Statistical Analysis of Yearly Consumption Expenditure of the Household on Durable, Non-Durable Goods and Services in Pakistan

Pamir Khan, Yasir Khan, and Said Farooq Shah
Department of Statistics, University of Peshawar, Peshawar, Pakistan
Corresponding Author: Pameerkhan18@gmail.com

Abstract
The purpose of the study was to explore the yearly consumption expenditure of household on durable, non-durable goods and services in Pakistan using partial data from Pakistan social and living standards measurement survey. From the hidden patterns in the data using different data mining techniques and statistical tools it has been observed that consumption quintiles in various regions, provinces, genders, and languages are changing. In addition, the mean consumption expenditures per household and mean consumption expenditures per-capita vary according to region, gender, province, and language. While summarizing consumption quintiles and people infected with different diseases, it has been seen that the number of infected people declines as consumption expenditure increases. The Chi-square test revealed that the number of people infected with malaria and hepatitis differs by consumption expenditure quintile. The number of days a person had malaria was negatively correlated with monthly consumption expenditure. The descriptive tables result that woman in rich households are more likely to make decisions in their households related to education and employment. In the regression analysis, women's involvement in decision-making impacts the monthly consumption expenditure per household except for medical decisions.

Keywords: Consumption expenditure, Durable, Non-durable, Goods, Services.
Introduction

Goods are the pillars of an economy, and the supply and demand of certain goods can be used as economic indicators to determine an economy’s strength. In economics, goods can be divided into two classes: durable goods and nondurable goods.

Durables, also identified as durable goods or consumer durables, are goods that do not need to be acquired very often and typically last for at least three years however non-durables are those goods that are purchased very often and can last for less than three years. (System of National Accounts, 2008)

Demand is generally categorized into demand for durable and non-durable goods. Because durable goods can be used repeatedly, their demand usually increases during economic growth which directly accounts for an increasing number of purchases. However, during the recession, the demand for durable goods leans to go down and so does the number of purchases. Hence, they are considered a good economic indicator. Nevertheless, the demand for non-durable goods remains unchanged throughout economic growth and setback. Consumers normally purchase the same quantity of non-durable goods as durable goods, during both recession and growth.

Goods and Services

Goods refer to the tangible consumable products, articles, commodities that are offered by the firms to the customers in exchange for money. They are the items that have physical features, i.e., shape, appearance, size and weight etc.

Services are the intangible economic product that is delivered by a person on the other person’s demand. It is an activity carried out for someone else. They can only be delivered at a particular moment, and hence they are perishable in nature. They lack physical identity. Services cannot be distinguished from the service provider.

Durable and Non-durable Goods

A durable good may be used frequently or continuously for more than a year, assuming a normal or average rate of physical usage. A consumer durable is a good that may be used for consumption frequently or continuously for a year or more. On the other hand, a non-durable good would be better defined as a single-use good, for example, food and drinks are used once only to satisfy hunger or thirst, heating oil, coal or firewood can be burnt once only but they are nevertheless extremely durable physically and can be stored. (System of National Accounts, 2008).

A durable good is the one whose life expectancy is one year or more such goods include furniture, clocks, wrist watches, television, and radio, etc. (Pakistan Bureau of Statistics, 2018-2019).
A non-durable goods as those goods whose life expectancy is less than one year. Non-
durable goods include food items, clothing, fuel, lighting, footwear, and medicines, etc. (Pakistan
Bureau of Statistics, 2018-2019).

**Consumption Expenditure**

An expenditure represents a payment with either cash or credit to purchase goods or services. An
expenditure is recorded at a single point in time (the time of purchase), compared to an expense
that is recorded in a period where it has been used up or expired.

Or according to the organization system of national accounts (SNS) Expenditures on goods
and services are defined as the values of the amount that buyers pay, or agree to pay, to sellers in
exchange for goods or services that sellers provide to them or other institutional units designated
by the buyers. (System of National Accounts, 2008)

**Literature Review**

Using quarterly data from Canada, the UK and the USA shows that inflation both in the short
and long run negatively affects durable and non-durable consumption, and positively influences
the current account balance. In particular, the impact of inflation is more noticeable on durable
relative to non-durable goods (Malick and Mohsim, 2016).

The importance of Consumer durable on consumption and saving over life found out that
the consumption expenditures on both durable and non-durable are hump-shaped: expenditures
are low early in life, then rise considerably and fall again. The average household in the survey
of consumer expenditures spends 65% more when the head of the household is 50 than when she
is 25 and around 80% more than when she is 65. Second, young households keep very few liquid
assets to hold most of their wealth in consumer durable. Later in life, then, families accumulate a
significant amount of financial assets for retirement. The importance of durable goods is also
mirrored in the aggregate composition of wealth: households hold 35% of their total assets in
real estate and other consumer durables and only 28% inequity. A nonparametric regression was
used to estimate the consumption life cycle profiles. The estimation model does not include the
cohort index or the quarter as a variable (Villaverde and Krueger, 2001).

Product category determinants of price knowledge for durable consumer goods estimate
the extent of variation in consumer price knowledge levels across an array of durable consumer
goods. A traditional approach for the study was used to conduct interviews with consumers in
which a sample of consumers are presented with products and asked to estimate their prices.
Data from the popular television game show THE PRICE IS RIGHT was used in the study which
rated as the fourth popular television game show in the United States at the time.

Across all products, the average price knowledge score was found to be 39, indicating that on
average, 39% of consumers were able to provide price estimates within 25% of the actual price.
However, it was also shown that significant variations in price knowledge scores can be
observed across categories, ranged from a low of 7 to a high of 74. Similar results were observed for signifying the percentage of respondents with price estimates within 10% of the actual price. To assess the statistical significance of price knowledge score variations across the 51 products, a chi-square test was conducted. The observed variations were found to be significant at the p < .05 level for both ($\chi^2 = 69.4; \Phi = 0.24$) and PKS ($\chi^2 = 71.3; \Phi = 0.24$) (Estelami and Maeyer, 2004).

**Methodology**

Computer and statistical tools useful to find out the hidden pattern in data. Methodology section covers tools and techniques that help us completing this study.

**Data**

Data used in this research comes from the Pakistan Social and Living Standard measurement or Household income integrated survey 2018-2019. Pakistan Social Living Standard Measure surveys collect information on a wide range of topics at both individual and household levels. Data is been modified and converted from one form to another form for research purpose and to achieve the goals of our research.

**Sample Design of PSLM’s Survey**

**Universe:** The universe of this survey contains of all urban and rural zones of the four provinces and Islamabad excluding military restricted areas.

**Sampling Frame:** In the year 2003 the Federal Bureau of Statistics Pakistan his developed its own Frame. They divided each city and town into blocks of 200-250 households. Then each block was classified into three categories of income groups that is low, middle and high.

**Stratification of Plan of PSLM’s Survey**

The Stratification plan of PSLM’s Survey was different for both Urban and rural areas.

**Urban Areas:** Large cities like Peshawar, Islamabad, Sialkot etc. were well-thought-out to be a separate stratum and further stratified according to low, middle and high income.

**Rural Areas:** Each District in Punjab, Sindh and Khyber Pakhtunkhwa was combined to form a stratum. While the outdated administrative division of Balochistan province was consider as a separate stratum.

**Sample Size and its Allocation**

To produce reliable results at provincial level and completing the objective of National and Provincial level the survey sample size was fixed at approximately 17600 households comprising 1252 sample villages or enumeration blocks. On the other hand, for the district level the survey sample was fixed into 79600 households comprising 5563 sample villages or enumeration blocks.
Sample Design: A two-stage stratified sample design was implemented in PSLM’s survey.

Primary Sampling Units Selection: In both urban and rural areas villages and enumeration blocks were considered as Primary Sampling Units (PSUs). Using Probability Proportional to Size (PPS) Samples were selection at Primary level.

Secondary Sampling Units Selection: Households within PSUs were acquired as Secondary Sampling Units (SSUs). A Specified number of households that is 12 and 16 in each sample PSU of rural and urban area were selected using systematic sampling with a random start.

Data Quality and Reliability
The built-in system of checking of field work by controllers in the field were used to guarantee the excellence of the data and also data was monitored through terms from Pakistan Bureau of Statistics headquarters. Regional workplaces ensure the data quality through initial editing in their department level. The entire data was then carried out to Pakistan Bureau of statistics headquarters in Islamabad and data entry program was used to check the number of in-built consistency checks. The Coefficient of Variation (CV’s) and Confidence Limit of important key indicators was also worked out to determine the reliability of the estimates.

Chi-square test for independence
Pearson’s $\chi^2$ statistic is used for measuring the association between variables in a contingency table and plays an important role in the construction of statistical tests, the $\chi^2$ statistic or $\chi^2$ coefficient for a $k \times l$ contingency table is given as

$$\chi^2 = \sum \frac{(\text{Observed Value} - \text{Expected Value})^2}{\text{Expected Value}}$$  \hspace{1cm} (1)

The idea behind the $\chi^2$ coefficient is that when the relationship between two variables is stronger, then the deviations between observed and expected frequencies are expected to be higher (because the expected frequencies are calculated assuming independence) and this indicates a stronger relationship between the two variables. If observed and expected frequencies are identical or similar, then this is an indication that the association between the two variables is weak and the variables may even be independent.

Independent Sample T-test
The independent-samples t-test evaluates the difference between the means of two independent or unrelated groups. That is, we evaluate whether the means for two independent groups are significantly different from each other. The independent-samples t-test is commonly referred to as a between-groups design, and can also be used to analyze control and experimental group. With an independent-samples t-test, each case must have scores on two variables, the grouping (independent) variable, and the test (dependent) variable.
Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

The Formula for ANOVA is

\[ F = \frac{\text{MST}}{\text{MSE}} \]  \hspace{1cm} (2)

Where in equation (2) F is ANOVA coefficient which is the ratio of MST and MSE

MST is Mean sum of square due to treatment.

MSE is Mean sum of square due to error.

The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

Regression Analysis

Regression analysis is a statistical method that helps us to analyze and understand the relationship between two or more variables of interest. The practice that is adapted to perform regression analysis helps to understand which factors are essential, which factors can be ignored and how they are influencing each other.

For the regression analysis to be a successful method, we know the following terms:

Dependent Variable: This is the variable that we are trying to understand or forecast.

Independent Variable: These are factors that influence the analysis or target variable and provide us with information regarding the relationship of the variables with the target variable.

In regression, we normally have one dependent variable and one or more independent variables. Here we try to “regress” the value of dependent variable “Y” with the help of the independent variables. In other words, we are trying to understand, how does the value of ‘Y’ change w.r.t change in ‘X’.

Linear Regression

The simplest of all regression types is Linear Regression where it tries to establish relationships between Independent and Dependent variables. The Dependent variable considered here is always a continuous variable.

\[ Y = a + bx \]  \hspace{1cm} (3)
In equation (3) ‘Y’ is our dependent variable, which is a continuous numerical and we are trying to understand how does ‘Y’ change with ‘X’.

**Simple Linear Regression model**

Below is the statistical or mathematical model for simple linear regression model.

\[
Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (4)
\]

Where in equation (4) is a simple linear regression model for modeling the \(i^{th}\) response in terms of the explanatory variables \(X\)’s and an unobservable error term. Further, we explain the terms in (3) as below:

- \(Y_i\) is Dependent variable
- \(\beta_0\) is Intercept
- \(\beta_1\) is Slope Coefficient
- \(X_i\) is Independent Variable
- \(\varepsilon_i\) is Random Error Term

**Multiple Linear Regression**

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable.

**Multiple Linear Regression model**

\[
Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_p x_{pi} + \epsilon \quad (5)
\]

Where in (3.5) is a multiple linear regression model for modeling the \(i^{th}\) response in terms of the explanatory variables \(X\)’s and an unobservable error term. Further, we explain the terms in (5) as below:

- \(Y_i\) is the dependent variable for \(i^{th}\) individual
- \(X_i\) is explanatory variables
- \(\beta_0\) is \(y\)-intercept (constant term)
- \(\beta_p\) is slope coefficients for each explanatory variable
- \(\epsilon\) is the model’s error term (also known as the residuals)
Analysis

Analysis section present the details findings of our research as a result of different statistical procedures used on the data of PSLM 2018-2019 survey. According to our study the average household size in urban region was 6.28 and in rural region it was 6.54.

Consumption Quintiles

Consumption quintiles are used to differentiate the population according to their benefit: poorest households are grouped together into the 1st quintile, those with higher consumption into the 2nd quintile, and so on. Five quintiles rank the population from the poorest 20 percent to the richest 20 percent. The main goal of quintile is to analyze how social and economic indicators change in relation to people’s welfare. Estimates by quintiles describe distributional differences, thus characterizing an important tool of analysis. Quintiles are calculated for the four provinces together (Punjab, Sindh, Khyber Pakhtunkhwa and Balochistan) so that the first quintile contains households from all provinces with the same welfare. However, if one province is relatively richer than others its population will not be evenly distributed in each quintile, but mostly concentrated in the higher quintiles. In fact, only at the overall level each quintile contains 20 percent of the population, but in urban areas, where people usually are richer, upper quintiles contain higher population percentages, and the opposite is true in rural areas.

Consumption Quintiles Computation

Consumption expenditure is used as a proxy to assess people’s welfare. Expenditure is calculated at the household level but it is adjusted by household size and its composition. Below table shows the ranges of per capita consumption expenditure for consumption quintile.

| Quintiles | 1st | 2nd | 3rd | 4th | 5th |
|-----------|-----|-----|-----|-----|-----|
| Ranges of Per Capita Consumption Expenditure | 0 Up to Rs.3271 | Rs. 3272 To Rs.4207 | Rs. 4208 To Rs.5402 | Rs.5403 To Rs.7508 | Rs.7509 and above |

Quintiles first started with consumption from 0 to Rs.3271 up to quintile 5th which is 7509 and above. Which means that a person who 0 to Rs.3271 is added in quintiles first and so on. The key rule of quintile is to examine in what way social and economic indicators change in relation to people’s welfare. Below is the area wise division by these quintiles.

Table 2: Average Household size by Quintiles and Region 2018-2019
Statistical Analysis of Yearly Consumption Expenditure of the Household On Durable, Non-Durable Goods And Services in Pakistan

| Region | QUINTILES | 1<sup>st</sup> | 2<sup>nd</sup> | 3<sup>rd</sup> | 4<sup>th</sup> | 5<sup>th</sup> |
|--------|-----------|----------------|----------------|----------------|----------------|----------------|
| Urban  |           | 8.22           | 7.29           | 6.49           | 5.71           | 4.82           |
| Rural  |           | 7.53           | 6.26           | 5.55           | 5.11           | 4.38           |

Table 2 shows the Average Household size by Quintiles and Region. The data of PSLM survey 2018-2019 shows that the average household size is decreasing from quintile 1<sup>st</sup> to quintile 5<sup>th</sup> which means that the richest household is smaller compare to the poor household. Also, Household size is larger in Urban Areas than in Rural Areas.

Table 1: Average Household Size by Provinces and Region 2018-2019

| Province            | Khyber Pakhtunkhwa | 7.54 |
|---------------------|--------------------|------|
|                     | Punjab             | 5.80 |
|                     | Sindh              | 6.28 |
|                     | Balochistan        | 8.07 |

Table 3 shows the average household size division across regions and provinces. We see that the average Household size in Balochistan is the largest among all the provinces. Similarly, province Punjab has the lowest household size.

Table 2: Average Monthly Household Consumption Expenditure by Quintiles & Region

| Region | Household Consumption Expenditure Quintiles | 1<sup>st</sup> | 2<sup>nd</sup> | 3<sup>rd</sup> | 4<sup>th</sup> | 5<sup>th</sup> |
|--------|--------------------------------------------|----------------|----------------|----------------|----------------|----------------|
| Urban  |                                            | 22118          | 28233          | 31576          | 36855          | 61760          |
| Rural  |                                            | 18871          | 24123          | 27374          | 33081          | 50282          |

Table 4 shows the average monthly household consumption, quintiles, and region-wise. consumption in the 1st quintile is the lowest in both urban and rural regions. While in 5th quintiles the consumption expenditures are the highest one. Similarly, it can be seen from the above table that the consumption in rural regions is less than the urban in all the quintiles.

Table 3: Per Capita Monthly Consumption Expenditure by Quintiles and Region
Table 5 shows the average monthly per capita consumption, quintiles, and region-wise. consumption in the 1st quintile is the lowest in both urban and rural regions. While in the 5th quintiles the per capita consumption expenditures are the highest one. Similarly, from the above table, it can be seen that the per capita consumption in rural regions is less than the urban in all the quintiles.

**Graphical Analysis**

Graphical representation of the data is another way to understand the ups and downs of the data. They are used for presentation and understanding. In this section, we are going to analyze some of the main categories of consumption expenditure through graphical analysis.

![Percentage Consumption expenditure on health province by Region](image)

**Figure 1:** Percentage Consumption expenditure on health in Provinces by Region
Figure 1 shows the percentage consumption expenditure on health province by region. From figure 1 we can see that the consumption expenditure on health is highest in both rural and urban areas of Khyber Pakhtunkhwa. On the other hand, both rural and urban areas of Balochistan have the lowest consumption expenditure on health.

Figure 2 shows the percentage consumption expenditure on education province-wise by region. From the figure, we can see that urban and rural both area of Punjab has the highest consumption expenditure on education. However, Balochistan on the other hand has the lowest consumption expenditure on education.

Chi-square Test of independence

The Chi-Square Test of independence determines to determine whether there is an association between categorical variables (i.e., whether the variables are independent or related). Hereby Chi-Square test we wish to find a statistically significant association between Consumption Quintiles and different demographic variables.

Table 4: Association between Consumption Expenditure Quintiles and Demographic variables.

| Variables | Household Consumption Expenditure Quintiles |
|-----------|--------------------------------------------|
|           | Chi-Square | D.F | P-value |
| Region    | 3994.85    | 4   | 0.00    |
| Gender    | 214.42     | 4   | 0.00    |
In this (Table 6) we have used a Chi-square test of independence to check the statistical association between the Household Consumption Quintiles and all the demographic variables. Here we have calculated the chi-square statistic, D.F, and the p-values. From the above table, it can be seen that all the P-values are statistically significant indicates that there is a statistical association between the Household Consumption Quintiles and the demographic variables i.e. (Region, Province, Gender, Language). Thus, we can say that Consumption Quintiles is not equally distributed through Language, Province, Gender, and Region wise. Or in other words, Consumption Quintiles are changing across Region, Province, Gender, and Language wise.

**Independent Sample T-test**

The independent-samples t-test evaluates the difference between the means of two independent or unrelated groups. That is, we evaluate whether the means for two independent groups are significantly different from each other. Here we wish to compare the mean consumptions per household (monthly) region and gender-wise. That is whether the mean consumption in rural and urban regions is statistically significant or not. Similarly, the mean consumption for males and females is statistically significant.

| Variables | Monthly Consumption Expenditures per household | T-test | D.F | P-value |
|-----------|-----------------------------------------------|--------|-----|---------|
| Region    |                                               | -43.44 | 12030 | 0.00    |
| Gender    |                                               | -2.38  | 24806| 0.017   |

In Table 7 we have calculated the t-statistic, D.F, and P-values. The p-values for both region and gender are statistically significant. This shows that the means differ both region and gender-wise. OR in other words, the monthly mean consumption per household for both urban and rural regions are different. Also, the monthly mean consumption per household for both males and females is different. But for 1% it is insignificant.
Statistical Analysis of Yearly Consumption Expenditure of the Household On Durable, Non-Durable Goods And Services in Pakistan

| Variables | Monthly per capita consumption expenditure |
|-----------|--------------------------------------------|
|           | T-test | D.F  | P-value |
| Region    | -44.05 | 11220 | 0.00    |
| Gender    | 10.28  | 24806 | 0.00    |

Also, from Table 8 we can observe that p-values for both region and gender are statistically significant. This shows that the means are differ both region and gender-wise. OR in other words, the monthly mean per capita consumption for both urban and rural regions are different. Also, the monthly mean per capita consumption for both males and females is different.

**Analysis of Variance**

The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

Here we used ANOVA to check whether the mean consumption per household (monthly) is statistically differ province and language wise or not.

**Table 7:** ANOVA test for comparison of mean of province, language with monthly consumption expenditures per household

| Variables | Monthly consumption Expenditures per household |
|-----------|-----------------------------------------------|
|           | F-test | Total D. F | P-value |
| Province  | 38.67  | 24807      | 0.00    |
| Language  | 187.70 | 24802      | 0.00    |

An ANOVA was performed to compare the mean of monthly consumption expenditures per household in Province and Language wise. In Table 9 we have our results as follows. For provinces, monthly consumption expenditures are statistically significant with F=38.67 and D. F = 24807. This means that the mean consumption per household for all the provinces is different. Similarly, for languages, the monthly consumption expenditures per household are also highly statistically significant with F=187.70 & D. F= 24802. This indicates that the consumption expenditures per household are different for different languages.
Table 10: ANOVA test for comparison of mean of province, language and consumption expenditure

| Variables | Monthly per capita consumption expenditure |
|-----------|---------------------------------------------|
|           | F-test | D.F | P-value |
| Province  | 90.02   | 24807 | 0.00    |
| Language  | 220.42  | 24802 | 0.00    |

Table 10 shows the results of the ANOVA test for Province, Language, and Monthly Per Capita Consumption Expenditure. For provinces, the per capita monthly consumption expenditures are statistically significant with $F=90.02$ and $D. F = 24807$. This means that the mean per capita consumption for all the provinces is not the same. Similarly, for languages, the per capita monthly consumption expenditures are also statistically significant with $F=220.42$ & $D. F= 24802$. This indicates that the consumption expenditures per household are different for different languages.

**Consumption expenditure of durable and non-durable**

Goods and services are broadly divided into two categories durable and non-durable. Consumption expenditure on both durable and non-durable varies across regions and provinces shown as follows. We will perform analysis on both combined and separately.

Table 8: Yearly average Consumption expenditure on durable & non-durable

| Region  | Province                | Average Consumption on Non-Durable | Average Consumption on Durable |
|---------|-------------------------|-----------------------------------|--------------------------------|
| Rural   | Khyber Pakhtunkhwa      | 325307                            | 9884                           |
|         | Punjab                  | 270389                            | 5702                           |
|         | Sindh                   | 249017                            | 4131                           |
|         | Balochistan             | 304117                            | 6030                           |
| Urban   | Khyber Pakhtunkhwa      | 474359                            | 12803                          |
|         | Punjab                  | 453236                            | 8213                           |
|         | Sindh                   | 432161                            | 7174                           |
Table 11 shows the yearly consumption expenditure on durable and non-durable goods. From the above table, it can be seen that both urban and rural regions of Khyber Pakhtunkhwa spend the highest on both durable and non-durable goods. In rural regions, Sindh has the lowest consumption expenditures. In urban areas, Sindh and Balochistan are with lowest figures in consumption expenditures on different durable and non-durable items.

**Statistical test on the consumption expenditure of non-durable**

Non-durable or soft goods are more likely to be purchased daily so they are more consumed than the durable. We are going to test whether the monthly consumption expenditure on non-durable varies region and gender-wise. Also, another test will be used to check if the monthly consumption expenditure is different province-wise and language-wise.

**Table 9: T-test for monthly consumption expenditure on non-durable and region, gender**

| Variables | Monthly consumption expenditure on non-durable |
|-----------|-----------------------------------------------|
|           | T-test | D.F | P-value |
| Region    | -44.64 | 11918 | 0.00  |
| Gender    | -2.571 | 24806 | 0.01  |

In Table 12 we have calculated the t-statistic, D.F, and P-values. The p-values for both region and gender are statistically significant. This shows that the average monthly per household consumption expenditures on non-durable goods & services are differ both region and gender-wise.

**Table 10: ANOVA Test for monthly consumption expenditure on non-durable and province, language**

| Variables | Monthly consumption expenditure on non-durable |
|-----------|-----------------------------------------------|
|           | F-test | D.F | P-value |
| Province  | 34.27  | 24807 | 0.00  |
| Language  | 192.89 | 24802 | 0.00  |
In Table 13 we have our results of the ANOVA test as follows. For provinces, monthly consumption expenditures are statistically significant with $F=34.27$ and $D. F = 24807$. This means that the mean consumption of non-durable goods & services per household for all the provinces is different. Similarly, for languages, the monthly consumption expenditures per household on non-durable is also highly statistically significant with $F=192.89$ & $D. F= 24802$. This indicates that the consumption expenditures per household on non-durable goods & services are different for different languages.

**Statistical test on the consumption expenditure of durable**

Durable goods are hard good are used several times and has life expectancy more than a year. Thus, they are not purchased more often in a year. A t-test and ANOVA test will be applied to check if the monthly consumption expenditure on durable goods varies across the region, gender, Province, and language-wise.

| Variables | Monthly consumption expenditure on durable | | |
|---|---|---|
| | T-test | D.F | P-value |
| Region | -5.571 | 17150 | 0.00 |
| Gender | -0.583 | 24807 | 0.56 |

In Table 14 we have calculated the t-statistic, D.F, and P-values. The p-values for the region are statistically significant. This indicates that the average consumption expenditures on durable goods & services are different across regions. But for gender, the p-value is insignificant shows that the average consumption expenditures on durable goods & services both for males and females are the same.

| Variables | Monthly consumption expenditure on durable | | |
|---|---|---|
| | F-test | D.F | P-value |
| Province | 22.29 | 24808 | 0.00 |
| Language | 13.69 | 24803 | 0.00 |
In Table 15 we have performed an ANOVA & our results are as follows. For provinces, monthly consumption expenditures are statistically significant with $F=22.29$ and $D. F = 24808$. This means that the mean consumption of durable goods & services per household for all the provinces is different. Similarly, for languages, the monthly consumption expenditures per household on durable is also highly statistically significant with $F=13.69$ & $D. F= 24803$. This indicates that the consumption expenditures per household on durable goods & services changing across different languages.

**Consumption Quintiles and various diseases**

Consumption quintiles are used to classify the population affording to their welfare: poorest households are grouped together into the 1st quintile, those with higher consumption into the 2nd quintile, and so on. Five quintiles rank the population from the poorest 20 percent to the richest 20 percent. In this section we are going to use these quintiles and showed how many people are infected with different diseases. Table 4.27 shows the number of people infected with different diseases and consumption quintiles.

**Table 13:** The number of infected people with different diseases and consumption quintiles

| Diseases          | Q1  | Q2  | Q3  | Q4  | Q5  |
|-------------------|-----|-----|-----|-----|-----|
| Malaria           | Yes | 399 | 182 | 152 | 109 | 63  |
|                   | No  | 8629| 4426| 3831| 3569| 3447|
| Hepatitis         | Yes | 281 | 187 | 159 | 202 | 267 |
|                   | No  | 8747| 4421| 3824| 3476| 3243|
| Tuberculosis (TB) | Yes | 119 | 46  | 41  | 37  | 28  |
|                   | No  | 8909| 4561| 3942| 3641| 3482|

In Table 16 we have shown the number of peoples who are infected with disease in each quintile. Number of infected peoples from disease is decreasing from 1st to 5th quintiles indicates that the peoples belonging to poor class are more infected with the disease as compare to those people who belongs to rich class.

**Chi-Square test for Consumption quintiles and different diseases**

The Chi-Square Test of independence determines to determine whether there is an association between categorical variables (i.e., whether the variables are independent or related). Hereby Chi-Square test we wish to find a statistically significant association between Consumption Quintiles and different diseases.
Table 14: Chi-Square test for consumption quintiles & different diseases

| Diseases            | Per capita consumption quintiles |  |
|---------------------|----------------------------------|--|
|                     | Chi-Square | D.F | P-value |
| Malaria             | 55.997     | 4   | 0.00    |
| Hepatitis           | 134.12     | 4   | 0.00    |
| Tuberculosis (TB)   | 12.23      | 8   | 0.141   |

Table 17 shows the results of a chi-square test of independence between per capita consumption quintiles and various diseases. Chi-square test for malaria and hepatitis is statistically significant indicates that the no of people infected with diseases are differ quintiles wise. But for Tuberculosis (TB) it is insignificant because the minimum expected cell count in this sub table is less than one. Chi-square result may be invalid for tuberculosis (TB). Thus, Fisher’s exact test may be used in such circumstances.

Table 15: Correlation between total monthly consumption expenditures & no. of days that a person had malaria

| Variables                         | Total Monthly consumption | How many days did a person had malaria? |
|-----------------------------------|---------------------------|----------------------------------------|
|                                   | r                         | 1                                      |
|                                   | P-value                   | 0.006                                  |
|                                   | N                         | 24808                                  |
| Total Monthly consumption         |                           |                                        |
| How many days did a person had malaria? | r                      | -0.91                                  |
|                                    | P-value                   | 0.006                                  |
|                                    | N                         | 905                                    |

In Table 18 we have calculated the correlation coefficient between the variable total monthly consumptions and the number of days that a person had malaria. The correlation coefficient(r) is -0.91 showing the strong negative relation between total monthly consumption expenditures and the number of days that a person had malaria. Also, the p-value is statistically is significant.
Women in decision making

In this section we are trying to determine how autonomous are women in Pakistan. Using data from PSLM survey we have formulated new variables by calculating the data from question like“Who in your household decides whether you can start or continue to get education? “. If any of the answer involved women the value of variable was changed into Yes otherwise No.

Descriptive Statistics WIDM & income consumption variables

WIDM is a variable computed from other variables in order to check if women in their household are allowed to take decisions or not. Following are the descriptive statistics of Women decision making in education and employment against consumption and income variables.

| WIDM in Education | WIDM in Employment |
|-------------------|-------------------|
| **Women in decision making** | **Average Monthly Consumption** | **Average Monthly Per Capita Consumption** | **Average Monthly Income** | **Average Monthly Per Capita Income** |
| Yes | 31463 | 5519 | 63291 | 10796 |
| No | 27402 | 4762 | 53502 | 9272 |
| Yes | 30229 | 5277 | 60577 | 10306 |
| No | 27933 | 4868 | 64611 | 9489 |

From the Table 19 we can clearly observe that women in household with high mean income & consumption are more likely to take decisions in their households i.e., in education and employment. OR in other words women in rich households are more autonomous than women in poor household.

Regression Analysis

Here we are performing a regression analysis to check the effect of different variables related to women decision making in purchase of food items, clothing and footwear, medical treatment and recreation & travel on monthly consumptions and income.

Assumption of regression Analysis

In order for our regression to produce unbiased estimates and for our statistical significance test to be valid, we must make sure our sample meets the Linear Regression assumptions.
**Linear Relationship**

In our study all the independent variables are dummy variables with two possible outcomes 0 and 1. Dummy variables meet the assumption of linearity by definition, because they create two data points, and two points define a straight line. There is no such thing as a non-linear relationship for a single variable with only two values.

**Normality**

Since the data used in this study is taken from PSLM survey which was collected from 24809 households using two stage stratified sampling. We can clearly assume that data used in this study is normally distributed.

**No or Little Multicollinearity**

To check if there is Multicollinearity between the independent variables, we used SPSS and run for collinearity diagnostics which shows VIF or Tolerance. VIF was less than 4.3 for every model so we can clearly say that we do not have the problem of multicollinearity.

**No Autocorrelation in errors**

To check for autocorrelation, we used Durbin-Watson test. The value of Durbin Watson test for model 4.31, 4.32, 4.33 and 4.34 was 1.643, 1.365, 1.6 and 1.6 respectively which was closed to 2 but we can say that we do not have the autocorrelation problem.

**Homoscedasticity**

To check for heteroscedasticity if there is any in our data there was not a direct method in SPSS software so we import our data in R studio and fit a linear regression model in R studio and then test the model with Breusch-Pagan test for heteroscedasticity. The test statistic is 187.03 and the corresponding p-value is 2.2e-16. Since the p-value is less than 0.05, we accept the null hypothesis. We have sufficient evidence to say that heteroscedasticity is present in the regression model. The result of Breusch-Pagan test was highly significant for consumption models where as it was not highly significant for income.

**Table 17:** Regression analysis for monthly consumption per household & other variables

| Model                      | Monthly consumption expenditure per household |
|----------------------------|-----------------------------------------------|
|                            | $\hat{\beta}$ | S. E ($\hat{\beta}$) | t-statistic | P-value |
| Constant                   | 10251.614     | 495.64                | 20.683      | .000    |
| Region                     | 14133.890     | 290.88                | 48.580      | .000    |
| WIDM in food items         | -2379.36      | 379.05                | -6.277      | .000    |
| WIDM clothing and footwear | 2883.99       | 495.29                | 5.823       | .000    |
Table 20: Linear regression for the effect of region & WIDM variables on per capita monthly consumption expenditure

| Model                          | Per capita monthly consumption expenditure |  |  |  |
|--------------------------------|--------------------------------------------|--|--|--|
|                                | \( \hat{\beta} \) | S. E (\( \hat{\beta} \)) | t-statistic | P-value |
| Constant                       | 831.003 | 84.92 | 9.78 | .000  |
| Region                         | 2609.89 | 49.84 | 52.36 | .000  |
| WIDM in food items             | 12.302 | 64.94 | .189 | .850  |
| WIDM clothing & footwear       | 293.14 | 84.86 | 3.45 | .001  |
| WIDM in medical treatment      | -105.54 | 109.34 | -0.974 | .330  |
| WIDM in recreation & travel    | 465.915 | 101.02 | 4.61 | .000  |

From the above Table 21 of a regression analysis, we want to check the impact of region and other WIDM variables on monthly per capita consumption expenditures. The bolded p-values are statistically significant indicating that the variable women decision (in clothing, footwear, recreation and travel) effect the per capita monthly consumption expenditure. On the other hand, the remaining variables are insignificant showing that these variables do not affect the per capita consumption expenditure.
Table 22 summarize the results of a linear regression analysis. The bolded p-values against various variables are statistically significant indicating that these variables are affecting the monthly income per household while other variables are insignificant indicating that these variables do not affect monthly income per household.

Table 20: Linear regression for the effect of region & WIDM variables on per capita monthly income

Table 22 shows the results of a linear regression analysis. From above results we can see that the two variables corresponding to the bolded p-values are significant which mean that these variables are affecting monthly per capita income while other variables are insignificant indicating that they are not affecting monthly per capita income.
Summary of the Regression models

The overall summary of the regression model represent weather a model is good fit or not. Below is the summary of models from table 4.31 to 4.34. and the ANOVA test is summarized in Table 23.

| Models | R   | R Square | S. E     | F-Test | P-value |
|--------|-----|----------|----------|--------|---------|
| 1      | 0.303 | 0.092    | 21107.15 | 495.29 | 0.00    |
| 2      | 0.335 | 0.112    | 3616.59  | 583.77 | 0.00    |
| 3      | 0.214 | 0.046    | 64445.87 | 223.44 | 0.00    |
| 4      | 0.233 | 0.054    | 11037.53 | 265.05 | 0.00    |

Table 1: ANOVA test results of regression models.

| Models | Sum of squares | D.F | Mean Square | F-Test | P-value |
|--------|----------------|-----|-------------|--------|---------|
| 1      | 1.1033E12      | 5   | 220658465892| 495.29 | 0.00    |
| 2      | 38178218992.   | 5   | 7635643798  | 583.77 | 0.00    |
| 3      | 4.6402E12      | 5   | 928033690233| 223.44 | 0.00    |
| 4      | 161456065646   | 5   | 32291213129 | 265.05 | 0.00    |

Conclusion

Consumption is the greatest part of a household’s expenses and for that matter, the nation’s expenses as well. It accounts for over half of GDP. It is an essential part of the well-being of households. Therefore, Consumption is divided into three buckets: services, non-durable goods, and durable goods. The main aim of this research was to analyze the consumption of durable goods, nondurable goods, and services to determine the consumption of each group of income on different durables and non-durables. The data was collected from the Pakistan Social Living Measurement survey 2018-2019. The sample size of the survey consisted of 28500 households. The data was analyzed with the help of SPSS, essential Chi-square, independent sample t-test,
and ANOVA tests were applied and the results are summarized in form of figures and tables. From the figure, we determine that Khyber Pakhtunkhwa has the highest consumption on health in both urban and rural areas while Balochistan has the lowest consumption expenditure on health. The figure shows that Punjab has the highest consumption expenditure on education and Balochistan has the lowest consumption expenditure on education. Chi-square tests to check associations between household consumption quintiles and different demographic variables i.e. (Region, Province, Gender, Language) were statistically insignificant indicating that there is an association between household consumption quintiles and demographics variables. An independent sample t-test was performed to check the difference between the mean consumption expenditure per household and per capita, region and gender-wise. This shows that both the mean consumption per household and mean consumption expenditure per capita vary both region and gender-wise. An ANOVA was performed to check the difference of mean monthly consumption expenditure per household and consumption expenditure per capita was differ both province and language-wise. Results show that Khyber Pakhtunkhwa had the highest yearly consumption expenditure on both durable and non-durable in their urban and rural areas. Also, rural areas of Sindh had the lowest consumption expenditure while urban areas of Sindh and Balochistan had the lowest consumption on both durable and non-durable. A T-test between the monthly consumption of non-durable goods was significant with both region and gender. An ANOVA test between monthly consumption expenditure and province/language was highly significant with both provinces and languages. The average consumption expenditure on durable goods and services is different across the region so the t-test results showed but consumption expenditure on durables goods is the same for the gender. The monthly consumption per household on durable goods and services is different throughout the provinces and it is also different for different languages. In the computation of the yearly consumption expenditure durable goods, the most expensive category was “Transport and traveling vehicles, etc.” with the mean consumption expenditure of Rs. 87459 on the other hand in the categories of non-durable goods the most expensive category was “Housing, Water Electricity, Gas and other Fuels” with the mean consumption expenditure of Rs. 83869 per year. The descriptive results of quintiles and people infected with different diseases showed that the number of infected people is decreasing as the consumption expenditure is increasing. Also, the chi-square test between malaria and hepatitis is statistically significant indicating that the number of infected people is of different consumption quintiles wise. Monthly consumption expenditure and the number of days that a person had malaria were strongly negatively correlated with coefficient -0.91. According to the summary of a descriptive table women in the rich household are more likely to take decisions related to education and employment. Also, the regression analysis between monthly consumption expenditure per household and different WIDM variables was statistically significant except for WIDM in medical it was insignificant. Women's decision-making in clothing, footwear, recreation, and travel effects the per capita monthly consumption however when we try to find out the impact of WIDM variables on monthly income per household it was significant for region and women to decision making in clothing and footwear. In the end, a
regression analysis of WIDM with per capita income shows that per capita monthly income is affected by region and when women are making the decision in travel.

Acknowledgement

We are thankful for the constructive cooperation and expert opinion of the reviewer and academic editor, which improves the quality of the manuscript.

References

Engel, C. and Wang, J. (2011). International trade in durable goods: Understanding volatility, cyclicity, and elasticities. Journal of International Economics, 83(1), 37–52. https://doi.org/10.1016/j.jinteco.2010.08.007

Erceg, C. J. and Levin, A. T. (2003). Optimal Monetary Policy with Durable and Non-durable Goods. SSRN Electronic Journal, 1–100. https://doi.org/10.2139/ssrn.366585

Estelami, H. and De Maeyer, P. (2004). Product category determinants of price knowledge for durable consumer goods. Journal of Retailing, 80(2), 129–137. https://doi.org/10.1016/j.jretai.2004.04.003

Fernández-Villaverde, J., and Krueger, D. (2011). CONSUMPTION AND SAVING OVER THE LIFE CYCLE: HOW IMPORTANT ARE CONSUMER DURABLES? Macroeconomic Dynamics, 15(5), 725–770. https://doi.org/10.1017/s1365100510000180

Ganti, A. (2021, January 27). Degrees of Freedom Definition. Investopedia. https://www.investopedia.com/terms/d/degrees-of-freedom.asp

Great Learning Team. (2020, December 21). Linear Regression in Machine Learning Definition, Advantage & uses. GreatLearning Blog: Free Resources What Matters to Shape Your Career! https://www.mygreatlearning.com/blog/linear-regression-in-machine-learning/

Heumann, C. and Schomaker, M. (2017). Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R (1st ed. 2016 ed.). Springer.

Kenton, W. (2021, February 8). How Analysis of Variance (ANOVA) Works. Investopedia. https://www.investopedia.com/terms/a/anova.asp

Mallick, S. K. and Mohsin, M. (2016). Macroeconomic Effects of Inflationary Shocks with Durable and Non-Durable Consumption. Open Economies Review, 27(5), 895–921. https://doi.org/10.1007/s11079-016-9405-0

S, S. (2018, July 26). Difference Between Goods and Services (with Comparison Chart). Key Differences. https://keydifferences.com/difference-between-goods-and-services.html

Beers, B. (2020, October 19). What P-Value Tells Us. Investopedia. https://www.investopedia.com/terms/p/p-value.asp

System of National Accounts 2008. (2008, January 1). Durable and Non-durable goods. https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf