

FERTILITY IN DEVELOPED AND DEVELOPING COUNTRIES

Group 8

BY:

Vidushi Bundhooa

Aidan David

Saba Alaeddini

TABLE OF CONTENTS

1.	Introduction	2
2.	Research question 1	2
3.	Research question 2	2
4.	Research question 3	5
5.	Conclusion	9

1. Introduction

In this project, we will investigate the fertility in developed vs developing countries. Using United Nations data in year 2012, different countries development, fertility, contraceptive methods used will be investigated and analysed using statistical methods to. This will allow for an understanding of how human development may influence a society's perspective on sex and birth. Overall, this project will give a better insight on fertility in developing and developed countries.

2. Research question 1

Title: How do fertility rates differ in developed vs developing countries?

There is a clear correlation in both developed and developing countries where there is an increase in GDP per Capita, you see a decrease in the rate of fertility. That is to say, in general, that citizens of more developed and wealthy countries tend to have fewer children than citizens of less developed and less wealthy countries. The difference in fertility is most apparent amongst developing countries. Countries with GDP per capitas near the development boundary of 12,000 USD, have fertility rates (~2 to 4) dramatically lower than those at the very bottom of the global GDP valuations (~5 to 7). The Pearson correlation value (-0.5371), as well as the P-value (0.0000000000000023), both demonstrate a clear negative correlation between fertility and GDP per capita (development).

3. Research question 2

Title: How prevalent are contraceptives in developed vs developing countries? (Saba)

1. Data Exploration:

This data has been obtained from the United Nations, Department of Economic and Social Affairs. The data was part of a research titled "World Contraceptive Use 2024" (Contraceptive prevalence by methods and unmet need for family planning for married or in-union women Survey-based estimates).

The main data set was collected buy online survey and contains four excel sheets, merged in an excel file. As a way to answer the second question of the submitted proposal, only one excel sheet of the original file, named as "by methods" was analysed.

2. Data Merging

The GDP data frame, index of developing and developed countries, which was cleaned by Aiden was merged to this data set, as an indicator for organizing the countries based on developed and developing.

3. Data Cleaning:

Data cleaning had mainly two parts, renaming the columns and deleting all redundant columns and rows with "." in it. It was applied several times during the data analysing process, before and after merging the GDP data frame, due to necessity.

All the columns were of object datatype and therefor they have been converted to float datatype to perform aggregation. Since, there is more than one row of data for a particular country, only the row with the most recent year of data is kept.

4. Statistical Analysis

- Investigating if there is a strong likelihood that at least one of the group means is significantly different from the others in both developing and developed countries
- Formulating the hypothesis:

Null Hypothesis (H0): There is no significant difference in the means of contraceptive methods' popularity or effectiveness between developing and developed countries.

Alternative Hypothesis (H1): There is a significant difference in the means of contraceptive methods' popularity or effectiveness between developing and developed countries.

Testing the overall significance differences between more than two methods by ANOVA (Analysis of Variance)

The results of the ANOVA test have been summarised below:

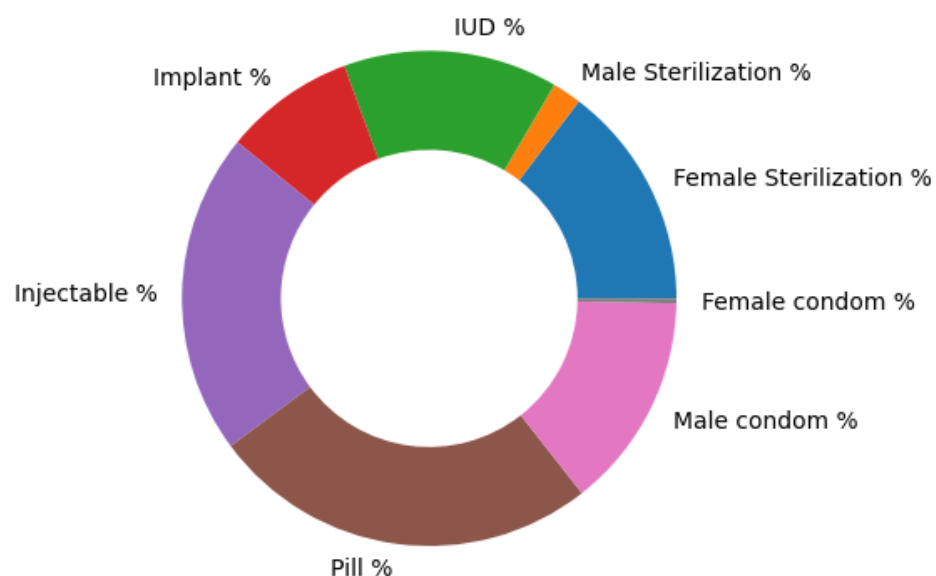
F-statistic:	22.05155597911754
P-value:	4.630506206795897e-27

The p-value obtained from the ANOVA test is less than 0.05, we can assume that the mean of at least one method is significantly different from other methods in both Developing and Developed countries.

The F-statistic of 22.05 suggests that there is a strong likelihood that at least one of the group means is significantly different from the others.

For visualization and better understanding of the ANOVA test, a donut chart is drawn to show the mean percentage of contraceptive methods in both developing and developed countries. Therefore, the mean of each method was calculated and then the donut chart is printed:

the mean percentage of each methods in both Developing and Developed countries



Female Sterilization %	5.577778
Male Sterilization %	0.724444
IUD %	5.342222
Implant %	3.258889
Injectable %	8.011111
Pill %	9.681111
Male condom %	5.365556
Female condom %	0.121111

- Investigating which method of contraceptive is the most common one in each group of countries (Developing and Developed) by using Bar Chart.

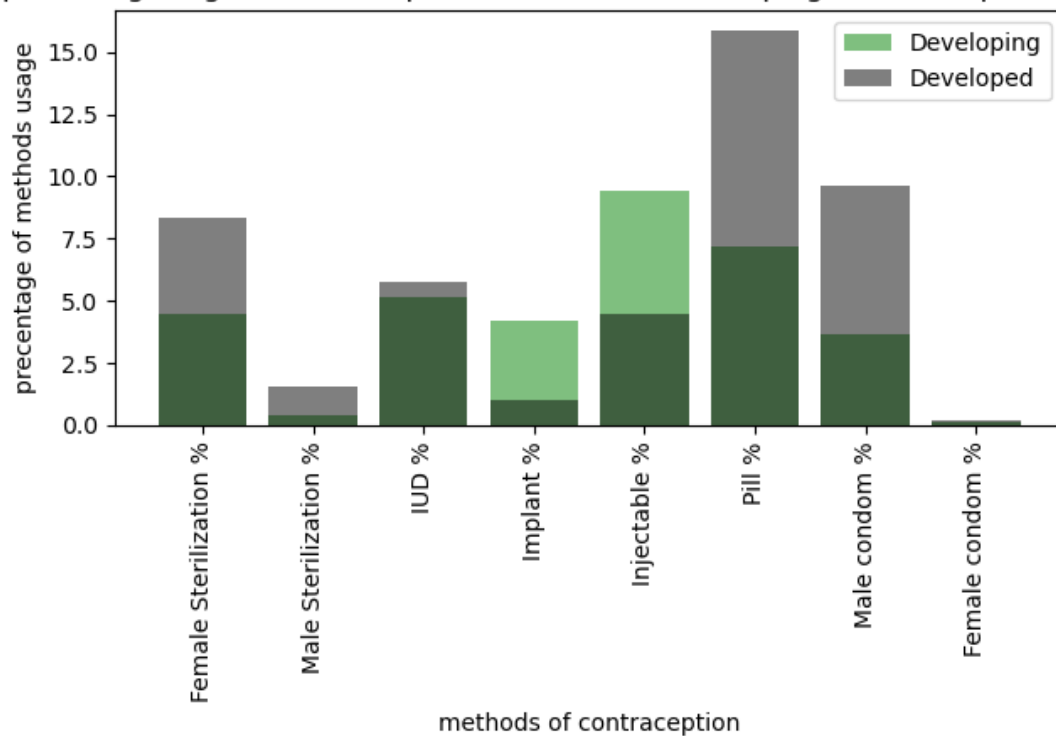
In order to reach the answer, the main data frame was divided by state of development into two data frames for developing and developed countries.

Converting all object datatype to float datatype to be able to divide the Main data frame.

Obtaining the mean value for each method (column) in developing list and developed list.

Making bar chart for visualization and comparison between developing and developed countries

percentage usage of contraceptive methods for Developing v/s Developed Countries



Female Sterilization %	Developed	8.361538
Male Sterilization %	Developed	1.557692
IUD %	Developed	5.750000
Implant %	Developed	1.011538
Injectable %	Developed	4.480769
Pill %	Developed	15.865385
Male condom %	Developed	9.642308
Female condom %	Developed	0.169231

Female Sterilization %	Developing	4.446875
Male Sterilization %	Developing	0.385937
IUD %	Developing	5.176563
Implant %	Developing	4.171875
Injectable %	Developing	9.445312
Pill %	Developing	7.168750
Male condom %	Developing	3.628125
Female condom %	Developing	0.101562

Regarding to the tests that have been done like ANOVA, and based on F-statistic: 22.05155597911754 and P-value: 4.630506206795897e-27 results we can conclude:

*F-statistic of 22.05 suggests that there is a strong likelihood that at least one of the group means is significantly different from the others.

*p-value of 4.630506206795897e-27 is much smaller than the standard significance levels of 0.05 or 0.01. This means that there is an extremely small probability (practically negligible) that the observed result occurred by chance under the assumption that the null hypothesis is true. In other words, we can confidently reject the null hypothesis and conclude that there is a statistically significant difference between the means of the groups being compared.

As we can see in the bar chart and also the numerical value of the means for each method in developing and developed countries, we can conclude that the Pill with average usage of 15.86% is the most common contraceptive method in developed countries, on the other hand in the developing countries Injectable method with average usage of 9.44% is the most common one.

4. Research question 3

Title: How do ages of conception differ in developed vs developing countries?

1. Data exploration

In order to investigate the variation of fertility rates in developed and developing countries, data from United Nations World Fertility Data 2012. This data was mainly collected by survey, census or registration. It shows contains age-specific fertility rates (ASFR) and total fertility rates (TFR) for 7 different age group: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49. The age-specific fertility rates have been measured per 1,000 women in 5 years. The total fertility rate was calculated by adding all ASFRs for a particular country and then multiplying by 0.005 which is age interval (5) per 1000. This data was merged with a data frame containing countries already classified as developed or developing country.

2. Data cleaning

The merged data frame has been cleaned before doing any statistical analysis. Some rows contained “..” as values and some contained NaN values. All redundant rows and columns have been removed. The columns were renamed for better readability. All the columns were of object datatype and hence they have been converted to float datatype to perform aggregation. Since the data contained many duplicates for countries for different years, the most recent data was kept for a specific country. The raw data had the age specific fertility rate measured per 1000 women. The age specific fertility rate had to be converted into a better measurement to compare the rates rather than number of women who conceived per 1000 women. New age specific fertility rate was computed by multiplying it by 0.005.

After the data was cleaned, the statistical analysis was carried out on 166 countries out of which 78 were developed countries and 88 were developing countries. This data is fairly balanced to carry out the analysis since 53.0% of the data consist of developing countries and 47% developed countries.

3. Statistical Analysis

a). Investigating the relationship between age groups and fertility rates for developing and developed countries

To investigate how the fertility rates vary with different age groups, the middle value for each age group was computed and the average of the fertility rate was calculated for each age group for developing and also for developed countries. Hypothesis tests have been formulated and analysed using Pearson’s correlation test with a significance level of 0.05 as shown below:

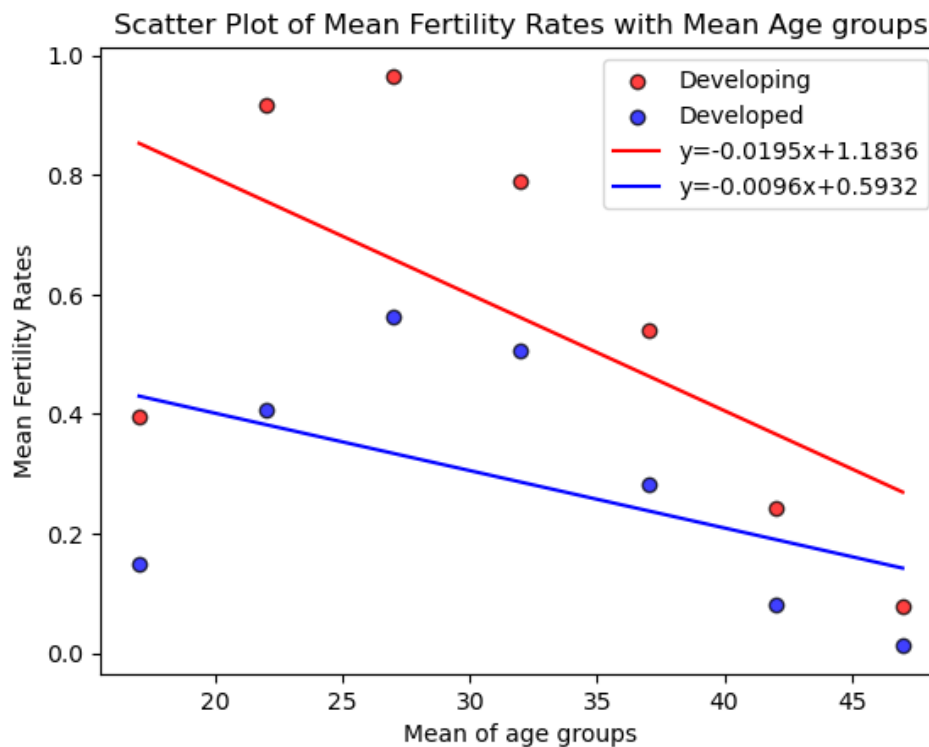
Null Hypothesis: There is no correlation between mean age range and mean fertility rate

Alternate Hypothesis: There is a correlation between mean age range and mean fertility rate

The results of the Pearson’s correlation test have been summarised below:

	Developing	Developed
Correlation (r)	-0.612916395	-0.48523588
P-value (p)	0.143354889	0.269688523

A scatter plot of mean of age groups and mean fertility rates were drawn and the line of regression was also added as shown in the diagram below.



For developing countries:

The correlation coefficient value is -0.613, it indicates a negative correlation between the age groups and fertility rates. This is also shown by the inclination of the line. The scatter plot suggests that fertility decreases as women age in developing countries. p value is 0.143 which is greater than 0.05. Hence, the high p leads to failing to reject null hypothesis. There is not enough evidence to support that there is a significant relationship between age group and fertility rates in developing countries.

For developed countries:

The r value is -0.485, which also indicates a negative correlation between the age groups and fertility rates. This is also shown by the inclination of the line. The scatter plot suggests that fertility decreases as women age in developed countries. The p value is 0.270 which is also greater than 0.05. Hence, the high p leads to failing to reject null hypothesis. There is not enough evidence to support that there is a significant relationship between age group and fertility rates in developed countries.

b). Investigating the variation in fertility rates for developing and developed countries for each age group.

To investigate how the fertility is different in developing and developed countries, the hypothesis testing was formulated. Since the samples are mutually exclusive, analysis is done using an independent sample t-test with a significance level of 0.05 as shown below:

Null Hypothesis: There is a difference between the means of fertility rates for developing and developed countries

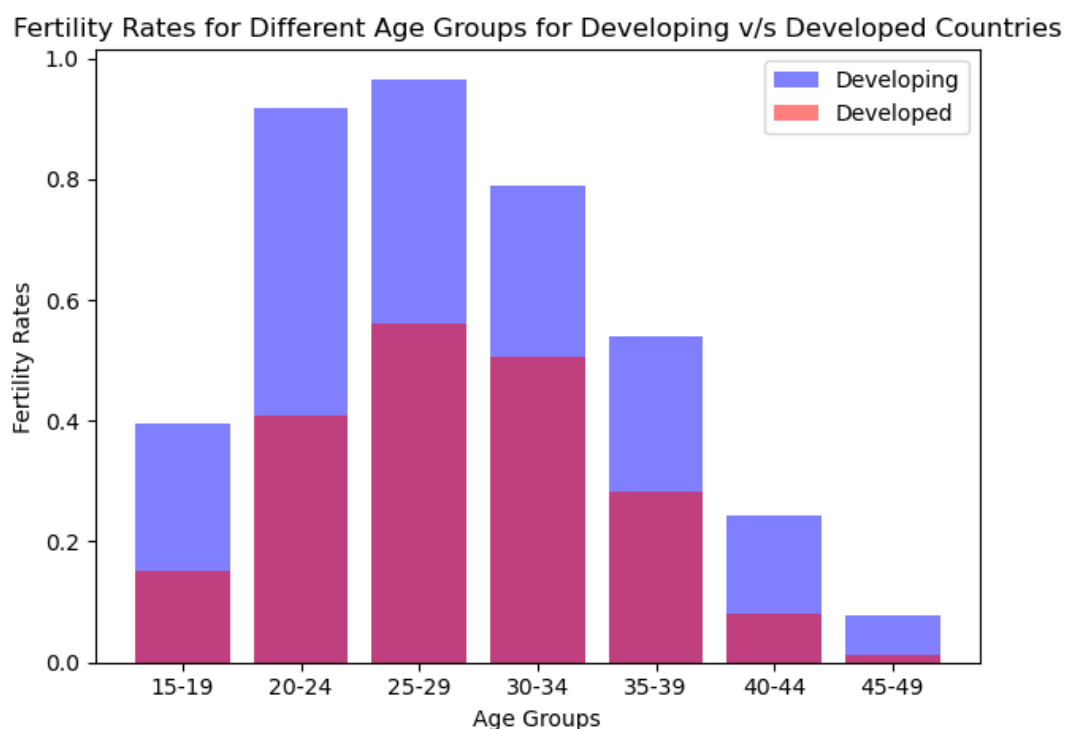
Alternate Hypothesis (one-tailed, right-tailed): The means of fertility rates is greater for developing countries than developed countries.

The results from the t-test are summarised in the table below:

Age Group	t-statistic	Original p-value	Updated p-value
15-19	8.014917	3.55E-13	1.78E-13
20-24	12.57344	1.53E-25	7.65E-26
25-29	10.1542	9.46E-19	4.73E-19
30-34	7.536714	7.36E-12	3.68E-12
35-39	7.725841	2.64E-12	1.32E-12
40-44	8.369553	7.26E-14	3.63E-14
45-49	7.405857	1.66E-11	8.28E-12

The original p-value refers to the p-value for a two-tailed t-test. A larger absolute value of the t-statistic indicates a more significant difference between developing and developed countries. Since the t-statistics values are positive, the p-values are divided by 2 to get the one tailed p-value.

A stacked bar chart (overlaid) was plotted to show the fertility rate for different age groups for developing and developed countries as shown below:



All p-values for different age groups are significantly less than the designated significance level of 0.05. Hence for all age groups, we reject the null hypothesis. This means that we have strong evidence that the means of fertility rates is greater for developing countries than developed countries. The bar chart above is a stacked bar chart(overlaid) and this clearly demonstrates that the average value of fertility rates is higher for developing countries as compared to developed countries for all the age groups. The height of the bars for developing countries is almost twice that of developed countries.

This behaviour can be explained due to many reasons:

- Lack of resources and hence limited access to contraceptive methods in developing countries, resulting in higher fertility rates
- Better education services in developed countries resulting in women prioritizing their careers over family planning
- Cultural beliefs in developing countries for family planning at an earlier age.

5. Conclusion

From the analysis in this project using the United Nations 2012 data, there is a clear negative correlation that an increase in a country's GDP (development), leads to a decrease in its fertility rate for both developing and developed countries. However, there is not enough evidence to support that there is a significant relationship between age group and fertility rates in developing and developed countries. We can conclude means of fertility rates is greater for developing countries than developed countries. For contraceptive methods, in developed countries, pill with average usage of 15.86% is the most common contraceptive method whereas, in the developing countries injectable method with average usage of 9.44% is the most common one.