*ARIEL GILGEOURS*

*oecd data Analysis*

**Table of Contents**

Fertility Rate\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2

Crop Production\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_6

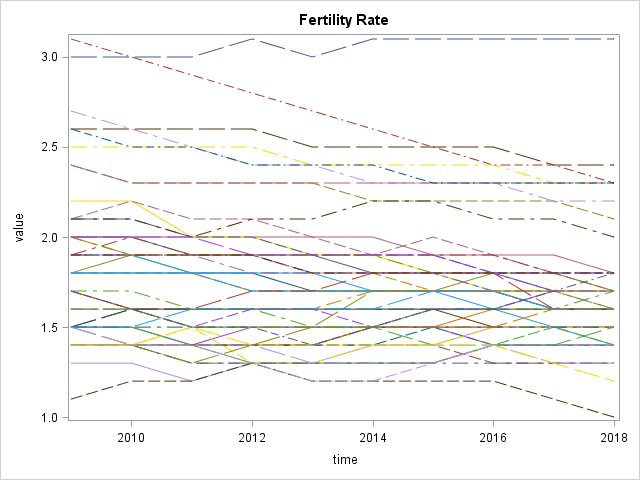
Education Performance \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_18

Appendix \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_23

**Fertility**

With mortality and migration, fertility is an element of population growth, which reflects the causes and effect of economic and social developments. In recent decade there has been a decline in birth rate due to the fact woman have postponed child-bearing and there is a decrease in desired family size. The total fertility rate is defined as the total number of children that would be born to each woman if she were to live to the end of child bearing years and given birth to children in alignment with prevailing age-specific fertility rates.

The research question that we are going to look at in this analysis; is there a significant difference between the fertility rates in the countries apart of the OECD?



When looking at the graphed data you can see that there is a variety overall differences in the fertility rates, some of them stayed the same within the 10-year period, while a few increased and others had a downward tick. However, the values stayed within the range of 1 to 3 total children/women. But while it might seem like a small range, that does not mean that the difference between countries’ fertility rates is not statistically significant.

Since we are looking at the fertility rate of the 37 OECD countries over 10 year period a repeated measure design would be best suited for this case. Repeated measures design uses time as the factor of the univariate ANOVA that assess individuals over a course of interventions. It is a within subjects design which looks at repeated measurements, a matched set such as twins, siblings or other attributes such as IQ, age, or gender. The observations across the intervals are dependent and correlated. It compares the sample means assuming that the covariances are equal to approximately equal. If the homogeneity assumption is not met the F statistic degree of freedom for the numerator and denominator need adjustment. Two methods used in this case are the Greenhouse-Gessier correction, which is a more common but also more conservative approach. While the second method of adjustment is the Huynh-Feldt correction which is more powerful but also less common approach to adjustment. The SAS program and data used in this analysis can be found below.

The GLM Procedure

Repeated Measures Analysis of Variance

Univariate Tests of Hypotheses for Within Subject Effects

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** | **Adj Pr > F** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **G - G** | **H - F** |
| **years** | 9 | 0.77855769 | 0.08650641 | 10.74 | <.0001 | <.0001 | <.0001 |
| **Error(years)** | 459 | 3.69844231 | 0.00805761 |  |  |  |  |

|  |  |
| --- | --- |
| **Greenhouse-Geisser Epsilon** | 0.2075 |
| **Huynh-Feldt Epsilon** | 0.2151 |

The adjusting probability found using the Greenhouse-Geisser has a probability of less than 0.1% and the Huynh-Feldt Epsilon also has a probability of less than 0.1%. In order to show that the means are the same, they would need to have a probability higher than 5%. Given the fact the probabilities of the two methods of adjustment are indictive that the means are not the same. We continue on the MANOVA test

| **MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no years Effect H = Type III SSCP Matrix for years E = Error SSCP Matrix  S=1 M=3.5 N=20.5** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Statistic** | **Value** | **F Value** | **Num DF** | **Den DF** | **Pr > F** |
| **Wilks' Lambda** | 0.31112243 | 10.58 | 9 | 43 | <.0001 |
| **Pillai's Trace** | 0.68887757 | 10.58 | 9 | 43 | <.0001 |
| **Hotelling-Lawley Trace** | 2.21416878 | 10.58 | 9 | 43 | <.0001 |
| **Roy's Greatest Root** | 2.21416878 | 10.58 | 9 | 43 | <.0001 |

Given the fact the Wilk’s lambda has a probability of less than .1 this shows that there is a statistical significance between the fertility rates in the countries. In order for the fertility rates to be comparable a probability of 5% would be need. Given the fact that the wilks lambda probability and the probability of the two methods of adjustment are indicative that the means are not the same. We reject the null hypothesis and accept the alternate that at least one of the means are not equal to one another.

| **Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DF = 51** | **t1** | **t2** | **t3** | **t4** | **t5** | **t6** | **t7** | **t8** | **t9** | **t10** |
| **t1** | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.991802 | | <.0001 | | |  | | --- | | 0.981963 | | <.0001 | | |  | | --- | | 0.974140 | | <.0001 | | |  | | --- | | 0.969444 | | <.0001 | | |  | | --- | | 0.946863 | | <.0001 | | |  | | --- | | 0.933676 | | <.0001 | | |  | | --- | | 0.915252 | | <.0001 | | |  | | --- | | 0.907260 | | <.0001 | | |  | | --- | | 0.893013 | | <.0001 | |
| **t2** | |  | | --- | | 0.991802 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.989290 | | <.0001 | | |  | | --- | | 0.983523 | | <.0001 | | |  | | --- | | 0.976860 | | <.0001 | | |  | | --- | | 0.953083 | | <.0001 | | |  | | --- | | 0.943144 | | <.0001 | | |  | | --- | | 0.923609 | | <.0001 | | |  | | --- | | 0.908007 | | <.0001 | | |  | | --- | | 0.891650 | | <.0001 | |
| **t3** | |  | | --- | | 0.981963 | | <.0001 | | |  | | --- | | 0.989290 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.988170 | | <.0001 | | |  | | --- | | 0.981826 | | <.0001 | | |  | | --- | | 0.961706 | | <.0001 | | |  | | --- | | 0.952832 | | <.0001 | | |  | | --- | | 0.937104 | | <.0001 | | |  | | --- | | 0.914015 | | <.0001 | | |  | | --- | | 0.896634 | | <.0001 | |
| **t4** | |  | | --- | | 0.974140 | | <.0001 | | |  | | --- | | 0.983523 | | <.0001 | | |  | | --- | | 0.988170 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.993304 | | <.0001 | | |  | | --- | | 0.979143 | | <.0001 | | |  | | --- | | 0.972755 | | <.0001 | | |  | | --- | | 0.952694 | | <.0001 | | |  | | --- | | 0.929303 | | <.0001 | | |  | | --- | | 0.911981 | | <.0001 | |
| **t5** | |  | | --- | | 0.969444 | | <.0001 | | |  | | --- | | 0.976860 | | <.0001 | | |  | | --- | | 0.981826 | | <.0001 | | |  | | --- | | 0.993304 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.988311 | | <.0001 | | |  | | --- | | 0.979422 | | <.0001 | | |  | | --- | | 0.958299 | | <.0001 | | |  | | --- | | 0.938069 | | <.0001 | | |  | | --- | | 0.916136 | | <.0001 | |
| **t6** | |  | | --- | | 0.946863 | | <.0001 | | |  | | --- | | 0.953083 | | <.0001 | | |  | | --- | | 0.961706 | | <.0001 | | |  | | --- | | 0.979143 | | <.0001 | | |  | | --- | | 0.988311 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.991431 | | <.0001 | | |  | | --- | | 0.977702 | | <.0001 | | |  | | --- | | 0.962262 | | <.0001 | | |  | | --- | | 0.943488 | | <.0001 | |
| **t7** | |  | | --- | | 0.933676 | | <.0001 | | |  | | --- | | 0.943144 | | <.0001 | | |  | | --- | | 0.952832 | | <.0001 | | |  | | --- | | 0.972755 | | <.0001 | | |  | | --- | | 0.979422 | | <.0001 | | |  | | --- | | 0.991431 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.986085 | | <.0001 | | |  | | --- | | 0.971135 | | <.0001 | | |  | | --- | | 0.955902 | | <.0001 | |
| **t8** | |  | | --- | | 0.915252 | | <.0001 | | |  | | --- | | 0.923609 | | <.0001 | | |  | | --- | | 0.937104 | | <.0001 | | |  | | --- | | 0.952694 | | <.0001 | | |  | | --- | | 0.958299 | | <.0001 | | |  | | --- | | 0.977702 | | <.0001 | | |  | | --- | | 0.986085 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.985301 | | <.0001 | | |  | | --- | | 0.972475 | | <.0001 | |
| **t9** | |  | | --- | | 0.907260 | | <.0001 | | |  | | --- | | 0.908007 | | <.0001 | | |  | | --- | | 0.914015 | | <.0001 | | |  | | --- | | 0.929303 | | <.0001 | | |  | | --- | | 0.938069 | | <.0001 | | |  | | --- | | 0.962262 | | <.0001 | | |  | | --- | | 0.971135 | | <.0001 | | |  | | --- | | 0.985301 | | <.0001 | | |  | | --- | | 1.000000 | |  | | |  | | --- | | 0.989296 | | <.0001 | |
| **t10** | |  | | --- | | 0.893013 | | <.0001 | | |  | | --- | | 0.891650 | | <.0001 | | |  | | --- | | 0.896634 | | <.0001 | | |  | | --- | | 0.911981 | | <.0001 | | |  | | --- | | 0.916136 | | <.0001 | | |  | | --- | | 0.943488 | | <.0001 | | |  | | --- | | 0.955902 | | <.0001 | | |  | | --- | | 0.972475 | | <.0001 | | |  | | --- | | 0.989296 | | <.0001 | | |  | | --- | | 1.000000 | |  | |

| **Sphericity Tests** | | | | |
| --- | --- | --- | --- | --- |
| **Variables** | **DF** | **Mauchly's Criterion** | **Chi-Square** | **Pr > ChiSq** |
| **Transformed Variates** | 44 | 0.0000149 | 531.08667 | <.0001 |
| **Orthogonal Components** | 44 | 0.0000149 | 531.08667 | <.0001 |

When looking at the correlation coefficient matrix. We look at the values on top right corner to see which is closer to 1. The value closer to 1 has a higher the correlation is between the variables. Since we are look at the time periods compared to each other and we can see that they are close to 1, we would say that this data is highly correlated meaning that there is a strong relationship between the one time period and the other.

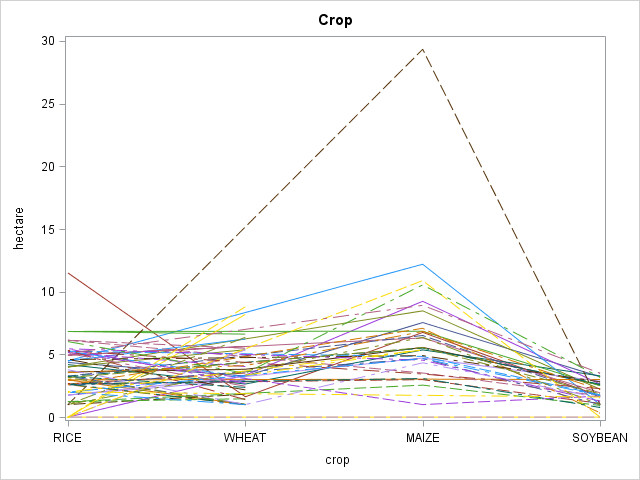
To see if the homogeneity assumption is met, we test for sphericity, which tests the variances of the different pairwise conditions across all of the conditions. In this case there is a significant statistical difference between the fertility rate of the countries that are involved in the OECD. The orthogonal component given the fact that the chi square value is less than .01%, compared to the 5% needed for it to be considered comparable. Therefore, in this case we reject the null hypothesis and accept the alternate hypothesis that at least one of the variances of the are not equal.

Based on the conclusions drawn from above, I can state that the fertility rate between the countries that are a part of the OECD are statistically different. There could be many reasons for the cause of the difference. However, with this information a country could use other practices from another country to increase or decrease fertility rate. For example, if one country would want to increase their fertility rate they might want to focus on the cultural aspects that might differ between a country with a higher rate. They could also look at a country who has had an increase in their fertility rate in recent years, since most countries have had a large decline in the past decades.

**Crop Production**

Crop production is essential for the survival of humans across the world. When it comes to crop production, there are many factors that influence the production such as availability of arable land, consumption patterns that influence yield and pricing. The important of understanding crop production is vital to global consumption. In this data the crop production was measured in tons per hectare.

The research question associated with this analysis is does mean crop production differ significantly among the countries apart of the OECD?



We now take a look at the data in order to see if it is a good fit for one of our ANOVA models. Given the means for the rice, wheat, maize and soybean for each country you can see that many of the countries seem to be producing similar quantities of each crop with some outliers, producing larger quantities of specific crops more than the other. To see if one country may be producing a larger quantity it would be interesting to see if there is or is not a statistical significance. Given the format of this data This data would be a good fit for the two-way factorial analysis.

The ANOVA Procedure

Dependent Variable: hectare

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 143 | 14406.11894 | 100.74209 | 112.06 | <.0001 |
| **Error** | 1296 | 1165.13462 | 0.89902 |  |  |
| **Corrected Total** | 1439 | 15571.25356 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **hectare Mean** |
| --- | --- | --- | --- |
| 0.925174 | 26.22352 | 0.948169 | 3.615718 |

| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **location** | 35 | 3207.272887 | 91.636368 | 101.93 | <.0001 |
| **crop** | 3 | 3054.005937 | 1018.001979 | 1132.34 | <.0001 |
| **crop\*location** | 105 | 8144.840118 | 77.569906 | 86.28 | <.0001 |

The Two -Way Factorial ANOVA Model above shows that there is a significant statistically difference between countries crop production given the fact that the probability is less than .01%. In order for the means of crop production to be comparable, a probability of 5% would be needed. You can also see above that the interactive effect crop and location have a significant effect on the amount of crop that will be produced since the probability is less than the .01%.

The crop and location have a significant effect on the amount of crop that will be produced. Not only do the main effects have an effect on the outcome but the interactive effect has an effect on the outcome. The statistically significant factors effecting the outcome hectare all of which contribute to nearly 92.5% of the crop production associated with these predictor variables.

|  |
| --- |
| The SAS System |

The ANOVA Procedure

t Tests (LSD) for hectare

|  |  |
| --- | --- |
| Note: | This test controls the Type I comparisonwise error rate, not the experimentwise error rate. |

|  |  |
| --- | --- |
| **Alpha** | 0.05 |
| **Error Degrees of Freedom** | 1296 |
| **Error Mean Square** | 0.899024 |
| **Critical Value of t** | 1.96180 |
| **Least Significant Difference** | 0.4159 |

| **Means with the same letter are not significantly different.** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **t Grouping** | | | | **Mean** | **N** | **location** |
|  |  | A |  | 7.9914 | 40 | ISR |
|  |  |  |  |  |  |  |
|  |  | B |  | 7.3116 | 40 | CHE |
|  |  |  |  |  |  |  |
|  |  | C |  | 5.7611 | 40 | EGY |
|  |  | C |  |  |  |  |
| D |  | C |  | 5.5443 | 40 | CHL |
| D |  | C |  |  |  |  |
| D |  | C |  | 5.5154 | 40 | USA |
| D |  |  |  |  |  |  |
| D |  | E |  | 5.1882 | 40 | AUS |
|  |  | E |  |  |  |  |
|  |  | E |  | 4.9324 | 40 | NZL |
|  |  | E |  |  |  |  |
|  |  | E |  | 4.8382 | 40 | TUR |
|  |  |  |  |  |  |  |
|  |  | F |  | 4.3491 | 40 | CHN |
|  |  | F |  |  |  |  |
|  |  | F |  | 4.3344 | 40 | ARG |
|  |  | F |  |  |  |  |
| G |  | F |  | 3.9743 | 40 | MEX |
| G |  |  |  |  |  |  |
| G |  | H |  | 3.8606 | 40 | CAN |
| G |  | H |  |  |  |  |
| G |  | H |  | 3.7871 | 40 | KOR |
| G |  | H |  |  |  |  |
| G |  | H |  | 3.7744 | 40 | UKR |
|  |  | H |  |  |  |  |
| I |  | H |  | 3.5470 | 40 | BRICS |
| I |  | H |  |  |  |  |
| I |  | H |  | 3.5167 | 40 | IRN |
| I |  | H |  |  |  |  |
| I |  | H |  | 3.4774 | 40 | BRA |
| I |  | H |  |  |  |  |
| I |  | H |  | 3.4681 | 40 | SAU |
| I |  |  |  |  |  |  |
| I |  |  |  | 3.3395 | 40 | PRY |
| I |  |  |  |  |  |  |
| I |  | J |  | 3.2070 | 40 | RUS |
| I |  | J |  |  |  |  |
| I |  | J | K | 3.1607 | 40 | PER |
|  |  | J | K |  |  |  |
| L |  | J | K | 2.9037 | 40 | ZAF |
| L |  | J | K |  |  |  |
| L |  | J | K | 2.8481 | 40 | JPN |
| L |  | J | K |  |  |  |
| L | M | J | K | 2.8279 | 40 | COL |
| L | M |  | K |  |  |  |
| L | M |  | K | 2.7601 | 40 | KAZ |
| L | M |  |  |  |  |  |
| L | M |  |  | 2.7424 | 40 | IDN |
| L | M |  |  |  |  |  |
| L | M | N |  | 2.6438 | 40 | VNM |
| L | M | N |  |  |  |  |
| L | M | N |  | 2.6266 | 40 | PAK |
| L | M | N |  |  |  |  |
| L | M | N |  | 2.5633 | 40 | MYS |
|  | M | N |  |  |  |  |
|  | M | N |  | 2.4289 | 40 | ETH |
|  |  | N |  |  |  |  |
| O |  | N |  | 2.2969 | 40 | IND |
| O |  | N |  |  |  |  |
| O |  | N |  | 2.2752 | 40 | THA |
| O |  |  |  |  |  |  |
| O |  | P |  | 1.9717 | 40 | GBR |
|  |  | P |  |  |  |  |
| Q |  | P |  | 1.7884 | 40 | PHL |
| Q |  |  |  |  |  |  |
| Q |  | R |  | 1.4364 | 40 | NGA |
|  |  | R |  |  |  |  |
|  |  | R |  | 1.1739 | 40 | NOR |

From the LSD Test above you can see for there are over 18 different groups that the countries are being grouped in. In order for countries to be grouped with one another they need have to a least significant difference of 0.4159. Highlighting the biggest group L, which highlights 8 of the 37 countries, which produce similar amounts of crop. The group includes countries such as Japan India and Vietnam, which is interesting given how much larger India is compared to Vietnam and Japan. But it is important to note that service is not the only factor in crop production. Now down to the smallest group which only consists of one country in group A, Israel which has the largest crop production among the countries. With the large amount of groups, you can see that there is statistical significance.

|  |
| --- |
| The SAS System |

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for hectare

|  |  |
| --- | --- |
| Note: | This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ. |

|  |  |
| --- | --- |
| **Alpha** | 0.05 |
| **Error Degrees of Freedom** | 1296 |
| **Error Mean Square** | 0.899024 |
| **Critical Value of Studentized Range** | 5.43973 |
| **Minimum Significant Difference** | 0.8155 |

| **Means with the same letter are not significantly different.** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tukey Grouping** | | | | | **Mean** | **N** | **location** |
|  |  |  | A |  | 7.9914 | 40 | ISR |
|  |  |  | A |  |  |  |  |
|  |  |  | A |  | 7.3116 | 40 | CHE |
|  |  |  |  |  |  |  |  |
|  |  |  | B |  | 5.7611 | 40 | EGY |
|  |  |  | B |  |  |  |  |
|  | C |  | B |  | 5.5443 | 40 | CHL |
|  | C |  | B |  |  |  |  |
|  | C |  | B |  | 5.5154 | 40 | USA |
|  | C |  | B |  |  |  |  |
|  | C |  | B |  | 5.1882 | 40 | AUS |
|  | C |  |  |  |  |  |  |
|  | C |  | D |  | 4.9324 | 40 | NZL |
|  | C |  | D |  |  |  |  |
|  | C |  | D |  | 4.8382 | 40 | TUR |
|  |  |  | D |  |  |  |  |
|  | E |  | D |  | 4.3491 | 40 | CHN |
|  | E |  | D |  |  |  |  |
|  | E |  | D |  | 4.3344 | 40 | ARG |
|  | E |  |  |  |  |  |  |
|  | E |  | F |  | 3.9743 | 40 | MEX |
|  | E |  | F |  |  |  |  |
|  | E |  | F |  | 3.8606 | 40 | CAN |
|  | E |  | F |  |  |  |  |
|  | E |  | F |  | 3.7871 | 40 | KOR |
|  | E |  | F |  |  |  |  |
|  | E |  | F |  | 3.7744 | 40 | UKR |
|  | E |  | F |  |  |  |  |
|  | E |  | F | G | 3.5470 | 40 | BRICS |
|  |  |  | F | G |  |  |  |
|  |  |  | F | G | 3.5167 | 40 | IRN |
|  |  |  | F | G |  |  |  |
|  |  |  | F | G | 3.4774 | 40 | BRA |
|  |  |  | F | G |  |  |  |
|  |  |  | F | G | 3.4681 | 40 | SAU |
|  |  |  | F | G |  |  |  |
|  | H |  | F | G | 3.3395 | 40 | PRY |
|  | H |  | F | G |  |  |  |
|  | H | I | F | G | 3.2070 | 40 | RUS |
|  | H | I | F | G |  |  |  |
|  | H | I | F | G | 3.1607 | 40 | PER |
|  | H | I |  | G |  |  |  |
|  | H | I | J | G | 2.9037 | 40 | ZAF |
|  | H | I | J | G |  |  |  |
|  | H | I | J | G | 2.8481 | 40 | JPN |
|  | H | I | J | G |  |  |  |
|  | H | I | J | G | 2.8279 | 40 | COL |
|  | H | I | J | G |  |  |  |
| K | H | I | J | G | 2.7601 | 40 | KAZ |
| K | H | I | J | G |  |  |  |
| K | H | I | J | G | 2.7424 | 40 | IDN |
| K | H | I | J |  |  |  |  |
| K | H | I | J |  | 2.6438 | 40 | VNM |
| K | H | I | J |  |  |  |  |
| K | H | I | J |  | 2.6266 | 40 | PAK |
| K | H | I | J |  |  |  |  |
| K | H | I | J | L | 2.5633 | 40 | MYS |
| K |  | I | J | L |  |  |  |
| K |  | I | J | L | 2.4289 | 40 | ETH |
| K |  |  | J | L |  |  |  |
| K |  |  | J | L | 2.2969 | 40 | IND |
| K |  |  | J | L |  |  |  |
| K |  |  | J | L | 2.2752 | 40 | THA |
| K |  |  |  | L |  |  |  |
| K |  |  | M | L | 1.9717 | 40 | GBR |
|  |  |  | M | L |  |  |  |
|  |  |  | M | L | 1.7884 | 40 | PHL |
|  |  |  | M |  |  |  |  |
|  |  |  | M |  | 1.4364 | 40 | NGA |
|  |  |  | M |  |  |  |  |
|  |  |  | M |  | 1.1739 | 40 | NOR |

Unlike the LSD test above the minimum significant difference is slightly higher at 0.8155. There are 13 groups instead of the 18 compared to the LSD test. In this test you can see that compared to the LSD test in this tukey test, Israel was grouped with Switzerland and is not in a group by itself given the high minimum significant difference, they are separated by a mean of 0.6798. Groups H and F both have 11 countries which decreased the number of groups compared to the LSD test. Given the fact that there are still a large number of groups. It can be said that the countries are statistically significant from one another.

In conclusion there is a statistically significant difference in crop production of maize, wheat, grain, rice and soybean. This could be due to many factors some of which could be adopted by another country in order for them to grow more crop. But it is also important to think about a saturated market for a specific type of crop, so a country might want to look into growing a crop that does not have such a larger production as a nearby country.

**Education Performance**

The reading performance for PISA takes measurement of a student’s the capacity to understand, use and reflect on written texts in order to develop knowledge, potential and participate in society. The mean score of each country is the measure, the mathematical performance for PISA measures the mathematical literacy of 15 year old boys and girl. It takes into account the interpretation of an array of contest to describe, predict and explain the role math plays in society. The mean score is the measure for each country.

The research question of this study Does Math Score influence Reading Score and Gender? Assuming that a student’s math score would likely increase the higher their reading score given the fact that solving open ended math problems does require having an understanding of reading comprehension.

Before performing an ANCOVA we need to check if it is a good fit by doing an exploratory analysis on the data in order to see that it is a good fit for this model.

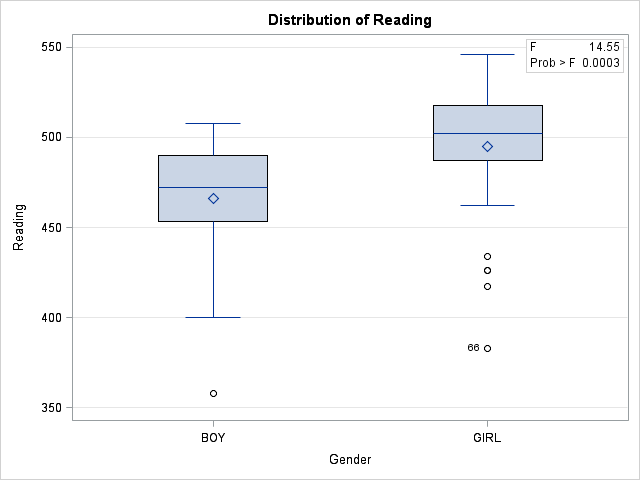
The ANOVA Procedure

Dependent Variable: Reading

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 1 | 16791.0125 | 16791.0125 | 14.55 | 0.0003 |
| **Error** | 78 | 90039.3750 | 1154.3510 |  |  |
| **Corrected Total** | 79 | 106830.3875 |  |  |  |

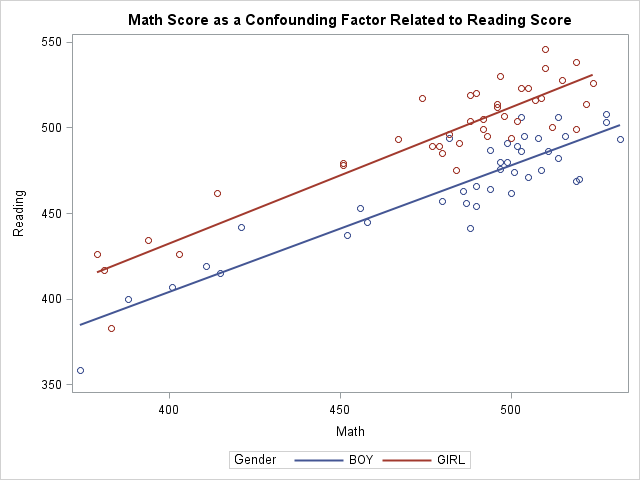
| **R-Square** | **Coeff Var** | **Root MSE** | **Reading Mean** |
| --- | --- | --- | --- |
| 0.157174 | 7.067788 | 33.97574 | 480.7125 |

| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Gender** | 1 | 16791.01250 | 16791.01250 | 14.55 | 0.0003 |
|  |  |  |  |  |  |
| **Level of Gender** | **N** | **Reading** | |
| **Mean** | **Std Dev** |
| **BOY** | **40** | 466.225000 | 32.4096278 |
| **GIRL** | **40** | 495.200000 | 35.4727776 |



We are exploring this model above by looking at reading scores by gender to get a sense of what the data looks like. From this ANOVA procedure you can see that the probability of this model is 0.03% which means that gender has a significant effect on the outcome. From the model you can see that girls score on average about 30 points higher than boys and that you can see that the standard deviations are approximately the same. From the box plot you can see the difference in the means and quartile values. Therefore, there is clearly an effect of gender on the outcome, reading score.

Next we want to look at math score as a confounding effect related to reading score. We want to see what the data look like before constructing the ANCOVA model. In order for ANCOVA to work the data needs to be parallel or roughly parallel and in the case below you can see when grouped by gender, it creates two parallel lines. The covariate should cause variables of interesting, gender, to create two parallel linea. In other words, the differences between the gender of the boys and girls is consistent no matter what the math score would be.



Now we want to check if truly in fact age is a significant factor, just looking at the data in an exploratory way in order to get an understanding of what the ANCOVA data would look like. A simple regression analysis of the math scores on the reading scores. Given the model you can see that math score is statistically significant on the outcome variable, reading score. When looking at the R square value you can see that math score is roughly 66% of the outcome measure, reading score. With a Beta value of .74196 which is near one shows that as math score increases, reading score increases as well.

|  |
| --- |
| Regression Analysis Reading Score by Math Score |

The REG Procedure

Model: MODEL1

Dependent Variable: Reading

|  |  |
| --- | --- |
| **Number of Observations Read** | 80 |
| **Number of Observations Used** | 80 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 70180 | 70180 | 149.36 | <.0001 |
| **Error** | 78 | 36651 | 469.88155 |  |  |
| **Corrected Total** | 79 | 106830 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 21.67675 | **R-Square** | 0.6569 |
| **Dependent Mean** | 480.71250 | **Adj R-Sq** | 0.6525 |
| **Coeff Var** | 4.50930 |  |  |

| **Parameter Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 123.20630 | 29.35334 | 4.20 | <.0001 |
| **Math** | **1** | 0.74196 | 0.06071 | 12.22 | <.0001 |

Now we look at the ANCOVA since we see that there is parallelism within the data above. Looking at the saturated model with main effects and interactive effects. Given the fact that this is a balanced data set, with data from 40 countries on both boys and girls therefore the Type I model fits best with this data. In regard to the full saturated model, gender turns out to be statistically significantly. Noticing that the R squared value has increased from what it previously was in the regression models. You can also see that the interactive effect math\*gender is not statistically significant with a probability of .4755 and therefore it should be eliminated from the model. Therefore, the analysis of covariance will only depend of reading score and gender as a main effects model. Gender is not significant in this model. We need to establish that the interactive term is not significant to the model

|  |
| --- |
| Reading Score, Gender and Math Score |

The GLM Procedure

Dependent Variable: Reading

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3 | 91857.2151 | 30619.0717 | 155.41 | <.0001 |
| **Error** | 76 | 14973.1724 | 197.0154 |  |  |
| **Corrected Total** | 79 | 106830.3875 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Reading Mean** |
| --- | --- | --- | --- |
| 0.859842 | 2.919878 | 14.03622 | 480.7125 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Gender** | 1 | 16791.01250 | 16791.01250 | 85.23 | <.0001 |
| **Math** | 1 | 74964.90137 | 74964.90137 | 380.50 | <.0001 |
| **Math\*Gender** | 1 | 101.30122 | 101.30122 | 0.51 | 0.4755 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Gender** | 1 | 4.37192 | 4.37192 | 0.02 | 0.8820 |
| **Math** | 1 | 74756.99808 | 74756.99808 | 379.45 | <.0001 |
| **Math\*Gender** | 1 | 101.30122 | 101.30122 | 0.51 | 0.4755 |

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 113.7403973 | B | 26.43466958 | 4.30 | <.0001 |
| **Gender Boy** | -5.6774316 | B | 38.11237933 | -0.15 | 0.8820 |
| **Gender Girl** | 0.0000000 | B | . | . | . |
| **Math** | 0.7959097 | B | 0.05496081 | 14.48 | <.0001 |
| **Math\*Gender Boy** | -0.0565165 | B | 0.07881670 | -0.72 | 0.4755 |
| **Math\*Gender Girl** | 0.0000000 | B | . | . | . |

|  |
| --- |
| Reading Score, Gender and Math Score |

The GLM Procedure

Dependent Variable: Reading

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 91755.9139 | 45877.9569 | 234.34 | <.0001 |
| **Error** | 77 | 15074.4736 | 195.7724 |  |  |
| **Corrected Total** | 79 | 106830.3875 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Reading Mean** |
| --- | --- | --- | --- |
| 0.858893 | 2.910652 | 13.99187 | 480.7125 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Gender** | 1 | 16791.01250 | 16791.01250 | 85.77 | <.0001 |
| **Math** | 1 | 74964.90137 | 74964.90137 | 382.92 | <.0001 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Gender** | 1 | 21576.28750 | 21576.28750 | 110.21 | <.0001 |
| **Math** | 1 | 74964.90137 | 74964.90137 | 382.92 | <.0001 |

| **Parameter** | **Estimate** |  | **Standard Error** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- | --- |
| **Intercept** | 126.9117405 | B | 18.95022826 | 6.70 | <.0001 |
| **Gender Boy** | -32.9131927 | B | 3.13514310 | -10.50 | <.0001 |
| **Gender Female** | 0.0000000 | B | . | . | . |
| **Math** | 0.7684279 |  | 0.03926900 | 19.57 | <.0001 |

In conclusion you can see that the main effects models is statically significant meaning that gender plays a role when it comes to reading scores.

Appendix

Abbreviated Data Sets

Fertility (<https://data.oecd.org/pop/fertility-rates.htm#indicator-chart> )

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | Fertility Rates | |  |  |  |  |
| Location | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| AUS | 2 | 2 | 1.9 | 1.9 | 1.9 | 1.8 | 1.8 | 1.8 | 1.7 | 1.7 |
| AUT | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| BEL | 1.8 | 1.8 | 1.8 | 1.8 | 1.7 | 1.7 | 1.7 | 1.7 | 1.6 | 1.6 |
| CAN | 1.7 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.5 | 1.5 | 1.5 |
| CZE | 1.5 | 1.5 | 1.4 | 1.5 | 1.5 | 1.5 | 1.6 | 1.6 | 1.7 | 1.7 |
| DNK | 1.8 | 1.9 | 1.8 | 1.7 | 1.7 | 1.7 | 1.7 | 1.8 | 1.8 | 1.7 |
| FIN | 1.9 | 1.9 | 1.8 | 1.8 | 1.8 | 1.7 | 1.7 | 1.6 | 1.5 | 1.4 |
| FRA | 2 | 2 | 2 | 2 | 2 | 2 | 1.9 | 1.9 | 1.9 | 1.8 |

……

Crop Production ( <https://data.oecd.org/agroutput/crop-production.htm#indicator-chart> **)**

|  |  |  |
| --- | --- | --- |
| LOCATION | SUBJECT | Value |
| ARG | RICE | 4.619 |
| ARG | RICE | 3.92 |
| ARG | RICE | 4.618 |
| ARG | RICE | 4.53 |
| ARG | RICE | 4.569 |
| ARG | RICE | 4.418 |
| ARG | RICE | 4.561 |
| ARG | RICE | 4.583 |
| ARG | RICE | 4.433 |
| ARG | RICE | 4.698 |
| ARG | WHEAT | 2.758 |
| ARG | WHEAT | 3.504 |
| ARG | WHEAT | 3.225 |
| ARG | WHEAT | 2.658 |
| ARG | WHEAT | 2.662 |
| ARG | WHEAT | 2.81 |
| ARG | WHEAT | 2.862 |
| ARG | WHEAT | 3.305 |
| ARG | WHEAT | 3.181 |
| ARG | WHEAT | 3.216 |
| ARG | SOYBEAN | 2.905 |
| ARG | SOYBEAN | 2.607 |
| ARG | SOYBEAN | 2.788 |
| ARG | SOYBEAN | 2.542 |

Education Performance ( <https://data.oecd.org/pisa/reading-performance-pisa.htm#indicator-chart> )

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Education | | |  |
| Location | Gender | Reading Score | Math Score | Science |
| **AUS** | **0** | 487 | 494 | 504 |
| **AUS** | **1** | 519 | 488 | 502 |
| **AUT** | **0** | 471 | 505 | 491 |
| **AUT** | **1** | 499 | 492 | 489 |
| **BEL** | **0** | 482 | 514 | 501 |
| **BEL** | **1** | 504 | 502 | 496 |
| **CAN** | **0** | 506 | 514 | 516 |
| **CAN** | **1** | 535 | 510 | 520 |
| **CZE** | **0** | 474 | 501 | 496 |
| **CZE** | **1** | 507 | 498 | 498 |
| **DNK** | **0** | 486 | 511 | 492 |
| **DNK** | **1** | 516 | 507 | 494 |
| **FIN** | **0** | 495 | 504 | 510 |
| **FIN** | **1** | 546 | 510 | 534 |

SAS Code

Fertility

**data** rmdata\_plot;

input location $ time value;

;

datalines;

…

**proc** **sgplot** data=rmdata\_plot;

title 'Fertility Rate';

x-axis type=time;

series x=time y=value /Group=location;

**Run**;

**data** rmdata;

input location $ t1 t2 t3 t4 t5 t6 t7 t8 t9 t10;

;

datalines;

…

**proc** **glm** data=rmdata plots=diagnostics;

title 'RM ANOVA for 10 test periods';

model t1-t10 = / nouni;

repeated years profile;

**run**;

**proc** **glm** data=rmdata;

title 'Oneway Repeated Measures with GLM: Fertility';

class Student;

model T1-T10 = / nouni;

repeated years **10** (**1** **2** **3** **4** **5** **6** **7** **8** **9** **10**) polynomial / summary printe;

**run**;

Crop Production

**data** twoway;

input location $ crop $ hectare;

datalines;

…

**proc** **sgplot** data=twoway;

title 'Crop';

x-axis type=time;

series x=crop y=hectare /Group=location;

**Run**;

**proc** **anova** data=twoway;

class crop location;

model hectare = location crop location\*crop;

Means location crop location\*crop / lsd tukey;

**run**;

ods graphics off;

Education Performance

**data** testscores;

input location $ Gender $ Reading Math

;

datalines;

…

**proc** **sgplot** data=testscores;

title 'Math Score as a Confounding Factor Related to Reading Score;

reg x= Reading y=Math / Group = Gender;

**run**;

**proc** **anova** data=testscores;

class gender;

model reading = gender;

means gender;

**run**;

**proc** **reg** data=testscores;

title 'Regression Analysis Reading Score by Math Score';

model Reading = Math;

plot Reading\*Math;

**run**;

**data** testscores;

input location $ Gender Reading Math

;

format Gender sex.;

datalines;

…

**proc** **sort**;

by Gender;

**proc** **glm** data=testscores;

title 'Reading Score, Gender and Math Score: Test for Parallelism';

model Reading = Gender\*Math/ solution;

**run**;

**proc** **sort**;

by Gender;

**proc** **glm** data=testscores;

class Gender;

model Reading = Gender Math Gender\*Math/ solution;

**run**;

**proc** **glm** data=testscores;

class Gender;

model Reading = Gender Math / solution;

lsmeans gender / stderr pdiff cov out=adjmeans;

**run**;

**proc** **print** data = adjmeans;

**run**;