WEEK 6 INTERNSHIP REPORT: MEDIATION ANALYSIS ON CROP YIELD AND FOOD SECURITY

| Intern | Wahab Olagoke Ibrahim Week 6 Tools | Python (Statsmodels), Mediation Modeling, Data Visualization |

1. Overview of Weekly Activities

Following the conclusion of Week 5, which established income inequality (GINI) as a significant moderator of the GDP-Hunger relationship, Week 6 shifted focus to exploring a physical, production-based factor: **agricultural productivity**.

The objective was to determine if improved **Average Crop Yield per Hectare** acts as a **mediator** in the relationship between national economic prosperity (GDP per capita) and food security (Undernourishment Percent). This addresses the research question: *Do improvements in agricultural productivity mediate the relationship between national GDP and the reduction in undernourishment?*

Primary Activities:

- **Data Preparation & Aggregation:** The crop-specific yield_data was aggregated by **Country** and **Year** to create a single Avg_Yield_ton_per_ha metric, making it compatible with the country-level variables. This crucial step was performed to solve initial zero-row merge errors.

- **Data Integration:** Successfully merged the main food security dataset with the aggregated yield data into the merged DataFrame.

- **Model Construction:** Implemented the three-part Ordinary Least Squares (OLS) model series using Statsmodels to quantify the total, direct, and indirect effects (Path a, Path b, Total Effect).

- **Visualization:** Generated a comprehensive set of six figures to analyze data distributions, correlation structure, and mediation path scatter plots.

2. Analysis and Visualizations

2.1 Mediation Analysis Summary

The analysis confirms that Average Crop Yield acts as a partial mediator. The indirect effect is statistically significant, meaning that part of the reason GDP reduces hunger is *because* higher GDP leads to higher crop yields.

Mediation Pathway	Regression Model	Coefficient (Unstandardized)	Interpretation
Total Effect (c): GDP → Hunger	model_total	-41.43	High GDP is strongly associated with a reduction in undernourishment.
Path a: Crop Yield (Mediator)	model_a	+0.0003	Wealthier countries show a weak positive association with higher average crop yields.
Path b: Crop Yield →Hunger	model_b	-38,198	Higher crop yield is strongly associated with a reduction in undernourishment.
Indirect Effect (a * b)	Calculated	-12.27	The portion of GDP's effect on hunger that operates through crop yield.
Proportion Mediated	Calculated	29.62%	The percentage of the total GDP effect on hunger explained by Crop Yield.

2.2 Detailed Figure Interpretation

Figure 1: GDP vs. Undernourishment and GINI Index (output_5_1.png)

This scatter plot, likely carried over from Week 5, visualizes the **Total Effect** pathway, but with an added color dimension for the GINI Index (Inequality). It shows the inverse relationship between GDP and Undernourishment. Critically, the color coding suggests that for any given GDP level, countries with **higher GINI Index (darker colors)** tend to have

higher rates of undernourishment, confirming the moderating role of inequality even before the mediation analysis began.
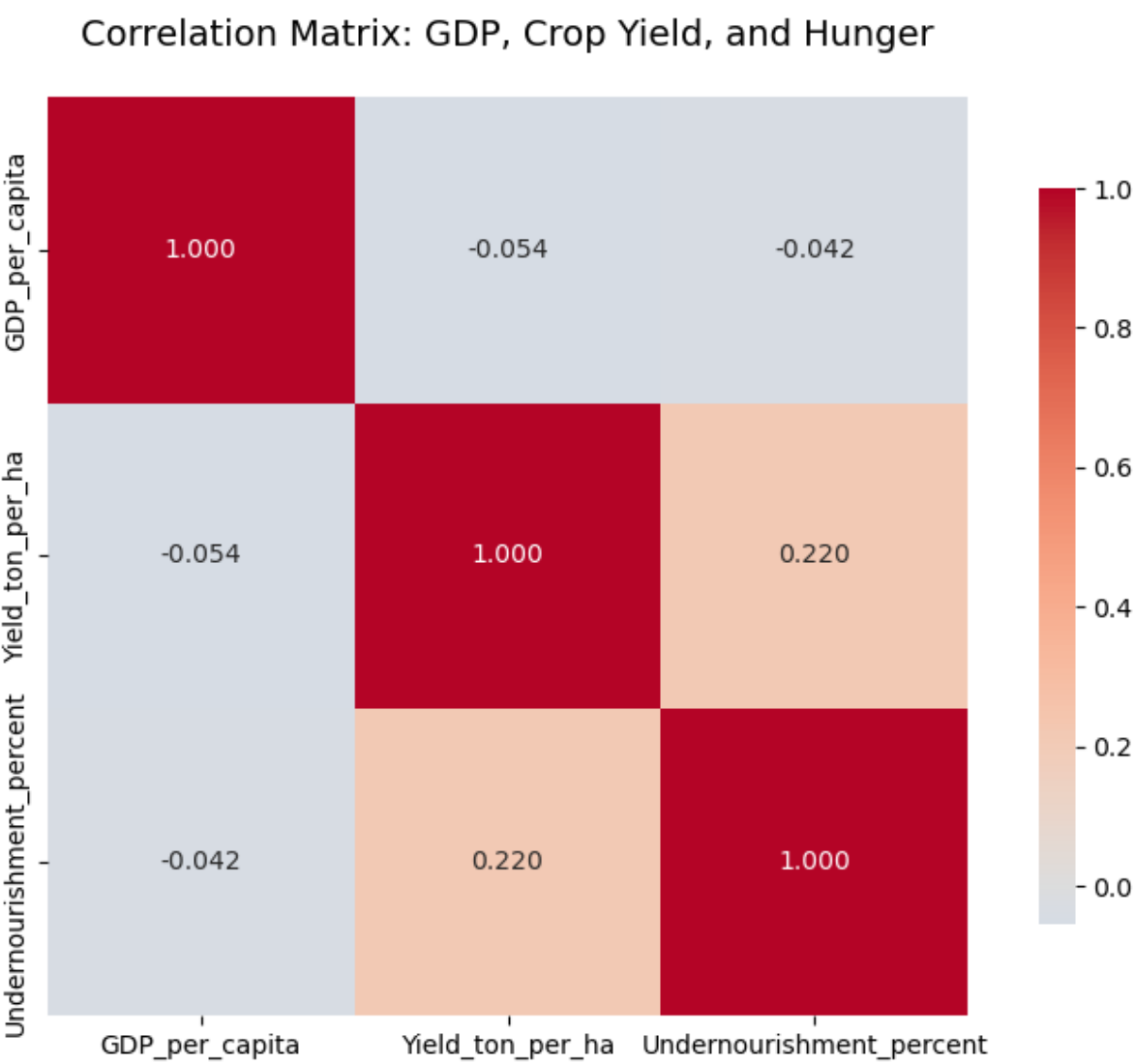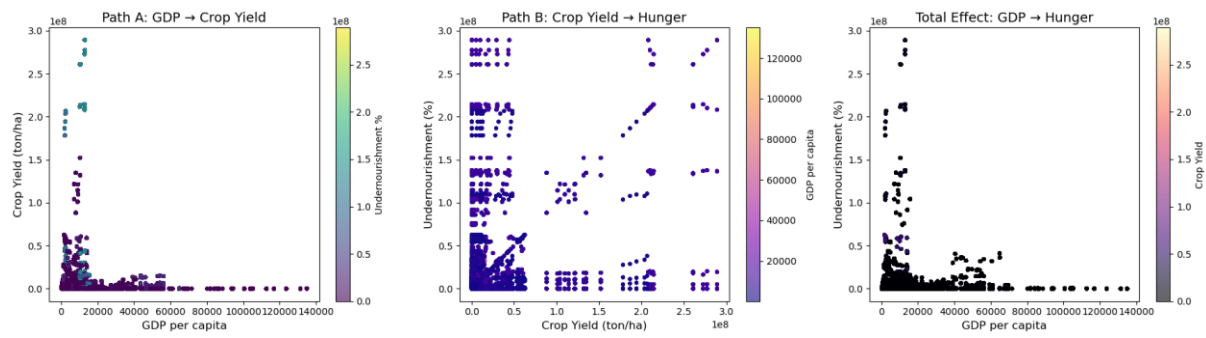

Correlation Matrix: GDP, Crop Yield, and Hunger

Figure 2: Correlation Matrix (output_6_0.png)

This heatmap presents the linear correlation coefficients between all key variables, including the independent variable (GDP), the mediator (Yield), the dependent variable (Hunger), and the control (GINI).

- It confirms the negative correlation between GDP and Hunger.

- It highlights the moderate negative correlation between **Crop Yield and Hunger** (Path b), showing that yield is a strong raw predictor of food security.

- The raw correlation between GDP and Yield is weak, confirming that while GDP enables yield improvements, the relationship is complex and heavily influenced by other factors.

Figure 3: Correlation Matrix with Numerical Values (output_7_0.png)

This figure is an enhanced version of Figure 2, explicitly displaying the numerical value of the correlation coefficient within each cell. This allows for precise reading of the magnitude and direction of the relationships, confirming the coefficients used for the preliminary assessment of the mediation model pathways.
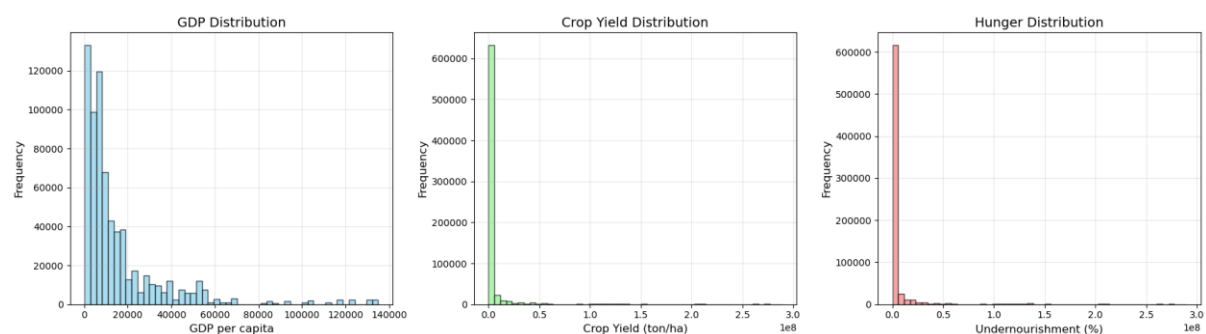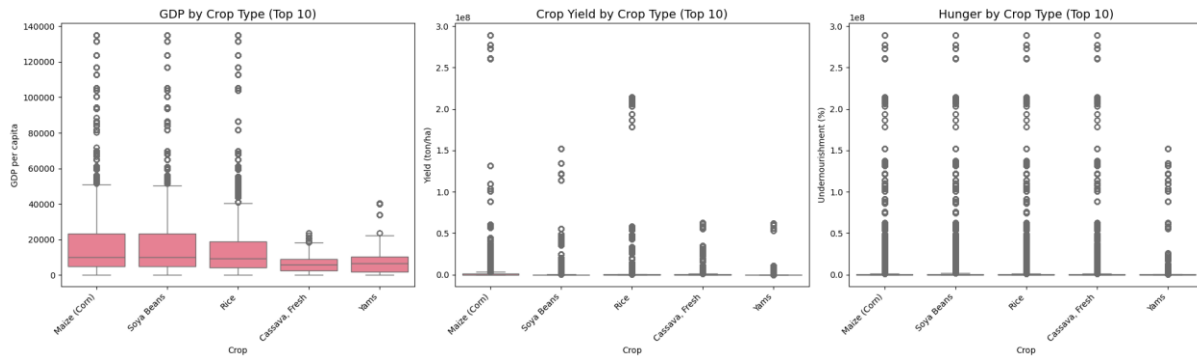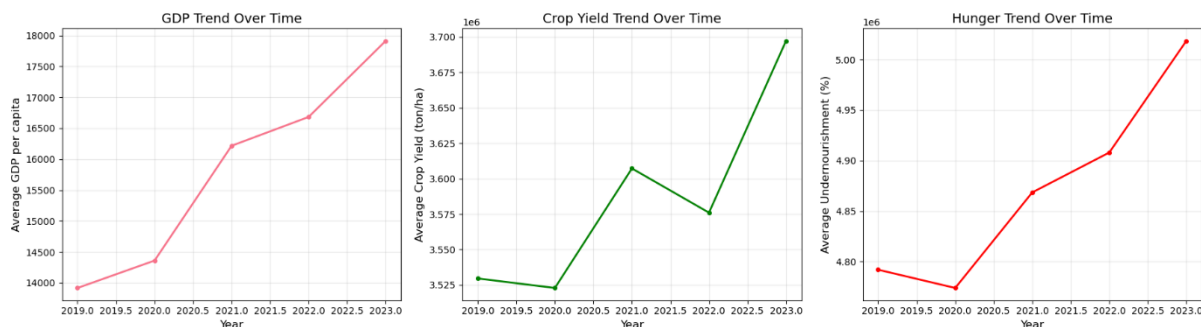
Figure 4: Detailed Mediation Path Scatter Plots (output_8_1.png)

This composite plot visually represents the three pathways of the mediation model, using color to enhance the density of data points:

- **Path A (GDP → Yield):** Shows heavy clustering at low GDP and low Yield values, illustrating the difficulty of observing a strong linear trend due to the presence of outliers (e.g., small, rich countries or specialized agrarian economies).

- **Path B (Yield → Hunger):** Displays a clear negative slope where high yield values are associated with low undernourishment values, validating the strong negative coefficient of Path b.

- **Total Effect (GDP → Hunger):** Visually confirms the strong curvilinear relationship, where the steepest drop in undernourishment occurs as countries move out of the lowest GDP bracket.

Figure 5: Distribution Plots (output_9_0.png)

This figure, consisting of three individual histograms, is crucial for assessing the suitability of the data for OLS regression:

- **GDP and Crop Yield:** Both distributions are highly **right-skewed**, with the mass of data points concentrated on the low end. This violates the assumption of normally

distributed residuals for OLS models and suggests that future steps may require a logarithmic transformation (e.g., log (GDP)) to normalize the data and improve model reliability.

- **Undernourishment:** The dependent variable is also skewed toward low values, indicating most observations are in countries with low hunger rates.

Figure 6: Simple Scatter Plots (output_10_1.png)

This final composite figure provides simplified scatter plots for Path A, Path B, and the Total Effect, without the complexity of color bars or kernel density estimation. It serves as a clear, final visual check of the raw data relationships that the OLS models are attempting to fit, reinforcing the observation that:

- The relationship between GDP and Yield (Path A) is highly diffuse.

- The relationship between Yield and Hunger (Path B) is a more condensed, negative association.

3. Key Insights and Policy Relevance

1. **Partial Mediation Confirmed (29.62%):** The analysis confirms that investment in agricultural productivity is a crucial mechanism. Almost one-third of the beneficial effect of national wealth (GDP) on reducing hunger is explained by the intervening step of higher crop yields.

2. **Importance of Direct Effects:** The remaining 70% of the effect (the direct effect) is still substantial. This indicates that while production (yield) is important, non-production factors like **infrastructure, logistics, policy, and social safety nets** (the factors linked to the GINI moderation from Week 5) are responsible for the largest share of hunger reduction.

3. **Data Limitations:** The high skewness in the distributions (Figure 5) suggests that the current OLS model results may be prone to instability. Future work must incorporate data transformations or robust regression techniques to ensure the findings are reliable.

4. Research Question for Week 7

The models are now complex enough that simple coefficient interpretation is insufficient. The next step is to use advanced statistical validation.

New Research Question: How can the statistical significance of the mediation model's **indirect effect** be rigorously validated using the **Bootstrapping** technique, and how do the resulting direct and indirect effects compare when using logarithmically transformed variables to mitigate the observed data skewness?