Jonathan Lee

4-17-15

STA4211

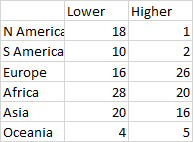
DATA ANALYSIS PROJECT: WRITTEN REPORT

**Objective**: The objective of my study was to determine if a country’s gross domestic product (GDP) per capita was effected by the continent it is located, duration (in years) of secondary school education, or an interaction of those two factors.

**Collection of data**: I found my data from the website <http://data.worldbank.org>. I found two datasets to use. The first included the GDP per capita in 2012 of 166 of the world’s countries, and the second listed the duration (in years) of a full secondary school education of those countries. Finally, I used a world map to determine the continent in which each country is located and recorded those by classifying continents as a numerical factor. For the several cases where a country spans two continents, I made my determination by where the majority of that country’s population resides, which led to me categorizing Russia within Europe and Turkey within Asia.

**Model and assumptions**: For this experiment, I divided each factor into a number of different levels. I broke down continent into the six continents that have a permanent population. As for duration of secondary school education, there were six distinct observations (four years, five years, six years, seven years, eight years, and nine years), so I chose to distinguish this factor into two levels: the first level being shorter duration (four years to six years) and the second level being longer duration (seven years to nine years).

The following table shows how many subjects had observations within each combination of the two factors:

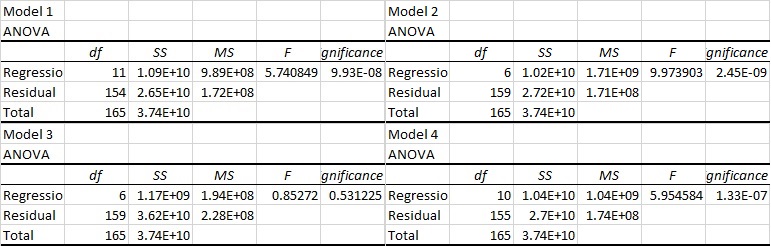


It is clear from the values in the table that this experiment is an unbalanced two-factor study, since the sample sizes are unequal. Thus, the model that I used to fit this dataset was:

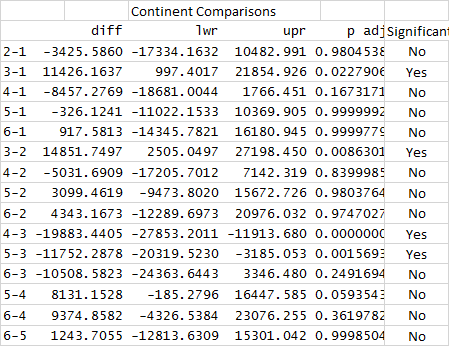
Yijk = µ**..** + ΣαmXijk,m + β6Xijk,6 + Σ(αβ)m,6Xijk,mXijk,6 + εijk, where m = 1,…,5, α = continent factor, β = secondary school education factor, (αβ) is the continent/secondary school education interaction effect, and the Xijk‘s are determined as follows: Xijk,1 = 1 if country is in North America, -1 if country is in Oceania, 0 otherwise; Xijk,2 = 1 if country is in South America, -1 if country is in Oceania, 0 otherwise; Xijk,3 = 1 if country is in Europe, -1 if country is in Oceania, 0 otherwise; Xijk,4 = 1 if country is in Africa, -1 if country is in Oceania, 0 otherwise; Xijk,5 = 1 if country is in Asia, -1 if country is in Oceania, 0 otherwise; and Xijk,6 = 1 if secondary school education is shorter duration, -1 if secondary school education is longer duration.

**Method of Analysis**: To test for main effects and for an interaction effect, I fit four different models to this dataset. In the first model, I included all continent effects, secondary school education effects, and interaction effects; in the second model, I included all continent effects and secondary effects but not interaction effects; in the third model, I included all secondary school education effects and interaction effects but not continent effects; finally, in the last model, I included all continent effects and interaction effects but not secondary school education effects. To test for the various effects, I used the General Linear F-test, which is as follows: , where SSE(R) = Error Sum of Squares of the reduced model, SSE(F) = Error Sum of Squares of the full model, dfe(R) = error degrees of freedom of the reduced model, and dfe(F) = error degrees of freedom of the full model.

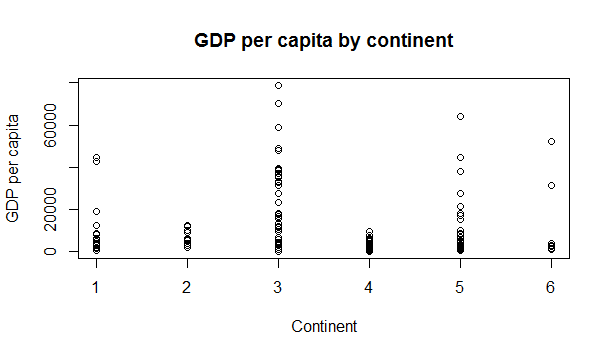
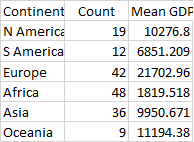
**Analysis**: For each factor being tested, model one was used as the full model. To test for an interaction effect, I set model two as the reduced model. The test statistic was F\* = 0.7538, the rejection region was F\* ≥ F(0.95,5,154) = 2.273, and the p-value was p = 0.5845, thus there is strong evidence that there is not a significant interaction effect. To test for a continent effect, I set model three as the reduced model. The test statistic was F\* = 11.276, the rejection region was F\* ≥ F(0.95,5,154) = 2.273, and the p-value was p ≤ 0.0001, thus there is extremely strong evidence that there is a significant continent effect. Then to test for a secondary school education effect, I set model four as the reduced model. The test statistic was F\* = 2.8809, the rejection region was F\* ≥ F(0.95,1,154) = 3.903, and the p-value was p = 0.092, therefore there is not a significant secondary school education effect. The ANOVA output for each of the four models is as follows:



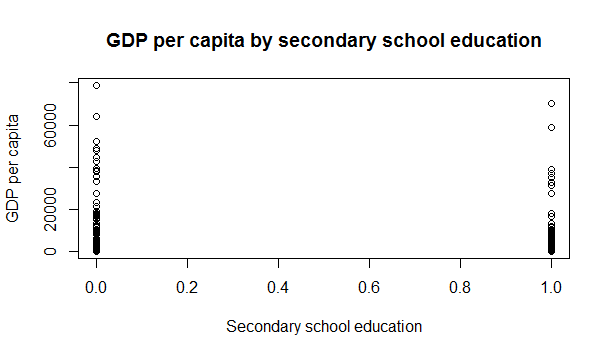
Because I found there to be a significant continent main effect, I then conducted Tukey’s post-hoc tests on the difference between the mean GDP per capita for each pair of two different continents to determine which differences were significant. A chart showing the Tukey’s comparisons is given on the next page. We can see that out of 15 pairs, only four of them are significantly different, but the significance of some of those differences is very high.



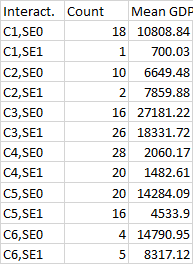
**Conclusions**: The three General Linear F-tests that I conducted demonstrate how a country’s continent of location, length of secondary school education, and an interaction between those two factors generally effect that country’s GDP per capita. The first test gives us enough evidence to conclude that there is a very strong difference between countries’ GDP per capita when considering continent. That is expected, since some continents have, as a whole, a much larger amount of economic activity and higher standards of living than other continents. For example, the continent with the largest mean GDP per capita was Europe at $21,702.96; however, the mean for Africa was $1,819.52, which demonstrates quite a large amount of variability. A chart and plot showing the variability of GDP per capita across different continents are given as:

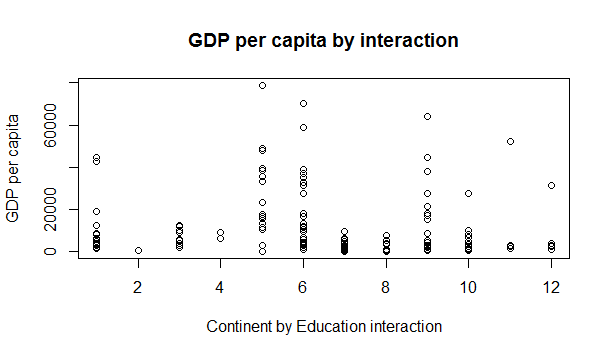


However, results were not significant according to the other two F-tests. The second F-test does not give enough evidence for any significant difference between countries’ GDP per capita when considering length of secondary school education. The mean GDP per capita of countries with a shorter length of secondary education (either four, five, or six years) was $11,442.54; however, the mean GDP per capita of countries with a longer length of secondary education (either seven, eight, or nine years) was $9,097.50. A chart and plot showing the variability of GDP per capita across the two groups of duration of secondary education are given as:



Finally, the third F-test did not give enough evidence for any significant difference between countries’ GDP per capita when considering the 12 different combinations of 6 continents and 2 groups of secondary school education length. A chart and a plot showing the variability of GDP per capita across these twelve groups is given as:





As for the Tukey’s post-hoc comparisons of continent means, there were four pairwise comparisons that were significantly different. These pairs were:

* Europe – North America, with a confidence interval of (997.40,21854.93)
* Europe – South America, with a confidence interval of (2505.05,27198.45)
* Africa – Europe, with a confidence interval of (-27853.20,-11913.68)
* Asia – Europe, with a confidence interval of (-20319.52,-3185.05)

No other comparison of two continent means was significantly different. What we can conclude from the post-hoc tests, with a large degree of confidence, is that the GDP per capita of a European country tends to be significantly higher than that of countries on most other continents.

**Dataset**: my complete data set is included below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country | Subject | Value | Continent | Secondary | Alpha | Beta |
| Afghanistan | 1 | 620.7733 | 5 | 6 | 5 | 0 |
| Albania | 2 | 3734.23 | 3 | 7 | 3 | 1 |
| Algeria | 3 | 4017.986 | 4 | 7 | 4 | 1 |
| Angola | 4 | 2640.926 | 4 | 6 | 4 | 0 |
| Argentina | 5 | 12072.65 | 2 | 6 | 2 | 0 |
| Armenia | 6 | 3109.578 | 3 | 7 | 3 | 1 |
| Australia | 7 | 52162.22 | 6 | 6 | 6 | 0 |
| Austria | 8 | 38997.03 | 3 | 8 | 3 | 1 |
| Azerbaijan | 9 | 4611.441 | 3 | 7 | 3 | 1 |
| Bahamas, The | 10 | 19274.02 | 1 | 6 | 1 | 0 |
| Bahrain | 11 | 15311.56 | 5 | 6 | 5 | 0 |
| Bangladesh | 12 | 728.8834 | 5 | 7 | 5 | 1 |
| Barbados | 13 | 12640.22 | 1 | 5 | 1 | 0 |
| Belarus | 14 | 5761.098 | 3 | 7 | 3 | 1 |
| Belgium | 15 | 35850.93 | 3 | 6 | 3 | 0 |
| Belize | 16 | 4121.915 | 1 | 6 | 1 | 0 |
| Benin | 17 | 665.2148 | 4 | 7 | 4 | 1 |
| Bhutan | 18 | 1824.405 | 5 | 6 | 5 | 0 |
| Bolivia | 19 | 1809.051 | 2 | 6 | 2 | 0 |
| Botswana | 20 | 6151.727 | 4 | 5 | 4 | 0 |
| Brazil | 21 | 9268.998 | 2 | 7 | 2 | 1 |
| Bulgaria | 22 | 5929.595 | 3 | 8 | 3 | 1 |
| Burkina Faso | 23 | 520.8371 | 4 | 7 | 4 | 1 |
| Burundi | 24 | 164.9385 | 4 | 7 | 4 | 1 |
| Cabo Verde | 25 | 3094.823 | 4 | 6 | 4 | 0 |
| Cambodia | 26 | 792.3901 | 5 | 6 | 5 | 0 |
| Cameroon | 27 | 1013.276 | 4 | 7 | 4 | 1 |
| Canada | 28 | 42590.25 | 1 | 6 | 1 | 0 |
| Central African Republic | 29 | 438.5544 | 4 | 7 | 4 | 1 |
| Chad | 30 | 709.3767 | 4 | 7 | 4 | 1 |
| Chile | 31 | 10149.56 | 2 | 6 | 2 | 0 |
| China | 32 | 5130.586 | 5 | 6 | 5 | 0 |
| Colombia | 33 | 5817.069 | 2 | 6 | 2 | 0 |
| Comoros | 34 | 727.8417 | 4 | 7 | 4 | 1 |
| Congo, Dem. Rep. | 35 | 249.6916 | 4 | 6 | 4 | 0 |
| Congo, Rep. | 36 | 470.5923 | 4 | 7 | 4 | 1 |
| Costa Rica | 37 | 8649.832 | 1 | 5 | 1 | 0 |
| Cote d'Ivoire | 38 | 1074.682 | 4 | 7 | 4 | 1 |
| Croatia | 39 | 10365.7 | 3 | 8 | 3 | 1 |
| Cyprus | 40 | 16898.03 | 3 | 6 | 3 | 0 |
| Czech Republic | 41 | 13587.06 | 3 | 8 | 3 | 1 |
| Denmark | 42 | 47920.08 | 3 | 6 | 3 | 0 |
| Dominica | 43 | 5905.51 | 1 | 5 | 1 | 0 |
| Dominican Republic | 44 | 4822.024 | 1 | 6 | 1 | 0 |
| Ecuador | 45 | 4026.476 | 2 | 6 | 2 | 0 |
| Egypt, Arab Rep. | 46 | 2714.834 | 4 | 6 | 4 | 0 |
| El Salvador | 47 | 3433.079 | 1 | 6 | 1 | 0 |
| Equatorial Guinea | 48 | 3962.519 | 4 | 6 | 4 | 0 |
| Eritrea | 49 | 395.5295 | 4 | 7 | 4 | 1 |
| Estonia | 50 | 13471.07 | 3 | 6 | 3 | 0 |
| Ethiopia | 51 | 368.7347 | 4 | 6 | 4 | 0 |
| Fiji | 52 | 3710.895 | 6 | 7 | 6 | 1 |
| Finland | 53 | 38357.36 | 3 | 6 | 3 | 0 |
| France | 54 | 35023.87 | 3 | 7 | 3 | 1 |
| Gabon | 55 | 5259.252 | 4 | 7 | 4 | 1 |
| Gambia, The | 56 | 423.0999 | 4 | 6 | 4 | 0 |
| Georgia | 57 | 3092.513 | 3 | 6 | 3 | 0 |
| Germany | 58 | 37147.03 | 3 | 9 | 3 | 1 |
| Ghana | 59 | 1162.686 | 4 | 7 | 4 | 1 |
| Greece | 60 | 17673.05 | 3 | 6 | 3 | 0 |
| Guatemala | 61 | 2757.583 | 1 | 5 | 1 | 0 |
| Guinea | 62 | 296.2043 | 4 | 7 | 4 | 1 |
| Guinea-Bissau | 63 | 387.8063 | 4 | 5 | 4 | 0 |
| Guyana | 64 | 2809.421 | 2 | 5 | 2 | 0 |
| Haiti | 65 | 700.0308 | 1 | 7 | 1 | 1 |
| Honduras | 66 | 1995.584 | 1 | 5 | 1 | 0 |
| Hungary | 67 | 9809.494 | 3 | 8 | 3 | 1 |
| Iceland | 68 | 31279.88 | 3 | 7 | 3 | 1 |
| India | 69 | 1279.415 | 5 | 7 | 5 | 1 |
| Indonesia | 70 | 3109.308 | 5 | 6 | 5 | 0 |
| Iran, Islamic Rep. | 71 | 4841.655 | 5 | 7 | 5 | 1 |
| Ireland | 72 | 33177.76 | 3 | 5 | 3 | 0 |
| Israel | 73 | 27594.61 | 5 | 6 | 5 | 0 |
| Italy | 74 | 27528.02 | 3 | 8 | 3 | 1 |
| Jamaica | 75 | 4798.357 | 1 | 5 | 1 | 0 |
| Japan | 76 | 38046.74 | 5 | 6 | 5 | 0 |
| Jordan | 77 | 4380.535 | 5 | 6 | 5 | 0 |
| Kazakhstan | 78 | 6530.68 | 5 | 7 | 5 | 1 |
| Kenya | 79 | 837.7386 | 4 | 6 | 4 | 0 |
| Kiribati | 80 | 2386.663 | 6 | 6 | 6 | 0 |
| Korea, Rep. | 81 | 21544.95 | 5 | 6 | 5 | 0 |
| Kyrgyz Republic | 82 | 917.2248 | 5 | 7 | 5 | 1 |
| Lao PDR | 83 | 1017.303 | 5 | 7 | 5 | 1 |
| Latvia | 84 | 11679.08 | 3 | 6 | 3 | 0 |
| Lebanon | 85 | 8659.23 | 5 | 6 | 5 | 0 |
| Lesotho | 86 | 1240.362 | 4 | 5 | 4 | 0 |
| Liberia | 87 | 235.2356 | 4 | 6 | 4 | 0 |
| Lithuania | 88 | 11986.52 | 3 | 8 | 3 | 1 |
| Luxembourg | 89 | 58690.61 | 3 | 7 | 3 | 1 |
| Macedonia, FYR | 90 | 3757.282 | 3 | 8 | 3 | 1 |
| Madagascar | 91 | 374.6392 | 4 | 7 | 4 | 1 |
| Malawi | 92 | 216.8438 | 3 | 6 | 3 | 0 |
| Malaysia | 93 | 8026.794 | 5 | 7 | 5 | 1 |
| Maldives | 94 | 4385.245 | 5 | 5 | 5 | 0 |
| Mali | 95 | 572.415 | 4 | 6 | 4 | 0 |
| Malta | 96 | 16578.3 | 3 | 7 | 3 | 1 |
| Mauritania | 97 | 530.4548 | 4 | 7 | 4 | 1 |
| Mauritius | 98 | 7845.695 | 4 | 7 | 4 | 1 |
| Mexico | 99 | 8095.488 | 1 | 6 | 1 | 0 |
| Micronesia, Fed. Sts. | 100 | 2989.547 | 6 | 6 | 6 | 0 |
| Moldova | 101 | 2082.334 | 3 | 7 | 3 | 1 |
| Mongolia | 102 | 2498.691 | 5 | 6 | 5 | 0 |
| Morocco | 103 | 2454.456 | 4 | 6 | 4 | 0 |
| Mozambique | 104 | 501.1062 | 4 | 5 | 4 | 0 |
| Namibia | 105 | 4971.109 | 4 | 5 | 4 | 0 |
| Nepal | 106 | 643.1855 | 5 | 7 | 5 | 1 |
| Netherlands | 107 | 39301.1 | 3 | 6 | 3 | 0 |
| New Zealand | 108 | 31231.83 | 6 | 7 | 6 | 1 |
| Nicaragua | 109 | 1587.973 | 1 | 5 | 1 | 0 |
| Niger | 110 | 342.81 | 4 | 7 | 4 | 1 |
| Nigeria | 111 | 2244.633 | 4 | 6 | 4 | 0 |
| Norway | 112 | 78812.51 | 3 | 6 | 3 | 0 |
| Oman | 113 | 18241.57 | 5 | 6 | 5 | 0 |
| Pakistan | 114 | 1217.076 | 5 | 7 | 5 | 1 |
| Panama | 115 | 8676.659 | 1 | 6 | 1 | 0 |
| Papua New Guinea | 116 | 1625.355 | 6 | 6 | 6 | 0 |
| Paraguay | 117 | 2903.537 | 2 | 6 | 2 | 0 |
| Peru | 118 | 5292.388 | 2 | 5 | 2 | 0 |
| Philippines | 119 | 2743.235 | 5 | 4 | 5 | 0 |
| Poland | 120 | 10462.27 | 3 | 6 | 3 | 0 |
| Portugal | 121 | 15798.31 | 3 | 6 | 3 | 0 |
| Qatar | 122 | 63936.47 | 5 | 6 | 5 | 0 |
| Romania | 123 | 6665.439 | 3 | 8 | 3 | 1 |
| Russian Federation | 124 | 11325.67 | 3 | 7 | 3 | 1 |
| Rwanda | 125 | 528.9222 | 4 | 6 | 4 | 0 |
| Samoa | 126 | 3023.336 | 6 | 7 | 6 | 1 |
| Sao Tome and Principe | 127 | 1221.214 | 4 | 5 | 4 | 0 |
| Saudi Arabia | 128 | 16966.9 | 5 | 6 | 5 | 0 |
| Senegal | 129 | 889.9781 | 3 | 7 | 3 | 1 |
| Seychelles | 130 | 9725.958 | 4 | 5 | 4 | 0 |
| Sierra Leone | 131 | 503.1122 | 4 | 6 | 4 | 0 |
| Singapore | 132 | 44606.86 | 5 | 4 | 5 | 0 |
| Slovak Republic | 133 | 13229.31 | 3 | 9 | 3 | 1 |
| Slovenia | 134 | 18083.09 | 3 | 7 | 3 | 1 |
| Solomon Islands | 135 | 956.9588 | 6 | 7 | 6 | 1 |
| South Africa | 136 | 5849.272 | 4 | 5 | 4 | 0 |
| Spain | 137 | 23208.4 | 3 | 6 | 3 | 0 |
| Sri Lanka | 138 | 2699.352 | 5 | 8 | 5 | 1 |
| St. Lucia | 139 | 6244.245 | 1 | 5 | 1 | 0 |
| St. Vincent and the Grenadines | 140 | 5615.702 | 1 | 5 | 1 | 0 |
| Sudan | 141 | 1423.606 | 4 | 5 | 4 | 0 |
| Suriname | 142 | 6450.753 | 2 | 7 | 2 | 1 |
| Swaziland | 143 | 2637.197 | 4 | 5 | 4 | 0 |
| Sweden | 144 | 48980.26 | 3 | 6 | 3 | 0 |
| Switzerland | 145 | 70071.26 | 3 | 7 | 3 | 1 |
| Tajikistan | 146 | 844.2911 | 5 | 7 | 5 | 1 |
| Tanzania | 147 | 508.7599 | 4 | 6 | 4 | 0 |
| Thailand | 148 | 4211.551 | 5 | 6 | 5 | 0 |
| Togo | 149 | 403.3457 | 4 | 7 | 4 | 1 |
| Trinidad and Tobago | 150 | 8549.821 | 1 | 5 | 1 | 0 |
| Tunisia | 151 | 3238.193 | 4 | 7 | 4 | 1 |
| Turkey | 152 | 9957.552 | 5 | 7 | 5 | 1 |
| Turkmenistan | 153 | 3327.402 | 5 | 7 | 5 | 1 |
| Uganda | 154 | 420.7007 | 4 | 6 | 4 | 0 |
| Ukraine | 155 | 3475.051 | 3 | 7 | 3 | 1 |
| United Arab Emirates | 156 | 27788.9 | 5 | 7 | 5 | 1 |
| United Kingdom | 157 | 32905.83 | 3 | 7 | 3 | 1 |
| United States | 158 | 44800.81 | 1 | 6 | 1 | 0 |
| Uruguay | 159 | 12263.56 | 2 | 6 | 2 | 0 |
| Uzbekistan | 160 | 1350.43 | 5 | 7 | 5 | 1 |
| Vanuatu | 161 | 2662.587 | 6 | 7 | 6 | 1 |
| Venezuela, RB | 162 | 9351.056 | 2 | 5 | 2 | 0 |
| Vietnam | 163 | 1372.308 | 5 | 7 | 5 | 1 |
| Yemen, Rep. | 164 | 1076.093 | 5 | 6 | 5 | 0 |
| Zambia | 165 | 1103.008 | 4 | 5 | 4 | 0 |
| Zimbabwe | 166 | 711.7976 | 4 | 6 | 4 | 0 |

**R code**

> GDP <- read.table("C:/Data/gdpeducationdata.txt",col.names=c("country","gdp","continent","secondary"))

> n <- 166

>

> for (i in 1:n) {

+ if (GDP$secondary[i] == 4 || GDP$secondary[i] == 5 || GDP$secondary[i] == 6) {

+ GDP$secondary[i] <- 0

+ } else if (GDP$secondary[i] == 7 || GDP$secondary[i] == 8 || GDP$secondary[i] == 9) {

+ GDP$secondary[i] <- 1

+ }

+ }

>

> secondary <- as.factor(GDP$secondary)

> continent <- as.factor(GDP$continent)

> data.table <- table(continent,secondary); data.table

secondary

continent 0 1

1 18 1

2 10 2

3 16 26

4 28 20

5 20 16

6 4 5

>

> value.matrix <- matrix(c(rep(0,12)),nrow=6,byrow=TRUE)

> for (i in 1:n) {

+ for (j in 1:6) {

+ if (GDP$continent[i] == j) {

+ if (GDP$secondary[i] == 0) {

+ value.matrix[j,1] <- value.matrix[j,1] + 1

+ } else if (GDP$secondary[i] == 1) {

+ value.matrix[j,2] <- value.matrix[j,2] + 1

+ }

+ }

+ }

+ }

>

> rowSums(value.matrix)

[1] 19 12 42 48 36 9

> colSums(value.matrix)

[1] 96 70

>

> x1 <- c(rep(0,n))

> x2 <- c(rep(0,n))

> x3 <- c(rep(0,n))

> x4 <- c(rep(0,n))

> x5 <- c(rep(0,n))

> x6 <- c(rep(0,n))

> for (i in 1:n) {

+ if (GDP$continent[i] == 1) {

+ x1[i] <- 1

+ } else if (GDP$continent[i] == 6) {

+ x1[i] <- -1

+ } else {

+ x1[i] <- 0

+ }

+

+ if (GDP$continent[i] == 2) {

+ x2[i] <- 1

+ } else if (GDP$continent[i] == 6) {

+ x2[i] <- -1

+ } else {

+ x2[i] <- 0

+ }

+

+ if (GDP$continent[i] == 3) {

+ x3[i] <- 1

+ } else if (GDP$continent[i] == 6) {

+ x3[i] <- -1

+ } else {

+ x3[i] <- 0

+ }

+

+ if (GDP$continent[i] == 4) {

+ x4[i] <- 1

+ } else if (GDP$continent[i] == 6) {

+ x4[i] <- -1

+ } else {

+ x4[i] <- 0

+ }

+

+ if (GDP$continent[i] == 5) {

+ x5[i] <- 1

+ } else if (GDP$continent[i] == 6) {

+ x5[i] <- -1

+ } else {

+ x5[i] <- 0

+ }

+

+ if (GDP$secondary[i] == 0) {

+ x6[i] <- 1

+ } else if (GDP$secondary[i] == 1) {

+ x6[i] <- -1

+ }

+ }

>

> x1x6 <- x1\*x6

> x2x6 <- x2\*x6

> x3x6 <- x3\*x6

> x4x6 <- x4\*x6

> x5x6 <- x5\*x6

>

> GDP.mod1 <- lm(GDP$gdp~x1+x2+x3+x4+x5+x6+x1x6+x2x6+x3x6+x4x6+x5x6)

> summary(GDP.mod1)

Call:

lm(formula = GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x6 + x1x6 + x2x6 +

x3x6 + x4x6 + x5x6)

Residuals:

Min 1Q Median 3Q Max

-26964 -6262 -1510 1973 51740

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9750 1696 5.748 4.70e-08 \*\*\*

x1 -3996 5761 -0.694 0.489009

x2 -2495 4484 -0.556 0.578691

x3 13006 2404 5.412 2.35e-07 \*\*\*

x4 -7979 2311 -3.453 0.000716 \*\*\*

x5 -341 2471 -0.138 0.890437

x6 2879 1696 1.697 0.091655 .

x1x6 2175 5761 0.378 0.706258

x2x6 -3484 4484 -0.777 0.438330

x3x6 1546 2404 0.643 0.521125

x4x6 -2590 2311 -1.121 0.263999

x5x6 1996 2471 0.808 0.420557

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

Residual standard error: 13130 on 154 degrees of freedom

Multiple R-squared: 0.2908, Adjusted R-squared: 0.2402

F-statistic: 5.741 on 11 and 154 DF, p-value: 9.931e-08

> anova(GDP.mod1)

Analysis of Variance Table

Response: GDP$gdp

Df Sum Sq Mean Sq F value Pr(>F)

x1 1 3.6696e+06 3669638 0.0213 0.88416

x2 1 1.2028e+08 120281570 0.6981 0.40472

x3 1 5.7808e+09 5780787587 33.5507 3.786e-08 \*\*\*

x4 1 3.1545e+09 3154483232 18.3081 3.288e-05 \*\*\*

x5 1 4.4358e+06 4435760 0.0257 0.87274

x6 1 1.1676e+09 1167558447 6.7763 0.01014 \*

x1x6 1 2.0374e+06 2037390 0.0118 0.91355

x2x6 1 9.6498e+07 96498263 0.5601 0.45538

x3x6 1 1.1418e+08 114175326 0.6627 0.41688

x4x6 1 3.2434e+08 324336188 1.8824 0.17206

x5x6 1 1.1238e+08 112381976 0.6522 0.42056

Residuals 154 2.6534e+10 172300229

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

>

> GDP.mod2 <- lm(GDP$gdp~x1+x2+x3+x4+x5+x6)

> summary(GDP.mod2)

Call:

lm(formula = GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x6)

Residuals:

Min 1Q Median 3Q Max

-25047 -6414 -3054 2496 53549

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9585.14 1261.74 7.597 2.45e-12 \*\*\*

x1 -1881.33 2832.33 -0.664 0.50750

x2 -4651.06 3350.73 -1.388 0.16706

x3 12802.51 2125.46 6.023 1.14e-08 \*\*\*

x4 -8244.90 1974.82 -4.175 4.90e-05 \*\*\*

x5 46.01 2169.37 0.021 0.98310

x6 2875.70 1100.42 2.613 0.00983 \*\*

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

Residual standard error: 13080 on 159 degrees of freedom

Multiple R-squared: 0.2735, Adjusted R-squared: 0.246

F-statistic: 9.974 on 6 and 159 DF, p-value: 2.448e-09

> anova(GDP.mod2)

Analysis of Variance Table

Response: GDP$gdp

Df Sum Sq Mean Sq F value Pr(>F)

x1 1 3.6696e+06 3669638 0.0215 0.883707

x2 1 1.2028e+08 120281570 0.7035 0.402856

x3 1 5.7808e+09 5780787587 33.8124 3.233e-08 \*\*\*

x4 1 3.1545e+09 3154483232 18.4509 3.025e-05 \*\*\*

x5 1 4.4358e+06 4435760 0.0259 0.872239

x6 1 1.1676e+09 1167558447 6.8292 0.009829 \*\*

Residuals 159 2.7184e+10 170966443

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

>

> GDP.mod3 <- lm(GDP$gdp~x6+x1x6+x2x6+x3x6+x4x6+x5x6)

> summary(GDP.mod3)

Call:

lm(formula = GDP$gdp ~ x6 + x1x6 + x2x6 + x3x6 + x4x6 + x5x6)

Residuals:

Min 1Q Median 3Q Max

-14691 -8887 -4959 2070 66725

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 10565.3 1270.7 8.314 3.89e-14 \*\*\*

x6 1026.1 1457.0 0.704 0.482

x1x6 -276.1 3270.7 -0.084 0.933

x2x6 -3838.4 3869.3 -0.992 0.323

x3x6 496.0 2454.4 0.202 0.840

x4x6 -2203.0 2280.4 -0.966 0.335

x5x6 3720.5 2505.1 1.485 0.139

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

Residual standard error: 15100 on 159 degrees of freedom

Multiple R-squared: 0.03117, Adjusted R-squared: -0.005384

F-statistic: 0.8527 on 6 and 159 DF, p-value: 0.5312

> anova(GDP.mod3)

Analysis of Variance Table

Response: GDP$gdp

Df Sum Sq Mean Sq F value Pr(>F)

x6 1 2.2262e+08 222619279 0.9765 0.3246

x1x6 1 1.6023e+07 16023410 0.0703 0.7913

x2x6 1 1.7681e+08 176814496 0.7756 0.3798

x3x6 1 1.2228e+07 12227709 0.0536 0.8172

x4x6 1 2.3586e+08 235858469 1.0346 0.3106

x5x6 1 5.0286e+08 502864222 2.2058 0.1395

Residuals 159 3.6248e+10 227977818

>

> GDP.mod4 <- lm(GDP$gdp~x1+x2+x3+x4+x5+x1x6+x2x6+x3x6+x4x6+x5x6)

> summary(GDP.mod4)

Call:

lm(formula = GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x1x6 + x2x6 +

x3x6 + x4x6 + x5x6)

Residuals:

Min 1Q Median 3Q Max

-26066 -7539 -1603 2266 52529

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 11303 1437 7.866 5.84e-13 \*\*\*

x1 1236 4897 0.252 0.8011

x2 -1175 4443 -0.264 0.7918

x3 11281 2191 5.149 7.85e-07 \*\*\*

x4 -9429 2160 -4.366 2.30e-05 \*\*\*

x5 -1804 2330 -0.774 0.4399

x1x6 -2528 5081 -0.498 0.6195

x2x6 -4915 4431 -1.109 0.2690

x3x6 3700 2054 1.802 0.0736 .

x4x6 -327 1898 -0.172 0.8635

x5x6 4067 2162 1.881 0.0619 .

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

Residual standard error: 13210 on 155 degrees of freedom

Multiple R-squared: 0.2775, Adjusted R-squared: 0.2309

F-statistic: 5.955 on 10 and 155 DF, p-value: 1.326e-07

> anova(GDP.mod4)

Analysis of Variance Table

Response: GDP$gdp

Df Sum Sq Mean Sq F value Pr(>F)

x1 1 3.6696e+06 3669638 0.0210 0.88485

x2 1 1.2028e+08 120281570 0.6897 0.40754

x3 1 5.7808e+09 5780787587 33.1484 4.445e-08 \*\*\*

x4 1 3.1545e+09 3154483232 18.0886 3.633e-05 \*\*\*

x5 1 4.4358e+06 4435760 0.0254 0.87349

x1x6 1 7.2979e+06 7297915 0.0418 0.83818

x2x6 1 7.2858e+07 72858222 0.4178 0.51900

x3x6 1 6.2202e+08 622017338 3.5668 0.06081 .

x4x6 1 1.4827e+06 1482691 0.0085 0.92665

x5x6 1 6.1695e+08 616948895 3.5377 0.06186 .

Residuals 155 2.7031e+10 174391082

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

>

> anova(GDP.mod2,GDP.mod1)

Analysis of Variance Table

Model 1: GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x6

Model 2: GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x6 + x1x6 + x2x6 + x3x6 +

x4x6 + x5x6

Res.Df RSS Df Sum of Sq F Pr(>F)

1 159 2.7184e+10

2 154 2.6534e+10 5 649429143 0.7538 0.5845

> anova(GDP.mod3,GDP.mod1)

Analysis of Variance Table

Model 1: GDP$gdp ~ x6 + x1x6 + x2x6 + x3x6 + x4x6 + x5x6

Model 2: GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x6 + x1x6 + x2x6 + x3x6 +

x4x6 + x5x6

Res.Df RSS Df Sum of Sq F Pr(>F)

1 159 3.6248e+10

2 154 2.6534e+10 5 9714237792 11.276 2.813e-09 \*\*\*

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

> anova(GDP.mod4,GDP.mod1)

Analysis of Variance Table

Model 1: GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x1x6 + x2x6 + x3x6 + x4x6 +

x5x6

Model 2: GDP$gdp ~ x1 + x2 + x3 + x4 + x5 + x6 + x1x6 + x2x6 + x3x6 +

x4x6 + x5x6

Res.Df RSS Df Sum of Sq F Pr(>F)

1 155 2.7031e+10

2 154 2.6534e+10 1 496382529 2.8809 0.09166 .

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

>

> qf(0.95,5,154)

[1] 2.272901

> qf(0.95,1,154)

[1] 3.902553

>

> plot(GDP$country,GDP$gdp,xlab="Country",ylab="GDP per capita",main="GDP per capita by country")

> plot(GDP$continent,GDP$gdp,xlab="Continent",ylab="GDP per capita",main="GDP per capita by continent")

> plot(GDP$secondary,GDP$gdp,xlab="Secondary school education",ylab="GDP per capita",main="GDP per capita by secondary school education")

>

> europe <- 0; e <- 0

> africa <- 0; a <- 0

> for (i in 1:n) {

+ if (GDP$continent[i] == 3) {

+ europe <- europe + GDP$gdp[i]

+ e <- e + 1

+ } else if (GDP$continent[i] == 4) {

+ africa <- africa + GDP$gdp[i]

+ a <- a + 1

+ }

+ }

> europe.mean <- europe/e; europe.mean

[1] 21702.96

> africa.mean <- africa/a; africa.mean

[1] 1819.518

> tapply(GDP$gdp,GDP$continent,mean)

1 2 3 4 5 6

10276.795 6851.209 21702.959 1819.518 9950.671 11194.376

> shorter <- 0; longer <- 0

> for (i in 1:n) {

+ if (GDP$secondary[i] == 0) {

+ shorter <- shorter + 1

+ } else if (GDP$secondary[i] == 1) {

+ longer <- longer + 1

+ }

+ }

> shorter; longer

[1] 96

[1] 70

> tapply(GDP$gdp,GDP$secondary,mean)

0 1

11442.538 9097.495

>

> g.sums <- c(rep(0,12))

> g.means <- c(rep(0,12))

> g.counts <- c(value.matrix[1,],value.matrix[2,],value.matrix[3,],value.matrix[4,],value.matrix[5,],value.matrix[6,])

> interaction.means <- function(cont,sec,iter) {

+ g.sum.iter <- 0

+ for (i in 1:n) {

+ if (GDP$secondary[i] == sec) {

+ if (GDP$continent[i] == cont) {

+ g.sum.iter <- g.sum.iter + GDP$gdp[i]

+ }

+ }

+ }

+ return(g.sum.iter)

+ }

>

> k <- 1

> for (i in 1:6) {

+ for (j in 0:1) {

+ g.sums[k] <- interaction.means(i,j,k)

+ k <- k + 1

+ }

+ }

> for (i in 1:12) {

+ g.means[i] <- (g.sums[i])/(g.counts[i])

+ }

> cbind(g.counts,round(g.means,2))

g.counts

[1,] 18 10808.84

[2,] 1 700.03

[3,] 10 6649.48

[4,] 2 7859.88

[5,] 16 27181.22

[6,] 26 18331.72

[7,] 28 2060.17

[8,] 20 1482.61

[9,] 20 14284.09

[10,] 16 4533.90

[11,] 4 14790.95

[12,] 5 8317.12

>

> interaction.plot <- function(cont,sec,iter,d) {

+ result <- c(rep(0,d))

+ t <- 1

+ for (i in 1:n) {

+ if (GDP$secondary[i] == sec) {

+ if (GDP$continent[i] == cont) {

+ result[t] <- iter

+ t <- t + 1

+ result[t] <- GDP$gdp[i]

+ t <- t + 1

+ }

+ }

+ }

+ return(result)

+ }

>

> g.interact <- c()

> s <- 1

> for (i in 1:6) {

+ for (j in 0:1) {

+ u <- 2\*(g.counts[s])

+ g.interact <- c(g.interact,interaction.plot(i,j,s,u))

+ s <- s + 1

+ }

+ }

>

> plot.matrix <- matrix(c(rep(0,2\*n)),nrow=n,byrow=TRUE)

> x <- 1

> for (i in 1:n) {

+ plot.matrix[i,1] <- g.interact[x]

+ x <- x + 1

+ plot.matrix[i,2] <- g.interact[x]

+ x <- x + 1

+ }

>

> plot(plot.matrix[,1],plot.matrix[,2],xlab="Continent by Education interaction",ylab="GDP per capita",main="GDP per capita by interaction")

>

> GDP.mod5 <- aov(GDP$gdp~factor(GDP$continent)+factor(GDP$secondary))

> anova(GDP.mod5)

Analysis of Variance Table

Response: GDP$gdp

Df Sum Sq Mean Sq F value Pr(>F)

factor(GDP$continent) 5 9.0637e+09 1812731557 10.6029 8.379e-09 \*\*\*

factor(GDP$secondary) 1 1.1676e+09 1167558447 6.8292 0.009829 \*\*

Residuals 159 2.7184e+10 170966443

---

Signif. codes: 0 ・\*\*・0.001 ・\*・0.01 ・・0.05 ・・0.1 ・・1

> TukeyHSD(GDP.mod5)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = GDP$gdp ~ factor(GDP$continent) + factor(GDP$secondary))

$`factor(GDP$continent)`

diff lwr upr p adj

2-1 -3425.5860 -17334.1632 10482.991 0.9804538

3-1 11426.1637 997.4017 21854.926 0.0227906

4-1 -8457.2769 -18681.0044 1766.451 0.1673171

5-1 -326.1241 -11022.1533 10369.905 0.9999992

6-1 917.5813 -14345.7821 16180.945 0.9999779

3-2 14851.7497 2505.0497 27198.450 0.0086301

4-2 -5031.6909 -17205.7012 7142.319 0.8399985

5-2 3099.4619 -9473.8020 15672.726 0.9803764

6-2 4343.1673 -12289.6973 20976.032 0.9747027

4-3 -19883.4405 -27853.2011 -11913.680 0.0000000

5-3 -11752.2878 -20319.5230 -3185.053 0.0015693

6-3 -10508.5823 -24363.6443 3346.480 0.2491694

5-4 8131.1528 -185.2796 16447.585 0.0593543

6-4 9374.8582 -4326.5384 23076.255 0.3619782

6-5 1243.7055 -12813.6309 15301.042 0.9998504

$`factor(GDP$secondary)`

diff lwr upr p adj

1-0 -5014.694 -9073.433 -955.9548 0.015779

>

> secondary.mean <- as.vector(tapply(GDP$gdp,GDP$secondary,mean)); secondary.mean

[1] 11442.538 9097.495

> continent.mean <- as.vector(tapply(GDP$gdp,GDP$continent,mean)); continent.mean

[1] 10276.795 6851.209 21702.959 1819.518 9950.671 11194.376