

Analysis of Tobacco Consumption in The U.S

Ketchina Duval

Statistical Methods 7100 – Dr. Ling Zu

December 16, 2024

Abstract

This analysis examines the consumption of tobacco products and analyzes trends and patterns in U.S. adult tobacco consumption from 2000 to 2023. A series of statistical methods were utilized to determine how different variables in the dataset relate to each other and the overall consumption. A T-test was conducted to compare total per capita consumption between combustible and noncombustible tobacco. The outcome showed a statistically significant difference between the two groups. A one-way ANOVA between the total per capita and different tobacco products revealed a significant difference between product types, proving that the type of products consumed has a measurable effect on per capita consumption. The third test performed was a linear regression analysis to determine how the variable year impacted tobacco consumption, the model revealed a statistical significance but weak negative relationship between year and consumption. All the tests were performed with a statistical significant level (α) of 0.05.

Introduction

1. Does tobacco consumption increase as the year increases?
2. Is there a significant difference in per capita consumption between domestic and imported tobacco products?
3. Is there a difference in the average of product type being consumed?

Objective

The goal is to understand the trend of tobacco consumption over the 23 years, to find out which tobacco product is contributing to the overall consumption, whether the year has an impact on the consumption of tobacco products, and how the consumption varies by product type. This analysis can influence public health initiatives, depending on how the consumption has trended over the years, they can pinpoint which years had the highest consumption and determine what

the cause was or vice versa. This analysis can help lawmakers pass tobacco control policies and launch campaigns toward smoking cessation. On the other hand, it can also help tobacco companies with market strategies about which products they should promote more.

Data Overview

This dataset was collected by the Center for Disease Control. It contains 312 observations and 14 variables. It was collected nationally from 2000 to 2023 and does not include any states or regions. It includes the total consumption of tobacco products and the per capita consumption of domestically produced and imported products, the datasets cover various types of combustible and noncombustible products. The analysis will analyze the relationship between categorical variables (measure, sub-measure, topic, location) and numerical variables (domestics, imports, year, per capita consumption).

SAS Studio 9.4 was used to clean the data by changing some of the variable names and sorting the variables by removing duplicates and missing data.

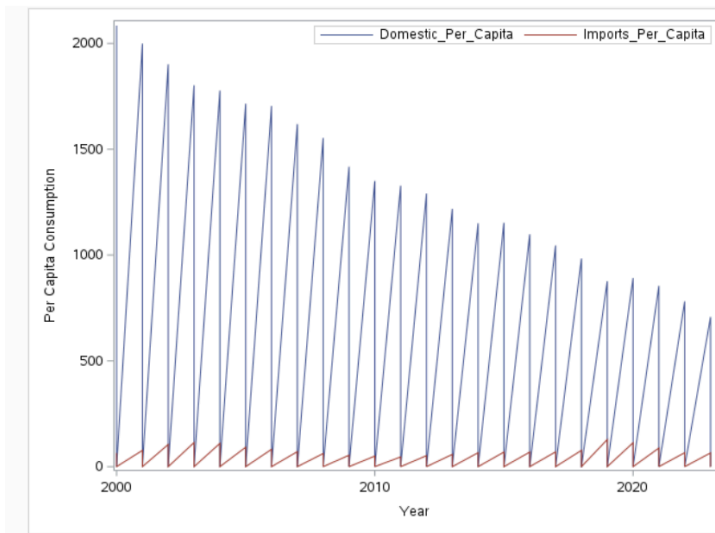
Exploratory data analysis

Summary Statistics

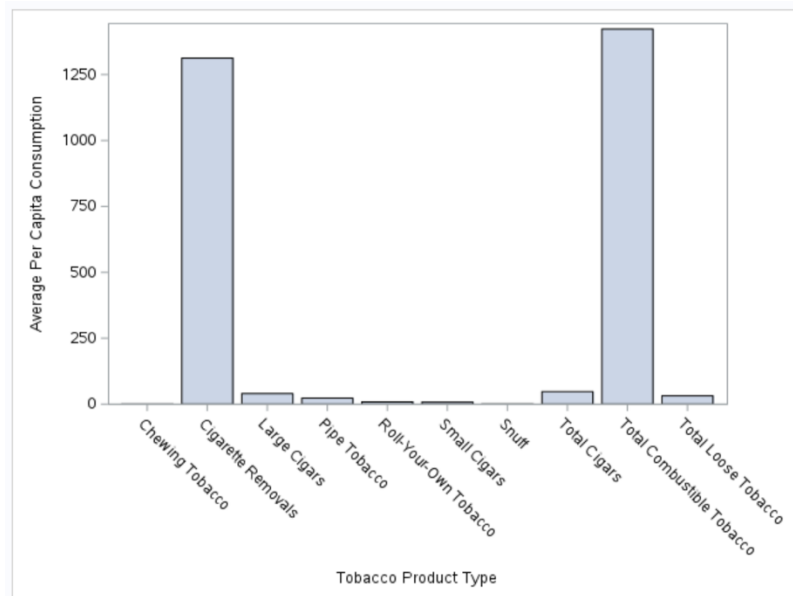
- Domestic Per Capita Consumption: Mean = 213.93, Std Dev = 492.12
- Imports Per Capita Consumption: Mean = 13.46, Std Dev = 25.18
- Total Per Capita Consumption: Mean = 227.43, Std Dev = 514.14

Visualizations

- Time-series line graphs of per capita consumption show domestic products dominate imports consistently and shows a downward trend in consumption.



- The bar chart of sub-measure (product type) categories highlights that cigarettes and total combustible tobacco are the most consumed product types.



Patterns and Anomalies

- Consumption of some product types fluctuates, suggesting changing preferences or cost.

Statistical Methods

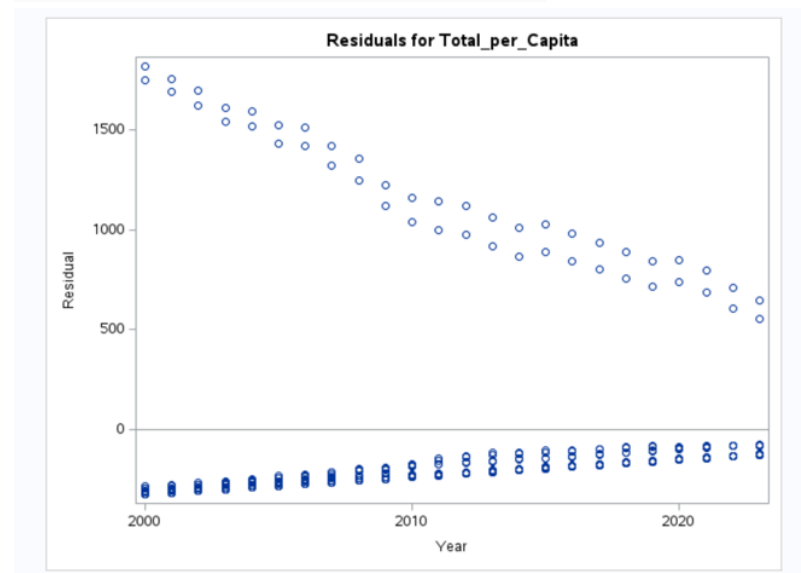
a) Hypothesis 1

H_0 : The null hypothesis is that per capita consumption has increased over the years.

H_1 : The alternative hypothesis is that there is no significant change in per capita tobacco consumption.

Test: Linear Regression ($R^2 = 0.0137$, p-value = 0.0385)

The REG Procedure					
Model: MODEL1					
Dependent Variable: Total_per_Capita					
Number of Observations Read		312			
Number of Observations Used		312			
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1129666	1129666	4.32	0.0385
Error	310	81081351	261553		
Corrected Total	311	82211017			
Root MSE					
		511.42228	R-Square	0.0137	
Dependent Mean		227.43242	Adj R-Sq	0.0106	
Coeff Var		224.86780			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	17713	8413.59598	2.11	0.0361
Year	1	-8.69270	4.18272	-2.08	0.0385

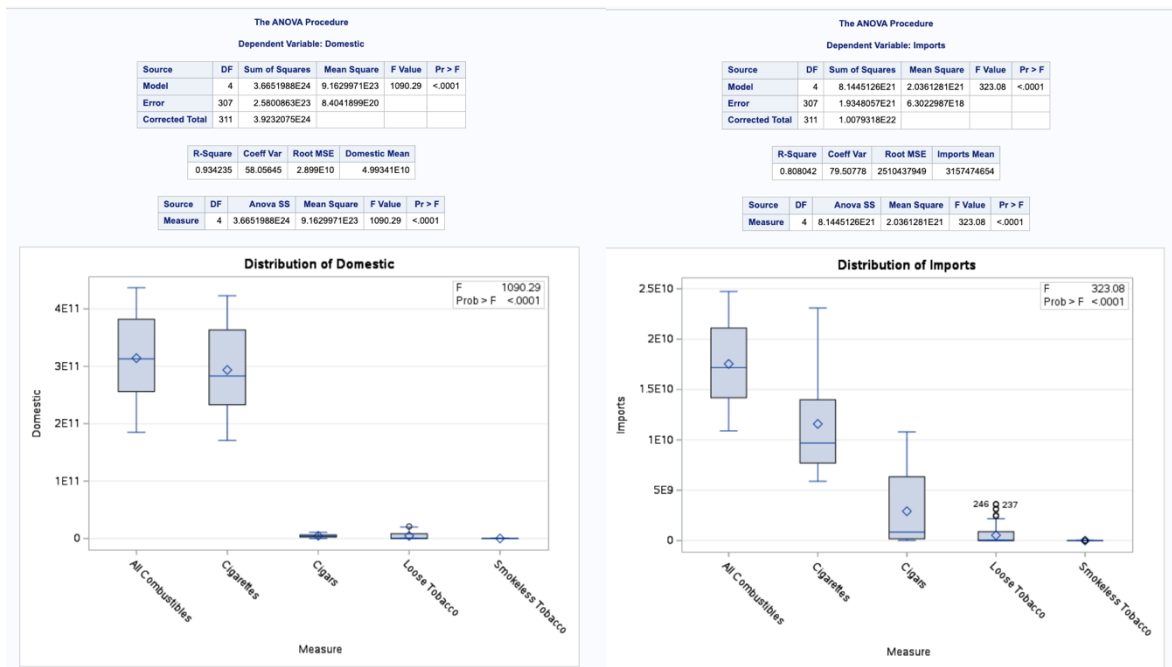


Hypothesis 2

H_0 : Mean per capita consumption is the same across all product types.

H_1 : Mean per capita consumption differs among product types.

Test: ANOVA (p-value of domestic consumption <0.0001, p-value of imported consumption <0.0001)



Hypothesis 3:

H_0 : There is no significant difference in the mean total per capita consumption between Combustible tobacco and noncombustible tobacco
 $H_0: \mu_{\text{combustible}} = \mu_{\text{noncombustible}}$

H_1 : There is a significant difference in the mean total per capita consumption between combustible tobacco and noncombustible tobacco.
 $H_1: \mu_{\text{combustible}} \neq \mu_{\text{noncombustible}}$

Test: T-test (pooled p-value = 0.0001, satterthwaite <0.0001)

T-Test: Comparison of Per Capita Consumption between Domestic and Imported Products

The TTEST Procedure

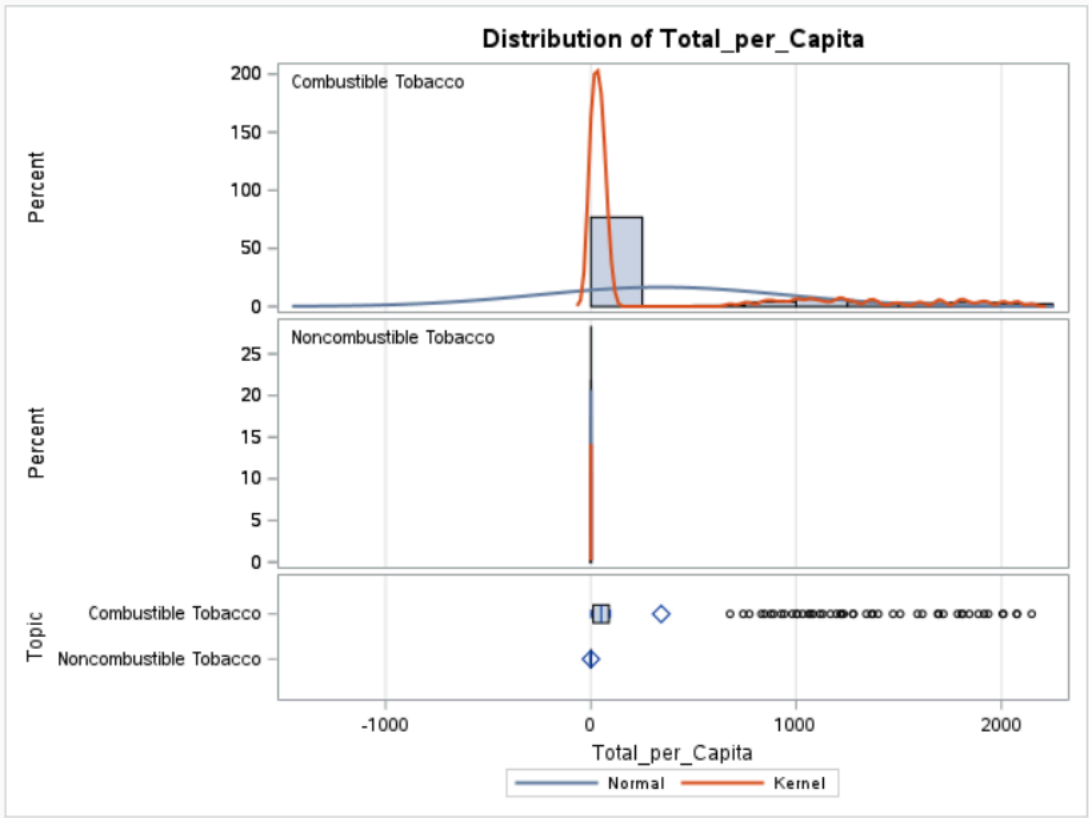
Variable: Total_per_Capita

Topic	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Combustible Tobacco		207	342.7	599.5	41.6673	0.00600	2148.0
Noncombustible Tobacco		46	0.2677	0.1545	0.0228	0.0500	0.4770
Diff (1-2)	Pooled		342.5	543.1	88.5265		
Diff (1-2)	Satterthwaite		342.5		41.6673		

Topic	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Combustible Tobacco		342.7	260.6 424.9	599.5	546.8 663.5
Noncombustible Tobacco		0.2677	0.2219 0.3136	0.1545	0.1281 0.1946
Diff (1-2)	Pooled	342.5	168.1 516.8	543.1	499.5 595.2
Diff (1-2)	Satterthwaite	342.5	260.3 424.6		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	251	3.87	0.0001
Satterthwaite	Unequal	206	8.22	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	206	45	1.506E7	<.0001



Interpretation of Result

Test 1: Linear Regression:

The p-value $0.0385 < 0.05$ means that the per capita consumption decline is statistically significant. Therefore, the null hypothesis was rejected. Though there is a relationship between the year and the consumption of tobacco, it is a weak relationship because $R^2 = 0.0137 \sim 1.37\%$. Meaning that the variable year had a small influence on the decline, it might have been due to other factors not included in the dataset.

Test 2: ANOVA

The p-value from our ANOVA for domestic and imported products is less than 0.0001 which is less than the alpha 0.05. Therefore, the null hypothesis was rejected. From the ANOVA tables, it is clear that there is a difference in the mean per capita consumption. The box plot showed more consistency in the domestic products that were consumed compared to imported products.

Test 3: T-Test

Based on the t-test result, the Satterthwaite p-value 0.0301 is less than 0.05, proving that there is a statistically significant between the mean of combustible and noncombustible products. Therefore, we reject the null hypothesis. Noncombustible Tobacco shows a wider spread and higher values, consistent with the mean.

Limitation

The datasets lack demographic breakdown; the dataset would have been even better had it included age, sex, income level, and consumption per state. The analysis would have been more informative and more detailed. I wish I could have studied the relationship between age and decrease in consumption.

Conclusion

The analysis shows that there has been a decline in tobacco consumption over the past 23 years, which could be due to a lot of different factors that were not included in the dataset. It also shows that domestically produced products had a higher consumption than imported products. While trends are consistent, future analysis should focus on demographic influences as well as trendy tobacco products like vapes and e-pens. Comparing younger adults versus older adults and which product they each consume the most. Furthermore, from a public health perspective, this is a good outcome. I would like to revisit this analysis with the demographics aforementioned and see if tobacco usage continues to trend downward.

Codebook

Variable Name	Label	Code	Format
Topic	Combustible/Noncombustible	3	1=poor
Submeasure/measure	Product Types	3	2=average
Domestic_per_Capita	Domestic	3	3=good
Imports_Per_Capita	Imports	3	4=excellent
Total_Per_capita	Per capita consumption	3	