GHS implementation status Logistic Regression Analysis

Steve Fick

Sep 20, 2017

# Introduction

The purpose of this exercise is to determine which variables in the GHS dataset best 'explain' variation GHS implementation status.

To do this we can use a multiple regression technique called *logistic regression*, in which the dependent response variable is modeled as '1' or '0' via a bionomial distribution.

This method essentially takes a regular multiple regression formula and performs a log transformation on it (thus 'logistic regression') so the response will always fall between 0 and 1.

One side-effect of this transformation is that interpretation of the effects become a bit different. Normally coefficients correspond to linear increases in the response variable ( eg in the formula y = mx + b , 1 unit increase in x results in m units increase in y). However in logistic regression, 1 unit increase in x results in a *nonlinear* increase in y which is related to m.

However, this is not a big problem for this analysis because we are mostly interested in identifying the importance of each variable, holding others constant, rather than the direct effect of each variable on the actual probability of GHS implementation occurring. Given this, it makes sense to first *scale* and *center* each variable so that comparisons of coefficients may be more general.

# Method

## 0. load data

library(knitr)  
library(Hmisc)  
library(car)  
d <- read.csv(file.path('..', outdir,'data.csv'), check.names=F, strings=F)

*Recode variables*

# ghs = 1 if partial or full implementation, 0 if no implementation  
d$ghs <- as.numeric(d$`GHS implementation` >0)  
  
# reduction to trade bariers  
d$wto <- d$`WTO membership`  
d$wto <- as.factor(d$wto)  
  
d$trade <- as.numeric(as.character(d$`Trade Open-ness for 2013`))

## Warning: NAs introduced by coercion

# Commitment to safety  
  
# ILO = 1 if \*ANY\* ratifications signed  
d$ILO <- rowSums(d[, grep('ILO Conv', names(d))])  
d$ILO <- as.factor(d$ILO > 0)  
  
# Commitment to environment  
  
# env = number of environmental documents ratified  
d$env <- rowSums(d[, c("Sthlm ratification", "Rotterdam ratificaiton","Basel ratification","SAICM focal point in place=1")])  
  
#globalization  
d$global <- d[['KOF sub-index C']]  
  
# GDP per person  
d$GDPP <- d$`GDP per Capita PPP for 2015`  
  
# GDP Total  
d$GDPT <- d$`Total GDP, PPP, for 2015`  
  
# Regulation   
d$regulation <- d$RegulatoryQuality  
  
# Governance  
d$governance <- d$GovernmentEffectiveness

*Scale continuous variables* to promote inter-comparisons

cc <- c('GDPP', 'GDPT', 'global', 'trade', 'env', 'regulation', 'governance')  
ds <-d   
ds[,cc] <- scale(d[,cc])  
rownames(ds) <- ds$`Country name`

## 1. Evaluate correlations

x <- rcorr(as.matrix(d[,c('governance', 'regulation', 'GDPP', 'GDPT','global', 'trade')]))

*Pearsons correlation*

kable(x$r)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | governance | regulation | GDPP | GDPT | global | trade |
| governance | 1.0000000 | 0.9271973 | 0.7409316 | 0.2163434 | 0.3984775 | 0.2217601 |
| regulation | 0.9271973 | 1.0000000 | 0.6943310 | 0.1454347 | 0.4242175 | 0.2735374 |
| GDPP | 0.7409316 | 0.6943310 | 1.0000000 | 0.1507982 | 0.2392583 | 0.3403252 |
| GDPT | 0.2163434 | 0.1454347 | 0.1507982 | 1.0000000 | 0.2767525 | -0.2119498 |
| global | 0.3984775 | 0.4242175 | 0.2392583 | 0.2767525 | 1.0000000 | -0.1904751 |
| trade | 0.2217601 | 0.2735374 | 0.3403252 | -0.2119498 | -0.1904751 | 1.0000000 |

Governance, regulation, and GDP per capita are highly correlated (all > .69). Next highest correlation is regulation and GD per capita (.42)

*Significance (P)*

kable(x$P)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | governance | regulation | GDPP | GDPT | global | trade |
| governance | NA | 0.0000000 | 0.0000000 | 0.0035347 | 0.0000000 | 0.0030971 |
| regulation | 0.0000000 | NA | 0.0000000 | 0.0514165 | 0.0000000 | 0.0002394 |
| GDPP | 0.0000000 | 0.0000000 | NA | 0.0433166 | 0.0014748 | 0.0000056 |
| GDPT | 0.0035347 | 0.0514165 | 0.0433166 | NA | 0.0002184 | 0.0055254 |
| global | 0.0000000 | 0.0000000 | 0.0014748 | 0.0002184 | NA | 0.0123216 |
| trade | 0.0030971 | 0.0002394 | 0.0000056 | 0.0055254 | 0.0123216 | NA |

All correlations are highly significant

Correlations could be problematic for our model, because it will be difficult to determine what the effect of regulation vs governance are 'independent of' GDP and vice versa, because we wont have (any) examples where GDP is is high but regulation is low to see the effects on GHS, for example.

## 2. Model fitting

Lets try to create a first model with all terms thrown in.

m0 <- glm(ghs ~ regulation + governance + GDPP + GDPT + global + env + ILO + trade + wto, data = ds , family = 'binomial')  
  
summary(m0)

##   
## Call:  
## glm(formula = ghs ~ regulation + governance + GDPP + GDPT + global +   
## env + ILO + trade + wto, family = "binomial", data = ds)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1926 -0.4067 -0.1090 0.1958 2.4921   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.6151 1.0008 -1.614 0.106552   
## regulation 0.6645 0.8299 0.801 0.423291   
## governance 1.9768 0.9805 2.016 0.043795 \*   
## GDPP -0.6330 0.5044 -1.255 0.209520   
## GDPT 0.3318 0.4104 0.809 0.418801   
## global 1.6557 0.4539 3.648 0.000264 \*\*\*  
## env 0.3218 0.5028 0.640 0.522201   
## ILOTRUE 1.7436 0.5753 3.031 0.002441 \*\*   
## trade 0.8266 0.3941 2.097 0.035957 \*   
## wto1 -0.9442 0.9683 -0.975 0.329521   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 218.322 on 165 degrees of freedom  
## Residual deviance: 92.807 on 156 degrees of freedom  
## (27 observations deleted due to missingness)  
## AIC: 112.81  
##   
## Number of Fisher Scoring iterations: 6

Ok, starting from the top we see our formula, then the deviance residuals. The deviance residuals appear approximately normal based on the quantiles, which is a good thing.

Next we see the coefficients. The intercept indicates the overall likelihood for a country to have GHS (without ILO ratification or wto ratification) The fact that this is negative means that overall, likelihood is small that a random country will have GHS implementation.

GDP per capita interestingly has a negative coefficient. This can happen, however when predictors are highly correlated -- one has to compensate for the other. In this case I think GDPP is compensating for Regulation and Governance.

GDPT is positive, but not significant

globalization is positive and significant

environmental ratification is positive but not significant

ILO is very strongly positive and significant

trade openness is large, positive and significant.

Interestingly WTO membership is slightly negative. However the standard error for this coefficient is high making interpretation difficult.

Lets explore this idea of correlation between GDPP and regulation, by examinging the Variance Inflation Factor (VIF) for each variable. This gives us a sense of how much the overall uncertainty in the model increases when adding a variable, compared to adding a completely uncorrelated variable. Any value greater than 2 is considered high, and > 5 is very high

vif(m0)

## regulation governance GDPP GDPT global env   
## 4.320626 5.579609 2.158644 1.322384 1.617785 1.191524   
## ILO trade wto   
## 1.158139 1.565025 1.206434

We see that regulation, governance, and GDPP have high VIF, likely due to their high correlation. This is a problem in linear modeling called multi-colinearity.

As a result it makes interpretation of the colinear variables problematic, because they have uncertain standard errors. This makes the model highly sensitive to the underlying data as well.

One simple solution to this problem is to simply remove a colinear variable. I think it is reasonable to drop governance because it provides similar information to regulation.

## 3. Refit model dropping government effectiveness

m <- glm(ghs ~ governance + GDPP+ GDPT + global + env + ILO + trade + wto, data = ds , family = 'binomial')

lets check the VIF

vif(m)

## governance GDPP GDPT global env ILO   
## 2.332341 2.138644 1.365968 1.605191 1.234090 1.160817   
## trade wto   
## 1.563667 1.161712

much more reasonable!

summary(m)

##   
## Call:  
## glm(formula = ghs ~ governance + GDPP + GDPT + global + env +   
## ILO + trade + wto, family = "binomial", data = ds)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2907 -0.3906 -0.1072 0.1953 2.5236   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.7500 0.9726 -1.799 0.071961 .   
## governance 2.5855 0.6310 4.098 4.17e-05 \*\*\*  
## GDPP -0.6541 0.5000 -1.308 0.190831   
## GDPT 0.3019 0.3962 0.762 0.446041   
## global 1.6638 0.4479 3.715 0.000204 \*\*\*  
## env 0.3555 0.4923 0.722 0.470275   
## ILOTRUE 1.7262 0.5740 3.007 0.002635 \*\*   
## trade 0.8499 0.3950 2.152 0.031413 \*   
## wto1 -0.7828 0.9301 -0.842 0.400031   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 218.322 on 165 degrees of freedom  
## Residual deviance: 93.465 on 157 degrees of freedom  
## (27 observations deleted due to missingness)  
## AIC: 111.47  
##   
## Number of Fisher Scoring iterations: 6

This looks better to me... The HDI effect is now much lower (although highly significant) trade and globalization are both significant Also, ILO ratification is now a large and significant explanatory variable

Outputting results to table.

kable(coef(summary(m)))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | -1.7500411 | 0.9725900 | -1.7993617 | 0.0719615 |
| governance | 2.5855292 | 0.6309614 | 4.0977616 | 0.0000417 |
| GDPP | -0.6541016 | 0.5000304 | -1.3081237 | 0.1908313 |
| GDPT | 0.3018828 | 0.3961548 | 0.7620323 | 0.4460407 |
| global | 1.6637662 | 0.4479084 | 3.7145234 | 0.0002036 |
| env | 0.3554566 | 0.4923012 | 0.7220309 | 0.4702755 |
| ILOTRUE | 1.7261887 | 0.5739759 | 3.0074237 | 0.0026347 |
| trade | 0.8498687 | 0.3949567 | 2.1518025 | 0.0314129 |
| wto1 | -0.7827705 | 0.9301349 | -0.8415666 | 0.4000306 |

Other statistics of interest: AIC (a measure of goodness of fit) -- decent, in this case

1- (Residual Deviance / Null deviance) -- analogous to R2. In this case it is 0.5718909 which is not bad.

## Omitted countries

Note: some countries were excluded from analysis because of missing data. Presented here...

i <- attr(m$model, 'na.action')  
kable(d[i,c(cc, 'ILO', 'wto', 'env')])

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | GDPP | GDPT | global | trade | env | regulation | governance | ILO | wto | env.1 |
| 1 | NA | NA | NA | NA | 1 | 0.8920590 | 1.7523551 | FALSE | 0 | 1 |
| 34 | 3513.6847 | 79766110568 | 75.75861 | NA | 4 | -0.5152425 | -0.6509632 | FALSE | 1 | 4 |
| 36 | 784.3652 | 60605402501 | NA | 103.41895 | 4 | -1.1939210 | -1.0227140 | FALSE | 1 | 4 |
| 53 | NA | NA | 27.46694 | NA | 4 | -2.1120157 | -1.5996221 | FALSE | 0 | 4 |
| 60 | 3496.7152 | 365266872 | 20.27766 | NA | 2 | -0.9924390 | -0.4171610 | FALSE | 0 | 2 |
| 80 | NA | NA | 66.32730 | 33.49550 | 4 | -1.2803912 | -0.2018129 | FALSE | 0 | 4 |
| 92 | 1994.7676 | 224257758 | 21.11161 | NA | 3 | -0.9065605 | -0.5021341 | FALSE | 0 | 3 |
| 99 | NA | NA | 50.75417 | 97.87311 | 4 | -2.2371929 | -1.6971939 | FALSE | 0 | 4 |
| 101 | NA | NA | 29.04045 | NA | 4 | 1.3837334 | 1.6654379 | FALSE | 1 | 4 |
| 108 | NA | NA | 32.16048 | NA | 3 | NA | NA | FALSE | 0 | 3 |
| 113 | 3910.9154 | 207251140 | 19.02752 | NA | 4 | -1.1236097 | -1.5973598 | FALSE | 0 | 4 |
| 114 | 14076.4831 | 29257308598 | 48.97169 | NA | 4 | 0.4466614 | 0.1261138 | TRUE | 1 | 4 |
| 121 | NA | NA | 66.98794 | 169.23622 | 4 | -0.8585302 | -1.0285877 | FALSE | 1 | 4 |
| 132 | 13064.5452 | 162980202 | NA | NA | 3 | -1.2394499 | -0.5658124 | FALSE | 0 | 3 |
| 139 | 15317.1662 | 326117786 | 39.12980 | NA | 3 | -0.8225325 | -0.5658124 | FALSE | 0 | 3 |
| 140 | NA | NA | 59.23830 | 96.29065 | 3 | -0.5541740 | -0.6209071 | FALSE | 1 | 3 |
| 142 | NA | NA | 25.84665 | NA | 4 | -2.3364401 | -1.6327326 | FALSE | 0 | 4 |
| 146 | 22124.3433 | 438400676870 | NA | 86.50117 | 4 | 0.5907927 | -0.0383196 | FALSE | 1 | 4 |
| 150 | 4387.8964 | 176546493705 | 54.64030 | NA | 4 | -1.5004314 | -1.4762483 | FALSE | 0 | 4 |
| 156 | NA | NA | 34.71026 | NA | 0 | NA | NA | FALSE | 0 | 0 |
| 157 | NA | NA | 32.17682 | NA | 3 | -2.1498590 | -2.2187388 | FALSE | 0 | 3 |
| 159 | 1853.7986 | 22875526538 | NA | NA | 0 | -1.6855018 | -2.1712434 | FALSE | 0 | 0 |
| 167 | NA | NA | 51.99130 | 53.21293 | 4 | -1.6325682 | -1.6338952 | TRUE | 0 | 4 |
| 172 | 16532.4859 | 88837346248 | 36.69389 | NA | 1 | -2.0912218 | -0.8743739 | FALSE | 0 | 1 |
| 173 | 2399.0144 | 2842268279 | NA | 32.78201 | 0 | -0.9665961 | -1.0453568 | FALSE | 0 | 0 |
| 178 | 3926.1735 | 38931936 | NA | 160.30894 | 2 | -0.7188845 | -0.9690896 | FALSE | 0 | 2 |
| 186 | NA | NA | 65.78635 | 43.11163 | 4 | -1.8577141 | -1.2201610 | TRUE | 1 | 4 |

## Top Residuals

r <- data.frame(country = names(resid(m)),residual = resid(m))  
r <- r[order(abs(r$residual), decreasing = TRUE),]  
r$rank <- 1:nrow(r)  
r$country <- NULL  
kable(r)

|  |  |  |
| --- | --- | --- |
|  | residual | rank |
| Kyrgyzstan | 2.5235839 | 1 |
| India | -2.2906843 | 2 |
| Ecuador | 2.2795724 | 3 |
| Georgia | 2.1677220 | 4 |
| Chile | -1.9745452 | 5 |
| South Africa | -1.9263548 | 6 |
| Indonesia | 1.7769344 | 7 |
| Kazakhstan | 1.6606529 | 8 |
| Zambia | 1.6474350 | 9 |
| Belarus | 1.6256151 | 10 |
| Philippines | 1.5748635 | 11 |
| United Arab Emirates | -1.5516116 | 12 |
| Armenia | 1.5491227 | 13 |
| Greece | 1.3745571 | 14 |
| Jordan | -1.3473063 | 15 |
| Montenegro | 1.3332059 | 16 |
| Malta | 1.3174853 | 17 |
| Ukraine | -1.2411813 | 18 |
| Bosnia and Herzegovina | 1.2111862 | 19 |
| Albania | 1.1236026 | 20 |
| El Salvador | -1.0914068 | 21 |
| Mexico | 1.0589297 | 22 |
| Colombia | -1.0565066 | 23 |
| Seychelles | -1.0360220 | 24 |
| Morocco | -1.0281394 | 25 |
| Ethiopia | -1.0150618 | 26 |
| Israel | -1.0105324 | 27 |
| Lebanon | -0.9846961 | 28 |
| Argentina | 0.9809499 | 29 |
| Tunisia | -0.9586882 | 30 |
| Algeria | -0.9534224 | 31 |
| Russia | 0.9492951 | 32 |
| Bulgaria | 0.9344875 | 33 |
| Brazil | 0.9267366 | 34 |
| Thailand | 0.9062204 | 35 |
| Mongolia | -0.8961741 | 36 |
| Mauritius | 0.8533769 | 37 |
| Republic of Moldova | -0.8526098 | 38 |
| Bahrain | -0.8323404 | 39 |
| Fiji | -0.7868485 | 40 |
| Ghana | -0.7840809 | 41 |
| Nigeria | -0.7690719 | 42 |
| Burkina Faso | -0.7444828 | 43 |
| Senegal | -0.7266342 | 44 |
| Namibia | -0.7151834 | 45 |
| Turkey | 0.7070749 | 46 |
| Iceland | 0.6929384 | 47 |
| Panama | -0.6927107 | 48 |
| Dominican Republic | -0.6839503 | 49 |
| Saudi Arabia | -0.6781731 | 50 |
| Niger | -0.6779438 | 51 |
| Jamaica | -0.6620478 | 52 |
| Uruguay | 0.6490184 | 53 |
| Bahamas | -0.6172218 | 54 |
| Kenya | -0.6127029 | 55 |
| Cuba | -0.5836680 | 56 |
| Viet Nam | 0.5685292 | 57 |
| Serbia | 0.5661635 | 58 |
| Sri Lanka | -0.5468243 | 59 |
| Italy | 0.5421683 | 60 |
| Latvia | 0.5217258 | 61 |
| Peru | -0.5168343 | 62 |
| Croatia | 0.5157410 | 63 |
| Samoa | -0.5051041 | 64 |
| Barbados | -0.4936524 | 65 |
| Lesotho | -0.4919022 | 66 |
| Trinidad and Tobago | -0.4904827 | 67 |
| Costa Rica | -0.4797565 | 68 |
| Guyana | -0.4662876 | 69 |
| Malaysia | 0.4655909 | 70 |
| Rwanda | -0.4593165 | 71 |
| United States | 0.3992883 | 72 |
| Gabon | -0.3912206 | 73 |
| United Republic of Tanzania | -0.3888607 | 74 |
| Cape Verde | -0.3851636 | 75 |
| Grenada | -0.3718668 | 76 |
| Botswana | -0.3714955 | 77 |
| Antigua and Barbuda | -0.3678875 | 78 |
| Lithuania | 0.3610572 | 79 |
| Zimbabwe | -0.3590229 | 80 |
| Cyprus | 0.3529207 | 81 |
| Pakistan | -0.3518222 | 82 |
| Qatar | -0.3439477 | 83 |
| Bolivia | -0.3127008 | 84 |
| Egypt | -0.3082853 | 85 |
| Liberia | -0.3078282 | 86 |
| Guatemala | -0.3068691 | 87 |
| Tajikistan | -0.3025477 | 88 |
| France | 0.2948087 | 89 |
| Honduras | -0.2915453 | 90 |
| Belize | -0.2897742 | 91 |
| Mozambique | -0.2873874 | 92 |
| Cambodia | -0.2851457 | 93 |
| Japan | 0.2840071 | 94 |
| Togo | -0.2823391 | 95 |
| Benin | -0.2767709 | 96 |
| Poland | 0.2738898 | 97 |
| Austria | 0.2707668 | 98 |
| Canada | 0.2667036 | 99 |
| Uganda | -0.2618184 | 100 |
| Azerbaijan | -0.2587482 | 101 |
| Brunei Darussalam | -0.2499204 | 102 |
| Paraguay | -0.2424964 | 103 |
| Bhutan | -0.2297575 | 104 |
| United Kingdom | 0.2222593 | 105 |
| Mali | -0.2183463 | 106 |
| Oman | -0.2142459 | 107 |
| Cameroon | -0.2100533 | 108 |
| Australia | 0.2079397 | 109 |
| Slovenia | 0.2041861 | 110 |
| Switzerland | 0.1968109 | 111 |
| Korea, Republic of | 0.1906032 | 112 |
| Malawi | -0.1902925 | 113 |
| Guinea | -0.1894420 | 114 |
| Bangladesh | -0.1863755 | 115 |
| Portugal | 0.1824325 | 116 |
| Spain | 0.1805895 | 117 |
| Hungary | 0.1778710 | 118 |
| China | 0.1777300 | 119 |
| Estonia | 0.1659890 | 120 |
| New Zealand | 0.1618531 | 121 |
| Slovakia | 0.1602992 | 122 |
| Dominica | -0.1589808 | 123 |
| Djibouti | -0.1575413 | 124 |
| Czech Republic | 0.1554398 | 125 |
| Maldives | -0.1533173 | 126 |
| Nicaragua | -0.1521965 | 127 |
| Gambia | -0.1476668 | 128 |
| Kuwait | -0.1468371 | 129 |
| Sao Tome and Principe | -0.1449988 | 130 |
| Norway | 0.1377577 | 131 |
| Nepal | -0.1273692 | 132 |
| Saint Lucia | -0.1247393 | 133 |
| Saint Vincent and the Grenadines | -0.1141528 | 134 |
| Luxembourg | 0.1094248 | 135 |
| Singapore | 0.1078106 | 136 |
| Swaziland | -0.1074876 | 137 |
| Sierra Leone | -0.1070083 | 138 |
| Suriname | -0.1033763 | 139 |
| Chad | -0.0992500 | 140 |
| Congo | -0.0954419 | 141 |
| Finland | 0.0952125 | 142 |
| Sweden | 0.0892355 | 143 |
| Ireland | 0.0848711 | 144 |
| Madagascar | -0.0847832 | 145 |
| Saint Kitts and Nevis | -0.0839100 | 146 |
| Denmark | 0.0816411 | 147 |
| Germany | 0.0808382 | 148 |
| Iraq | -0.0792519 | 149 |
| Burundi | -0.0784936 | 150 |
| Lao People's Democratic Republic | -0.0719395 | 151 |
| Angola | -0.0711612 | 152 |
| Vanuatu | -0.0680865 | 153 |
| Belgium | 0.0629415 | 154 |
| Uzbekistan | -0.0626847 | 155 |
| Tonga | -0.0577514 | 156 |
| Yemen | -0.0474433 | 157 |
| Central African Republic | -0.0473464 | 158 |
| Afghanistan | -0.0468961 | 159 |
| Netherlands | 0.0417520 | 160 |
| Guinea-Bissau | -0.0397411 | 161 |
| Myanmar | -0.0326600 | 162 |
| Equatorial Guinea | -0.0249049 | 163 |
| Solomon Islands | -0.0223857 | 164 |
| Comoros | -0.0214854 | 165 |
| Haiti | -0.0094770 | 166 |