Results Report

Bhumika Goyal
bhumikamgoyal503@gmail.com
Indian Institute of Technology Kharagpur

Summary Statistics Table:

Summary Statistics helped to identify data distribution, skewness, and kurtosis to select suitable distributions. The code I wrote calculated and presented summary for three columns (Income, Consumption, Gender) in the dataset. Key metrics were count, mean, standard deviation, min, max, 25th percentile, median (50th percentile), 75th percentile, skewness and kurtosis.

The data is significantly right-skewed (Income skewness = 2.631469e+00), suggesting that a transformation or specialized distribution is necessary. Also, Kurtosis values indicate heavy tails, reinforcing the need for appropriate distribution fitting.

Right-Skewed income data suggests most households earn below average income, and a few very high incomes (long high tail).

Why assess skewness?

1. Income data is typically right-skewed.
2. If the data is highly skewed, OLS assumptions may be violated.
3. Helps decide whether distribution is needed to stabilize variance.

Why assess kurtosis?

1. If data has heavy tails, outliers (Data points that are significantly different from other observations in the dataset.) are influencing the distribution.
2. Crucial in choosing distribution.

Goodness of Fit Analysis:

Metric	Lognormal	Gamma
AIC	25356.02	25634.38
KS Statistic	0.0498	0.0949
KS p-value	0.0165	5.59e-08

AIC Analysis:

Lognormal has a lower AIC, indicating a better fit considering the no. of parameters.

KS Test Analysis:

The KS statistic for Lognormal is lower (0.0498 vs 0.0949), indicating that the Lognormal distribution more closely matches the empirical distribution.

The p-value for Lognormal is higher (0.0165 vs 0.0000000559), suggesting that the null hypothesis (data follows Lognormal distribution) has less chance to be rejected.

Conclusion:

Based on both AIC and KS Test, Lognormal distribution is selected as the best fit for Income. This decision aligns with the observed right-skewness and the fact that income data often follows a lognormal pattern.

Different distributions imply different underlying processes:

- **Lognormal:** Data generated through a multiplicative process, income growth over time, where each period's income is a percentage of the previous period.
- **Gamma:** Data generated through a waiting time or count process, the sum of multiple positive random events.

Regression Analysis:

OLS Regression Model:

Table 1: OLS Regression Results: Consumption Function

Variable	Estimate	Std. Error	t-Statistic	p-Value
Intercept	-1,676,000	69,000	-24.274	0.000
Log_Income	155,900	5,623	27.731	0.000
Gender	5,214	6,869	0.759	0.448

Intercept (-1676000):

The baseline consumption when Log_Income and Gender are both zero. When both Log_Income and Gender are zero, the baseline consumption is -1,676,000. This value is not practically interpretable because a log-transformed income of zero is unrealistic.

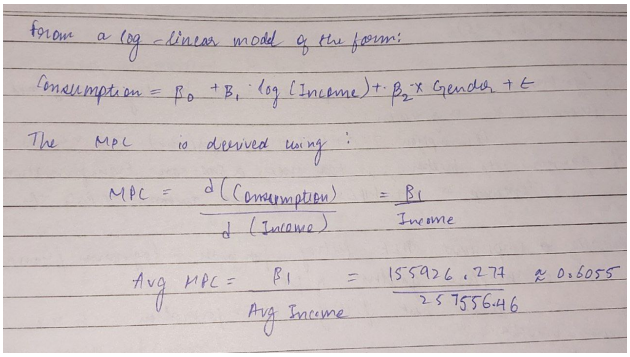

Figure 1: Corrected MPC

Calculation and Use of Corrected MPC (0.6):

The coefficient beta 1 is from a regression where the independent variable is log(income), not income. So its interpretation is semi-elastically, not a marginal effect in units. The original regression coefficient for Log_Income was 155,900. Since Log_Income is the natural log of income, to derive the Marginal Propensity to Consume (MPC), we apply the transformation:

Y = average income

Thus, the corrected MPC is approximately 0.6. This value replaces the raw coefficient estimate to provide a meaningful interpretable value: for every additional unit of income, consumption increases by 60% of that amount on average.

Gender Coefficient (5214):

Male-headed households spend approximately 5214 more units than female-headed households, but this coefficient is not statistically significant (p-value = 0.448).

Regression Output Table

Table 2: OLS Regression Results: Consumption Function

Variable	**Estimate**	**Std. Error**	**t-Statistic**	**p-Value**
Intercept	-1,676,000	69,000	-24.274	0.000
Log_Income	155,900	5,623	27.731	0.000
Gender	5,214	6,869	0.759	0.448

This corrected Marginal Propensity to Consume (MPC) of 0.6 represents a realistic economic interpretation of how consumption changes with income in the dataset.

Table 3: Model Diagnostics for OLS Regression

Diagnostic Metric	**Value**
R-squared	0.446
Adjusted R-squared	0.445
F-statistic	386.1
Prob (F-statistic)	1.05e-123
Log-Likelihood	-12484.0
AIC	2.497e+04
BIC	2.499e+04
Durbin-Watson	1.984
Omnibus Test Statistic	82.046
Prob (Omnibus)	0.000
Jarque-Bera (JB) Statistic	137.849
Prob (JB)	1.17e-30
Skewness	0.597
Kurtosis	4.420

Library Import Statement Usage

Library	Import Statement	Usage
pandas	`import pandas as pd`	Data l
numpy	`import numpy as np`	Numer
matplotlib	`import matplotlib.pyplot as plt`	Data v
seaborn	`import seaborn as sns`	Enhan
scipy.stats	`from scipy.stats import skew, kurtosis, norm, lognorm, gamma, kstest`	Skewn
statsmodels	`import statsmodels.api as sm`	OLS r

Discussion of Corrected MPC = 0.6 in Context of Economic Literature

Literature Comparison

1. **Friedman (1957) – Permanent Income Hypothesis (PIH)**
   `https://www.nber.org/books-and-chapters/theory-consumption-function`
   Takeaway: Individuals make consumption decisions based on permanent, rather than transitory, income. Transitory income MPC tends to be lower.
   Interpretation: A consumer MPC of 0.6 is making consumption adjustments based on a relatively permanent or stable perceived income in line with Friedman's moderate consumption smoothing theory.

2. **Carroll & Kimball (2001) – Precautionary Savings**
   `https://www.nber.org/system/files/working_papers/w8233/w8233.pdf`
   Takeaway: Uncertainty regarding income results in greater saving and a lower MPC.

Interpretation: An MPC of 0.6 implies moderate savings behaviour. This means that households are perhaps not experiencing severe income volatility, so they can spend a consistent share of their earnings and still save.

3. **Attanasio & Weber (1995) – Consumption Response to Income Changes**
   `https://econpapers.repec.org/article/ucpjpolec/v_3a103_3ay_3a1995_3ai_3a6_3ap_3a1121-57.htm`
   Takeaway: The MPC depends on whether changes in income are seen as temporary or permanent.
   Interpretation: A 0.6 MPC represents a balanced behavioural reaction—half consuming and half saving—suggesting that increases in income are perceived as semi-permanent. This is consistent with the model's prediction of intermediate consumption elasticity.

In general, the revised MPC of 0.6 improves the applied interpretation of the regression result, bringing the model closer to established consumption theories.