CONTRACEPTIVE PREVELANCE AND STUDY OF FACTORS THAT MIGHT AFFECT IT

Abhishek Angadi

December 6, 2018

WHAT IS CONTRACEPTIVE PREVALANCE?

Contraceptive prevalence is the percentage of women who are currently using, or whose sexual partner is currently using, at least one method of contraception, regardless of the method used. It is usually reported for married or in-union women aged 15 to 49.

"(# of women 15-49 using a contraceptive method)*100" /"total # of women married or in-union aged 15-49)"

IMPORTANCE?

Contraceptive use slows population growth. This is important because overpopulation puts pressure on the environment, the economy and services such as education an d health.

Contraceptive prevalence is the most widely reported measure of outcome for family planning programs at the population level. Hence tracking it and more importantly understanding different factors that affect it is extremely important.

WHY IS THE DEMONITATOR AS ALL WOMEN?

To obtain a true contraceptive use rate, the denominator should reflect the population at risk (of pregnancy), i.e., sexually active women who are not infecund, pregnant. The numerator should reflect the number of contraceptive users from that population. Secondly, an additional filter is that the women should be married and in union. Since quite a few countries are more conservative, the data reliability of this measure outside marriage for women would be low.

OBJECTIVE?

Primary: Whilst keeping the importance of Contraceptive prevalence in mind, the objective of this study was to explore the broad scope of factors that affect and do not affect the measure.

Specific Questions in Interest:

Q) GDP growth is an important parameter to access a country's economic health. How does it affect Contraceptive prevalence .

R) Does changing the Government policies (on fertility) in a country affect CP.

S) Tourism tends to boost the economy of a country. Does it have any effect on CP.

T) Is Literacy rate of a country an important measure in increasing it's CP.

VARIABLE DEFINITIONS?

ECONOMIC:

GDP annual: Gross domestic product (GDP) is the monetary value of all the finished goods and services produced within a country's borders in one year. GDP per Capita: Annual GDP divided by its average population for the year. GDP Growth per capita: one quarter of the country's gross domestic product to the previous quarter. PPP GNI: Purchasing Power Parity gross national income (GNI) converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GNI as a U.S. dollar has in the United States. GNI using Atlas Method: It is similar to PPP GNI. It instead uses a three-year average of exchange rates to smooth effects of transitory exchange rate fluctuations. HDI: The Human Development Index (HDI) is a statistic composite index of life expectancy, education, and per capita income indicators HDI Rank: Countries ranked as per HDI Development Category: The UN has categorized nations to two categories, less developed and more developed countries.

FERTILITY POLICIES: Categories - Lower, Maintain, No Intervention, no official Policy and Raise.

GEOGRAPHIC LOCATION: Categories – Africa, Asia, Europe, Latin Amer. & Caribbean, Northern America, Oceania

POPULATION STATISTICS:

Population Density: The number of people living in per square km Total Population: Self-explanatory

Literacy Rate: percentage of the population aged 15 years and over who cannot both read and write Urbanization Rate: Average Annual Rate of Change of the Percentage Urban by Country

TOURISM: Number of International Flight Arrivals: Self-explanatory

POPULATION STATISTICS:

Gender Inequality Index: an index for measurement of gender disparity that was introduced in the 2010 Human Development Report 20th anniversary edition by the United Nations Development Program. Sex ratio: Population men/population female

CREATION OF DATASET: PROBABLY THE HARDEST BIT OF THE STUDY

SUMMARY OF FINAL DATA:

```
summary(working_data)
```

```
##  Country_or_area      Any_method    Female\r\nsterilization
##  Length:172         Min.   : 5.70   Min.   : 0.000
##  Class :character   1st Qu.:34.62   1st Qu.: 0.900
##  Mode  :character   Median :55.90   Median : 3.200
##                     Mean   :52.48   Mean   : 6.777
##                     3rd Qu.:68.10   3rd Qu.: 8.800
##                     Max.   :88.40   Max.   :40.700
##                                     NA's   :23
##  Male\r\nsterilization      Pill              ID              GDP
##  Min.   : 0.000        Min.   : 0.40   Min.   :  1.00   Min.   :   0.214
##  1st Qu.: 0.000        1st Qu.: 4.05   1st Qu.: 45.75   1st Qu.:   1.192
##  Median : 0.100        Median : 8.90   Median : 90.50   Median :   5.283
##  Mean   : 1.396        Mean   :12.22   Mean   : 92.58   Mean   : 225.013
##  3rd Qu.: 0.500        3rd Qu.:17.55   3rd Qu.:140.25   3rd Qu.:  90.041
##  Max.   :21.000        Max.   :54.40   Max.   :188.00   Max.   :5558.089
##  NA's   :65            NA's   :5
##  sex_ratio_2015   pop_density_2015    UR_2015_2018     Literacy_Rate
##  Min.   : 84.83   Min.   :   1.916   Min.   :-1.1143   Min.   : 29.00
##  1st Qu.: 96.61   1st Qu.:  34.240   1st Qu.: 0.1882   1st Qu.: 71.00
##  Median : 98.85   Median :  80.151   Median : 0.4958   Median : 90.00
##  Mean   :102.05   Mean   : 235.344   Mean   : 0.6885   Mean   : 81.96
##  3rd Qu.:101.08   3rd Qu.: 188.652   3rd Qu.: 1.0508   3rd Qu.: 97.75
##  Max.   :306.56   Max.   :7907.517   Max.   : 2.5319   Max.   :100.00
##  NA's   :3                                             NA's   :38
##       Policy_on_fertility_level                    Region
##  Lower             :74          Africa                 :47
##  Maintain          :25          Asia                   :44
##  No intervention   :14          Europe                 :36
##  No official policy:12          Latin Amer. & Caribbean:32
##  Raise             :46          Northern America       : 2
##  NA's              : 1          Oceania                :10
##                                 NA's                   : 1
##         Development_level men_legal_marriage_age
##  Less dev. region:130     Min.   :12.00
##  More dev. region: 41     1st Qu.:18.00
##  NA's            :  1     Median :18.00
##                           Mean   :18.41
##                           3rd Qu.:18.00
##                           Max.   :22.00
##                           NA's   :7
##  Gender_Inequality_Index_2014 total_pop_2015           HDI
##  Min.   :0.01642              Min.   :     21.3   Min.   :0.3539
##  1st Qu.:0.20271              1st Qu.:   2905.7   1st Qu.:0.5904
```

```
##  Median :0.38605              Median :   9057.6   Median :0.7350
##  Mean   :0.36682              Mean   :  41529.9   Mean   :0.7081
##  3rd Qu.:0.52189              3rd Qu.:  29173.0   3rd Qu.:0.8132
##  Max.   :0.74396              Max.   :1397028.6   Max.   :0.9525
##  NA's   :26
##     HDI_rank      number_of_arrivals gni_pp_latest
##  Min.   :  1.00   Min.   :    5700   Min.   :   663.1
##  1st Qu.: 50.00   1st Qu.:  348000   1st Qu.:  3947.2
##  Median : 95.50   Median : 1662000   Median : 11266.0
##  Mean   : 94.97   Mean   : 7056188   Mean   : 17208.3
##  3rd Qu.:140.25   3rd Qu.: 6782000   3rd Qu.: 24109.4
##  Max.   :188.00   Max.   :82570000   Max.   :117896.1
##                   NA's   :7          NA's   :6
##  gni_atlas_latest    gdp_per_capita    gdp_growth_per_capita
##  Min.   :2.722e+08   Min.   :  320.1   Min.   :-9.962
##  1st Qu.:8.835e+09   1st Qu.: 1856.0   1st Qu.: 0.701
##  Median :3.693e+10   Median : 5334.9   Median : 2.179
##  Mean   :4.503e+11   Mean   :12772.8   Mean   : 2.073
##  3rd Qu.:2.159e+11   3rd Qu.:15408.4   3rd Qu.: 3.837
##  Max.   :1.898e+13   Max.   :80189.7   Max.   :25.059
##  NA's   :4           NA's   :4         NA's   :4
```

INITIAL IMPRESSIONS?

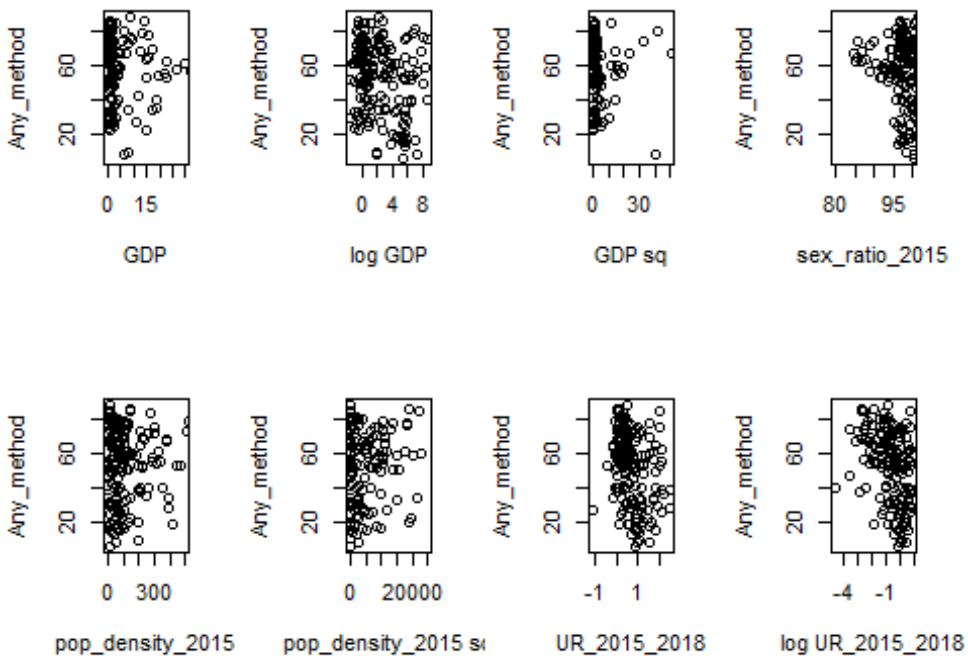
 172 Countries in total. Islands not picked due to lack of data. Literacy Rate has 22% and Gender Inequality Index (2014) has 15% missing values. Since its lower than 30%, imputation model not used. The rest of them have missing values <5%. Not keen on keeping GDP. Large countries have higher GDPs usually while smile countries have smaller. Scientifically, GDP should not have a high correlation with contraceptive prevalence. Concerned about correlation between GDP per capita and growth rate. Will watch them closely. Many variables have outliers. But these are true values and hence it was decided not to remove them. E.g. India & China clear outliers for total_pop_2015. HDI & HDI Rank due to their nature, would have high multicollinearity. Dropping HDI rank. men_legal_marriage_age 1st and 3rd quadrant have the same value. Dropping the legal marriage variable. Just two countries in North America region category. But not combining the groups, since USA and Canada are indeed very different than Latin America. Looking at the first and third quadrant values, they mostly seem symmetrically distributed around the mean which probably signifies that they are normally distributed. Again, Graphs would explain better.

SCATTER PLOTS

```r
#gdp
par(mfrow = c(2,4))
plot(y = working_data$Any_method, x = (working_data$GDP), xlab = "GDP", ylab
= "Any_method",xlim=c(0,30))
#not great.
```

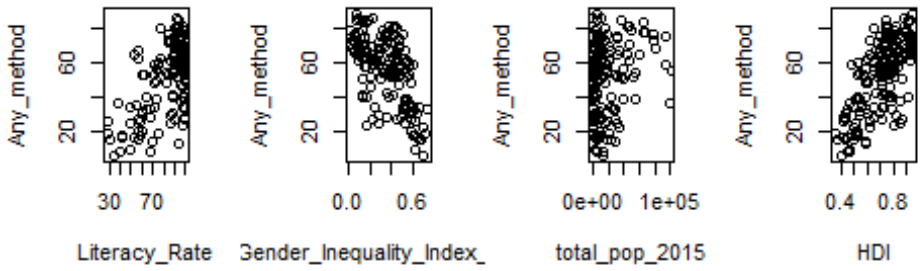
```r
#might require transformation
plot(y = working_data$Any_method, x = log(working_data$GDP), xlab = "log GDP"
, ylab = "Any_method")
#not to happy. flat trend.
working_data$gdp_sq<-(working_data$GDP)^2
plot(y = working_data$Any_method, x = working_data$gdp_sq, xlab = "GDP sq", y
lab = "Any_method",xlim=c(0,50))
#not to happy. we could try

#sex_ratio_2015 some kind of trend
plot(y = (working_data$Any_method), x = (working_data$sex_ratio_2015), xlab =
"sex_ratio_2015", ylab = "Any_method",xlim=c(80,100))

#pop_density_2015. not great. not happy about constant variance. may be sqaur
ing is better. can try.
plot(y = working_data$Any_method, x = (working_data$pop_density_2015), xlab =
"pop_density_2015", ylab = "Any_method",xlim=c(0,500))
plot(y = working_data$Any_method, x = (working_data$pop_density_2015)^2, xlab
= "pop_density_2015 sq", ylab = "Any_method",xlim=c(0,25000))


#UR_2015_2018 sad graph. no trend. log helps significantly.
plot(y = working_data$Any_method, x = (working_data$UR_2015_2018), xlab = "UR
_2015_2018", ylab = "Any_method")
plot(y = working_data$Any_method, x = log(working_data$UR_2015_2018), xlab =
"log UR_2015_2018", ylab = "Any_method")
```

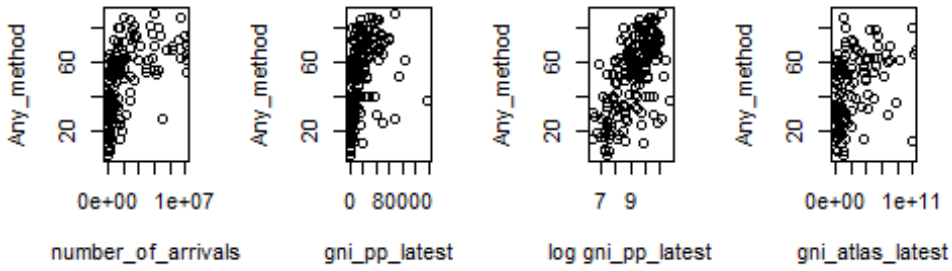
```
## Warning in log(working_data$UR_2015_2018): NaNs produced
```




```r
#Literacy_Rate. good trend. good constant variance.
plot(y = working_data$Any_method, x = (working_data$Literacy_Rate), xlab = "L
iteracy_Rate", ylab = "Any_method")

#Gender_Inequality_Index_2014 good trend. good constant variance.
plot(y = working_data$Any_method, x = (working_data$Gender_Inequality_Index_2
014), xlab = "Gender_Inequality_Index_2014", ylab = "Any_method")

#total_pop_2015 not bad trend. good constant variance. India and China the cl
ear outliers.
plot(y = working_data$Any_method, x = (working_data$total_pop_2015), xlab = "
total_pop_2015", ylab = "Any_method",xlim=c(0,100000))

#HDI beutiful trend. good constant variance.
plot(y = working_data$Any_method, x = (working_data$HDI), xlab = "HDI", ylab
= "Any_method")

#Interesting: number_of_arrivals decent trend. good constant variance. Proble
m: higher values
plot(y = working_data$Any_method, x = (working_data$number_of_arrivals), xlab
= "number_of_arrivals", ylab = "Any_method",xlim=c(0,10000000))

# gni_pp_latest ok trend. good constant variance
plot(y = working_data$Any_method, x = (working_data$gni_pp_latest), xlab = "g
ni_pp_latest", ylab = "Any_method")
#beautiful
```

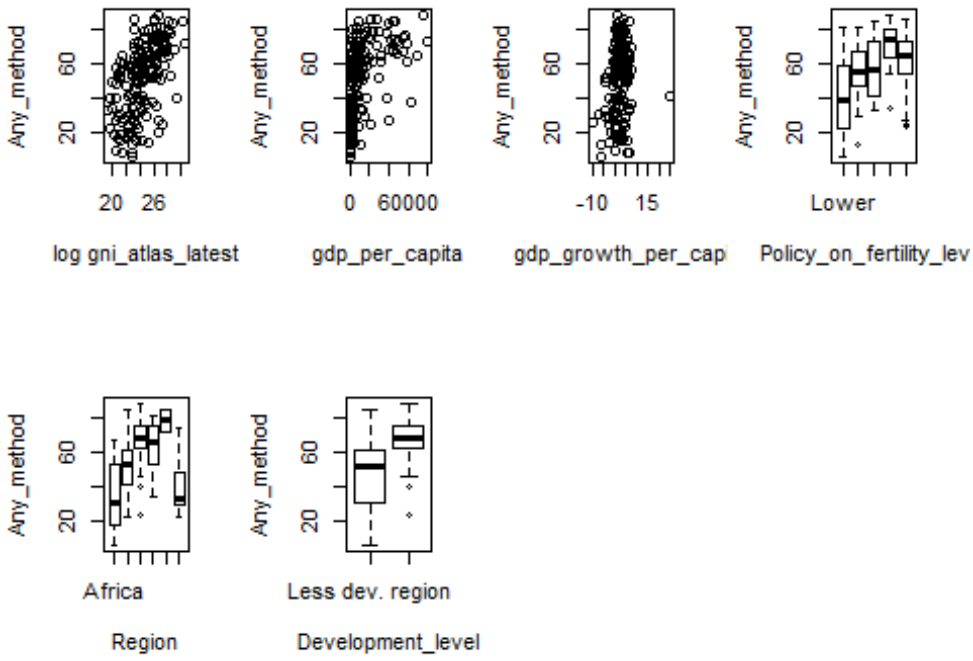
```r
plot(y = working_data$Any_method, x = log(working_data$gni_pp_latest), xlab =
"log gni_pp_latest", ylab = "Any_method")

# gni_atlas_latest ok trend. good constant variance
plot(y = working_data$Any_method, x = (working_data$gni_atlas_latest), xlab =
"gni_atlas_latest", ylab = "Any_method",xlim=c(0,100000000000))
```






```r
#log so much better
plot(y = working_data$Any_method, x = log(working_data$gni_atlas_latest), xla
b = "log gni_atlas_latest", ylab = "Any_method")

#gdp_per_capita
plot(y = working_data$Any_method, x = (working_data$gdp_per_capita), xlab = "
gdp_per_capita", ylab = "Any_method")

#gdp_growth_per_capita. Not good. transformations do not help.
plot(y = working_data$Any_method, x = (working_data$gdp_growth_per_capita), x
lab = "gdp_growth_per_capita", ylab = "Any_method")




#Policy_on_fertility_level good trend.  constant variance ok. #highlight this
! no official policy reacts weird, low in count.
plot(y = working_data$Any_method, x = (working_data$Policy_on_fertility_level
```

```
), xlab = "Policy_on_fertility_level", ylab = "Any_method")

#Region beautiful trend.  constant variance ok. #highlight this!
plot(y = working_data$Any_method, x = (working_data$Region), xlab = "Region",
ylab = "Any_method")

#trend seems to be there.  constant variance might be an issue. #highlight th
is! can drop this variable. HDI seems to be better variable.
plot(y = working_data$Any_method, x = (working_data$Development_level), xlab
= "Development_level", ylab = "Any_method")
```



PLOTS HIGHLIGHTS

GDP shows a flat trend. Dropping the variable. HDI shows a beautiful trend and has good constant variance. Interesting: number of tourist arrivals has a decent trend. good constant variance. Problem at: higher values. GNI pp and GNI atlas has okay trends. The log transformation has so much better trends. pop density has not so great trends. Others have a decent or an okay trend. Nothing remarkable stands out.

```
corr<-working_data[,c(-3,-4,-5,-6)] %>%
  select_if(is.numeric) %>%
  cor(.,use = "pairwise.complete.obs") %>%
  round(3)

temp<-na.omit(working_data[,c(2,13,14)])
```

```r
chisq.test(working_data$Development_level,working_data$Policy_on_fertility_le
vel, correct=F,simulate.p.value = TRUE)

##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  working_data$Development_level and working_data$Policy_on_fertility
_level
## X-squared = 66.088, df = NA, p-value = 0.0004998

chisq.test(working_data$Development_level,working_data$Region, correct=F,simu
late.p.value = TRUE)

##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  working_data$Development_level and working_data$Region
## X-squared = 156.86, df = NA, p-value = 0.0004998

chisq.test(working_data$Policy_on_fertility_level,working_data$Region, correc
t=F,simulate.p.value = TRUE)

##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  working_data$Policy_on_fertility_level and working_data$Region
## X-squared = 132.03, df = NA, p-value = 0.0004998

temp<-na.omit(working_data[,c(2,13,14)])
temp<-temp[grepl("Northern",x = temp$Region,ignore.case = TRUE)==FALSE,]
temp<-temp[grepl("Oceania",x = temp$Region,ignore.case = TRUE)==FALSE,]

chisq.test(temp$Development_level,temp$Region, correct=F)

##
##  Pearson's Chi-squared test
##
## data:  temp$Development_level and temp$Region
## X-squared = 153.53, df = 3, p-value < 2.2e-16
```

MULTICOLLINEARITY

Testing performed using plots, correlation matrix , Chi square test & analysis of variance tests.

The tests turned out to be an extremely crucial part of the study.

Literacy Rate, HDI and Gender Inequality Index have extremely high correlation amongst themselves. Similarly, HDI, GNI PPP and GDP Per Capita have high correlation amongst themselves.

As per the Chi square test, all categorical variables are correlated to each other.

As per the analysis of variance test, all the categorical variables are correlated to HDI, Urbanization Rate, Number of international flight arrivals, GNI PPP and GNI Atlas.

DECISION: Dropped all categorical Variables. Dropped GDP per capita and kept GDP growth. Kept HDI (scientific reason) and hence dropped Gender Inequality Index and Literacy Rate. Kept GNI PPP since corr between it and HDI is 0.75 while keeping it on watch.

RESULTS NOT SHOWN:

chisq.test(temp$Development_level$, $temp$Region, correct=F)

summary(aov(working_data$HDI\ working_data$Policy_on_fertility_level))
summary(aov(working_data$UR_{2015_{2018}}\ working_data$Policy_on_fertility_level))
summary(aov(working_data$number_of_arrivals\ working_data$Policy_on_fertility_level))
summary(aov(working_data$gni_ppp_latest\ working_data$Policy_on_fertility_level))
summary(aov(working_data$gni_atlas_latest\ working_data$Policy_on_fertility_level))

summary(aov(working_data$HDI\ working_data$Region))
summary(aov(working_data$UR_{2015_{2018}}\ working_data$Region))
summary(aov(working_data$number_of_arrivals\ working_data$Region))
summary(aov(working_data$gni_ppp_latest\ working_data$Region))
summary(aov(working_data$gni_atlas_latest\ working_data$Region))

summary(aov(working_data$HDI\ working_data$Development_level))
summary(aov(working_data$UR_{2015_{2018}}\ working_data$Development_level))
summary(aov(working_data$number_of_arrivals\ working_data$Development_level))
summary(aov(working_data$gni_ppp_latest\ working_data$Development_level))
summary(aov(working_data$gni_atlas_latest\ working_data$Development_level))

categorical vs categorical

temp<-na.omit(working_data[,c(2,13,14)])
chisq.test(working_data$Development_level$, $working_data$Policy_on_fertility_level, correct=F,simulate.p.value = TRUE)
chisq.test(working_data$Development_level$, $working_data$Region, correct=F,simulate.p.value = TRUE)
chisq.test(working_data$Policy_on_fertility_level$, $working_data$Region, correct=F,simulate.p.value = TRUE)

temp<-na.omit(working_data[,c(2,13,14)]) temp<-temp[grepl("Northern",x = temp$Region,ignore.case = TRUE)==FALSE,] temp<-temp[grepl("Oceania",x = temp$Region,ignore.case = TRUE)==FALSE,]

chisq.test(temp$Development_level$, $temp$Region, correct=F)

INTERACTION

The multicollinearity has seriously limited the possibilities of interactions.

Plots are similar, so there is no strong evidence of interaction

The Interactions to be tried: GDP growth per capita & Development level GDP growth per capita & Region

PLOTS NOT SHOWN

no trend

Region:

xyplot(Any_method ~ sex_ratio_2015 | Region, data = working_data)#no trend xyplot(Any_method ~ UR_2015_2018 | Region, data = working_data) #may be xyplot(Any_method ~ HDI | Region, data = working_data)#no trend xyplot(Any_method ~ gdp_growth_per_capita | Region, data = working_data)#no trend xyplot(Any_method ~ number_of_arrivals | Region, data = working_data) bwplot(Any_method~Region | Policy_on_fertility_level, data = working_data)#no trend bwplot(Any_method~Region | Development_level, data = working_data)#no trend

Policy_on_fertility_level:

xyplot(Any_method ~ sex_ratio_2015 | Policy_on_fertility_level, data = working_data)#no trend xyplot(Any_method ~ UR_2015_2018 | Policy_on_fertility_level, data = working_data) #no trend xyplot(Any_method ~ HDI | Policy_on_fertility_level, data = working_data)#no trend xyplot(Any_method ~ gdp_growth_per_capita | Policy_on_fertility_level, data = working_data)#no trend xyplot(Any_method ~ number_of_arrivals | Policy_on_fertility_level, data = working_data)#no trend bwplot(Any_method~Policy_on_fertility_level | Development_level, data = working_data)#no trend

Development_level:

xyplot(Any_method ~ sex_ratio_2015 | Development_level, data = working_data)#no trend
xyplot(Any_method ~ UR_2015_2018 | Development_level, data = working_data) #no trend
xyplot(Any_method ~ HDI | Development_level, data = working_data)#no trend
xyplot(Any_method ~ gdp_growth_per_capita | Development_level, data = working_data)#may be xyplot(Any_method ~ number_of_arrivals | Development_level, data = working_data)#no trend

cAny_method ~ csex_ratio_2015 + cpop_density_2015 + cUR_2015_2018 + ctotal_pop_2015 + cHDI + cnumber_of_arrivals + cgni_pp_latest + cgni_atlas_latest + cgdp_growth_per_capita

MEAN CENTERING DONE

DIFFERENT REGRESSION MODELS TRIED

1. A model without transformations & interactions
2. A model with log on GNI pp
3. A model with log on GNI atlas
4. A model with pop density squared
5. A model with log on UR 2015 2018
6. A model with interaction between GDP growth per capita & Country Development level
7. A model with interaction between GDP growth per capita & Region

```
#We will first run the models and then compare them.
reg1 = lm(Any_method ~ csex_ratio_2015 + cpop_density_2015 + cUR_2015_2018
+ ctotal_pop_2015 + cHDI + cnumber_of_arrivals + cgni_pp_latest + cgni_atlas_
latest  + cgdp_growth_per_capita, data = working_data)

working_data$log_gni_pp_latest<-log(working_data$gni_pp_latest)
working_data$clog_gni_pp_latest = working_data$log_gni_pp_latest - mean(worki
ng_data$log_gni_pp_latest,na.rm = TRUE)
reg2 = lm(Any_method ~ csex_ratio_2015 + cpop_density_2015 + cUR_2015_2018
+ ctotal_pop_2015 + cHDI + cnumber_of_arrivals + clog_gni_pp_latest + cgni_at
las_latest  + cgdp_growth_per_capita, data = working_data)

working_data$log_gni_atlas_latest<-log(working_data$gni_atlas_latest)
working_data$clog_gni_atlas_latest = working_data$log_gni_atlas_latest - mean
(working_data$log_gni_atlas_latest,na.rm = TRUE)
reg3 = lm(Any_method ~ csex_ratio_2015 + cpop_density_2015 + cUR_2015_2018
+ ctotal_pop_2015 + cHDI + cnumber_of_arrivals + cgni_pp_latest + clog_gni_at
las_latest  + cgdp_growth_per_capita, data = working_data)

working_data$sqr_cpop_density_2015<-(working_data$cpop_density_2015)^2
reg4 = lm(Any_method ~ csex_ratio_2015 + sqr_cpop_density_2015 + cUR_2015_201
8   + ctotal_pop_2015 + cHDI + cnumber_of_arrivals + cgni_pp_latest + cgni_at
```

```
las_latest  + cgdp_growth_per_capita, data = working_data)

working_data$log_UR_2015_2018<-log(working_data$UR_2015_2018+2)
working_data$clog_UR_2015_2018 = working_data$log_UR_2015_2018 - mean(working
_data$log_UR_2015_2018,na.rm = TRUE)
reg5 = lm(Any_method ~ csex_ratio_2015 + cpop_density_2015 + clog_UR_2015_201
8   + ctotal_pop_2015 + cHDI + cnumber_of_arrivals + cgni_pp_latest + cgni_at
las_latest  + cgdp_growth_per_capita, data = working_data)

reg6 = lm(Any_method ~ csex_ratio_2015 + cpop_density_2015 +Development_level
*cgdp_growth_per_capita   + ctotal_pop_2015, data = working_data)

reg7 = lm(Any_method ~ csex_ratio_2015 + cpop_density_2015 +Region*cgdp_growt
h_per_capita   + ctotal_pop_2015, data = working_data)
```

Comparing and Selecting the models:

The models are compared based upon 4 things: 1. Standarized Residual Plots and Q-Q norm plots 2. R square 3. Nested F test 4. Interpretabilty # 1. All residual plots show random pattern which is good. Except the 6th one. # However, the q-q plot for 1,2,3,4,5 have a better trend, with 4 and 5 being brilliant.

par(mfrow = c(4,4)) plot(reg1,which=1) plot(reg1,which=2)

plot(reg2,which=1) plot(reg2,which=2)

plot(reg3,which=1) plot(reg3,which=2)

par(mfrow = c(3,2))

plot(reg4,which=1) plot(reg4,which=2)

plot(reg5,which=1) plot(reg5,which=2)

plot(reg6,which=1) plot(reg6,which=2)

plot(reg7,which=1) plot(reg7,which=2)

2. R squares of the respective models:

1. A model without transformations and interactions - 61.54
2. A model with log on gni_pp - 61.52
3. A model with log on gni_atlas - 0.6335
4. A model with pop_density sqaured - 61.90
5. A model with log on UR_2015_2018 - 61.59

6. A model with interaction between gdp_growth_per_capita & Development_level - 0.237
7. A model with interaction between gdp_growth_per_capita & Region - 0.51

3. Nested F test

Model 1, 2, 3, 4, 5 till now have proved to better models. Running Nested F test to further ensure the same

For all models, with P-values significantly > 0.05, we cannot reject the null hypothesis the two models are different.

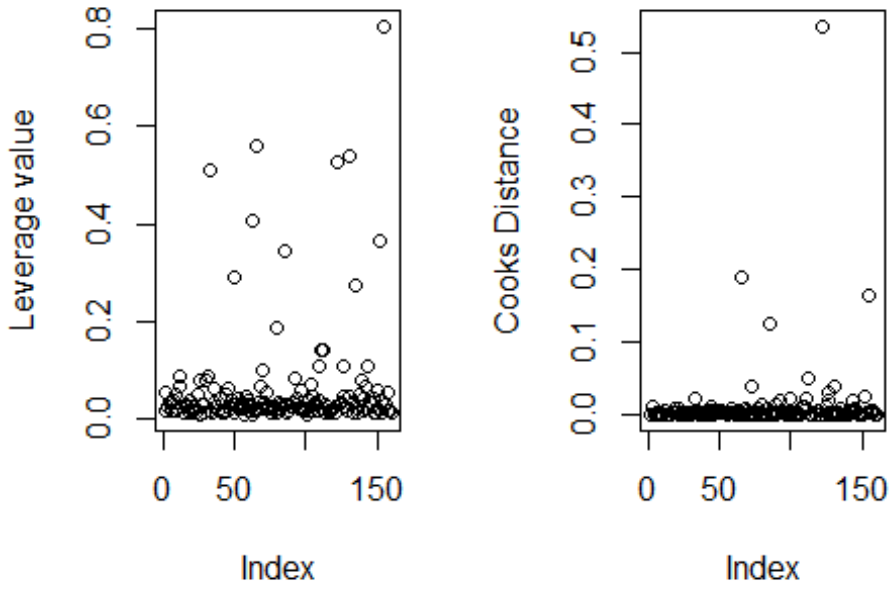
The variables do not add significant value in comparision to MODEL 1

this evaluation made the task of choosing easier.

4. Interpretation: defintely model 1 has the easiest interpretation. While model 2,3,5 have a easier interpretabilty than model 4. Model 6 and 7 have interactions, which make the interpretation harder.

Final selection: It's an easy choice: Model 1. It has a good resiual plot, qq-plot, high R sqaure and an easiest interpretabilty.

```
# Outlier Check:

par(mfrow = c(1,2))
lev = hatvalues(reg1)
cooks = cooks.distance(reg1)

plot(lev, ylab = "Leverage value")
plot(cooks, ylab = "Cooks Distance")
```



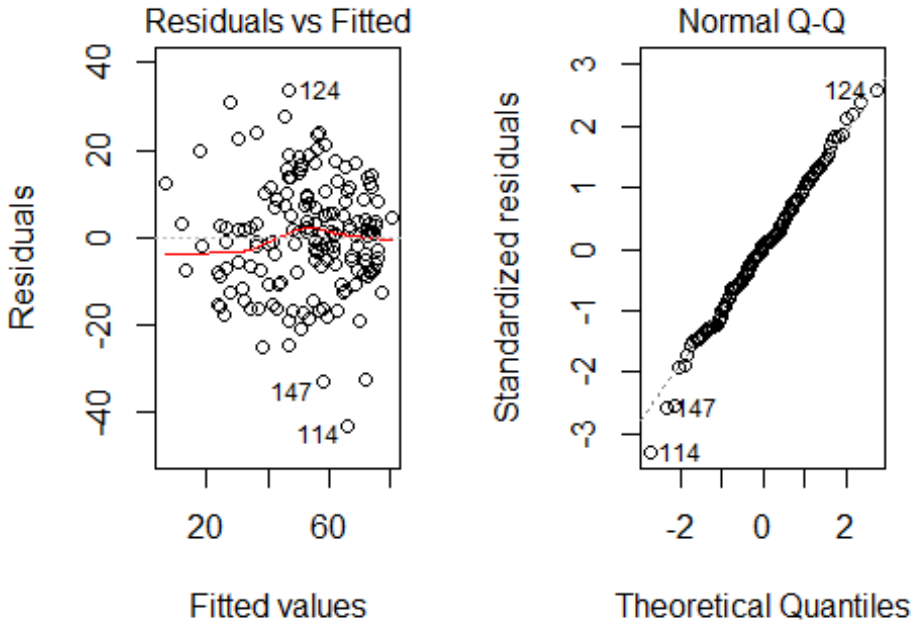
just a couple of outliers. The plots look fine. we can keep them. Would not affect the model too much if we remove them.

```
summary(reg1)
```

```
##
## Call:
## lm(formula = Any_method ~ csex_ratio_2015 + cpop_density_2015 +
##     cUR_2015_2018 + ctotal_pop_2015 + cHDI + cnumber_of_arrivals +
##     cgni_pp_latest + cgni_atlas_latest + cgdp_growth_per_capita,
##     data = working_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.992  -7.800   0.576   8.432  33.595
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             5.237e+01  1.045e+00  50.142  < 2e-16 ***
## csex_ratio_2015        -1.564e-01  6.267e-02  -2.495  0.01368 *
## cpop_density_2015      -1.176e-04  1.414e-03  -0.083  0.93381
## cUR_2015_2018           5.873e+00  2.210e+00   2.658  0.00872 **
## ctotal_pop_2015         1.266e-05  8.694e-06   1.456  0.14747
## cHDI                    1.321e+02  1.447e+01   9.131 4.32e-16 ***
## cnumber_of_arrivals     4.854e-08  1.131e-11   0.429  0.06685
## cgni_pp_latest         -1.674e-04  1.305e-04  -1.283  0.20133
```

```
## cgni_atlas_latest      -5.631e-13  8.646e-13  -0.651  0.51587
## cgdp_growth_per_capita  3.397e-01  3.320e-01   1.023  0.30783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.17 on 150 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.6154, Adjusted R-squared:  0.5923
## F-statistic: 26.66 on 9 and 150 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(1,2))
plot(reg1,which=1)
plot(reg1,which=2)
```



CONCLUSIONS & INTERPRETATION

Intercept Interpretation: The mean contraceptive prevalence is 50.31 at mean values of all country's sex ratio, population density, Urbanization rate, total population, Human development index, GNI PP, GNI Atlas and GDP growth rate per capita.

GDP growth rate per capita is not an important variable! (The Standard error is high and the confidence interval includes 0). This initially was perplexing. GDP is high for countries which are developing while low for countries both developed and under developed.

Human Development index is the most important variable (t value=9.13). #Keeping all else constant, if the HDI increases by 0.1 units, the y variable increases by 10.35 units.

Number of international flight arrivals is fairly a significant variable too! This suggests that countries with higher tourism tend to have better contraceptive prevalence. #Keeping all else constant, if the arrival flights increases by 1 unit, the y variable increases by 4.8x10-8 units.

The other important variables are: sex ratio and urbanization rate. #Keeping all else constant, the Contraceptive prevalence decreases by -0.15 if sex ratio is increased by one unit while increases by 5.8 if urbanization rate increases by one unit.

CONCLUSIONS 2

To see the effect of omitted factors, two things were done: 1) The individual effect of the variable 2) The difference between the actual and predicted values plotted against the omitted variables

Policies on Fertility: Individually, it has a clear trend. Lowering decreases CV while Raising increases it. But when plotted against the difference, the trend is flat, indicating that the policies do not add value over and above the variables in model. Region: The CV increases in order: Africa, Asia, Europe, Latin America. For the 2nd analysis, it has the same effect like that of the Policies. Gender Inequality Index: As the disparity ↑,CV ↓. The second analysis, same conclusion.

Literacy Rate: Individually, it has a good trend. But when plotted against the difference, there trend is present, indicating that the Literacy Rate does add value over and above the variables in mode! GDP per Capita: Individually, it does have a decent trend. And when plotted against the difference, there is a trend present as well!