**Impact of Women in Country’s Development**

**Team E:**

**Anusha Kokkinti (Z1856785)**

**Aishwarya Vijay Saswade (Z1855087)**

**Neha Agarwal (Z1861955)**

**Tejaswini Mummadi (Z1859635)**

**OMIS 645: Applied Statistics for Business Analytics using SAS**

**Professor: Dr. Kishen Iyenger**

Table of Contents

[Introduction: 3](#_Toc26901941)

[Executive Summary: 4](#_Toc26901942)

[Data Cleaning: 4](#_Toc26901943)

[Data Directory: 5](#_Toc26901944)

[Data Glimpse: 7](#_Toc26901945)

[Analysis: 8](#_Toc26901946)

[Analysis-1 8](#_Toc26901947)

[Analysis -2 10](#_Toc26901948)

[Analysis-3 13](#_Toc26901949)

[Analysis-4 15](#_Toc26901950)

[Analysis-5 18](#_Toc26901951)

[Analysis -6: 20](#_Toc26901952)

[Analysis -7: 25](#_Toc26901953)

[Conclusions: 29](#_Toc26901954)

[References: 29](#_Toc26901955)

# Introduction:

United Nations Secretory-General Antanio Guterres stressed on Women’s empowerment saying that women empowerment and gender equality are “essential to global progress”. The autonomy of women and the improvement of their political, social, economic and health status is highly important for growth of the country.

The term "Women Empowerment" alludes to the way toward giving capacity to women to face the control of the others and help them to lead a prosperous and an effective life. Verifiably, Women have been viewed as comprising more fragile area in the general public. Be that as it may, the start of the twentieth century achieved a progression of changes in the status of women.

Women establish half of total populace and are verifiably liable for creating people in the future. These two are enough to express the massive requirement for strengthening. At the point when a lady gets socially, mentally and monetarily enabled, she assembles the family and network too and becomes a good example for her kids and others. Thus, it makes for better financial aspects and better society to enable women,

Therefore, hereby we are doing the statistical analysis of relationship between Women’s Progress with respect to various aspects like health and other socio-economic factors which are considered in the analyzing Country’s development and growth.

# Executive Summary:

The main goal of this project is to analyze the impact of women in country’s development. Indicators such as Gross Domestic Product, Human Development Index, Gender Development Index, Level of Urbanization that provides us a glimpse on the country’s development are considered for the analysis. Various Socio-economic factors like Education, Employment, Health, especially related to Women are used in the analysis.

Women’s Progress in education is measured using the indicators such as percentage of women’s enrollment in primary, secondary, tertiary level schools. These indicators represent the level of education attained by them. By using these indicators, we would like to perform analysis and see how their progress is impacting towards country’s development.

Infant Mortality Rate of a country reflects the condition of a country how serious they’re for the health of their women and mothers. An increase in infant mortality in a country is negatively correlated with the development of that country. Development may be narrowed down to the state of sound medical services. Thus, it is expected that in a country with good medical services, the rate of infant mortality should be low. We are interested in knowing if mortality rate is same across all the geographically divided regions and is Gender Inequality a reason behind it. We think Education brings more awareness among people and thus reducing mortality rate. So, we would like to check if this theory is correct or not.

As education alone cannot be deciding factors for county’s progress, we would like to analyze the impact of the progress of women in employment and the need for all-round development of women for country’s development.

# Data Cleaning:

Most of our data is collected from the website [data.worldbank.org](https://data.worldbank.org/indicator)/indicator. The site has the data for various socio- economic indicators for a period of 1960 to 2018. From this data, we selected few parameters like education, employment, health etc related to women that are relevant to our analysis for the year 2017. As each indicator data was provided as different datasets, we chose our level of analysis as country and merged the various indicators into one file using vlookup. Then we filtered out the inappropriate data and finally the dataset for our analysis was built with 151 countries.

# Data Directory:

**1. GDP per capita** –As per the site data.worldbank.org, it stands for Gross Domestic Product (GDP) per capita (per person). It is derived from a straightforward division of total GDP by the population. Per capita GDP is typically expressed in local current currency, local constant currency or a standard unit of currency in international markets, such as the U.S. dollar (USD).

**2. Gender Development Index (GDI)-** As per the site data.worldbank.org, the GDI measures the gender gaps in human development achievements for disparities between women and men in three basic dimensions of human development- health, knowledge and living standards.

**3. Human Development Index**- As per the site data.worldbank.org ,the Human Development Index (HDI) is a statistic composite index of [life expectancy](https://en.wikipedia.org/wiki/Life_expectancy), [education](https://en.wikipedia.org/wiki/Education), and [per capita income](https://en.wikipedia.org/wiki/Per_capita_income) indicators, which are used to rank countries into four tiers of [human development](https://en.wikipedia.org/wiki/Human_development_(humanity)).

**4.** **Country’s Economic Status (Developed/ Developing)-** As per the site data.worldbank.org, standard criteria for evaluating a country's level of development are [income per capita](https://www.investopedia.com/terms/i/income-per-capita.asp) or per capita gross domestic product, the level of industrialization, the general standard of living, and the amount of technological infrastructure.

**5. Primary School Enrollment-** As per the site data.worldbank.org, Primary Education is the initial stage of education and has as its basic aim to create, establish and offer opportunities to all children, regardless of age, gender or country of origin, to achieve a balanced cognitive, emotional and psychomotor development.

**6. Secondary School Enrollment-** As per the site data.worldbank.org, Secondary education completes the provision of basic education that began at the primary level and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers.

**7. Tertiary School Enrollment**- As per the site data.worldbank.org, Gross enrollment ratio of total enrollment, regardless of age, to the population of the group that officially corresponds to the level of education shown. Tertiary education, whether to an advanced research qualification, normally requires, as a minimum condition of admission, the successful completion of education at the secondary level.

**8. Employment in Agriculture-** As per the site data.worldbank.org, Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The agriculture sector consists of activities in agriculture, hunting, forestry and fishing.

**9. Employment in Industry-** As per the site data.worldbank.org, Employment in Industry sectors like Manufacturing, services, public utilities, construction, Information Technology etc.

**10. Infant Mortality Rate-** As per the site data.worldbank.org, the infant mortality rate is the number of deaths under one year of age occurring among the live births in a given geographical area during a given year per 1000 live births occurring among the population of the given geographical area during the same year.

**11. Urbanization-** As per the site data.worldbank.org, when populations of people grow, the population of a place may spill over from city to nearby areas. This is called urbanization.

**12. Self-Employment-** As per the site data.worldbank.org, being self-employed means running your own business, but that business can be set up in a variety of ways. The IRS says that a self-employed individual is someone who "owns an unincorporated business," like a sole proprietor or independent contractor or a sole owner of an LLC.

**13. Wages and Salaried Employed:** As per the site data.worldbank.org, Wage and salaried workers (employees) are those workers who hold the type of jobs defined as "paid employment jobs," where the incumbents hold explicit (written or oral) or implicit employment contracts that give them a basic remuneration that is not directly dependent upon the revenue of the unit for which they work.

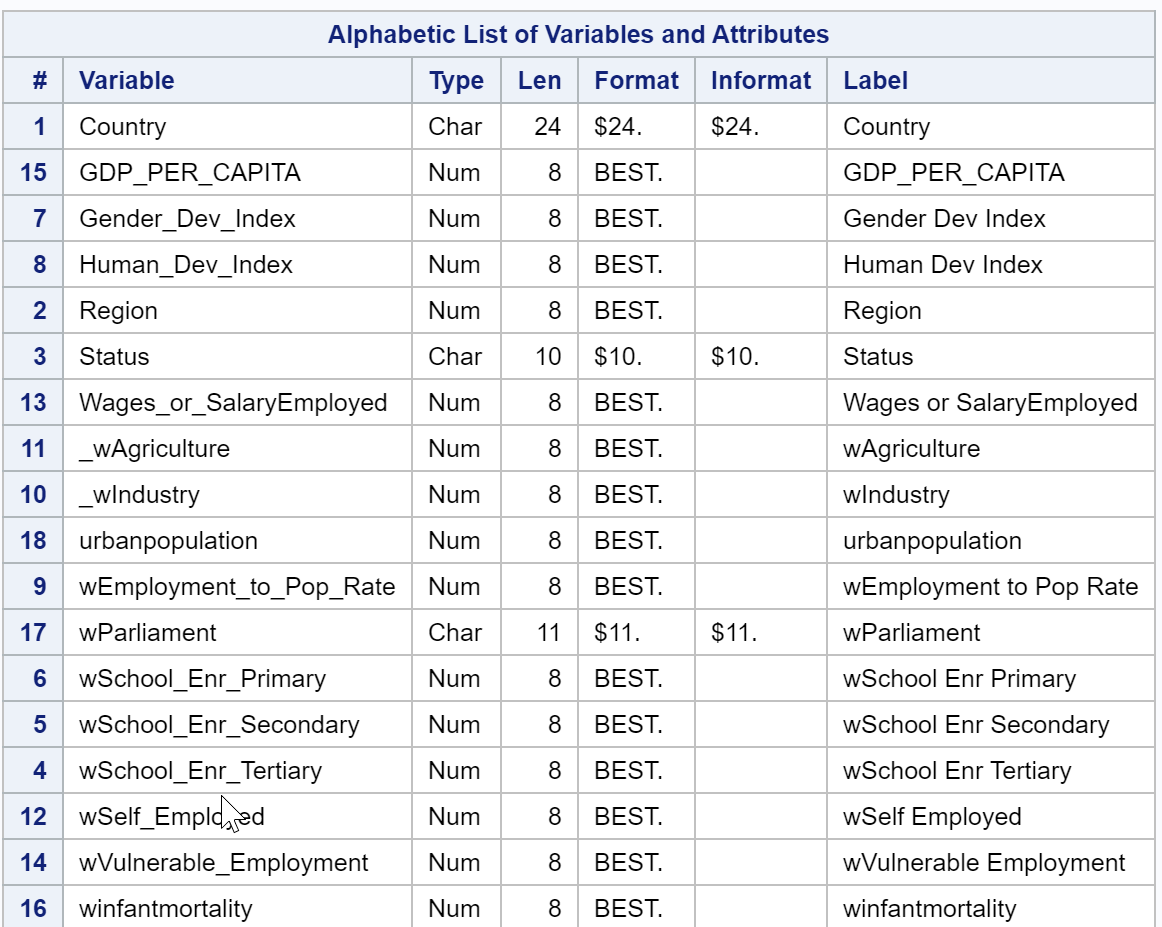
**14. Vulnerable Employment**- As per the site data.worldbank.org, the percentage of women contributing family workers and own-account workers as a percentage of total employment. They are less likely to have formal work. arrangements, and are more likely to lack decent working conditions, adequate social security.

**15.Women in Parliament-** As per the site data.worldbank.org, the percentage of women representatives in legislative houses like parliament and Congress.

**16.Employment to Population Rate**: As per the site data.worldbank.org, Employment to population ratio is the proportion of a country's population that is employed. Ages 15 and older are generally considered the working-age population.

# Data Glimpse:

|  |  |
| --- | --- |
| Region |  |
| 0 | Western Europe |
| 1 | Central and Eastern Europe |
| 2 | Commonwealth of Indep States |
| 3 | Southeast Asia |
| 4 | South Asia |
| 5 | East Asia |
| 6 | Latin America |
| 7 | Northern America and ANZ |
| 8 | Middle East and North Africa |
| 9 | Sub-Saharan Africa |



# Analysis:

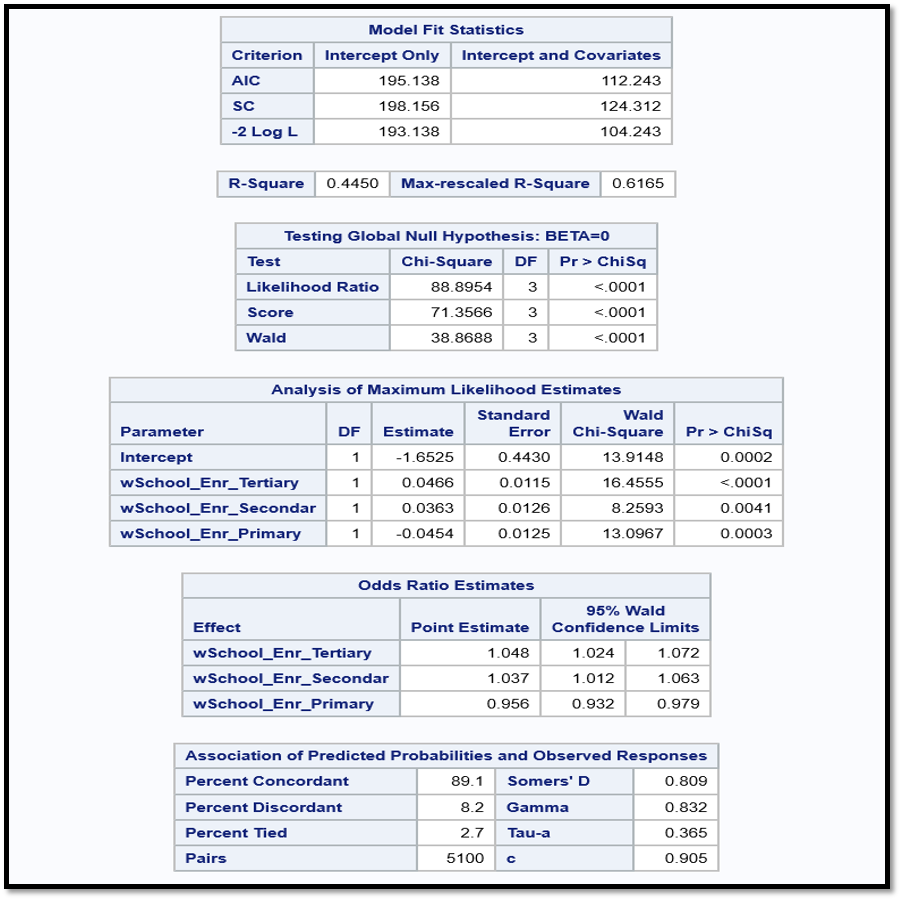
## Analysis-1

**Question: How the Women Education Impacting the "Development" Status of the Country?**

Model used**:** Logistic regression

Dependent Variable:Status [Developed, Developing]

Independent Variables**:** wschoolenrollmentprimary, wschoolenrollmentsecondary, wschoolenrollmenttertiar



**Interpretations:**

1. From the results, -2 log L with the parameters (104.243) is significantly less than the model with only intercepts (193.138). Variance is around 44%.
2. The Wald’s Chi-Square test static’s p-value is less than 0.05. Hence the model is significant.
3. The log odds of country to gain a Developed status is decreasing by 44% for a percent increase in the women’s enrollment in the primary level schools. (Hence If women are just having primary education it doesn't make any difference, to dig deeper we ran separately (with just primary enrollment parameter) and found it's not a significant variable.
4. The log odds of country to gain a Developed status is Increasing by 37% for a percent increase in the women’s enrollment in the secondary level schools.
5. The log odds of country to gain a Developed status is Increasing by 48% for a percent increase in the women’s enrollment in the tertiary level schools.
6. C = 0.905, means around 90% rows correctly predicted the Status of the Countries.

**Inferences**:

Country Status whether being a Developed / developing is significantly impacted by % of women enrollment in schools for higher level of education. Encouraging women to attain progress towards higher levels of education will have positive results on country’s growth.

## Analysis -2

**Analysis-2a:**

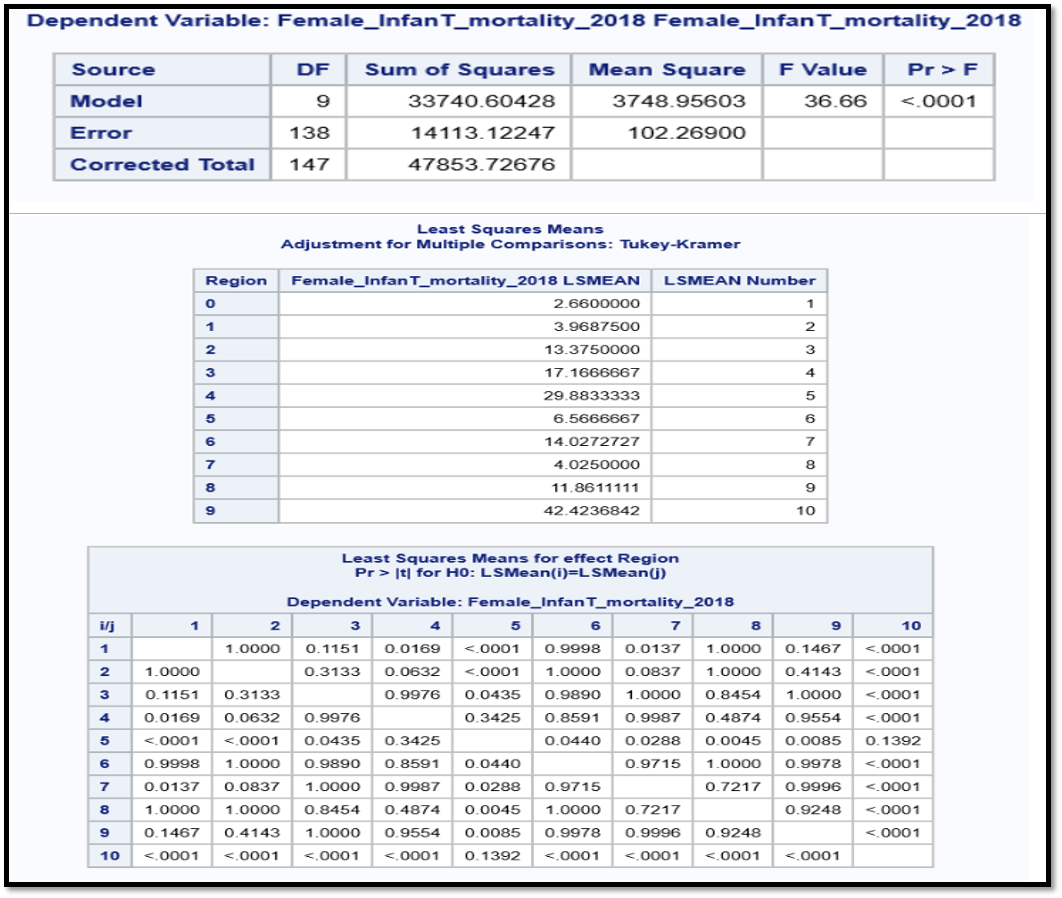
**Question: Is Infant mortality rate different for different regions?**

Test: ANOVA

**Hypothesis:**

**H0:** mu1 = mu2 = mu3 = mu4...

**H1:** mu1! = mu2 or mu2! = mu3 or mu3! = mu4 ….



**Interpretations:**

1. P-value < 0.0001 It means the model is significant. And there is at least one pair of regions for which the infant mortality rate is significantly different.
2. To know which regions, have significantly different means we performed the Tukey HSD (Honestly significant different) test. It is clear from the table that there are many pairs of regions for which infant mortality rate is significantly different.
3. Example P-value for Region 3 and Region 0 is 0.0169 which is smaller than 0.05 and hence for this pair of regions infant mortality rate is significantly different. Another pair of regions for which infant mortality rate is significantly different includes Region 4 & 0, Region 4 & 1, Region 4 & 2, Region 5 & 4, Region 6 & 0, Region 6 & 4, Region 7 & 4, Region 8 & 4, Region 9 & 0, Region 9 & 1, Region 9 & 2, Region 9 & 3, Region 9 & 5, Region 9 & 6, Region 9 & 7, Region 9 & 8 and Region 9 & 9.

**Analysis-2b:**

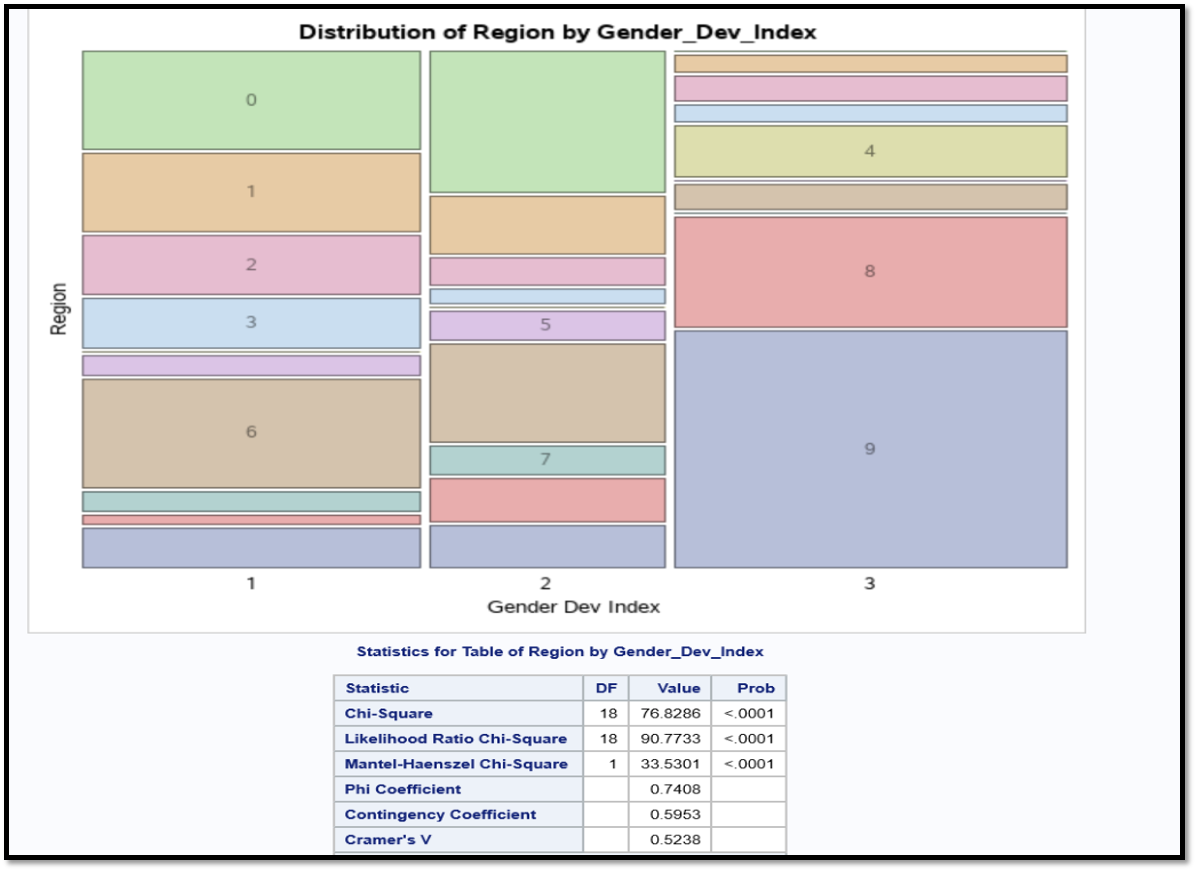
**Question: Is there any relationship between region and gender development index?**

Model Used**:** CHI SQUARE

**Hypothesis:**

**H0:** Region and GDI are independent

**H1:** Region and GDI are dependent

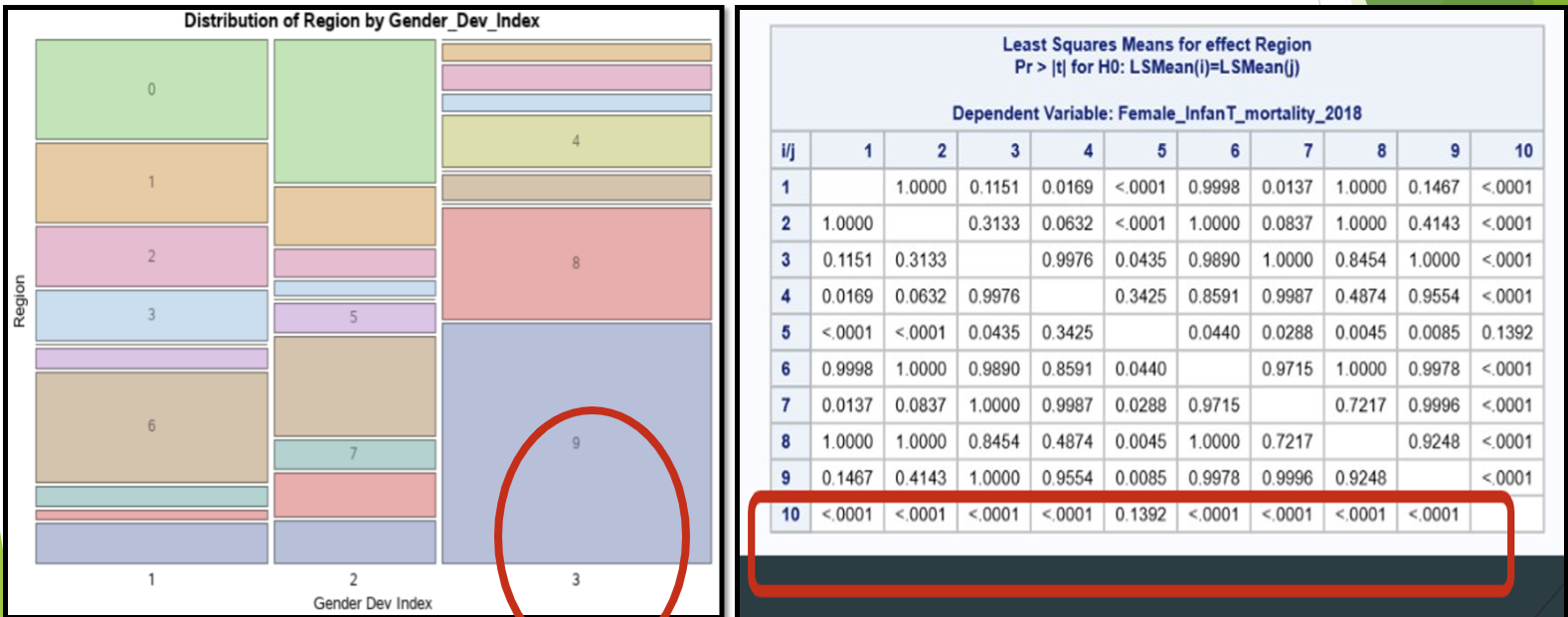


**Interpretations:**

1. P-value < 0.0001 -> Reject Null Hypothesis
2. It means we have enough evidence to support Alternative Hypothesis.
3. Region & GDI are dependent.
4. From above “Distribution of Region by GDI” figure it is clear GDI (=1) is high for Region 6, GDI (=2) is high for Region 0 and GDI (=3) is high for Region 9.

**Inferences:**

It is evident that if GDI (=3, gender inequality is more) for certain region is high, the infant mortality rate for that region will also High.



## Analysis-3

**Question: Is infant mortality rate decreasing with the increase of women’s education across various levels?**

Model Used: Multi-Linear Regression

**Hypothesis:**

**For Slope-**

**H0:** Slope for each independent variable is equal to zero

**H1:** Slope for each independent variable is not equal to zero

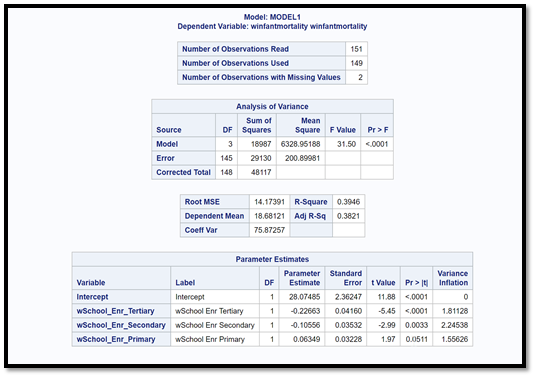
**For Intercept-**

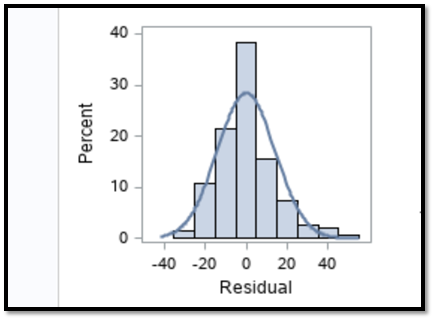
**H0:** Intercept is equal to zero

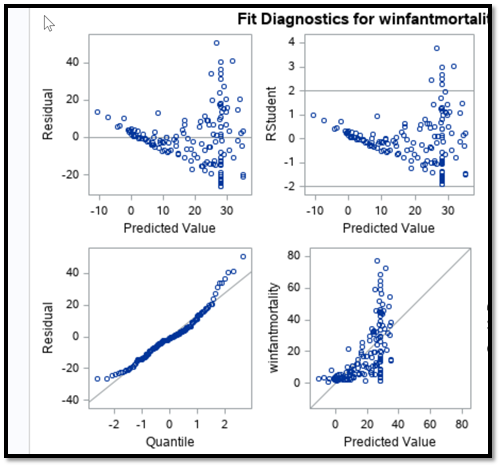
**H1:** Intercept not equal to zero

Dependent Variable: female\_infant mortality rate

Independent Variables: wschoolenrollmentprimary, wschoolenrollmentsecondary, wschoolenrollmenttertiary







From above figure, all the assumptions for the linear regression model holds true.

**Interpretations:**

1. From the result, the model (with the parameters %womensenrollmentprimary, %womenenrollmentsecondary, %womenenrollmenttertiary to impact female infant mortality rate) is significant and explains around 38% of variation.
2. Women’s enrollment in primary level school education is not significant in impacting the female infant mortality rate.
3. A percentage increase in the women’s enrollment in secondary level education is decreasing female infant mortality rate approximately by 1 per 10000 female births.
4. A percentage increase in the women’s enrollment in Tertiary level education is decreasing female infant mortality rate approximately 2 per 10000 female births.

**Inferences:**

Increasing in the % of enrollments of women into higher levels of education will decrease the female infant mortality rate.

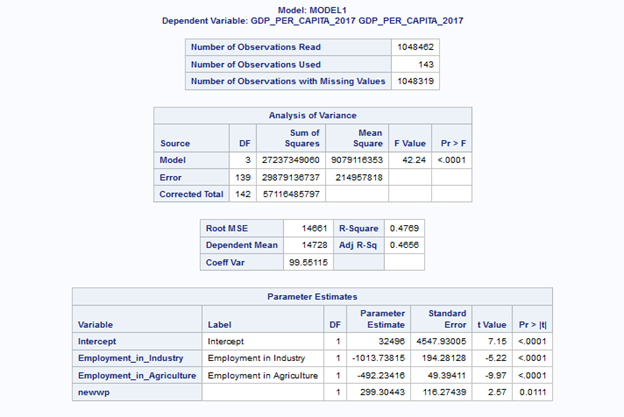
## Analysis-4

**Question: How is the impact of women’s employment in areas like Industry, Agriculture, parliament on country’s GDP Per Capita. (Country development indicator)**

Model Used**:** Linear Regression

Dependent Variable: GDP\_Per\_Capita

Independent Variable: %Employment in Industry, %Employment in Agriculture, % %Employment in Parliament



**Results**: From the above screenshot, we can see that the f-static p-value is less than 0.05. Hence the model with parameters Employment\_in\_Industry, Employment\_in\_agriculture, Employment in \_parliament for predicting GDP Per Capita of a country is significant and explaining 46 % of variation.

**Interpretations:**

1. A percent increase in Employment of Women in agriculture is Decreasing Country GDP per Capita by 492.2 dollars
2. A percent increase in Employment of Women in Industry is Decreasing Country’s GDP per Capita by –1013.738 dollars
3. A percent increase of women’s in parliament is increasing the country’s GDP per Capita by 299.30 dollars

**Inferences:**

1. Increase in % of Women’s in parliament will positively impact country’s Economy.

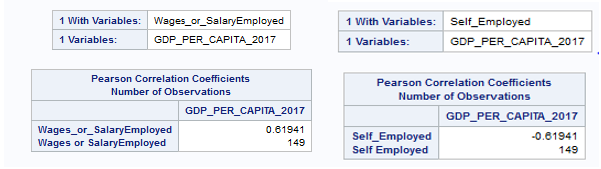
Evaluation: International growth center Research has shown that there has been a phenomenal global increase in the proportion of women in politics in the last two decades. The constituencies that elect women experience significantly higher growth in economic activity through the electoral term than similar constituencies that elect men.

1. Increase in % of women’s employment in agriculture is not showing a positive impact on country’s Economy.

Evaluation: According to Food and Agriculture Organization report of United Nation (FAO)on gendered analysis of agriculture trade, though women’s contribution to agricultural production in most developing countries is significant, women are generally associated with non-economic and unpaid work. So, their contribution is not registered in the system of national accounts and therefore no market value is given to the labor involved. As indicated by FAO, women’s agricultural activities are limited by a lack of financial capital as well as constrained by inadequate access to productive resources. Women tend to present low levels of mechanization and technological inputs, which translates into low productivity.

1. Increase in % of women’s employment in Industry is not showing a positive impact on country’s Economy.

Evaluation: As per the world development report 2012, Economic development is positively correlated with the share of female workers in wage or salary employment and negatively correlated with the share of women in unpaid work, self-employment, and entrepreneurship. It holds true for our analysis as well. Below are the results:



From the above screenshot, we can see that the Wages\_or\_SalaryEmployed is positively correlated with the GDP\_Per\_Capita as its correlation coefficient is positive (0.6) and Self\_Employed is negatively correlated with the GDP\_Per\_Capita as its correlation coefficient is negative (-0.6)

Adding to it, a research on women by Catalyst in workforce states that Women’s work force participation rate declining globally. It means that women are more interested into the self-employed leading to negative impact on economy.

## Analysis-5

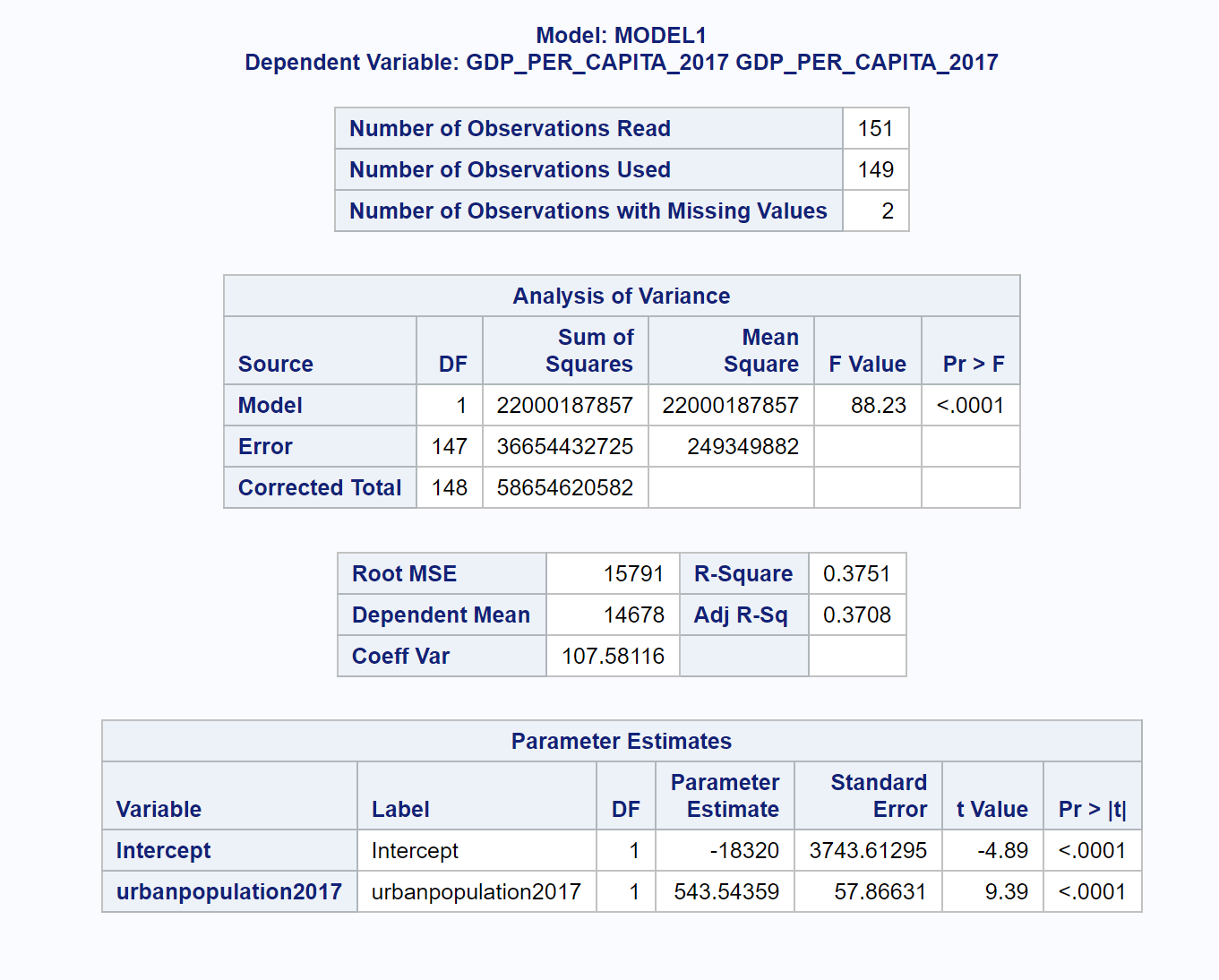
**Question: Testing the moderation effect of women self-employment in the urban population on country’s economy.**

Model Used: Moderation Analysis – Linear Regression

1. First checking the relationship between Urbanization and GDP Per capita

Dependent variable: GDP\_Per\_Capita

Independent Variable: Urban Population

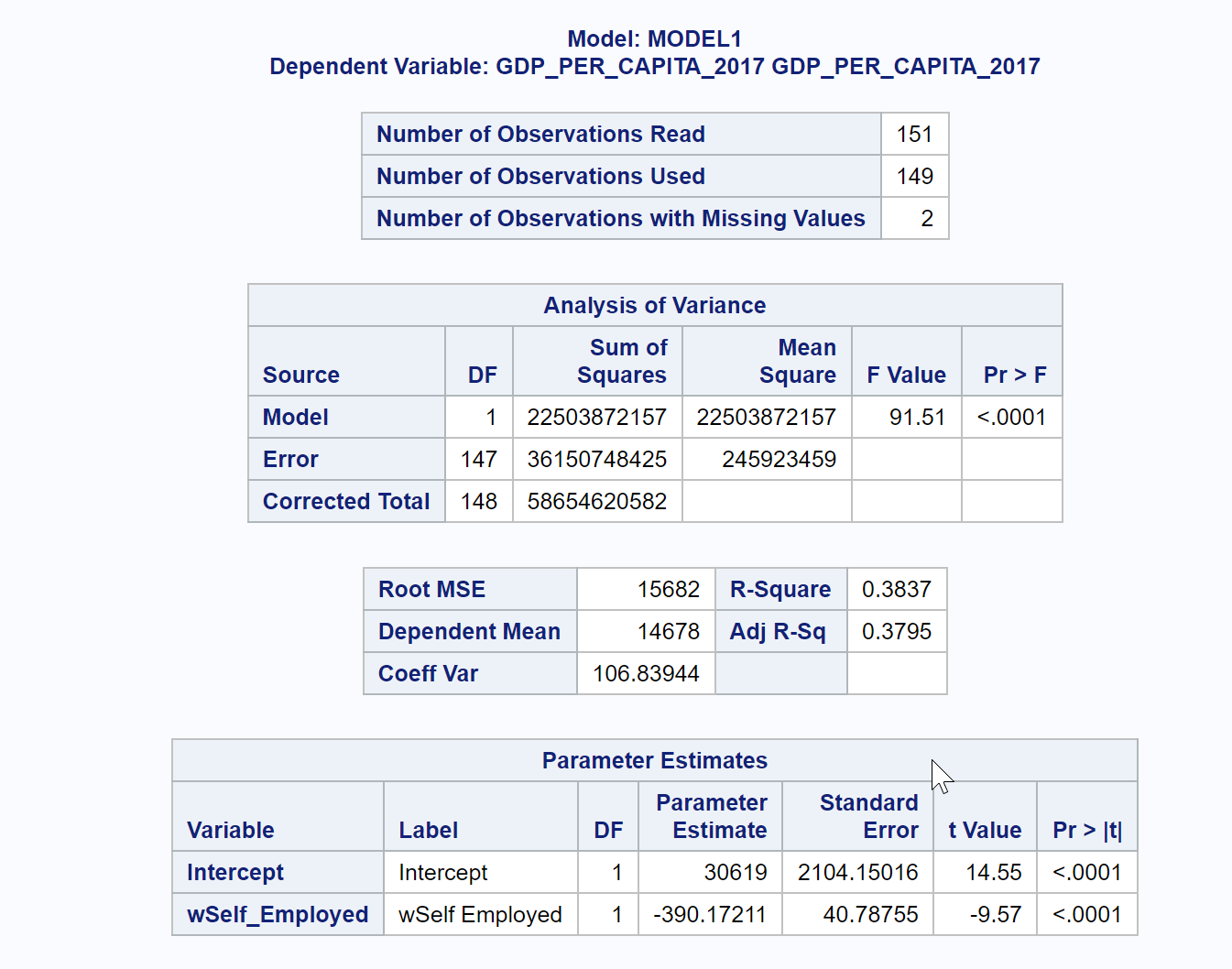


**Result:** It is Significantly shown that there is a positive relationship between urbanization and country’s GDP\_per\_capita.

1. Second checking the relationship between Women Self- employment and GDP Per Capita

Dependent Variable: GDP\_Per\_Capita

Independent Variable: %WomenSelfEmployed



**Results**: It is Significantly shown that there is a negative relationship between Percent of Women’s being self-employed and country’s GDP\_per\_capita.

1. Checking the interaction effect of urbanization with Women being self- Employed as moderator on country’s economy.



**Result**: The p value is less than 0.05. Hence the model is significant. The p-value for the interaction effect between self-employment and urban population is also less than 0.05 and hence significant

**Inferences:**

It is evident, though urbanization have positive impact on country’s economy, it will be hampered, if it sees more of a self-employed woman.

## Analysis -6:

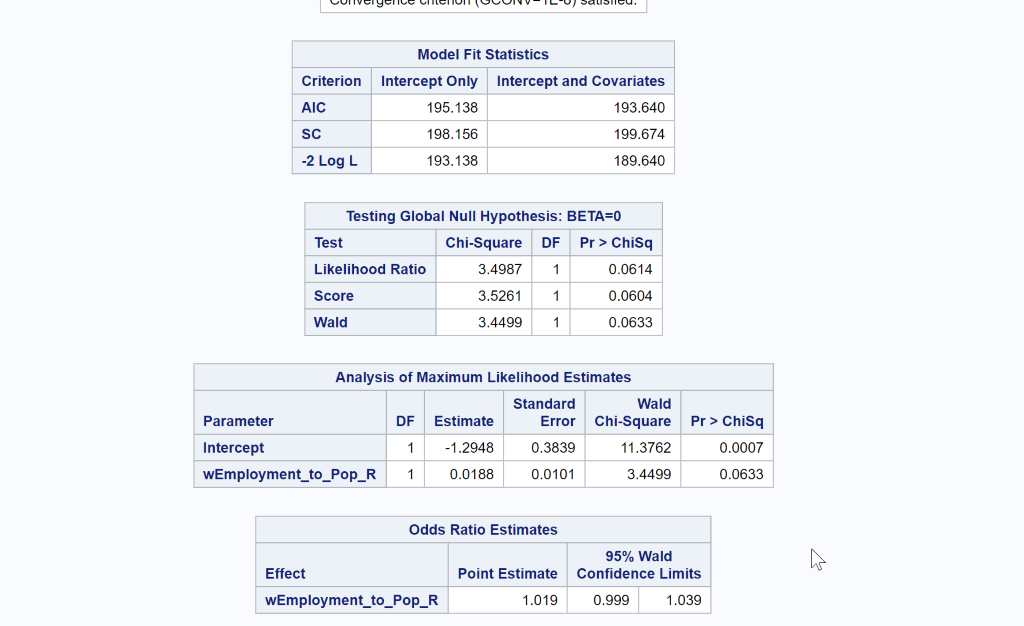
In general employment to population ratio is a positive factor for country’s development. More the ratio, it shows that a greater number of people are working towards country's growth, hence contributing towards country’s Developed status. But employment to population rate alone is not enough to decide the country’s status because it may consider varied type of employments which might have negative impact. Among such variables, Vulnerable employment is one that have negative impact. The model was not significant when only employment to population ratio is taken but turned to be significant when vulnerable employment is considered. Hence, we would like to check if there is any interaction effect between these two variables in deciding the country’s status.

**Question 6a:**  **Can the Log odds of country gaining Developed status be decided by Employment to population ratio? ​**

Model Used: Logistic Regression

Dependent Variable: Status

Independent Variable: Employment-to-Population Ratio



**Result:** The p-value corresponding to Wald’s Chi-Square is 0.0633 which is greater than 0.05. Hence the model is not Significant.

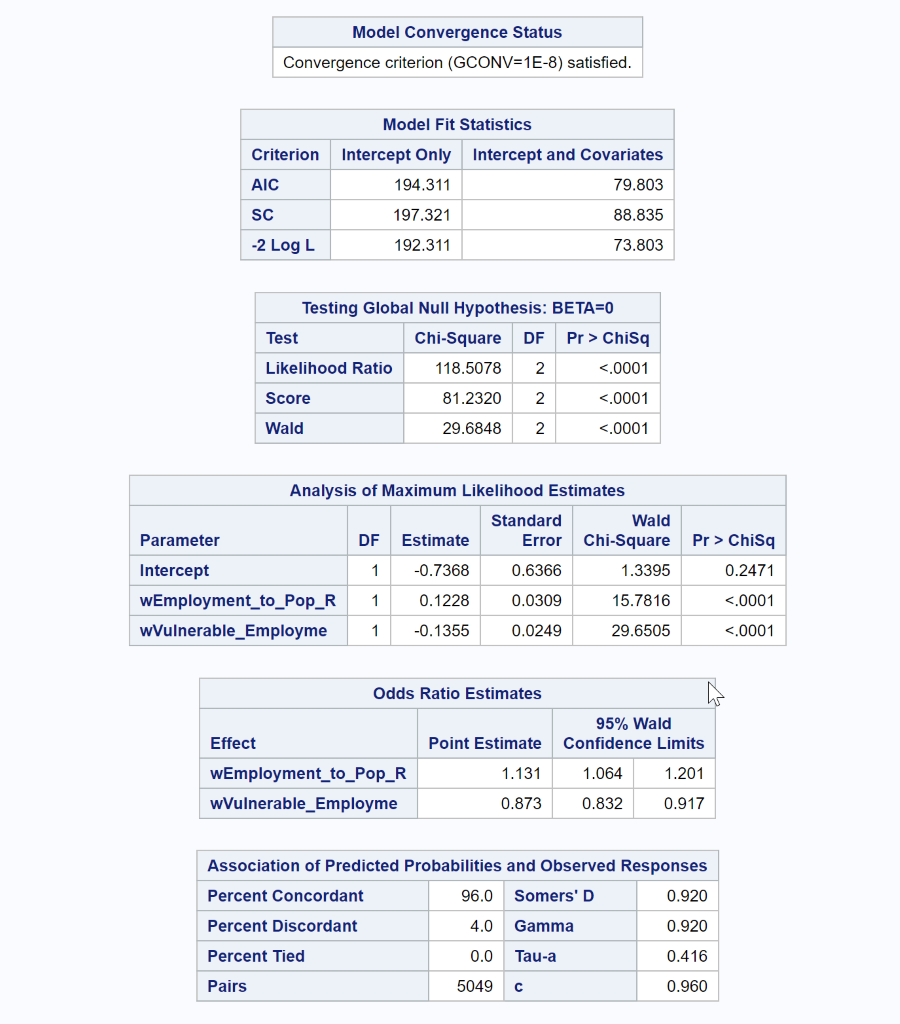
**Inference:** Log odds of country gaining Developed status cannot be decided by Employment to population ratio alone.

**Question 6b:** **Can the Log odds of country gaining Developed status be decided by Employment to population ratio and % of vulnerable Employment (women)​**

Model used: Logistic Regression​

Dependent Variable: Status​

Independent Variable: Employment-to-PopulationRatio, % of Vulnerable Employment(women)**.** ​



**Result:**  From the above results, -2 Log L value for model with parameters (73) is significantly less than the model without parameters (192). The Wald’s Chi-Square test static p-value is less than 0.05. Hence the model is significant.

**Interpretations:**

1. The p-value corresponding to the maximum likelihood estimates of Employment-to-Population ratio and % of vulnerable Employment are less than 0.0001.

2. Hence the estimate values are significantly different than zero and hence can be used in the analysis pertaining to predict country status.

3. The log odds of country gaining developed status is increasing by 13% by a unit increase in Employment to population ratio.

4. The log odds of country gaining developed status is decreasing by 12.7% by a unit increase in percentage of Vulnerable Employment(women)

5. The c-value is 0.96, which shows that the model predicted the values correctly 96% of the time.

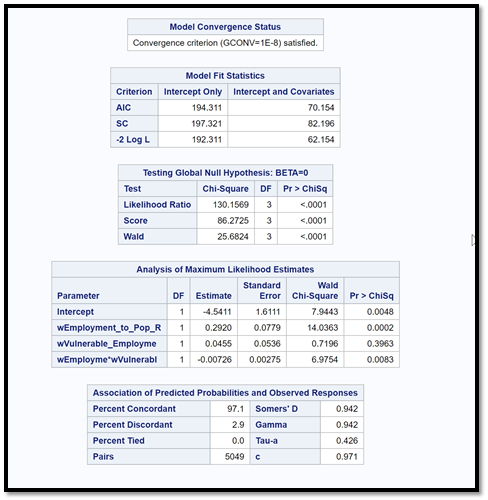
**Inference:** Model turned significant with the addition of Vulnerable employment (women)

**Question 6c:** **Is there any interaction effect between variables Employment-to-Population ratio and Vulnerable Employment​?**

**Model used**: Logistic Regression​ (Interaction Effect)

Dependent Variable: Status​

Independent Variable: Employment-to-PopulationRatio, % of Vulnerable Employment(women).



**Result:** From the above results, -2 Log L value for model with parameters (62) is significantly less than the model without parameters (192). The Wald’s Chi-Square test static p-value is less than 0.05. Hence the model is significant.

**Interpretations:**

1. The p-value corresponding to the maximum likelihood estimates of Interaction element between Employment-to-Population ratio and % of vulnerable Employment is less than 0.0001.

2. Hence the estimate value is significant from zero. Hence there is Interaction Effect between the variables.

3. The c-value is 0.971, which shows that the model predicted the values correctly 97% of the time which is greater than the model with our interaction element added.

**Inference:**

It is evident that the Effect of Employment to population ratio on country’s development is dependent on % of Vulnerable Employment (Women)​

## Analysis -7:

* As Vulnerable Employment is playing significant role in determining Country’s Status, Will it have same relationship with HDI?
* What is HDI?

HDI expands to Human Development Index is higher when  [lifespan](https://en.wikipedia.org/wiki/Life_expectancy_at_birth) is higher, the [education](https://en.wikipedia.org/wiki/Education) level is higher, and the gross national income [GNI (PPP) per capita](https://en.wikipedia.org/wiki/GNI_(PPP)_per_capita) is higher.

* Is % of Vulnerable Employment (women) considered for HDI? We will check this by testing if there is any correlation between given % of Vulnerable Employment values and HDI.

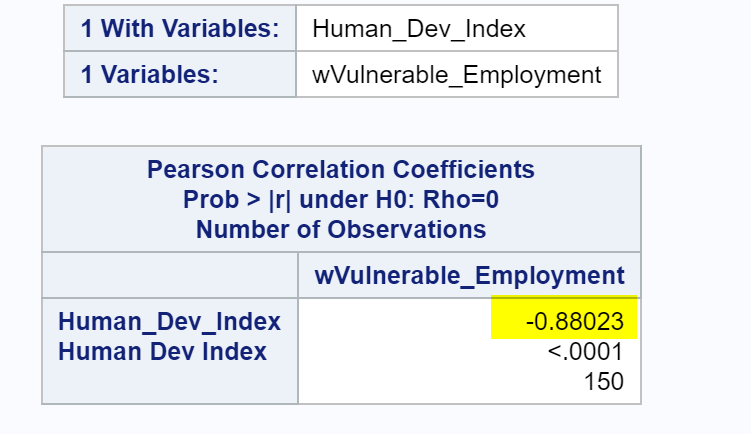
**Question 7a**: **Is there any correlation between %Vulnerable Employment(women) and HDI.**

Model Used: Correlation Analysis

Variables:

a. Human Development Index,

b. % Vulnerable Employment (women)



**H0:** There is no correlation between Human\_Development \_Index and %Vulnerable\_Employment(women)

**H1**: Correlation exists between Human\_Development \_Index and %Vulnerable\_Employment(women)

**Result**: The p-value is <0.0001 which conveys that we can reject null Hypothesis.

**Interpretations:**

1. Correlation exists between Human Development Index and Vulnerable Employment.

2. Human Development Index is negatively correlated with % of Vulnerable Employment.

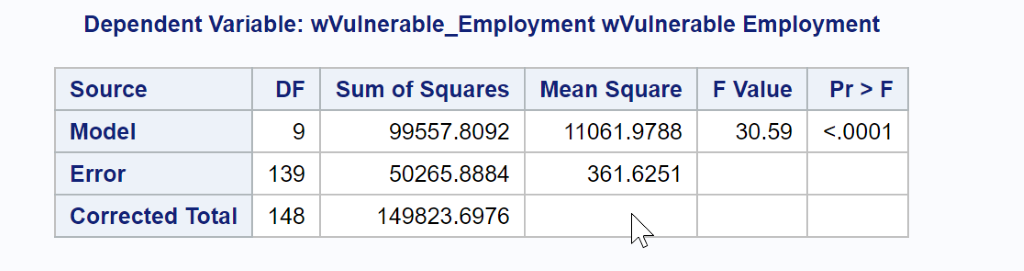
**Inference:**

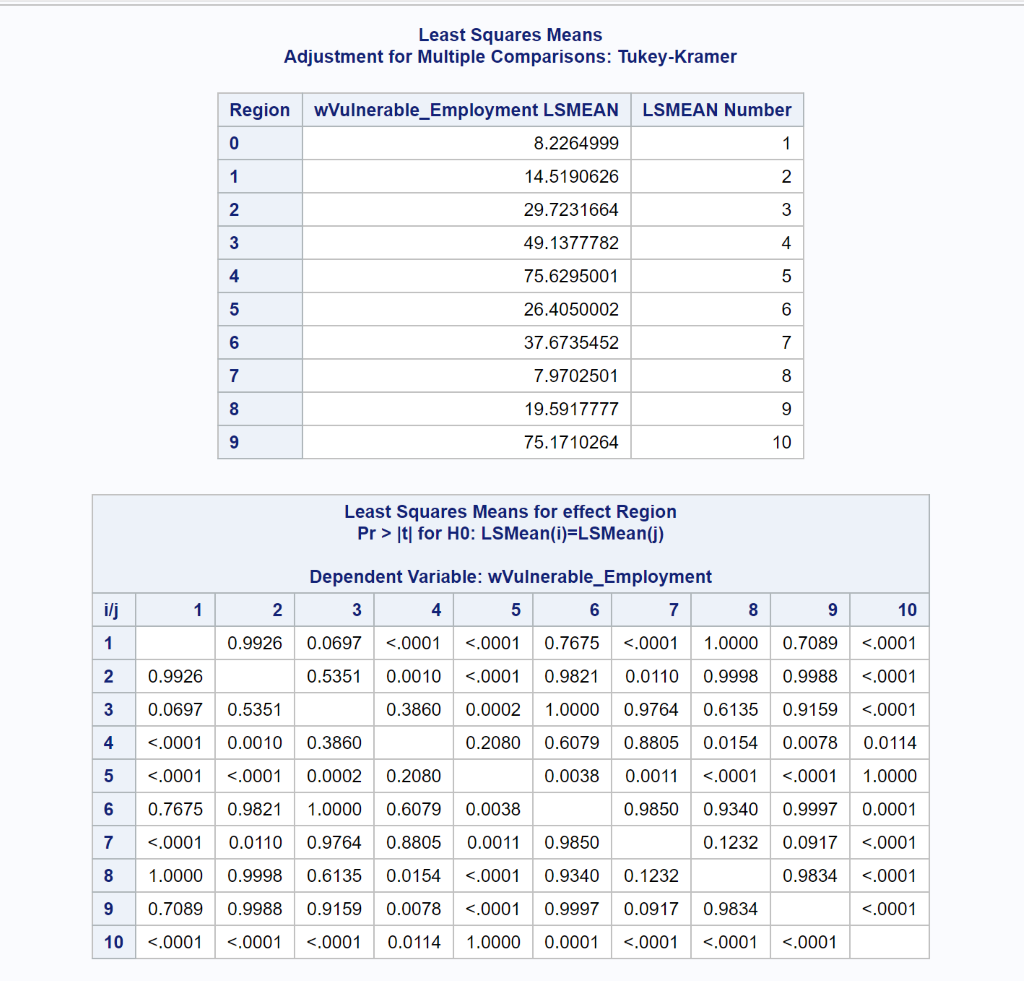
% of Vulnerable Employment of Women can impact country’s Human Development Index.

**Question 7b**.: **Is there any significant variation in % Vulnerable Employment of women across the different geographic regions?**

Model Used: ANOVA

Variables used: Region, % of Vulnerable Employment (women)





**H0**: The mean % of Vulnerable Employment (women) is same across all regions

**H1:** The mean % of Vulnerable Employment(women) is not same at least in one of the regions.

**Result**: P-value corresponding to the F-static value for Analysis of variances is less than 0.0001.

Hence the model is significant.

**Interpretations:**

1. Following regions are significantly different in terms of their mean % Vulnerable Employment.

a. 0 is different from 3,4,6,9.

b. 1 is different from 3, 4, 6,9.

c. 2 is different from 4 ,9

d. 3 is different from 0,1,7,8,9

e. 4 is different from 0,1, 2, 5,6,7,8

f. 5 is different from 4,9

g. 6 is different from 0,1, 4, 9

h. 7 is different from 4,9

i. 8 is different from 3,4,9

j. 9 is different from 0,1,2,3,5,6,7,8

**Inference:**

Region 4, 9 are significantly differing from most of the regions. But both does not have any significant difference among them.

Region 4 is South Asia

Region 9: sub Saharan Africa

**Question 7c: Why these areas are particularly different in terms of vulnerable Employment. Is there any dependence on gender Development Index?**

The Chi-Square test of Independence ran before shows the relationship between regions and Gender Development Index.

**Interpretation:** It shows that there are a greater number of countries in the regions South Asia and Sub Saharan Africa with More Gender Inequality (More suppression towards women).

**Inference:** Country’s with more suppression towards Women have significant difference in their % of Vulnerable Employment.

# Conclusions:

1.Suppression of women through gender equality is impacting the country’s development.

This is tested through mortality rate and vulnerable employment.

2. Progress in women’s education has a positive impact on country’s development.

3. Encouraging women for higher education levels might have positive impact in such cases.

4. While urbanization can be positive factor for country’s economy, having more self-employed people lead to the negative impact on the country’s economy.

5. More women in parliament should be encouraged for a country’s positive development and gender equality must be improved for the contribution of women in agriculture and industry to improve economy.

6. % Vulnerable Employment affects the country’s development (tested through various variables like country status and employment-to-population ratio and compared with HDI).

# References:

Data from:

* The World Bank Group. Retrieved from <https://data.worldbank.org/indicator>

Theory behind the evaluation of the results from:

* Women in the Workforce-Global: Quick Take. Retrieved from <https://www.catalyst.org/research/women-in-the-workforce-global/>
* A gendered analysis of agricultural trade. Retrieved from <http://www.fao.org/3/a0493e/a0493e06.htm>
* International Growth Centre. Retrieved from <https://www.theigc.org/blog/are-women-politicians-good-for-economic-growth/>

Few other References for Context:

# Women’s empowerment ‘essential to global progress’ says Guterres, marking International Day, UN News. Retrieved from <https://news.un.org/en/story/2019/03/1034241>