Probability and Statistics Project

GLOBAL INFLUENCE OF NUTRITIONAL PATTERNS ON COVID-19

Group 6

TABLE OF CONTENTS

“HEALTH REQUIRES HEALTHY FOOD." ROGER WILLIAMS (1603 – 1683)

45,429,824

Total reported tests in the US

4,146,543

Positive tests reported in the US

Disclaimer

The data used in this analysis is compiled from publicly available sources including the Food and Agriculture Organization of the United Nations, the Population Reference Bureau, the Johns Hopkins Center for Systems Science and Engineering, and ChooseMyPlate.gov. While we strive to ensure the accuracy of the data, we make no warranties or representations of any kind, express or implied, about the completeness, accuracy, reliability, or suitability of the data for any purpose. This data is provided for educational and informational purposes only and should not be construed as professional advice or a substitute for consultation with qualified public health experts. The analysis and conclusions drawn are solely those of the author. Any medical information is general and should not be used to diagnose or treat a health condition without consulting a qualified healthcare professional.

Introduction

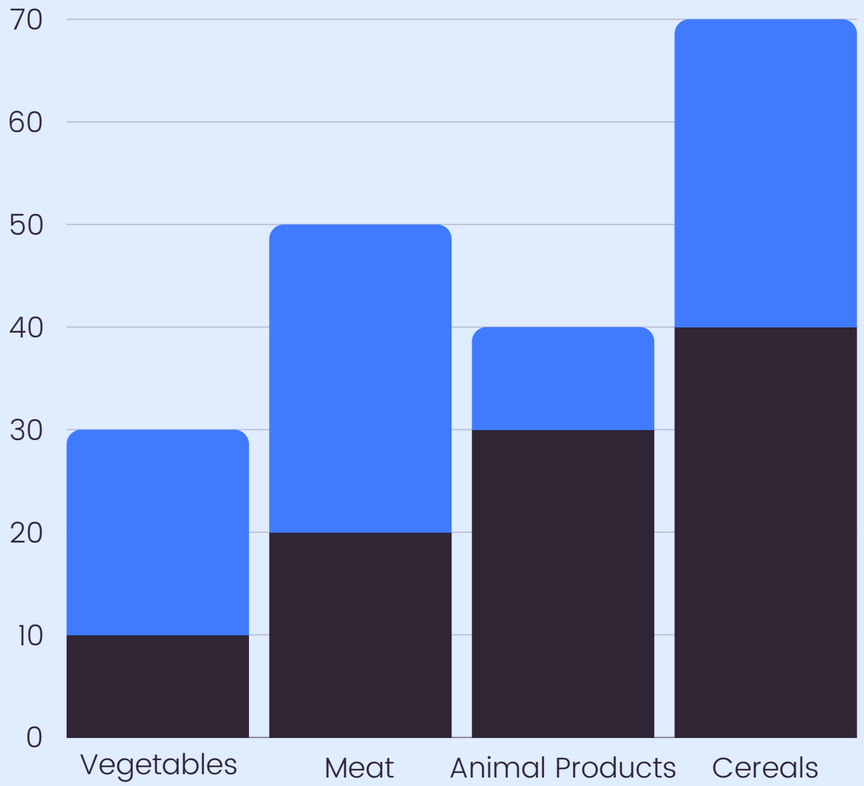
- This project investigates the link between diet patterns and COVID-19 outcomes on a global scale.
- By analyzing international data on food supplies, obesity/hunger levels, and COVID-19 cases/deaths per country, the goal is to identify correlations that can inform dietary recommendations for building immunity and resilience against future pandemics.
- The study aims to guide research on lifestyle risk factors for infectious diseases, offering valuable insights into the role of healthy diets in combating diseases like COVID-19 with less severity.

Problem Statement

The project aims to analyze the impact of diets on COVID-19 recovery in 170 countries using 2020 data. It will use descriptive statistics and hypothesis testing to identify key food groups and their intake to enhance resistance against COVID-19.

Data Collection

- Data collected from 4 major international organizations
- Covers 170 countries
- Food consumption data analyzed with COVID-19 statistics



Data Source - Kaggle

Data Preparation

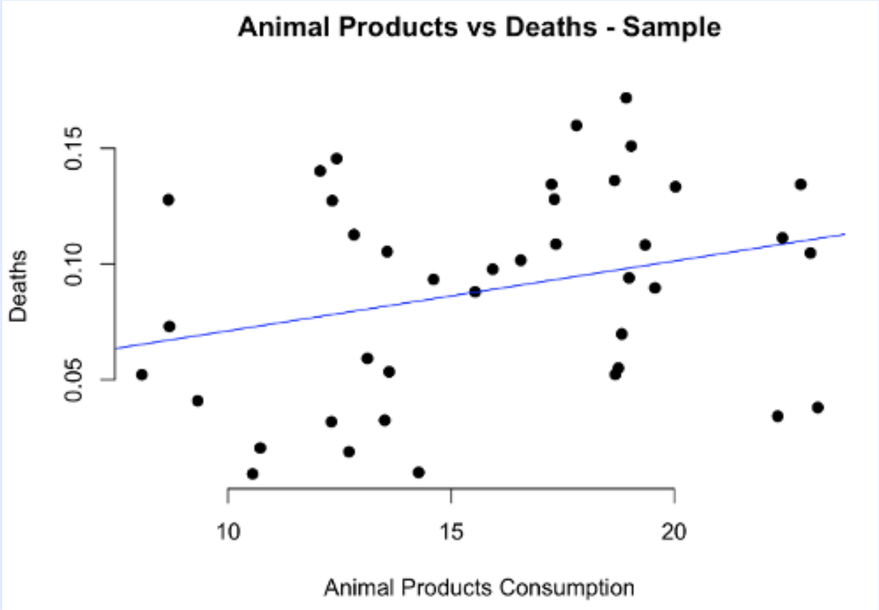
	Country	Animal.Products	Cereals..	Meat	Sugar..	Vegetables	Confirmed	Deaths	Recovered	Population
2	Montenegro	22.8328	5.0447	3.3604	6.0418	5.0001	10.4081993569132	0.134405144694534	9.03987138263666	622000
3	Czechia	17.8065	5.6937	4.7618	3.8081	4.5289	9.61284061216872	0.159845091452034	8.55532848077641	10716000
4	Luxembourg	14.6038	5.1366	4.0469	8.1121	4.8536	8.1506329113924	0.0933544303797468	7.66155063291139	632000
5	Slovenia	18.9196	7.6345	4.161	2.4915	5.002	8.2359010936757	0.171754636233952	7.31293390394674	2103000
6	Georgia	15.5347	15.6253	2.6217	3.311	4.0993	7.04226110363392	0.0879946164199192	6.80524899057873	3715000
7	Panama	12.3346	14.2627	4.98	4.2277	1.948	7.62232080317535	0.127363997198226	6.77385010506654	4283000
8	Israel	18.7463	7.6277	5.5003	1.7747	8.8586	7.43905164930556	0.0550238715277778	6.60963541666667	9216000
9	Portugal	18.6614	6.2984	4.5168	1.7316	7.3366	7.42960507069722	0.136070209653827	5.84683568990736	10255000
10	Croatia	17.311	5.7726	3.6945	3.088	12.5188	5.84890547263682	0.127960199004975	5.64870646766169	4020000
11	Armenia	13.564	7.2982	2.2675	2.6579	16.7019	5.68122462787551	0.105345060893099	5.3984100135318	2956000
12	Lithuania	23.049	6.4732	3.7899	4.502	4.4045	6.66707229778096	0.104760200429492	4.93962061560487	2794000
13	Austria	19.5654	6.2116	4.681	2.6094	5.1098	4.73998205070675	0.0896791563832174	4.49687009199013	8914000
14	Slovakia	18.9841	5.4026	4.3096	5.5166	4.5601	4.7926400585866	0.094013181984621	4.36142438667155	5462000
15	Colombia	12.821	8.1055	3.554	4.7518	3.0981	4.35079483860529	0.112638540571151	4.09014440579241	49444000
16	North Macedonia	12.0639	7.3937	2.0683	2.9104	14.5512	4.54802884615385	0.140192307692308	4.02572115384615	2080000
17	Brazil	17.347	8.4292	6.4784	2.7755	3.307	4.46016514645063	0.108602912016316	3.97946858534927	211812000
18	Argentina	19.3454	8.4102	7.0421	3.0536	4.3503	4.35614738744298	0.108226634638694	3.90519205765035	45377000
19	Switzerland	22.4227	5.9818	3.6314	2.9737	4.9478	6.16021542738013	0.111257817929117	3.67848042622191	8634000
20	Romania	15.9287	8.0412	3.0499	1.4126	9.3641	3.86332830933943	0.0977547944493529	3.58335845330284	19241000

Data Preparation

- Step 1 - Importing Datasets
- Step 2 - Handling Missing Values
- Step 3 - Removing rows with missing values
- Step 4 - Removing rows with duplicate data

Data Visualization

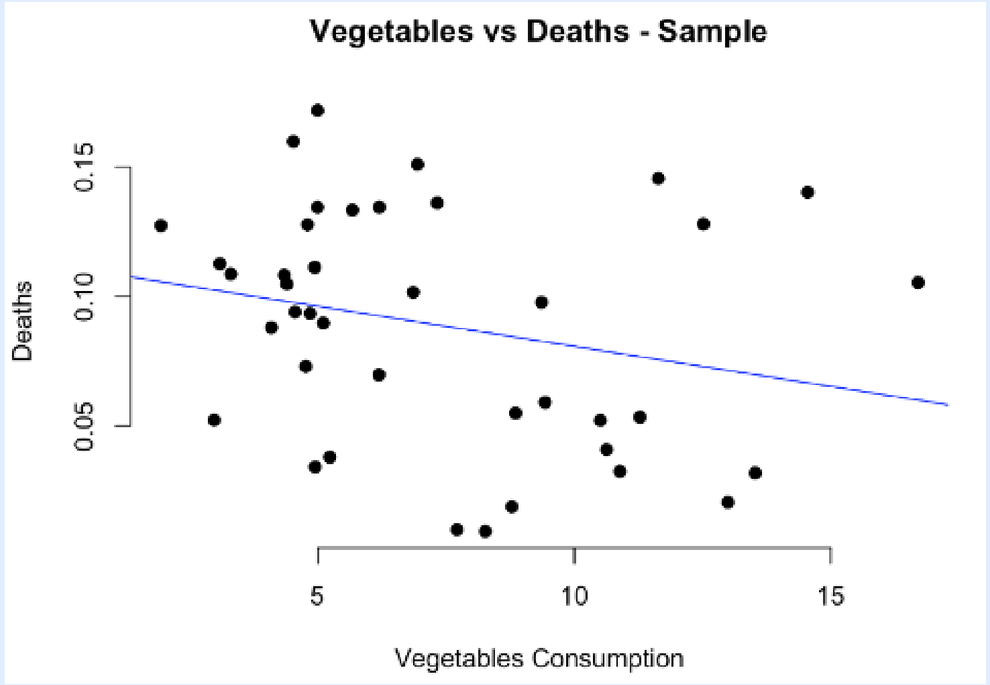
Scatter Plots - Animal Products



The scatter plot reveals a weak positive correlation between COVID-19 death percentages and animal product consumption. As consumption increases, death percentages also rise, with a regression value around 0.28 suggesting a potential impact on fatality.

Data Visualization

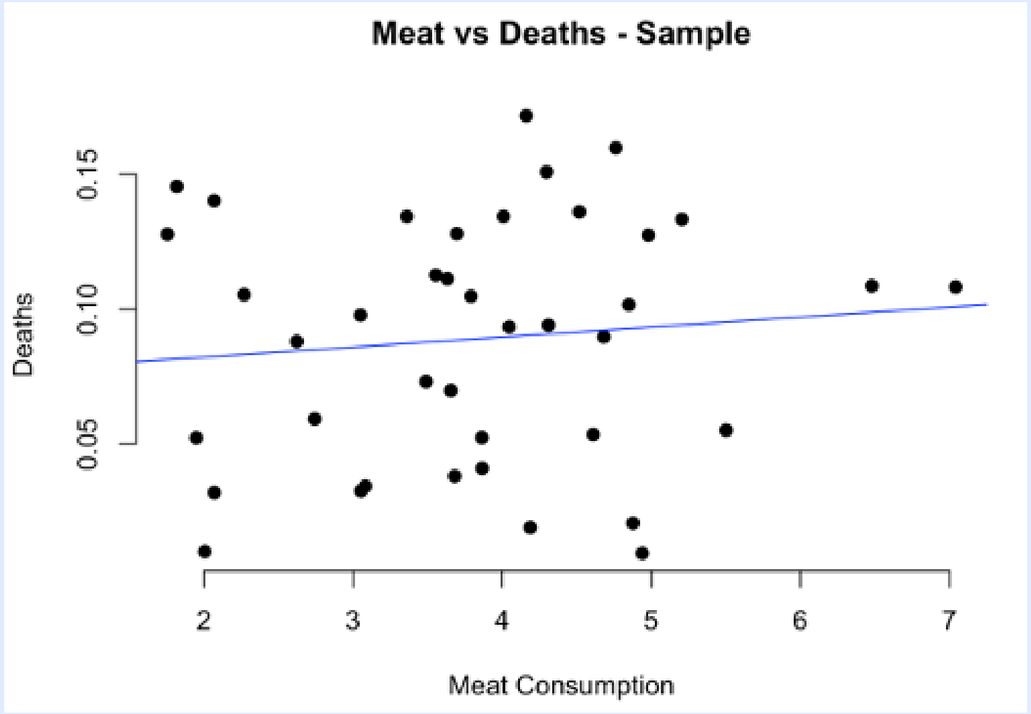
Scatter Plots - Vegetables



The scatter plot shows a moderately negative correlation with a regression value around 0.24 between COVID-19 death percentages and vegetable consumption, with increased vegetable consumption resulting in a decrease in death percentages, although outliers exist, suggesting potential benefits.

Data Visualization

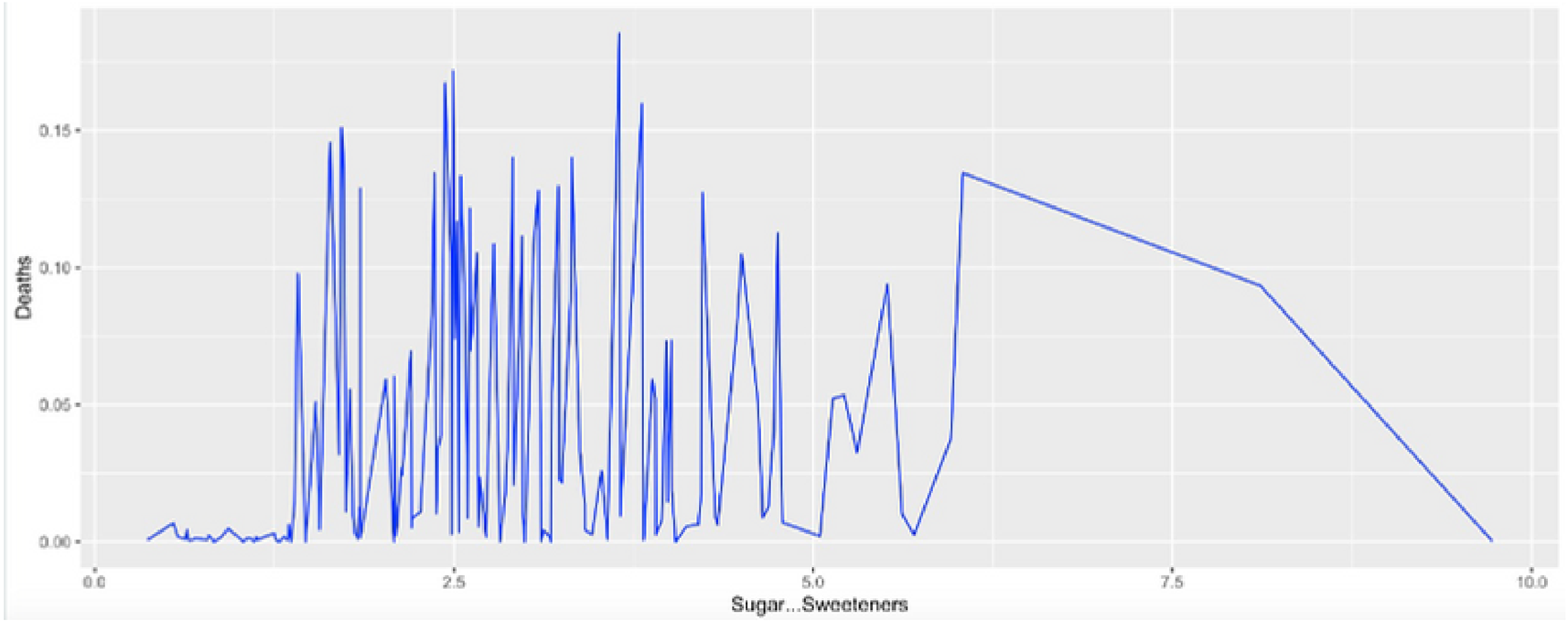
Scatter Plots - Meat



The scatter plot reveals a moderate positive correlation with a regression value around 0.1 between COVID-19 death percentages and meat consumption, with higher meat consumption correlated with higher death rates, suggesting a potential impact on fatality, with noticeable outliers and spread in the scatter.

Data Visualization

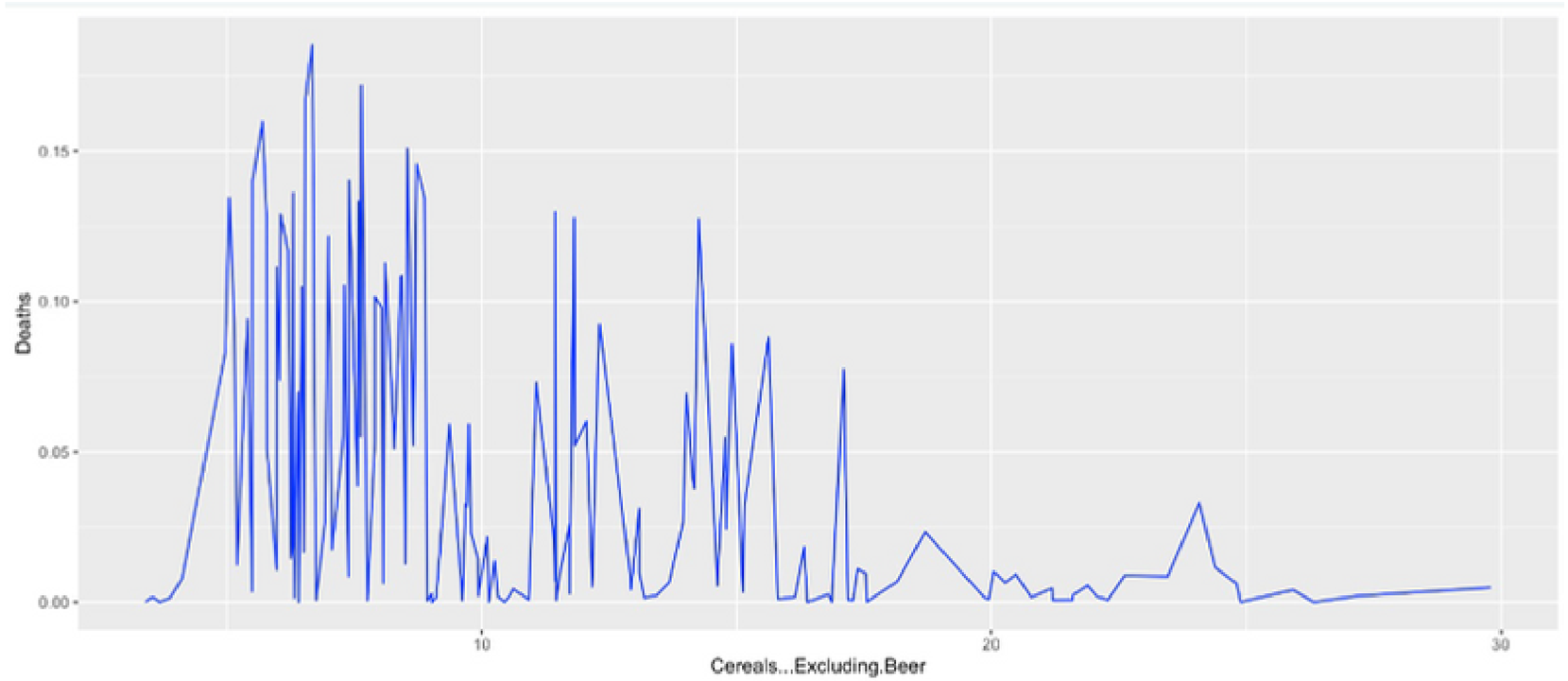
Line Charts - Sugar and Sweetners



The graph reveals a positive correlation with a regression value of 0.117 between sugar and sweetener intake and COVID-19 deaths in 2020, with Belgium showing the highest mortality rate at 0.18%, suggesting a possible link that requires further study.

Data Visualization

Line Charts - Cereals



The graph shows negative correlation with a regression value of 0.31 indicating that higher cereal consumption corresponds with lower COVID-19 death rates, with Belgium having the highest death rate at 0.18%. However, further analysis is needed to confirm this and determine if cereal consumption actually protects against COVID-19 deaths.

Hypothesis Testing - Z-test

Is a particular food group affecting death/recovery rates from COVID-19?

NULL HYPOTHESIS H0:

Consumption of the particular food group doesnt affect death/recovery rates from COVID-19

Alternative Hypothesis Ha:

Consumption of the particular food group above the true mean consumption which either decrease/increase death rates from COVID-19.

Alternative Hypotheses

We interpret scatter plots with regression lines to formulate the hypothesis

ANIMAL PRODUCTS

If the consumption percentage of animal products exceeds the true mean consumption(>12.1218%) this would increase the death rates from COVID-19.

SUGAR

If the consumption rate of sugar exceeds the true mean consumption(>2.7324), this would positively affect the immunity and cause an decrease in death rates.

MEAT

If the consumption rate of meat exceeds the true mean consumption(>5.5697), this would adversely affect the immunity and cause an increase in death rates.

VEGETABLES

If the consumption rate of vegetables exceeds the true mean consumption(>5.9715%), this would boost the immunity and recovery rate, thus reducing death rates..

CEREALS:

If the consumption rate of cereals exceeds the true mean consumption(>12.0050), this would boost the immunity and recovery rate, thus reducing death rates.

Alternative Hypotheses

If p>0.05 -> accept Null Hypothesis
If p<0.05 -> accept Alternative Hypothesis

MEAT

Accept Null Hypothesis
Consumption less than true mean -> doesn't affect COVID-19 Death Rate.

ANIMAL PRODUCTS

Accept the Alternative Hypothesis
Increase in consumption -> Decrease in immunity -> Increase in Death Rate

VEGETABLES

Accept the Alternative Hypothesis
Increase in consumption -> Increase in immunity -> Decrease in Death Rate

SUGAR

Accept the Alternative Hypothesis
Increase in consumption -> Increase in immunity -> Decrease in Death Rate

CEREALS:

Accept Null Hypothesis
Consumption less than true mean -> doesn't affect COVID-19 Death Rate..

1 SAMPLE T -TEST

WORKSHEET 3

One-Sample T: Animal.Products, Meat, Sugar...Sweeteners, Vegetables, Cereals...Excluding.Beer

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean	95% CI for μ
Animal.Products	20	17.544	3.317	0.742	(15.991, 19.096)
Meat	20	4.141	1.260	0.282	(3.551, 4.730)
Sugar...Sweeteners	20	3.473	1.665	0.372	(2.694, 4.253)
Vegetables	20	6.574	3.948	0.883	(4.727, 8.422)
Cereals...Excluding.Beer	20	7.669	2.748	0.614	(6.383, 8.955)

μ: mean of Animal.Products, Meat, Sugar...Sweeteners, Vegetables, Cereals...Excluding.Beer

Animal Products:
A comprehensive analysis of the 20 most COVID-19-recovered countries reveals that the overall consumption of animal products varies between 15.991% and 19.096%.

Meat:
A global analysis of the 20 most recovered countries indicates that the overall consumption of meat products ranges from 3.551% to 4.73%.

Sugar Sweeteners:
Across 20 most COVID-19-recovered countries, the combined consumption of sugar sweeteners from animal products varies between 2.694% and 4.253%.

Vegetables:
In the examination of the 20 most COVID-19-recovered countries, the overall consumption of animal products within vegetables ranges from 4.727% to 8.422%.

Cereals...Excluding Beer:
A holistic assessment of the 20 most COVID-19-recovered shows that the collective consumption of animal products in cereals (excluding beer) ranges from 6.383% to 8.955%.

2 SAMPLE T -TEST

WORKSHEET 5

Two-Sample T-Test and CI: Vegetables (most recovered), Vegetables (least rec)

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Vegetables (most recovered)	20	6.57	3.95	0.88
Vegetables (least rec)	20	5.43	3.69	0.82

Estimation for Difference

Difference	95% CI for Difference
1.14	(-1.31, 3.59)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
0.95	37	0.350

WORKSHEET 5

Two-Sample T-Test and CI: Meat (most recovered countries), Meat (least recovered)

Method

μ_1: mean of Meat (most recovered countries)

μ_2: mean of Meat (least recovered)

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Meat (most recovered countries)	20	4.14	1.26	0.28
Meat (least recovered)	20	3.21	1.70	0.38

Estimation for Difference

Difference	95% CI for Difference
0.933	(-0.027, 1.892)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.97	35	0.056

Vegetables :

Highest recovery rates -> average consumption of vegetables is 6.57%.
Lowest recovery rates -> average consumption of vegetables is 5.43%.
There is a notable and significant difference in vegetable consumption.

Meat :

Highest recovery rates -> average consumption of meat is 4.14%.
Lowest recovery rates -> average consumption of meat is 3.21%.
There is a notable and significant difference in meat consumption.

2 SAMPLE T -TEST

WORKSHEET 5

Two-Sample T-Test and CI: Animal.Products (Most recoverd), Animal.Products (Least

Method

μ_1: mean of Animal.Products (Most recoverd)
μ_2: mean of Animal.Products (Least rec)
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Animal.Products (Most recoverd)	20	17.54	3.32	0.74
Animal.Products (Least rec)	20	10.35	6.29	1.4

Estimation for Difference

Difference	95% CI for Difference
7.19	(3.94, 10.45)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
4.52	28	0.000

WORKSHEET 5

Two-Sample T-Test and CI: Sugar...Sweeteners, Sugar...Sweeteners (least rec)

Method

μ_1: mean of Sugar...Sweeteners
μ_2: mean of Sugar...Sweeteners (least rec)
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sugar...Sweeteners	20	3.47	1.67	0.37
Sugar...Sweeteners (least rec)	20	2.59	2.09	0.47

Estimation for Difference

Difference	95% CI for Difference
0.885	(-0.327, 2.097)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.48	36	0.147

Animal products:

Highest recovery rates -> average consumption of animal products is 17.54%.
Lowest recovery rates -> average consumption of animal products is 10.35%.
There is a notable and significant difference in animal product consumption.

Sugar and Sweeteners:

Highest recovery rates -> average consumption of sugar and sweeteners is 3.47%.
Lowest recovery rates -> average consumption of sugar and sweetners is 2.59%.
There is a discernible difference in sugar and sweetener consumption.

2 SAMPLE T –TEST

WORKSHEET 5

Two-Sample T-Test and CI: Cereals (most recovered), Cereals (Least r

Method

μ_1: mean of Cereals (most recovered)
μ_2: mean of Cereals (Least rec)
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Cereals (most recovered)	20	7.67	2.75	0.61
Cereals (Least rec)	20	13.30	7.64	1.7

Estimation for Difference

Difference	95% CI for Difference
-5.63	(-9.39, -1.88)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-3.10	23	0.005

Cereals:

Highest recovery rates -> average consumption of cereals is 7.67%.

Lowest recovery rates -> average consumption of cereals is 13.30%.

There is a discernible difference in cereals consumption.

ANOVA-Test

We have used to assess if there are any statistically significant differences between the means of the Death Rate, Recovery Rate and Confirmed Cases among the countries in the continents of Asia, Africa, Australia, Europe, North America, and South America.

Null Hypothesis H0: Mean Death rate/ Recovery Rate / Confirmed Cases of all continents are equal.

Alternative Hypothesis Ha: Not all means of Death Rate/ Recovery Rate/ Confirmed Cases are equal.

1. Death Rate

Method

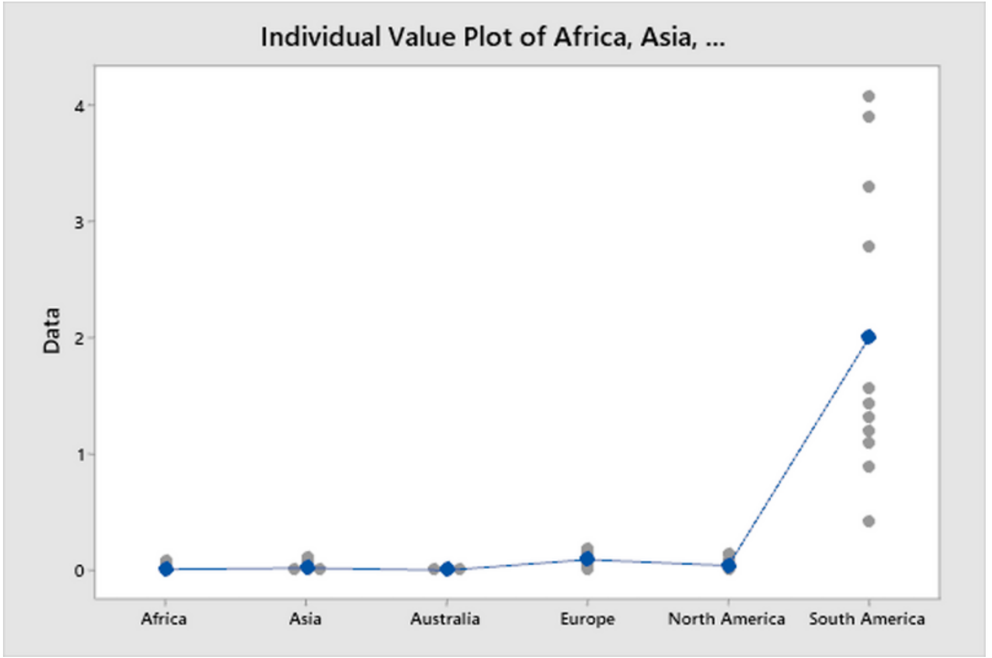
Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	α = 0.05
Rows unused	116

Equal variances were assumed for the analysis.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	5	39.55	7.9096	70.21	0.000
Error	148	16.67	0.1127		
Total	153	56.22			

Factor	N	Mean	StDev	95% CI
Africa	45	0.00864	0.01617	(-0.09024, 0.10751)
Asia	36	0.01968	0.02592	(-0.09087, 0.13022)
Australia	6	0.000709	0.001396	(-0.270063, 0.271481)
Europe	39	0.09205	0.04901	(-0.01415, 0.19826)
North America	17	0.0402	0.0484	(-0.1207, 0.2011)
South America	11	2.001	1.285	(1.801, 2.201)



We reject the Null Hypothesis and accept the Alternative Hypothesis

2. Recovery Rate

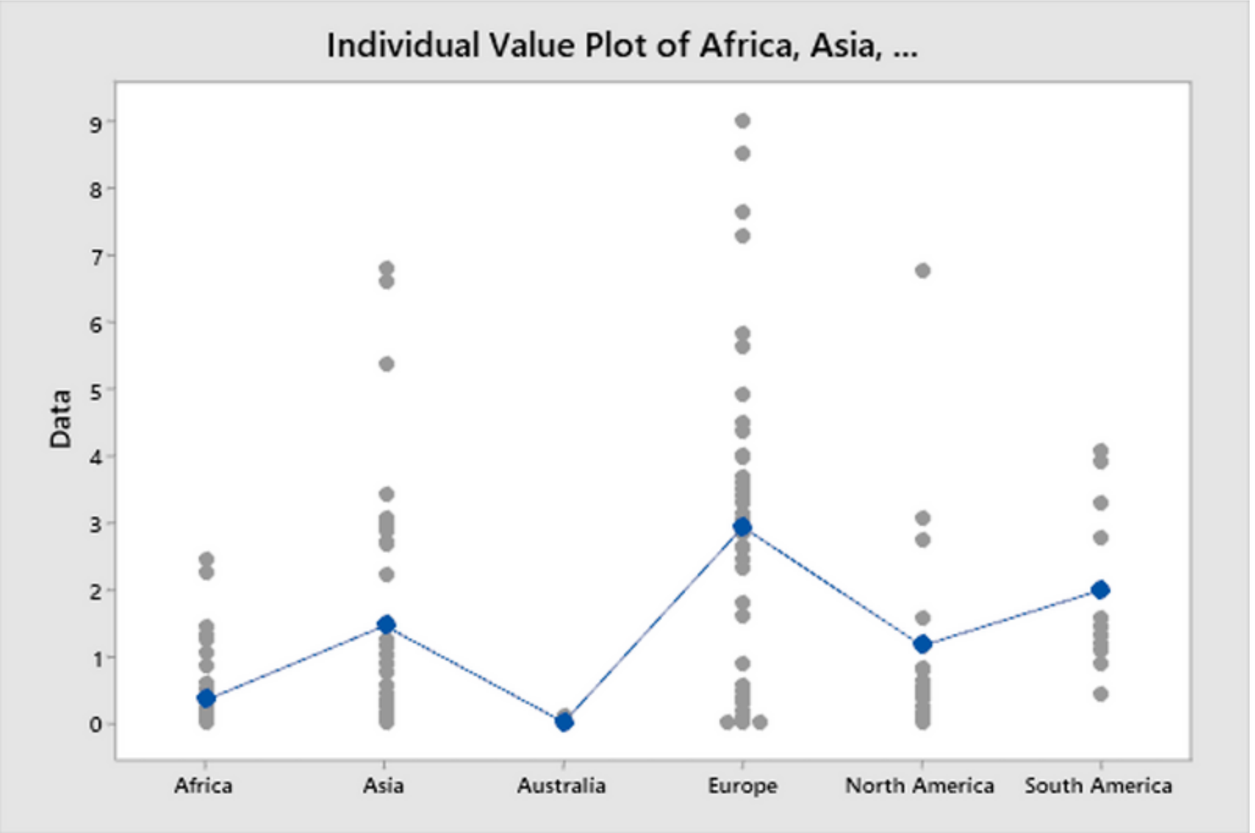
Means

Factor	N	Mean	StDev	95% CI
Africa	45	0.3547	0.5678	(-0.1443, 0.8537)
Asia	36	1.467	1.844	(0.910, 2.025)
Australia	6	0.0258	0.0408	(-1.3407, 1.3922)
Europe	39	2.942	2.454	(2.406, 3.478)
North America	17	1.168	1.695	(0.356, 1.980)
South America	11	2.001	1.285	(0.992, 3.010)

Pooled StDev = 1.69379

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	5	157.7	31.538	10.99	0.000
Error	148	424.6	2.869		
Total	153	582.3			



We reject the Null Hypothesis and accept the Alternative Hypothesis

3. Confirmed Cases

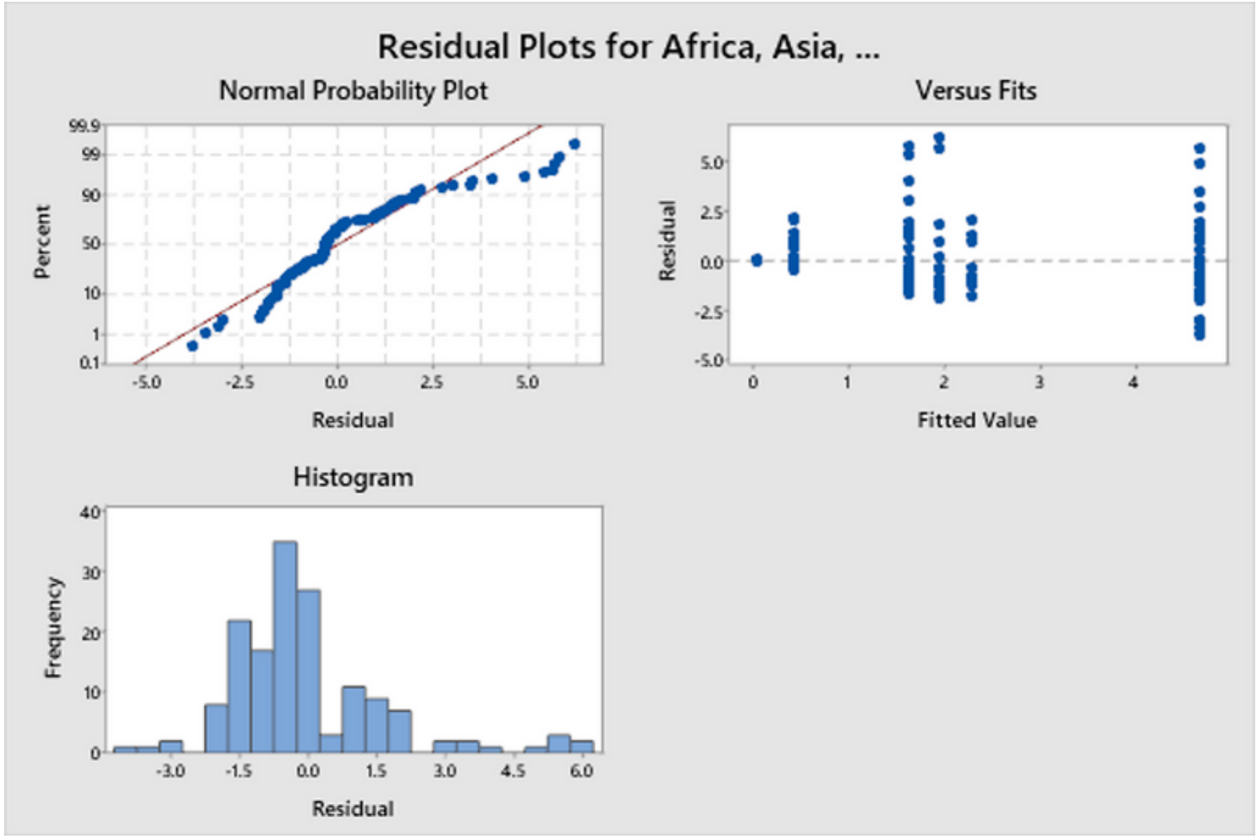
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	5	419.9	83.988	27.09	0.000
Error	148	458.8	3.100		
Total	153	878.7			

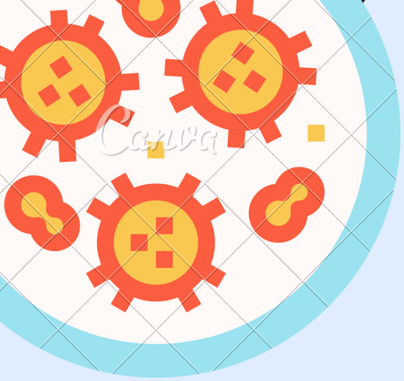
Means

Factor	N	Mean	StDev	95% CI
Africa	45	0.4169	0.6280	(-0.1017, 0.9356)
Asia	36	1.618	2.023	(1.038, 2.198)
Australia	6	0.0281	0.0448	(-1.3923, 1.4485)
Europe	39	4.671	2.191	(4.114, 5.228)
North America	17	1.937	2.462	(1.093, 2.781)
South America	11	2.271	1.370	(1.222, 3.320)

Pooled StDev = 1.76068



We reject the Null Hypothesis and accept the Alternative Hypothesis



Results and Conclusion

- Comparing countries with the most and least recovery rates,
- 95% confidence intervals showed - Positive differences in consumption for vegetables, meat, animal products, and sugar for the higher recovery countries.
- Cereals showed lower consumption for higher recovery countries.
- Hypothesis tests found that increased consumption of animal products and sugar above the population average significantly increased COVID-19 death risk.
- Increased vegetable consumption above the population average lowered COVID-19 death risk.
- Meat and cereal consumption below their respective population averages showed no significant impact on COVID-19 immunity or deaths.
- ANOVA tests found statistically significant differences in mean COVID-19 death rate, recovery rate and confirmed cases between different continents.

In summary, the analysis suggests potential benefits of diets higher in vegetables and lower in animal products/sugar for resilience against COVID-19, while higher meat and cereals showed mixed impacts.

Limitations

Figuring out if what we eat directly causes COVID-19 outcomes is tricky due to many factors, like

- Long-term effects
- Individual differences
- Ethical concerns in studying patients

Proposed Next Steps and Future Work

- Enhanced Data Precision
- Extended Analysis Duration
- Inclusive Variable Consideration
- Practical Intervention Testing
- Experiment with interventions, such as public health communications, for practical insights

GROUP 6

Probablity and Statistics Final Project

Shalini Dutta

Grandhi Venkata Kishan Madhav

Siddhi Yeshwant Sonwalkar

Navisha Shetty

Jonna Jaidhitya

Thank you!