Women of the world - A statistical approach

Introduction

Mankind always searched for answers on the great questions. Where are we coming from? Where are we going to? Why does attractiveness and easiness of women around the world differ?

This document tries to answer the last question with a statistical approach. For this, 134 countries were investigated on the relationship between attractiveness / easiness and a set of further variables.

Disclaimer:
The following theoretical models do not reflect the real world in its full accuracy and might contain statistical errors based on data selection, wrong source data or other mistakes. Furthermore, only models for attractiveness and easiness are calculated. "Enjoyability" of time spent with women from a certain country isn't taken into account.

Variables explained

Easiness: Easiness refers to the amount of effort, men have to put in to sleep with women of a certain country. Range 1 - 5, with 1 being very difficult and 5 being very easy.

Attractiveness: The beauty of women in a certain country is described by attractiveness. Range 1 - 5, with 1 being very ugly and 5 being very sexy.

GDP per Capita: The GDP of a country divided by its population.

Economical Gender Gap: The difference in economic opportunities between men and women. Range 0 - 1, with 0 being perfect inequality and 1 perfect equality.

Birth Rate: Number of births per 1000 people.

Median Age: The median age splits the population in two equally large groups, which are divided by age. Therefore, half of the population is younger and half of the population is older than the median age.

Fragile States Index: Annual index of the US think tank Fund for Peace. The Fragile States Index measures the stability of a state and considers 12 indicators in its computation such as rule of law or state legitimacy. The higher the value of the FSI, the lower the stability of the country.

Culture: The author of this paper decided that 7 different culture groups could be relevant. Protestant; Roman; Orthodox; Islamic; African; Latin American; Sino. Every country that fits into one of the previously mentioned groups, got assigned to one. Value 0 or 1, with 0 describing the fact of not being part of a group and 1 being part.
Important note: A country can only be member of one culture group.

How to read

Plot of Fitted Model
Fragile States Index = 138,467 - 2,27854*Median Age
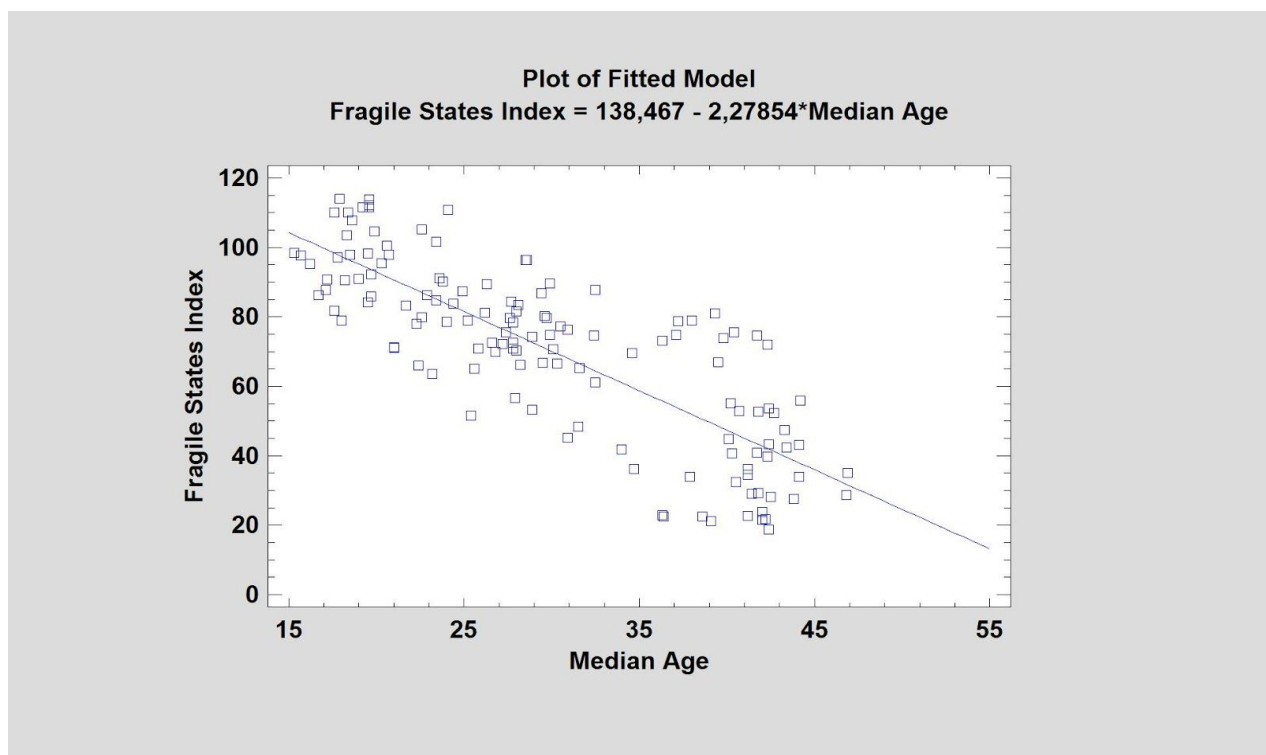
The plot above shows the relationship between the Fragile States Index (FSI) and the Median Age. As one can clearly see, the relationship is a negative one. The higher the Median Age, the lower the FSI in a certain country. This makes sense, since older populations are more risk averse and in almost all cases higher developed. The line,

which goes through the whole plot, is called regression line and gives the predicted value of FSI for each median age based on our prediction model.

The statistics program also gives us some detailed numbers:

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	138,487	4,5192	30,6397
Slope	-2,27854	0,14438	-15,7816

Every variable (in this case only one, the slope aka median age), is significant for the relationship, if its T-Statistic has an <u>absolute</u> <u>value</u> greater than 2. An absolute T-Statistic value of 2 is equivalent with a 95% confidence in the variable's significance. The Intercept is a constant and the value by which the regression line intercepts the Y-axis (the starting point of the line). Only the Intercept is allowed to have an absolute T-Statistic below 2.

This leads us to our prediction model:

Constant   Slope   R²-Statistic
⇩   ⇩   ⇩
Fragile States Index = 138,467 - 2,27854*Median Age
                              (-15,7816)   R² = 65,5316%
                                 ⇧
                           T-Statistic

Interpretation:

The T-Statistic of the variable Median Age has an absolute value greater than 2. This means there is a significant correlation between median age and FSI.

The FSI has a value of 138,467 in a country with a theoretical Median Age of 0. Since the relationship between FSI and the median age is a negative one (indicated by the negative sign), the FSI decreases by 2,27854 for each unit of Median Age.

The R²-Statistic indicates how much of the variability is explained by the model. Therefore, the model as fitted explains 65,5316% of the variability in Fragile States Index.

Simple regressions

In the following paragraph, the relationships between one variable and attractiveness and easiness will be shown. These relationships are called simple regressions.

GDP per Capita:

Attractiveness = 2,821 + 0,00000315*GDP per Capita

                (0,634) $R^2 = 0,30\%$

Easiness = 3,582 - 0,0000139*GDP per Capita

                (-2,480) $R^2 = 4,42\%$

Economical Gender Gap:

Attractiveness = 2,052 + 1,320*Economical Gender Gap

                (2,646) $R^2 = 5,04\%$

Easiness = 1,591 + 2,949*Economical Gender Gap

                (5,536) $R^2 = 18,84\%$

Birth Rate:

Attractiveness = 3,671 - 0,0838*Birth Rate

                (-5,631) $R^2 = 19,25\%$

Easiness = 3,0998 + 0,01488*Birth Rate

                (1,723) $R^2 = 2,18\%$

Median Age:

Attractiveness = 1,578 + 0,0427*Median Age

                (5,215) $R^2 = 16,98\%$

Easiness = 3,837 - 0,014*Median Age

                (-1,365) $R^2 = 1,38\%$

Fragile States Index:

Attractiveness = 3,563 - 0,00993*Fragile States Index

$\quad\quad\quad\quad\quad\quad\quad\quad$ (-3,218) $\quad\quad\quad\quad\quad\quad$ R^2 = 7,27%

Easiness = 3,233 + 0,00253*Fragile States Index

$\quad\quad\quad\quad\quad\quad\quad\quad$ (0,683) $\quad\quad\quad\quad\quad\quad$ R^2 = 0,35%

Protestant:

Attractiveness = 2,885 - 0,2699*Protestant

$\quad\quad\quad\quad\quad\quad\quad\quad$ (-0,992) $\quad\quad\quad\quad\quad\quad$ R^2 = 0,74%

Easiness = 3,451 - 0,3739*Protestant

$\quad\quad\quad\quad\quad\quad\quad\quad$ (-1,194) $\quad\quad\quad\quad\quad\quad$ R^2 = 1,06%

Roman:

Attractiveness = 2,848 + 0,152*Roman

$\quad\quad\quad\quad\quad\quad\quad\quad$ (0,495) $\quad\quad\quad\quad\quad\quad$ R^2 = 0,18%

Easiness = 3,472 - 0,772*Roman

$\quad\quad\quad\quad\quad\quad\quad\quad$ (-2,218) $\quad\quad\quad\quad\quad\quad$ R^2 = 3,57%

Orthodox:

Attractiveness = 2,735 + 0,765*Orthodox

$\quad\quad\quad\quad\quad\quad\quad\quad$ (3,688) $\quad\quad\quad\quad\quad\quad$ R^2 = 9,28%

Easiness = 3,389 + 0,156*Orthodox

$\quad\quad\quad\quad\quad\quad\quad\quad$ (0,622) $\quad\quad\quad\quad\quad\quad$ R^2 = 0,29%

Islamic:

Attractiveness = 2,991 - 0,683*Islamic

$\quad\quad\quad\quad\quad\quad\quad\quad$ (-3,497) $\quad\quad\quad\quad\quad\quad$ R^2 = 8,42%

Easiness = 3,716 - 1,562*Islamic

$\quad\quad\quad\quad\quad\quad\quad\quad$ (-8,109) $\quad\quad\quad\quad\quad\quad$ R^2 = 33,08%

African:

Attractiveness = 2,912 - 0,341*African

                                 (-1.549) $R^2 = 1{,}77\%$

Easiness = 3,211 + 1,313*African

                                 (5,723) $R^2 = 19{,}76\%$

Latin American:

Attractiveness = 2,793 + 0,470*Latin American

                                 (2,063) $R^2 = 3{,}10\%$

Easiness = 3,345 + 0,497*Latin American

                                 (1,887) $R^2 = 2{,}61\%$

Sino:

Attractiveness = 2,84 + 0,382*Sino

                                 (1,190) $R^2 = 1{,}06\%$

Easiness = 3,392 + 0,275*Sino

                                 (5,723) $R^2 = 0{,}41\%$

Multiple regressions - Explanation

By combining multiple of the above variables into one model, the attractiveness and easy of a country's women can be better predicted. So, while simple regressions explain the relationship between two variables, multiple regressions illustrate the relationship between one variable explained by multiple variables. With multiple regressions, problems with overlapping explanatory value can occur.

As an example, let's look at GDP per Capita explained by FSI, Median Age and Birth Rate. First the 3 simple regressions:

GDP per Capita = 47.710,3 - 510,458*Fragile States Index

                                 (-15,051) $R^2 = 63{,}18\%$

GDP per Capita = 29.127,2 - 807,95*Birth Rate

          (-7,218) R^2 = 28,17%

GDP per Capita = -22.000 + 1.132,55*Median Age

          (9,279) R^2 = 39,30%

All variables above have a T-Statistic of an absolute value greater than 2, and therefore a significant relationship with GDP per Capita.

Now all variables are put together in a single model.

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	21.729,0	13.463,9	1,61387
Fragile States Index	-548,254	56,3455	-9,73023
Birth Rate	597,04	203,195	2,93826
Median Age	531,94	267,737	1,9868

In the combined model, Median Age has an absolute T-Statistic value below 2. Although the T-Statistic indicates insignificance, <u>Median Age can't be declared insignificant</u>, since the previous simple regression has clearly shown a significant correlation between GDP per Capita and Median Age. More probable, Birth Rate and Median Age share the same explanatory value. This effect is called multicollinearity.

In a next step, every variable with an absolute T-Statistic below 2 is removed from the model, starting with the ones with the lowest absolute T-Statistic. Afterwards the multiple regression is being recalculated.

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	48.026,0	2.495,22	19,2472
Fragile States Index	-597,231	51,2338	-11,657
Birth Rate	271,52	121,529	2,2342

The final model:

GDPpC = 48.026,5 - 597,231*FSI + 271,52*Birth Rate

        (-11,657)   (2,234) R^2 = 63,99%

Interpretation:

The interpretation of a multiple regressions differs slightly from a simple regression ones. The explained variable's value changes by the amount of an explaining variable's factor for each unit added, <u>if all other variables stay constant</u>.

For each additional unit of Fragile States Index, GDP per Capita decreases by 596,738, if Birth Rate doesn't change.

In a country with a given amount of Fragile States Index, the GDP per Capita increases by 270,266 for each additional unit of Birth Rate.

United States:
GDP per Capita = $55.836,79
Fragile States Index = 34
Birth Rate = 13

GDP per Capita in US = 48.026,5 - 597,231 * 34 + 271,52 * 13 = 31.250,41

Because the model explains only 64,01% of GDP per Capita, no perfect result for the US example is being predicted.

Multiple regressions - Results

1.) Attractiveness multiple regression

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	1,34207	1,38051	0,972157
Fragile States Index	0,00479772	0,00667613	0,718638
GDP per Capita	-0,00000241038	0,00000942519	-0,255738
Birth Rate	-0,0317512	0,0211556	-1,50084
Median Age	0,0480761	0,028747	1,67239
Easiness	0,0280935	0,100707	0,278963
Economical Gender Gap	-0,0229356	0,67943	-0,0337571
Roman	-0,162841	0,353262	-0,460965

Protestant	-0,320147	0,393348	-0,813902
Orthodox	0,391165	0,307396	1,27251
Islamic	0,273682	0,321372	0,851605
African	0,92176	0,331545	2,7802
Latin American	0,859965	0,277354	3,1006
Sino	0,511462	0,33365	1,53293

Adjusted:

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	1,04119	0,314504	3,31057
Median Age	0,048966	0,00962238	5,08876
Orthodox	0,586086	0,205076	2,8579
African	0,594523	0,227708	2,6109
Latin American	0,928442	0,210743	4,40556
Sino	0,64187	0,278174	2,30744

The final model:

A = 1,041 + 0,049*MA + 0,586*Ort + 0,595*Afr + 0,928*LatA + 0,642*S
                    (5,089)   (2,858)   (2,611)   (4,406)   (2,307)   R² = 29,75%

A = Attractiveness
MA = Median Age
Ort = Orthodox
Afr = African
LatA = Latin American
S = Sino

2.) Easiness multiple regression

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	-0,281131	1,25563	-0,223896
Fragile States Index	0,00385625	0,00605248	0,637135
GDP per Capita	-0,000021309	0,00000831874	-2,56157
Birth Rate	0,0438027	0,018932	2,31368
Median Age	0,0556699	0,0258568	2,15301
Attractiveness	0,0230688	0,0826949	0,278963
Economical Gender Gap	2,52015	0,571085	4,41291
Roman	-1,01059	0,30683	-3,29363
Protestant	-0,514769	0,35432	-1,45283
Orthodox	-0,764798	0,271596	-2,81594
Islamic	-1,44113	0,260794	-5,52595
African	0,172765	0,30956	0,558098
Latin American	-0,0696668	0,261127	-0,266793
Sino	-0,480998	0,302116	-1,5921

Adjusted:

	Least Squares Estimate	Standard Error	T-Statistic
Intercept	0,226472	0,997431	0,227055
GDP per Capita	-0,0000279936	0,00000590861	-4,73777
Birth Rate	0,0513196	0,0159642	3,21468
Median Age	0,0493432	0,023826	2,07098
Economical Gender Gap	2,23372	0,540796	4,13043
Roman	-0,723594	0,259485	-2,78858
Orthodox	-0,557573	0,228411	-2,4411
Islamic	-1,4669	0,204889	-7,15947

The final model:

$$E = 0{,}226 - 0{,}000028 \cdot GDP + 0{,}051 \cdot BR + 0{,}049 \cdot MA + 2{,}234 \cdot EGG$$
$$\qquad\quad (-4{,}738) \qquad (3{,}215) \qquad (2{,}071) \qquad (4{,}130)$$

$$\qquad - 0{,}724 \cdot Roman - 0{,}558 \cdot Orthodox - 1{,}467 \cdot Islamic$$
$$\qquad\quad (-2{,}789) \qquad\qquad (-2{,}441) \qquad\qquad (-7{,}159) \qquad\qquad R^2 = 57{,}12\%$$

E = Easiness
GDP = GDP per Capita
BR = Birth Rate
MA = Median Age
EGG = Economical Gender Gap

Sources

GDP per Capita:
http://data.worldbank.org/indicator/NY.GDP.PCAP.CD

Easiness:
http://www.targetmap.com/viewer.aspx?reportId=19491

Attractiveness:
http://www.targetmap.com/viewer.aspx?reportId=32066

Gender Gap:
http://www.genderindex.org/countries
http://reports.weforum.org/global-gender-gap-report-2015/the-global-gender-gap-index-results-in-2015/

Median Age:
https://www.cia.gov/Library/publications/the-world-factbook/fields/2177.html

Fragile States Index:
http://fsi.fundforpeace.org/rankings-2016

Birth Rate:
http://data.worldbank.org/indicator/SP.DYN.CBRT.IN

Culture:
http://www.p12.nysed.gov/ciai/socst/grade3/geoimages/Image12.gif
https://en.wikipedia.org/wiki/Clash_of_Civilizations

Datasets

Statgraphics
http://s000.tinyupload.com/index.php?file_id=52305765601025250202

Excel
http://s000.tinyupload.com/index.php?file_id=02258942656880853120

XML
http://s000.tinyupload.com/index.php?file_id=77845412545075682868