.

Conclusion of RQ2:

We found that there is a significant interaction between the state excise tax on cigarettes and % smoking population for a given level of age group. This led us to determine what the difference of impact in each groups is. Below are our findings:

The maximum variation in % smoking population based on State Excise Tax on per pack of cigarette is observed in Age group 25 to 44 years.

The least variation in % smoking population based on State Excise Tax on per pack of cigarette is observed in Age group 65 years and older.

6. Limitations of findings:

1. We used the state excise tax per pack of cigarettes data to look for its relation with % smoking population. However, tax on cigarette also includes federal level tax. There can be scenarios where state excise tax is reduced and federal excise tax increased and the total tax still increasing which could lead to reduced smoking percentage. But for our test such scenario will be seen reduced state tax and reduced smoking percentage. Leading to errors in the test.

2. Restricted Range: The best fit line as shown in the scatterplots in the research question analysis only applies to the range of data we have used for analysis. This poses certain limitations on the study because we do not have data of smoking population percentage and excise tax on the same before year 1995. If the data is made available, it can influence the best fit line and the predictions can differ based on the same.

3. Causation: We can see that the data is negatively correlated which means if the Excise Tax is more, then smoking population percentage is less. Also if the excise tax is more, then the smoking population percentage is seen on the higher side. Here, we cannot exactly say that this is only due to the Excise Tax rates. There can be other factors too.

7. Test Summary Data and Codes

R-code included for the entire project included in the file embedded below.

Descriptive Statistics and Hypothesis testin

8. References

http://www.cdc.gov/tobacco/data_statistics/surveys/index.htm

http://www.cdc.gov/brfss/data_documentation/pdf/userguidejune2013.pdf