MG-220 Course Project

From: Rachel Vlassis

To: Dr. Thomas Craig

Date: April 26th, 2022

Subject: Analysis of the Effect of Geographic Location on Skin Cancer Rates

---

Introduction and Overview

Cancer is the second leading cause of death behind cardiovascular diseases. In 2020 alone, over 18 million cases of cancer were diagnosed. Of those 18 million, melanoma of the skin accounted for roughly 325,000 diagnoses, or 1.8%. For decades, trillions of dollars have gone towards cancer research in order to gain a better understanding of the disease and develop pharmaceuticals, therapies, and various other treatment options. An addition to this, there are many preventative activities that individuals can partake in to help reduce their own active risk of developing cancer. According to the CDC, these activities include: avoiding tobacco, protecting the skin, limiting alcohol intake, keeping a healthy weight, and being vaccinated.

However, what about those factors that exist beyond the control of an individual? To help understand this, this paper will analyze the relationship between geographical location and rate of skin cancer. To accomplish this, 158 countries have been categorized into the following geographic regions: Americas, Asia-Pacific, Europe, Middle East and Northern Africa, and Sub-Saharan. To understand how these different areas play a role in the development of skin cancer, the following variables will be observed: frequency of skin cancer, photovoltaic power potential as an indicator of UV radiation, GDP per capita in U.S dollars as a measure of the standard of living, and STC indices as a measure of healthcare efficiency. These variables were selected because they are contingent upon where an individual lives and thus uncontrollable without intervention beyond the individual level. The goal of this analysis is to determine the relationship between these variables and geographic location to determine if an individual's country of residence is impacting their likelihood of developing skin cancer.

This analysis will use the following tests: ANOVA with Tukey comparison, regression, and multiple regression. An ANOVA test allows for the comparison of more than two groups at the same time to determine whether a relationship exists between them. The ANOVA Tukey comparison allows the data to be grouped together by statistical similarity. Regression analysis is used to determine the strength of a relationship between a dependent and an independent variable. With regression analysis is the r-squared value, which indicates how much of the variation in the dependent variable can be explained by the independent variable. The multiple regression is similar to the regression analysis, except it allows for the use of multiple independent variables to be used and helps to understand their explanatory power.

Analysis One: Frequency of Skin Cancer Diagnoses (2020)

The data for the frequency of skin care diagnoses in 2020 was obtained from the World Health Organization's Internal Agency for Research on Cancer. It includes the region, number of cases, and a crude rate which will be used as the point of analysis. The crude rate can be defined as the total number of melanoma cases reported, divided by the mid-year total population, and multiplied by a constant of 100,000. An analysis of the crude rate by geographic region will provide a baseline to see if differences in frequency exist.

Hypotheses:

Null Hypothesis (H_0): There is no difference in the frequencies of skin cancer in 2020 of the defined geographic regions.

Alternative Hypothesis (H_1): There is a difference in the frequencies of skin cancer in 2020 of the defined geographic regions.

Data:

Grouping Information Using the Tukey Method and 95% Confidence

Region	N	Mean	Grouping	
Europe	40	18.39	A	
Asia-Pacific	31	4.92		B
Americas	26	3.94		B
Middle East and North Africa	17	1.191		B
Sub-Saharan Africa	44	0.5639		B

Means that do not share a letter are significantly different.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Region	4	7933	1983.2	19.31	0.000
Error	153	15717	102.7		
Total	157	23650			

Evaluation:

The data provides evidence that we can reject the null hypothesis and support the alternative hypothesis, indicated by the p-value of 0.000 and the Tukey comparisons. The analysis shows that Europe falls in a different grouping than the other regions and thus there is a significant statistical difference in the crude rate (cancer rate) between Europe and the other regions. From this, it is understood that cancer rates do vary by region. To help understand why these variances exist, the following tests will analyze the remaining factors and their variance by region as well as their relationship as a predictor variable of crude rate.

<u>Analysis Two: Global Photovoltaic Power Potential (2020)</u>

The data for the photovoltaic power potential was obtained from The World Bank's Data Catalog. Photovoltaic power potential is a measure of direct sunlight being turned into electricity. This measure gives an idea of the amount of solar radiation that each country is experiencing. Various research has shown a relationship between UV exposure and development of skin cancer, hence the CDC's recommendation to use spf and skin protection. An analysis of the relationship between photovoltaic radiation and geographic location will provide insight as to what regions experience the highest UV exposure. From there, the relationship between photovoltaic radiation and skin cancer will be analyzed to further indicate to what extent photovoltaic power potential can be used as an indicator for skin cancer.

Hypotheses:

1. **Photovoltaic power potential vs Geographic region**

Null Hypothesis (H_0): There is no difference in the photovoltaic power potential in 2020 for the defined geographic regions.

Alternative Hypothesis (H_1): There is a difference in the photovoltaic power potential in 2020 for the defined geographic regions.

2. Photovoltaic power potential vs Crude rate

Hypothesis: If a region has a higher voltaic power potential, it will have a higher crude rate.

Data:

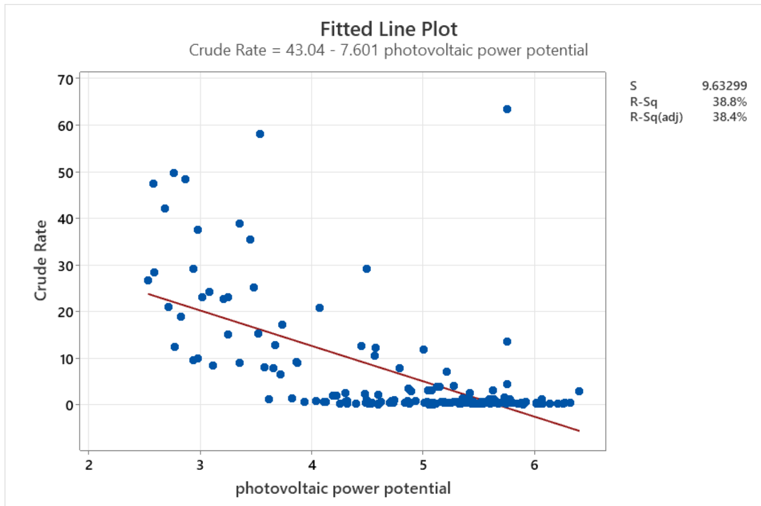
Grouping Information Using the Tukey Method and 95% Confidence

Region	N	Mean	Grouping			
Middle East and North Africa	17	5.8003	A			
Sub-Saharan Africa	44	5.5180	A			
Americas	26	5.117		B		
Asia-Pacific	31	4.603			C	
Europe	40	3.530				D

Means that do not share a letter are significantly different.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Region	4	107.99	26.9965	81.33	0.000
Error	153	50.78	0.3319		
Total	157	158.77			



Model Summary

S	R-sq	R-sq(adj)
9.63299	38.79%	38.40%

Evaluation:

The data provides evidence that we can reject the null hypothesis and support the alternative hypothesis, indicated by the p-value of 0.000 and the Tukey comparisons. Based on the Tukey comparisons, the Middle East and North Africa and Sub-Saharan Africa are statistically similar and have the highest photovoltaic power potential. The Americas, Asia-Pacific, and Europe are all statistically different from the Middle East and Africa as well as from each other. Using photovoltaic power potential as an indicator for UV exposure, this indicates that the Middle East and Africa experience the highest amount of UV radiation. Additionally, the relationship between photovoltaic power as an indicator of crude rate was analyzed. The test does not support the hypothesis as there is a negative relationship between photovoltaic power potential and crude rate, meaning that as photovoltaic power potential increases, the crude rate decreases. However, an r-squared value of 38.79% was observed. This indicates that roughly 38% of the variation in skin cancer rates can be explained by the amount of UV exposure that region receives.

Analysis Three: GDP Per Capita(2020)

The data for the GDP per capita was obtained from The World Bank's Data Catalog. GDP per capita is the GDP of a country divided by the total population, or the amount of money earned per person in a country. GDP per capita is typically used as a measure of standard living and indicates prosperity and overall well being in a country. A GDP per capita that is low usually indicates that a country is struggling to supply its inhabitants with necessities. Using this as a frame of reference, an analysis of the relationship between GDP per capita and geographic region will provide insight as to how GDP may differ between regions. From there, the relationship between GDP per capita and crude rate will be analyzed to further indicate to what extent GDP per capita can be used as an indicator for skin cancer.

Hypotheses:

1. **GDP per capita vs Geographic region**

Null Hypothesis (H_0): There is no difference in the GDP per capita in 2020 for the defined geographic regions.

Alternative Hypothesis (H_1): There is a difference in the GDP per capita in 2020 for the defined geographic regions.

2. **GDP per capita vs Crude Rate**

Hypothesis: If a region has a lower GDP per capita, it will have a higher crude rate.

Data:

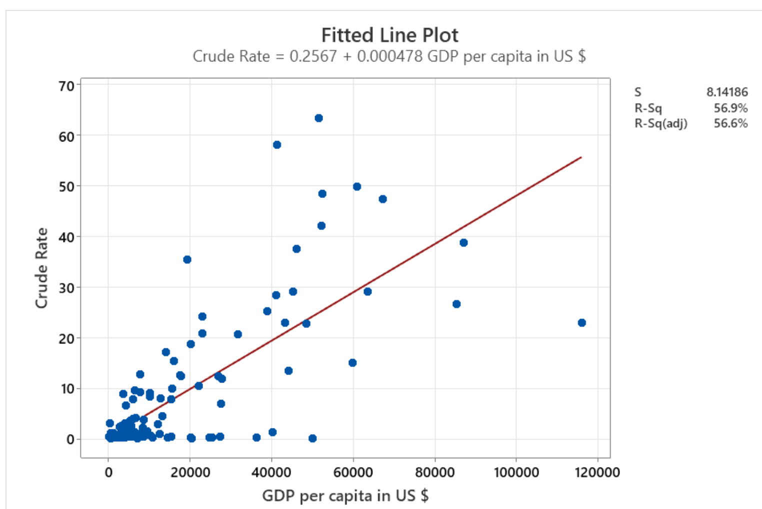
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Region	4	15820301553	3955075388	13.88	0.000
Error	150	42748086941	284987246		
Total	154	58568388494			

Grouping Information Using the Tukey Method and 95% Confidence

Region	N	Mean	Grouping
Europe	40	29032	A
Middle East and North Africa	16	15191	B
Americas	26	10683	B
Asia-Pacific	30	10250	B
Sub-Saharan Africa	43	2085	B

Means that do not share a letter are significantly different.



Model Summary

S	R-sq	R-sq(adj)
8.14186	56.91%	56.63%

Evaluation:

The data provides evidence that we can reject the null hypothesis and support the alternative hypothesis, indicated by the p-value of 0.000 and the Tukey comparisons. Based on the Tukey comparisons, Europe is statistically different from the other four regions which are statistically similar. For the context of this analysis, this indicates that Europe has a relatively higher GDP per capita than that of all the other regions. Additionally, the relationship between GDP per capita as an indicator of crude rate was analyzed. The test does not support the hypothesis as there is a positive relationship between GDP per capita and crude rate, meaning that as GDP per capita increases, the crude rate decreases. However, an r-squared value of 56.91% was observed. This indicates that roughly 57% of the variation in skin cancer rates can be explained by differences in the GDP per capita by region.

Analysis Four: STC Health Index (2020)

The STC Health Index was obtained from Hudon's Global Residence Index in conjunction with data obtained from the World Health Organization's 2020 Health Statistics Monitoring. The STC Health Index is a country based index made by Hudson using 13 different variables that factor into the efficiency of a country's healthcare system. Some of these variables include physician rate per 10,000 inhabitants, cancer death rate from 30-70 years, life expectancy, etc. Hudson uses the information provided by the WHO to curate the index, giving every country a rating from 0.00 to 1.00. For this analysis, the STC Health Index will be used as an indicator for the quality of the healthcare system in each country. It can be expected that a country with a lower STC health index has inadequate access to medical resources and understanding of healthy life practices and thus have a decreased understanding of proper prevention techniques. Using this as a frame of reference, an analysis of the relationship between the STC Health Index and geographic region will provide insight on how different regions measure on the index. From there, the relationship between the STC Health Index and the crude rate will be analyzed to further indicate to what extent a country's STC score can be used as an indicator for skin cancer.

Hypotheses:

1. **STC Health Index vs. Geographic region**

    Null Hypothesis (H_0): There is no difference in the relative STC Health Index values in 2020 for the defined geographic regions.

    Alternative Hypothesis (H_1): There is a difference in the relative STC Health Index values in 2020 for the defined geographic regions.

2. **STC Health Index vs. Crude rate**

    Hypothesis: If a region has a lower STC health index, it will have a higher crude rate.

Data:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Region	4	9.337	2.33426	67.13	0.000
Error	153	5.320	0.03477		
Total	157	14.657			

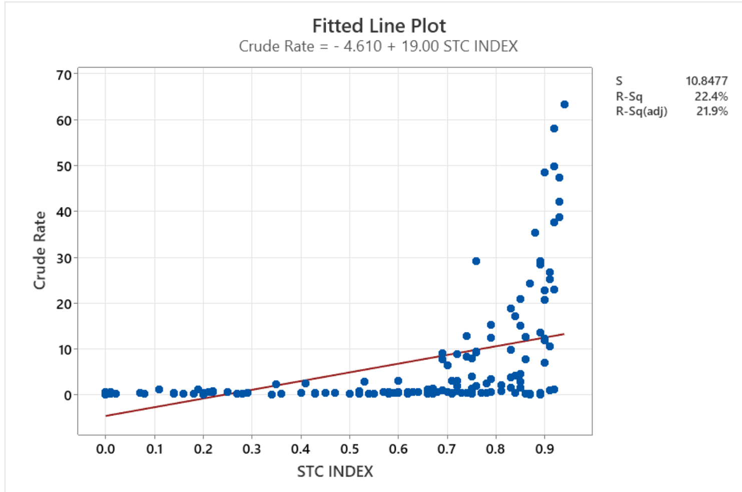
Grouping Information Using the Tukey Method and 95% Confidence

Region	N	Mean	Grouping		
Europe	40	0.8275	A		
Middle East and North Africa	17	0.7582	A	B	
Americas	26	0.7162	A	B	
Asia-Pacific	31	0.6048		B	
Sub-Saharan Africa	44	0.2152			C

Means that do not share a letter are significantly different.

Model Summary

S	R-sq	R-sq(adj)
10.8477	22.38%	21.88%



Evaluation:

The data provides evidence that we can reject the null hypothesis and support the alternative hypothesis, indicated by the p-value of 0.000 and the Tukey comparisons. Based on the Tukey comparisons, Europe, the Middle East and North Africa, and the Americas are statistically similar. The Middle East and North Africa, the Americas, and Asia-Pacific are also statistically similar to each other. Sub-Saharan Africa shares no statistical similarities. For the context of this analysis, this indicates that Europe has the highest relative STC index value but it is not significantly different from the Middle East and the Americas. Sub-Saharan Africa has the lowest STC index value. Additionally, the relationship between STC index as an indicator of crude rate was analyzed. The test does not support the hypothesis as there is a positive relationship between STC value and crude rate, meaning that as the STC index value increases, the crude rate increases. However, an r-squared value of 22.38% was observed. This indicates that roughly 22% of the variation in skin cancer rates can be explained by the quality of the healthcare system of the region.

Analysis Five: Multiple Regression of all Variables as Indicators of Crude Rate

In addition to all variables being looked at individually, they can also be looked at together to determine how much of the variation in crude rate can be explained by all three factors combined. This provides a better understanding of how all of the variables measure as factors that vary by region and can impact an individual's likelihood of developing cancer.

Data:

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.53830	63.54%	62.82%	58.09%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	14956.7	4985.58	87.73	0.000
photovoltaic power potential	1	1432.0	1432.03	25.20	0.000
STC INDEX	1	0.1	0.06	0.00	0.974
GDP per capita in US $	1	5021.7	5021.68	88.37	0.000
Error	151	8580.7	56.83		
Total	154	23537.5			

Evaluation:

Overall, all three variables account for 63.5% in the variation of the crude rates for the different regions. However, photovoltaic power potential and GDP per capita are the only variables that have significant explanatory power for the variance in crude rate. This is indicated by the p-value of 0.00. The STC index does not have significant explanatory power for the variances in crude rate.

Conclusion and Summary of Findings

This analysis served to better understand how different external variables, variables that are outside of an individual's control, may impact their likelihood of developing skin cancer based on their geographic location. First, it was determined using ANOVA that Europe experienced the largest number of skin cancer diagnoses in 2020 and was statistically significant as compared to the other regions. This indicates that the frequency of cancer diagnoses in 2020 did vary by region. Second, it was determined using ANOVA that the Middle East and North Africa had the highest photovoltaic power potential and thus the highest amount of UV exposure. It was hypothesized that the region with the highest UV exposure would have the highest number of skin cancer diagnoses, however this was not supported by the data. Additionally, it was determined using regression that UV exposure accounts for approximately 39% of the variance in cancer rate by region. Third, it was determined using ANOVA that Europe has the highest collective GDP per capita of all of the regions in 2020 and is statistically different from the other regions. It was hypothesized that the region with the lowest GDP per capita would have the highest number of skin cancer diagnoses, however this was not supported by the data. Additionally, it was determined using regression that GDP per capita accounts for approximately 57% of the variance in cancer rate by region. Fourth, it was determined using ANOVA that Europe has the highest relative STC index value, however its value is statistically significant to that of the Americas and the Middle East and North Africa. It was hypothesized that the region with the lower STC index value would have a higher number of skin cancer diagnoses, however this was not supported by the data. This could be due to the inability to get a proper diagnosis in countries with a low STC, meaning that countries with a higher STC may only have a higher crude value because they have access to proper medical care. Additionally, it was determined using regression that the STC value accounts for approximately 22% of the variance in cancer rate by region. Lastly, it was determined using multiple regression that photovoltaic power potential and GDP per capita account for approximately 64% of the variance in cancer rate by region.

It is evident that the data and findings did not fully support the hypothesized claims. Based on the multiple regression, there are other geographic variables that can significantly impact the likelihood of developing skin cancer that fall outside of the scope of this analysis. It is important to understand that despite the numerous resources, this analysis examined the entire globe and there are limitations that caused not every region to be accurately surveyed and displayed. However, with the data that was available, it can be concluded that skin cancer rates in 2020 did vary by region in that each region experienced different levels of UV radiation, had a different GDP per capita, and had varying amounts of health care resources. These are just three variables of many that individuals have no control over, yet can greatly impact their health. Cancer takes thousands of lives every year and it is important that there are continued efforts towards understanding how and what causes cancer and what can be done to counteract it.

Works Cited

Centers for Disease Control and Prevention. (2020, July 29). *Healthy choices to lower your cancer risk*. Centers for Disease Control and Prevention. Retrieved May 1, 2022, from https://www.cdc.gov/cancer/dcpc/prevention/other.htm

Solargis. (n.d.). Global solar atlas. Retrieved May 1, 2022, from https://globalsolaratlas.info/map?c=26.194877%2C6.943359%2C3

STC Health Index - Country Ranking of healthcare in the world. Global Residence Index. (2021, October 12). Retrieved May 1, 2022, from https://globalresidenceindex.com/hnwi-index/health-index/

The World Bank. (n.d.). *GDP per capita (current US$) - middle east & north africa*. Data. Retrieved May 1, 2022, from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=ZQ&most_recent_year_desc=false

World Health Organization. (n.d.). *Cancer Today*. Global Cancer Observatory. Retrieved May 1, 2022, from https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=population&mode_population=regions&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=0&include_nmsc_other=1

\