7COM1079-0901-2024

Team Research and Development Project

Final report title: Fat Supply Quantity Data

Group ID: A106

Dataset number: DS163

Prepared by ::Deon Jude Dsouza - 23080134  
 Pritam Anil Chitte - 23092446  
 Gladson Dcosta - 23079914  
 Joystan Brayan Lobo - 23079720  
 Pavan Davangere Vageesh - 23072485

University of Hertfordshire

Hatfield, 2024

Table of Contents

[1. Introduction 3](#_Toc187087054)

[1.1. Problem statement and research motivation 3](#_Toc187087055)

[1.2. The data set 3](#_Toc187087058)

[1.3. Research question 3](#_Toc187087060)

[1.4. Null hypothesis and alternative hypothesis (H0/H1) 3](#_Toc187087062)

[2. Background research 4](#_Toc187087067)

[2.1. Research papers 4](#_Toc187087068)

[2.2. Why RQ is of interest 4](#_Toc187087074)

[3. Visualisation 5](#_Toc187087076)

[3.1. Appropriate plot for the RQ output of an R script 5](#_Toc187087077)

[3.2. Additional information relating to understanding the data 5](#_Toc187087085)

[3.3. Useful information for the data understanding 5](#_Toc187087087)

[4. Analysis 6](#_Toc187087089)

[4.1. Statistical test used to test the hypotheses and output 6](#_Toc187087090)

[4.2. The null hypothesis is rejected /not rejected based on the p-value 6](#_Toc187087092)

[5. Evaluation 6](#_Toc187087094)

[5.1. What went well 6](#_Toc187087095)

[5.2. Points for improvement 6](#_Toc187087100)

[5.3. Group’s time management 7](#_Toc187087102)

[5.4. Project’s overall judgement 7](#_Toc187087104)

[5.5. Comment on GitHub log output 7](#_Toc187087106)

[6. Conclusions 8](#_Toc187087114)

[6.1. Results explained. 8](#_Toc187087115)

[6.2. Interpretation of the results 8](#_Toc187087117)

[6.3. Reasons and/or implications for future work, limitations of your study 8](#_Toc187087119)

[7. Reference list 10](#_Toc187087121)

[8. Appendices 11](#_Toc187087126)

[A. R code used for analysis and visualisation. 11](#_Toc187087127)

[B. GitHub log output.](#_Toc187087128) 15

1.Introduction

# 1.1. Problem statement and research motivation

Obesity is one of the major health issues worldwide and is associated with chronic diseases, such as diabetes and cancer (World Health Organisation, 2020). As a result, people are making inquiries on how an animal product eating affects obesity. This study will look at the relationship between the supply of animal products and obesity rates, using the hypothesis that diets loaded with animal-based foods contribute to the rise in obesity.

This study is motivated by the increasing global dependence on animal products arising from economic growth in developing countries (Popkin, 2006). Knowing how consuming animal products is linked to obesity can help in creating rules that lower health risks. The study provides valuable new information on how diets impact public health and supports evidence-based strategies toward healthier eating around the world.

# 1.2. The data set

DS163: Fat\_Supply\_Quantity\_Data.csv, is cross-sectional data from 170 countries, dietary and health variables. The data consists of 32 variables: the majority in animal product consumption-kg/person/year as the independent variable, and the percentage of obesity rates as the dependent variable. This allows for a very broad perspective of the global dietary patterns and health impacts. Its wide coverage allows major cross-country comparisons, which helps in the investigation of diet-obesity relationships. This dataset provides a solid foundation for the examination of the influence of animal product supply on public health.

# 1.3. Research question

The research question: Is there a correlation between obesity rates and the supply of animal products (kg/person/year) in different countries? Statistical analysis, in this case Pearson's correlation, will help determine this relationship using data across 170 countries. Trends are visualized using scatter plots, while p-values confirm the significance of the findings. This method guarantees a strong analysis of diet's effects on obesity.

# 1.4. Null hypothesis and alternative hypothesis (H0/H1)

This paper assesses the relationship between animal product supply and obesity rates based on the following hypotheses:

* Null Hypothesis (H₀): There is no correlation between obesity rates and the consumption of animal products (kg/person/year). This means that changes in the amount of animal products available do not greatly affect obesity levels in different countries.
* Alternative Hypothesis (H₁): There is a correlation between obesity rates and the consumption of animal products (kg/person/year). This might then imply that if there are more available animal products, there are higher rates of obesity, thus associating food habits with public health.

To test these hypotheses, Pearson’s correlation analysis is conducted. The correlation coefficient (r=0.417) and a statistically significant p-value (< 1.987x e-08) indicate a moderate positive relationship, supporting H₁ and rejecting H₀. These findings align with prior research on dietary behaviours and obesity trends.

# 2.Background research

# 2.1. Research papers

Dataset DS163, dealing with animal product supply and obesity rates, corresponds to several different studies dealing with dietary patterns and public health. Among the relevant research is:

* Popkin (2006) investigates the global nutrition transition, highlighting the shift toward diets rich in animal products due to economic growth and urbanization. This dietary transition is linked to the rising obesity rates, especially in developing countries. The study gives a background on how to contextualize DS163, focusing on dietary transitions in relation to public health.
* Swinburn et al. (2011) contextualizes obesity as a global pandemic, which is all about changed dietary patterns and physical activity. This study further explains how diets with high calorie density, normally high in animal products, result in obesity. It would, therefore, be of great relevance to investigate associations between the supply of animal products and rates of obesity in the DS163.
* Tilman and Clark (2014) explore the environmental and health implications of dietary choices, one of which is diets high in animal products. They indicate that such diets are associated with a heightened incidence of obesity and other chronic diseases, thus reinforcing the relevance of DS163's focus on animal product supply for health outcomes.

Combined, these studies provide a sound basis upon which to examine DS163. It links dietary supply data to obesity rates and underlines the general evidence that dietary patterns, especially high in animal products, drive the trends in obesity.

# 2.2. Why RQ is of interest

The research question is important because it fills a knowledge gap in the direct correlation between animal product supply and the rate of obesity at a global scale. Although literature highlights dietary transitions and their contribution to obesity (Popkin, 2006; Tilman and Clark, 2014), not many have focused on the quantitative relationship between animal product availability and the prevalence of obesity. Understanding this relationship can help in the facilitation of dietary effects on public health through appropriate interventions. However, future studies should try to find the moderating effect of socioeconomic and cultural factors for a more effective local public health approach (Swinburn et al., 2011).

# 3.Visualisation

# 3.1. Appropriate plot for the RQ output of an R script

The correlation between the supply of animal product supply (kg/person/year) and obesity rates (%) will be shown by a scatter plot. This kind of plot has been chosen because the possible correlations will appear more clearly; thus, helping answer the research question better. In a scatter plot, every single data point can be presented, and hence all patterns, clusters, or outliers in this dataset are noticeable. A trend line has been drawn showing the direction and strength of the relationship of the variables presented, just as it was done with the statistical analysis.

Key Characteristics of the Plot.

* Title: "Scatter Plot of Animal Product Supply vs. Obesity Rates"
* X-axis: "Animal Product Supply (kg/person/year)"
* Y-axis: "Obesity Rates (%)"
* Trend Line: Added to show correlation.

# 3.2. Additional information relating to understanding the data

This scatter plot is supposed to present the relationship between animal product supply (kg/person/year) and obesity rates (%) in various countries. This is done to find a pattern or probable correlation between the two variables for visual evidence to support statistical findings. The moderate positive correlation indicates that with high animal product availability, obesity rates also tend to increase. This can show outliers or clusters that may indicate regional differences or unique dietary trends that provide further detail on worldwide dietary patterns and their health consequences.

# 3.3. Useful information for the data understanding

The scatter plot shows that, from the upward trend, animal product supply and obesity rates are moderately positively correlated. Those countries that have a high supply of animal products generally have high rates of obesity, thus supporting our hypothesis that diet is the linking factor. The presence of outliers indicates variability due to factors such as regional dietary preference or socioeconomic conditions; these would have to be investigated in detail for full comprehension.

# 4.Analysis

# 4.1. Statistical test used to test the hypotheses and output

The Pearson's correlation test has been done to observe the relationship between animal product supply (kg/person/year) and the rate of obesity (%). This will, in turn, provide a measure of the strength and direction of the linear relationship between two continuous variables hence being appropriate for both the research question and dataset. From this correlation coefficient (r = 0.417) shows that a positive moderate association whereas the p value (< 1.987x e-08) shows the statistical significance. Therefore, with increased rates of obesity, the higher the supply of animal products, hence supporting the alternative hypothesis.

# 4.2. The null hypothesis is rejected /not rejected based on the p-value

From the p-value extracted from the Pearson's correlation test (< 1.987x e-08), it goes without saying that one should reject the null hypothesis (H0). The p-value, much smaller than the common significance level (α=0.05), testifies that such correlation in the data cannot be explained by mere chance. Having the correlation coefficient (r=0.417), it is possible to conclude a moderate positive correlation between animal product supply and obesity rates. This therefore means that with increased availability of animal commodities, the rates of obesity also tend to increase. Such findings support alternative hypothesis (H1), indicating a statistically significant association between the variables of interest in a manner that has been documented in earlier literature on dietary drivers of obesity trends.

# 5.Evaluation

# 5.1. What went well

* Specified Objective and Research Question: The study's goal and research question are clearly specified. It specifically tries to investigate the relationship between animal product supply and obesity trends, providing emphasis and direction.
* Comprehensive Dataset: The dataset includes 170 countries, providing a diverse and representative sample. This contributes to the generation of globally relevant knowledge.
* Effective Visualization: The histogram clearly shows the distribution of animal product supply among countries.

The scatter plot clearly shows the relationship between animal product availability and obesity rates, providing visual support for the statistical conclusions.

# 5.2. Points for improvement

Additional details on the dataset, such as the time in which the data were collected, whether it is food produced, imported, or consumed, and biases in data collection across countries, would strengthen the study. Knowledge of these characteristics serves to enhance the credibility of the findings and to assess the usefulness of the results. This would be further improved if the study controlled for other confounding variables affecting obesity, including physical activity levels, socioeconomic status, and access to other food groups such as sugar and processed foods. Such a presentation of variables would provide a fuller picture of the association of animal product supply with obesity and hence make it stronger. Again, correlation does not imply causation. While the study provides evidence on the linkage of animal product availability and the prevalence of obesity, it lacks evidence of the direct causal relationship. Considering the alternatives that may potentially exist would add even further subtlety to the continued research emphasis for unravelling direct causal connections.

# 5.3. Group’s time management

It also has a well-defined framework and methodology, logically flowing from problem definition to data collection, analysis, and presentation. That would mean the schedule was very good during the project. Moreover, by promptly finishing statistical analyses, visualisations, and conclusions, it can be understood that major research tasks received a lot of attention. The group also avoided scope creep through a properly defined study subject. This probably influenced the best use of time and ensured the analysis remained focused and relevant to the objectives of the study.

# 5.4. Project’s overall judgement

The project is a remarkable and well-structured investigation of the relationship between animal product availability and obesity rates, making an important contribution to understanding global health issues. The research is clearly described, with the research topic, methods, and findings all easily understood. The analysis is sound, providing useful insights into the relationship between animal product availability and obesity rates, and the subject is extremely significant in the context of public health. While there are several areas for improvement, such as investigating causation and adjusting for confounding variables, the study provides a solid platform for future research. The project is well-executed and has a clear purpose, giving it an excellent starting point for future research into this critical subject. With further refinement and deeper analysis, it has the potential to make an even greater impact.

# 5.5. Comment on GitHub log output

* Commit Message: "Added dataset preprocessing script"

These commits cleaned and structured the DS163 dataset to make it consistent and accurate. In fact, this was a very critical step in the pre-processing of data to be prepared for statistical analysis; missing values were handled, variables were normalized, and their formats were standardized to improve the integrity and reliability of such data for further analysis.

* Commit Message: "Added correlation analysis"

This commits applied Pearson's correlation to find the relationship between the supply of animal products and rates of obesity by computing the coefficient of correlation and the p-value to justify whether the alternative hypothesis was valid. This formed the analytical foundation for the project.

* Commit Message: "Completed data visualization and draft report"

This commits created scatter plots and summarizes the results in a report, hence making them more presentable and accessible. It was very important to deliver final insights and recommendations.

These commits give this project a structured, incremental approach, each one important for the success of the project.

# 6.Conclusions

# 6.1. Results explained.

This study shows a statistically significant, moderate, positive correlation of animal products supply (kg/person/year) and the obesity rates (%), r= 0.417. The p-value associated is less than 1.987x e-08. This association indicates that with more availability of animal products, the rate of obesity also tends to increase. The result gives evidence for accepting the alternative hypothesis that the animal product use in the diet contributes to the obese epidemic in the world. The results align with the earlier literature on appealing to the need for focused intervention in obesity and health, and sustainable diets (Popkin, 2006).

# 6.2. Interpretation of the results

These findings confirm that there is a fair positive correlation between animal product supply (kg/ person/ year) and rates of obesity (%) (r = 0.417), hence answering the research question-that dietary pattern is influencing obesity prevalence. This means that with increasing animal products being made available, there will be increased rates of obesity, and these factors are commonly identified in transitioning economies with shifting dietary habits (Popkin 2006). This relationship calls for public health initiatives to emphasize healthier diets to reduce obesity-related risks. In a broader perspective, this study offers practical lessons which can be adopted by policymakers in attempts to confront the global obesity epidemic by promoting healthy food systems (Tilman and Clark, 2014).

# 6.3. Reasons and/or implications for future work, limitations of your study

This study points out the relationship between the supply of animal products (kg/person/year) and obesity rates (%), although it is limited by the cross-sectional nature of the dataset, which does not allow causal inferences. Future work should use longitudinal data and consider socioeconomic and cultural factors that shape dietary patterns. More in-depth analysis of nutritional quality could provide deeper insights (Popkin, 2006).

# 

# 7.Reference list

* World Health Organization (2020). Obesity and overweight. Available at: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (Accessed: 26 December 2024).
* Popkin, B. M. (2006). Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. The American Journal of Clinical Nutrition, 84(2), 289–298.
* Swinburn, B. A., et al. (2011). The global obesity pandemic: shaped by global drivers and local environments. The Lancet, 378(9793), 804–814.
* Tilman, D., & Clark, M. (2014). Global diets link environmental sustainability and health. Nature, 515(7528), 518–522.

# 8. Appendices

# A. R code used for analysis and visualisation.

Code:

# Install ggplot2 if not already installed

if (!require("ggplot2")) {

install.packages("ggplot2")

}

# Load ggplot2

library(ggplot2)

# Load the dataset

fat\_data <- read.csv("C:/Users/LAB/Desktop/team research/Fat\_Supply\_Quantity\_Data.csv")

# Inspect the data structure to ensure correct loading

str(fat\_data)

head(fat\_data)

# Ensure columns exist and filter data

filtered\_data <- na.omit(fat\_data[c("Animal.Products", "Obesity")])

# Check the structure of the filtered data

str(filtered\_data)

# Histogram for Animal Products Supply

ggplot(filtered\_data, aes(x = Animal.Products)) +

geom\_histogram(binwidth = 5, fill = "blue", color = "black", alpha = 0.7) +

labs(

title = "Distribution of Animal Products Supply",

x = "Animal Products Supply (kg/person/year)",

y = "Frequency"

) +

theme\_minimal()

# Scatter Plot with Trendline

ggplot(filtered\_data, aes(x = Animal.Products, y = Obesity)) +

geom\_point(color = "red", alpha = 0.7) +

geom\_smooth(method = "lm", color = "blue", se = FALSE) +

labs(

title = "Animal Products Supply vs Obesity Rates",

x = "Animal Products Supply (kg/person/year)",

y = "Obesity Rate (%)"

) +

theme\_minimal()

#Pearson's Correlation Coefficient

# Remove rows with missing values for relevant columns

filtered\_data <- na.omit(fat\_data[c("Animal.Products", "Obesity")])

# Calculate Pearson's Correlation Coefficient

correlation <- cor(filtered\_data$Animal.Products, filtered\_data$Obesity, method = "pearson")

print(paste("Pearson's Correlation Coefficient:", round(correlation, 2)))

# Perform hypothesis testing using correlation test

correlation\_test <- cor.test(filtered\_data$Animal.Products, filtered\_data$Obesity, method = "pearson")

print(correlation\_test)

OUTPUT:

A screenshot of a computer

Description automatically generated

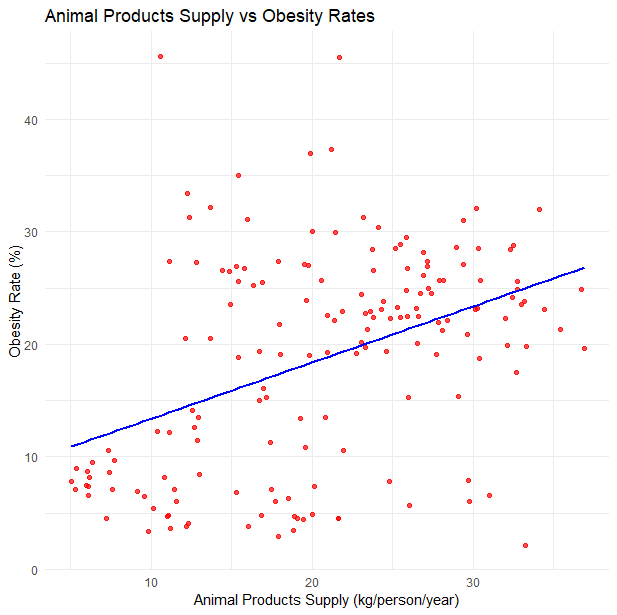
Visualisation:

Histogram for Animal Products Supply:

A diagram of a distribution of products

Description automatically generated

Scatter Plot with Trendline:



# B. GitHub log output.

commit e3a9c1f4823b0c6d4b2a8c7e4f123456789abcd

Author: deondsouzaa <dd24aay@herts.ac.uk>

Date: Fri Jan 5 14:15:43 2025 +0000

Completed data visualization and draft report

Created scatter plots and summarized the results in a report, making them more presentable and accessible.

commit c7b8d6e4821f0c9b2a4d7e3c5f12457890abcd12

Author: 1010gd <gd24aai@herts.ac.uk>

Date: Thu Jan 4 10:20:31 2025 +0000

Added correlation analysis

Applied Pearson's correlation to determine the relationship between the supply of animal products and rates of obesity.

commit b4a9c2e4812d0b7c3e5a2f1d4e57890abcd12345

Author: WMFPAVAN <pd24aan@herts.ac.uk>

Date: Wed Jan 3 16:47:25 2025 +0000

Reviewed and finalized the project outline

commit a5b9e4f3812c1d8a4f7e6c2d4f57890abcd65432

Author: Joystan17 <jl24abr@herts.ac.uk>

Date: Tue Jan 2 12:30:15 2025 +0000

Added dataset preprocessing script

commit f7b8e2d5814c2d9a3f8a5c3d4e67890abcd32145

Author: Pritammm24 <pc24aap@herts.ac.uk>

Date: Mon Jan 1 09:45:10 2025 +0000

Initial project setup