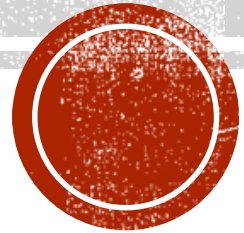
HUL-315

LINEAR REGRESSION ANALYSIS OF POPULATION GROWTH RATES IN 580 DISTRICTS OF INDIA



Amit Choudhary - 2016CH10077

Rudraksh Agarwal – 2016PH10549

Varunesh Kumar – 2016PH10573

OBJECTIVE & ITS RELEVANCE

- **Objective:** Analysis of dependence of Population Growth on various socio-economic factors.
- **Relevance:** Population Growth Affects –
  - Age structure of a country's population,
  - International Migration,
  - Economic Inequality,
  - Size of country's workforce,
  - Overall Economic Growth.[1]
- Population Growth varies across communities, classes and geographies.
  - Will assist in better policies directed towards masses.

LITERATURE REVIEW

J. Bongaarts (1978)[2] measured the impact of socio-economic factors on fertility.

Indirect determinants.

- Socio-economic, cultural, environmental variables.

Direct determinants.

- Intermediate fertility variables like the prevalence of contraception, induced abortions.

Urbanization: Increases childrearing financial cost directly & the opportunity cost. (Findley-1980)[3]

Rise of non-agricultural employment.

More Exposure to Economic downturns.

Education: Female education, a strong influencer of fertility. (M. Nag-2019).[4]

Acceptance of family planning

Delayed Age of Marriage.

SC/ST Section: Higher fertility than non-SC/ST females. (M.M. Gandotra-1998) [5]

Lower socioeconomic status

Poor family planning services.

FACTS ABOUT DATA

- Data is collected from Kaggle
  - India Socio-Economic Data - 2011 India Census Data (2015-2016)
    - Comprises of 580 districts across India.
  - Omitted-
    - Nagaland- Complex behavior and it lies mainly as outlier in many analysis.
    - Data not sufficient for some districts, specially of North East.

EXPLANATORY VARIABLES

X1-
Total Population of District

X2-
Percentage Urban Population(%)

X3-
SC Population(%)

X4-
ST Population(%)

X5-
Sex Ratio(No. of female per 1000 male)

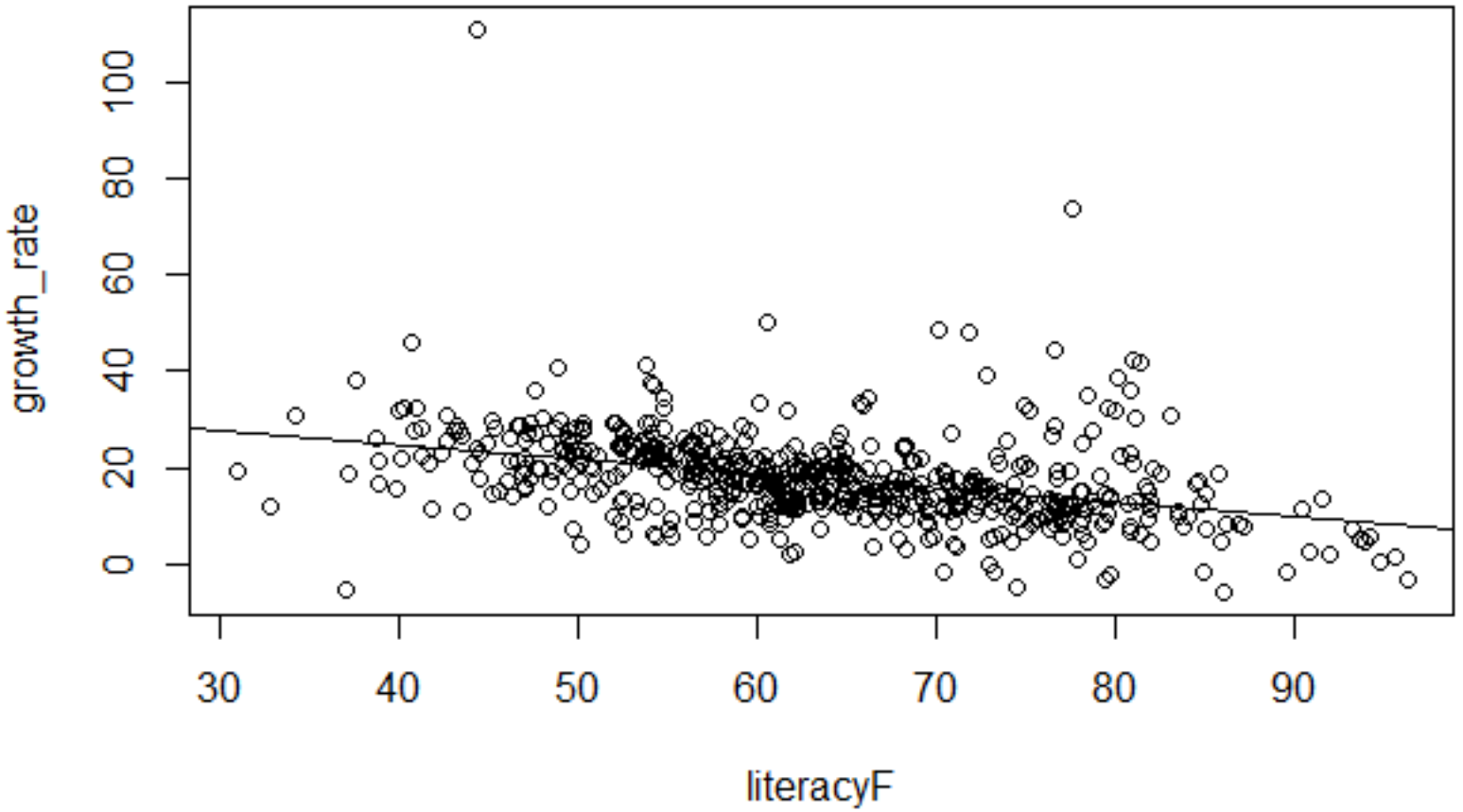
X6-
Overall Literacy (%)

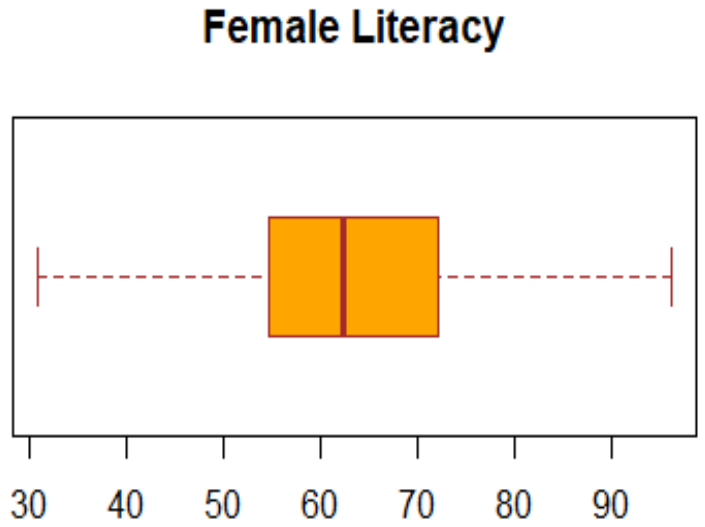
X7-
Female Literacy (%)

X8-
Male Literacy (%)

Dependent Variable - Population Growth Rate(Y)

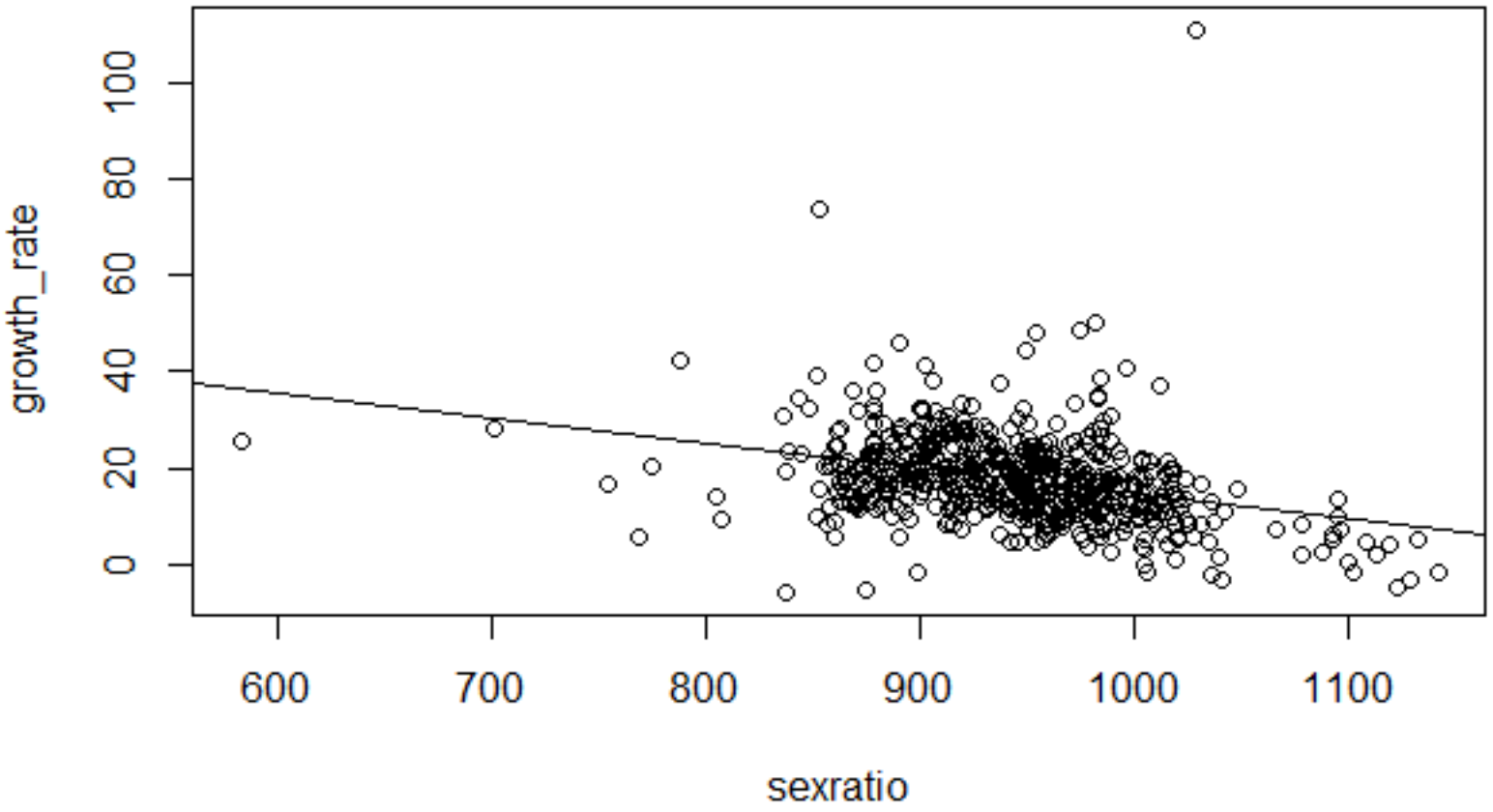
SCATTER PLOT: FEMALE LITERACY

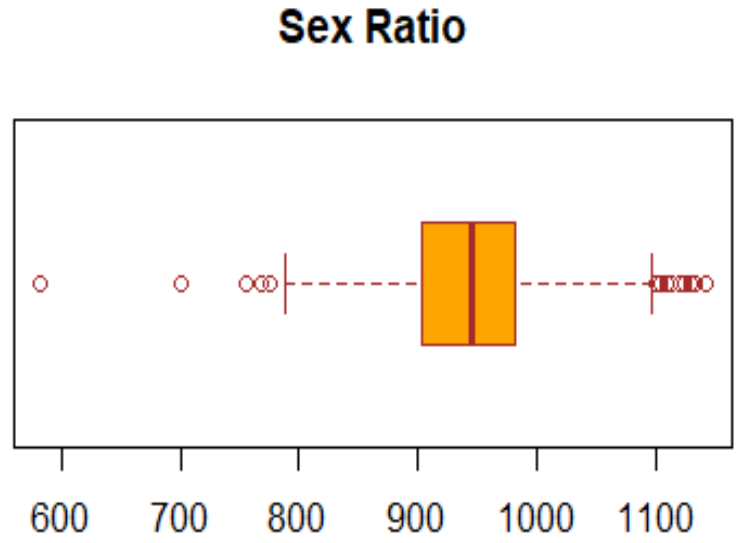




Minimum	30.97
Q1	54.80
Median	62.48
Mean	63.45
Q3	72.13
Maximum	96.26

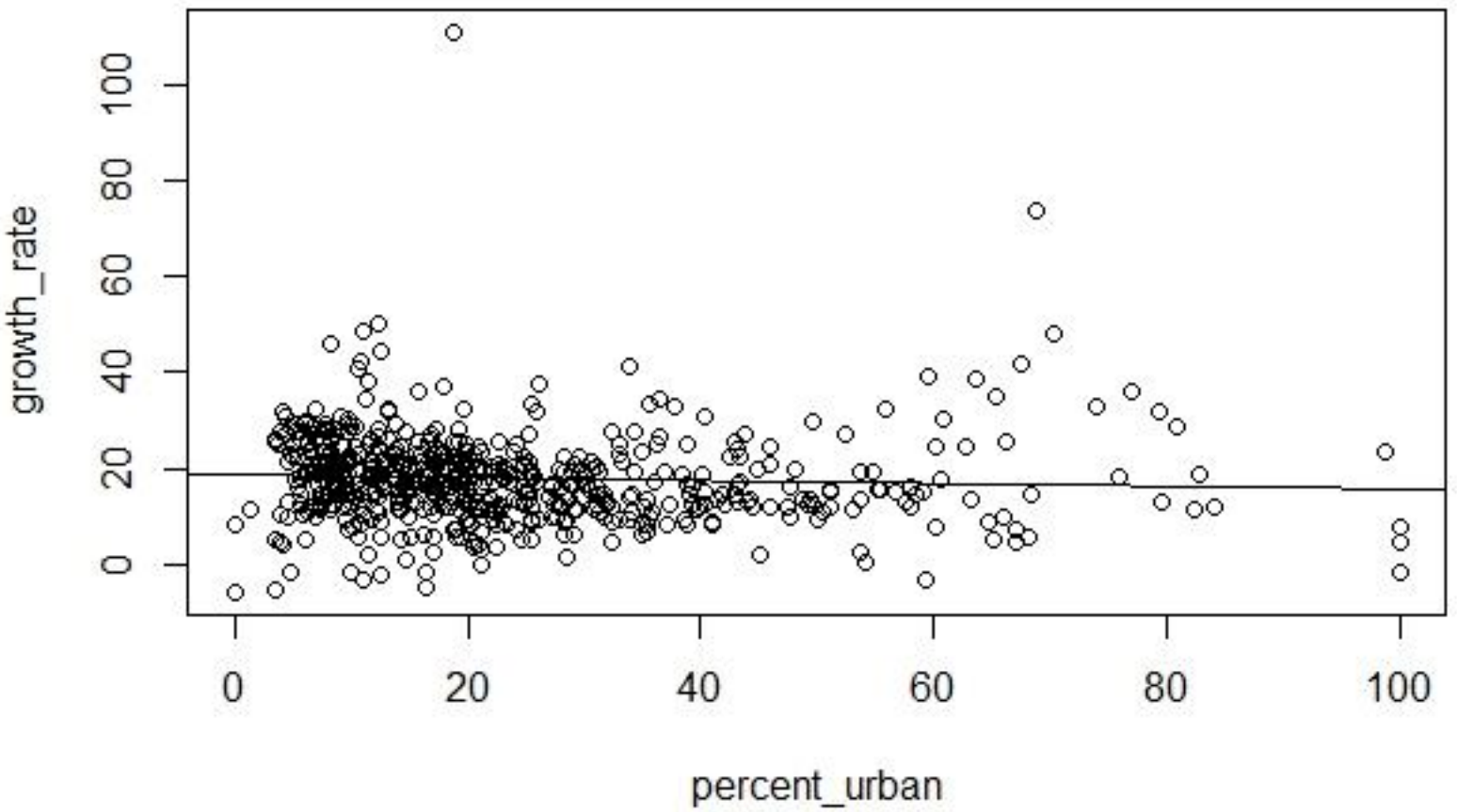
SCATTER PLOT: SEX RATIO

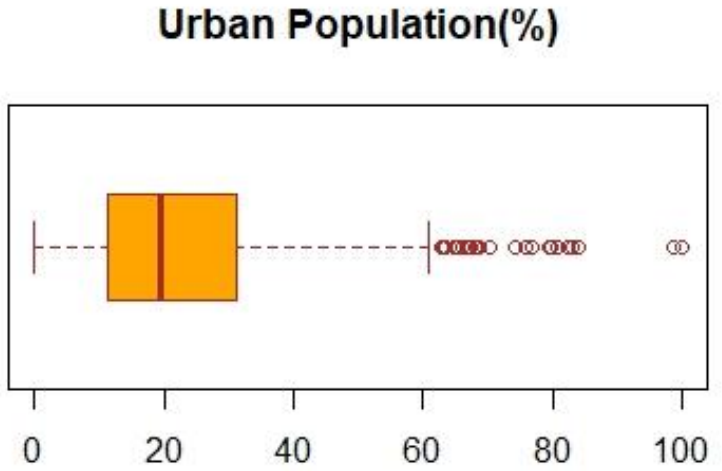




Minimum	583
Q1	903
Median	946
Mean	942.8
Q3	981
Maximum	1142

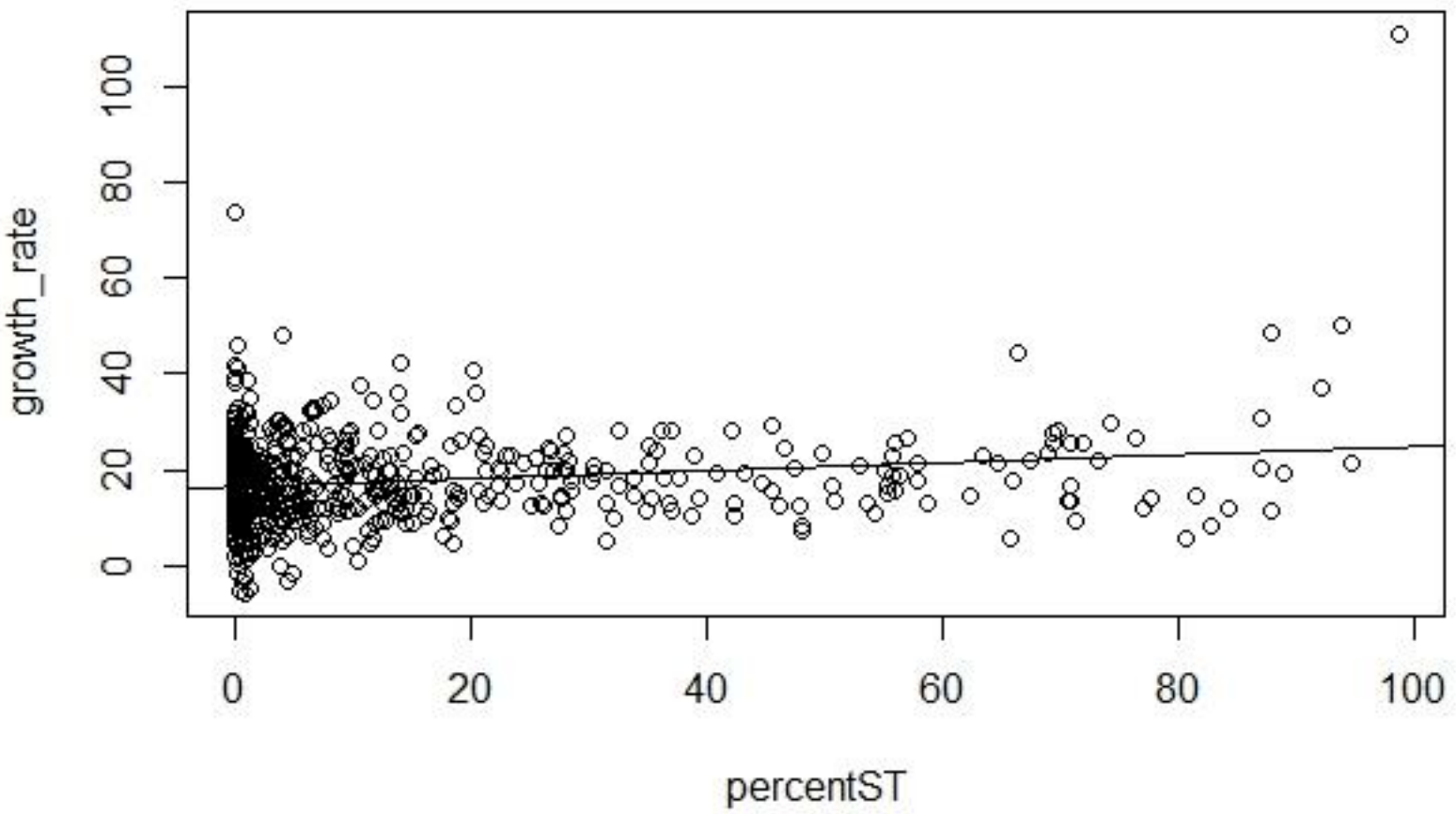
SCATTER PLOT: URBAN POPULATION

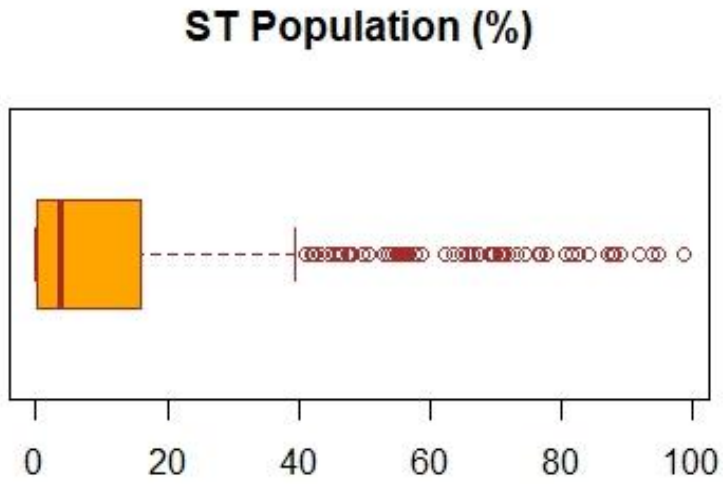




Minimum	0.00
Q1	11.25
Median	19.34
Mean	23.98
Q3	31.08
Maximum	100

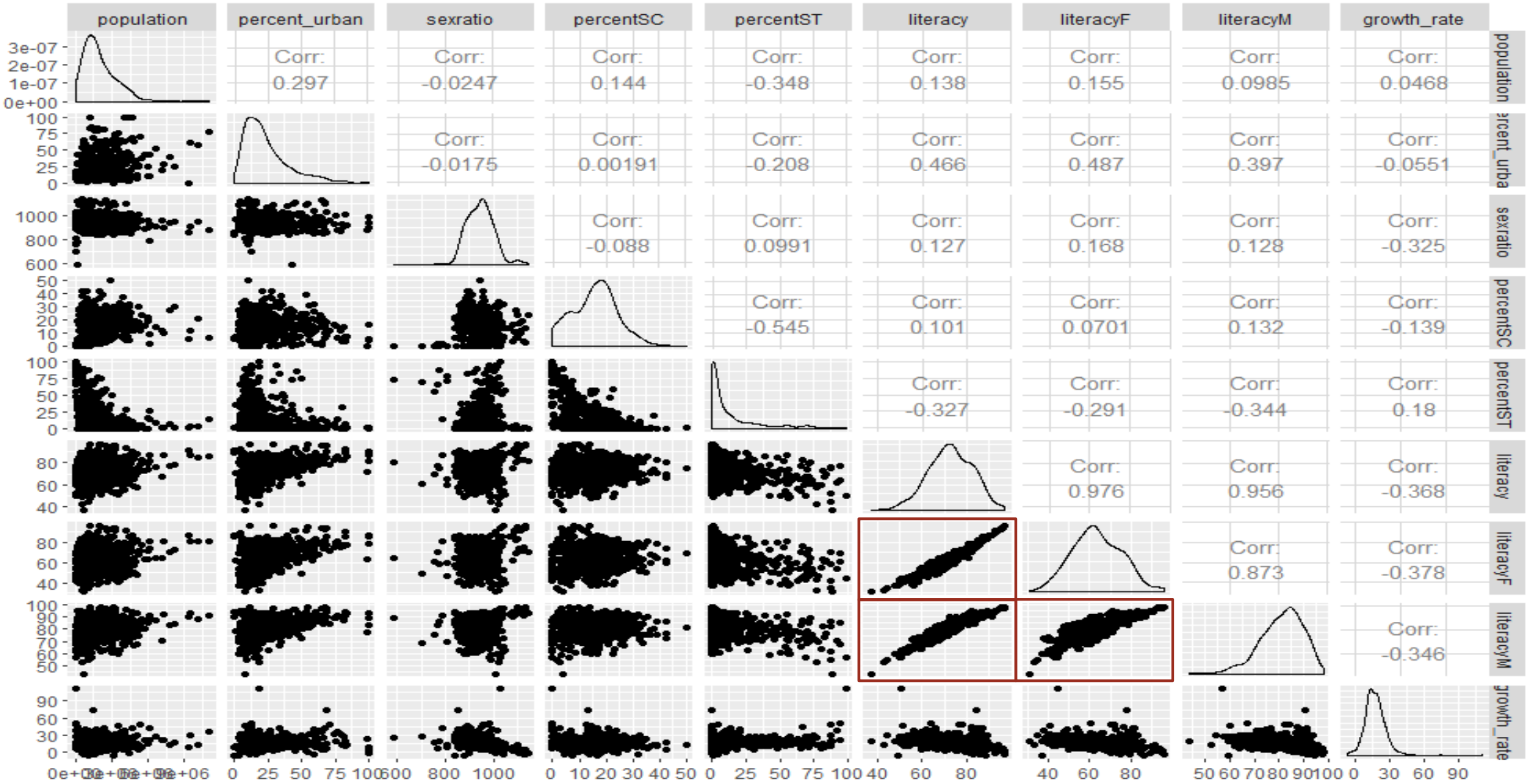
SCATTER PLOT: ST POPULATION





Minimum	0.00
Q1	0.33
Median	3.76
Mean	13.58
Q3	15.97
Maximum	98.58

CORRELATION MATRIX



```
lm(formula = growth_rate ~ percent_urban + sexratio + percentSC +
    percentST + literacyF + literacyM, data = new_data)

Residuals:
    Min      1Q  Median      3Q     Max
-31.792  -4.442  -0.696   3.361  87.078

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  78.428942   6.858585  11.435  < 2e-16 ***
percent_urban  0.072894   0.022750   3.204  0.00143 **
sexratio      -0.044341   0.006006  -7.383 5.48e-13 ***
percentSC     -0.106282   0.048181  -2.206  0.02779 *
percentST      0.033594   0.020997   1.600  0.11015
literacyF     -0.281188   0.062761  -4.480 9.00e-06 ***
literacyM     -0.018996   0.085845  -0.221  0.82495
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.31 on 573 degrees of freedom
Multiple R-squared:  0.2489,    Adjusted R-squared:  0.241
F-statistic: 31.65 on 6 and 573 DF,  p-value: < 2.2e-16
```

MODEL-1: ALL VARIABLES

- Model –
  - $\mathbf{Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_7 X_7 + \beta_8 X_8}$
- From t-test-
  - **%Urban Population, Sex Ratio, %SC, Female Literacy** statistically significant.
- F-stat depicts that the model is significant.
  - $H_o = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_7 = \beta_8 = 0$
  - p-value < 0.05
    - Reject the null Hypothesis.

```
lm(formula = growth_rate ~ percent_urban + sexratio + percentSC +
    percentST + literacyF, data = new_data)

Residuals:
    Min      1Q  Median      3Q     Max
-31.666  -4.428  -0.688   3.354  87.249

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  77.58571    5.69788  13.617  < 2e-16 ***
percent_urban  0.07330    0.02266   3.235  0.00129 **
sexratio     -0.04432    0.00600  -7.387 5.34e-13 ***
percentSC    -0.10672    0.04810  -2.219  0.02691 *
percentST     0.03425    0.02077   1.649  0.09962 .
literacyF    -0.29276    0.03467  -8.445 2.50e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.303 on 574 degrees of freedom
Multiple R-squared:  0.2488,    Adjusted R-squared:  0.2423
F-statistic: 38.03 on 5 and 574 DF,  p-value: < 2.2e-16
```

MODEL-2:

- Model –
  - $\mathbf{Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_7 X_7}$
- From t-test-
  - **%Urban Population, Sex Ratio, %SC, Female Literacy** statistically significant.
- F-stat depicts that the model is significant.
  - $H_o = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_7 = 0$
  - p-value < 0.05
    - Reject the null Hypothesis.

```
lm(formula = growth_rate ~ percent_urban + sexratio + percentSC +
    percentST, data = new_data)

Residuals:
    Min      1Q  Median      3Q     Max
-29.517  -5.013  -0.701   4.039  90.224

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  71.201171   5.982837  11.901  < 2e-16 ***
percent_urban -0.014267   0.021341  -0.669 0.504076
sexratio     -0.055930   0.006187  -9.040  < 2e-16 ***
percentSC    -0.087686   0.050902  -1.723 0.085489 .
percentST     0.075113   0.021395   3.511 0.000482 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.797 on 575 degrees of freedom
Multiple R-squared:  0.1555,    Adjusted R-squared:  0.1497
F-statistic: 26.47 on 4 and 575 DF,  p-value: < 2.2e-16
```

MODEL-3: DEMOGRAPHIC ANALYSIS

- Model –
  - $\mathbf{Y} = \beta_1 + \beta_2\mathbf{X}_2 + \beta_3\mathbf{X}_3 + \beta_4\mathbf{X}_4 + \beta_5\mathbf{X}_5$
- The model depicts that the growth rate in case of demographic analysis have significant dependence on
  - Sex Ratio
    - Decreases with higher sex ratio.
  - % ST population.
    - Increases with higher ST population.
  - While, % Urban is not statistically significant.

```
lm(formula = growth_rate ~ percent_urban + sexratio + percentSC +
    literacyF, data = new_data)

Residuals:
    Min      1Q  Median      3Q     Max
-32.426  -4.507  -0.582   3.236  89.103

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  78.663579   5.668743  13.877  < 2e-16 ***
percent_urban  0.069312   0.022561   3.072 0.002226 **
sexratio      -0.043239   0.005973  -7.239 1.45e-12 ***
percentSC     -0.150198   0.040294  -3.728 0.000212 ***
literacyF     -0.306083   0.033764  -9.065  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.316 on 575 degrees of freedom
Multiple R-squared:  0.2453,    Adjusted R-squared:  0.24
F-statistic: 46.72 on 4 and 575 DF,  p-value: < 2.2e-16
```

MODEL-4:

- Model –
  - $\mathbf{Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 + \beta_7 X_7}$
- From t-test-
  - **%Urban Population, Sex Ratio, %SC, Female Literacy** statistically significant.
- F-stat depicts that the model is significant.
  - $H_o = \beta_2 = \beta_3 = \beta_5 = \beta_7 = 0$
  - p-value < 0.05
    - Reject the null Hypothesis.

```
lm(formula = growth_rate ~ literacyM + literacyF, data = new_data)

Residuals:
    Min      1Q  Median      3Q     Max
-31.347  -4.358  -0.748   3.162  86.575

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 40.02115    4.26896   9.375  < 2e-16 ***
literacyM   -0.07607    0.08921  -0.853    0.394
literacyF   -0.25343    0.06254  -4.052 5.77e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.84 on 577 degrees of freedom
Multiple R-squared:  0.1441,    Adjusted R-squared:  0.1411
F-statistic: 48.58 on 2 and 577 DF,  p-value: < 2.2e-16
```

MODEL-5: LITERACY ANALYSIS

- Model –
  - $\mathbf{Y} = \beta_1 + \beta_7 \mathbf{X}_7 + \beta_8 \mathbf{X}_8$
- The model depicts that the growth rate in case of literacy analysis have significant dependence on
  - Female Literacy
    - Decreases with higher female literacy.
  - While, Male literacy is not statistically significant.
    - Large correlation between Male & Female Literacy Rate.

TEST FOR OLS CONDITION SATISFACTION

- Model-2:
  - $\mathbf{Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_7 X_7}$
  - **Adjusted R^2 = 0.2423**
  - **Multiple R^2 = 0.2488**
- Correlation of Residuals with the various variables.

```
> cor(residuals(model2),new_data[,c(2,3,4,5,7,9)])
     percent_urban      sexratio    percentSC    percentST    literacyF
[1,]  6.635037e-17 -4.116378e-16 9.299986e-17 1.980413e-16 1.760023e-17
```

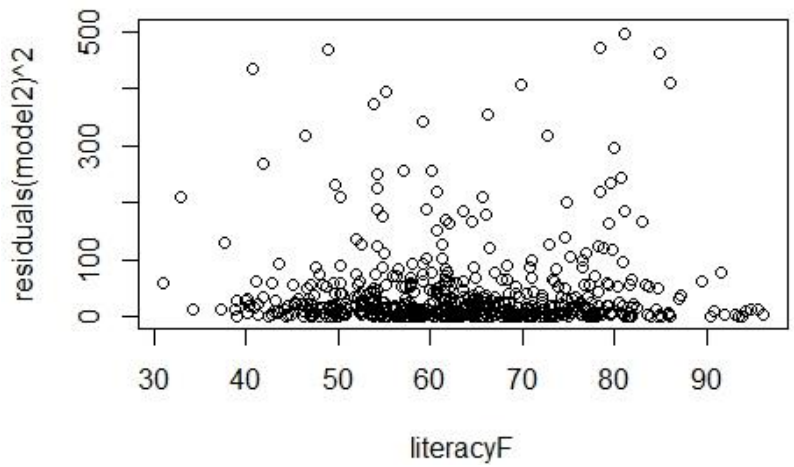
  - All the correlation values are quite small, thus telling almost no correlation between explanatory variables and residuals.

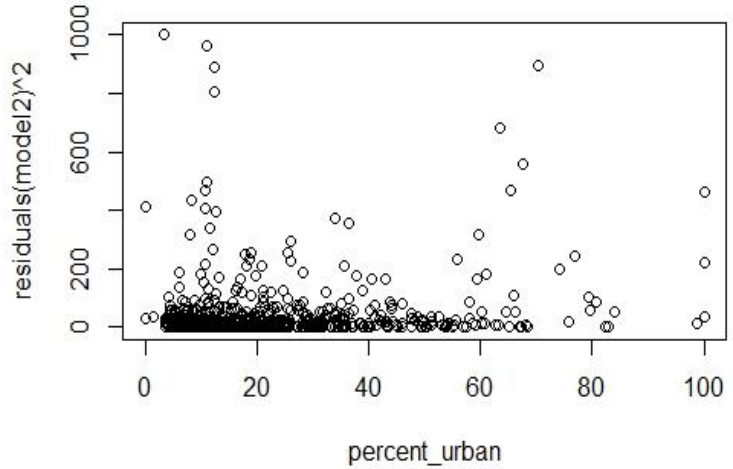
MULTICOLLINEARITY

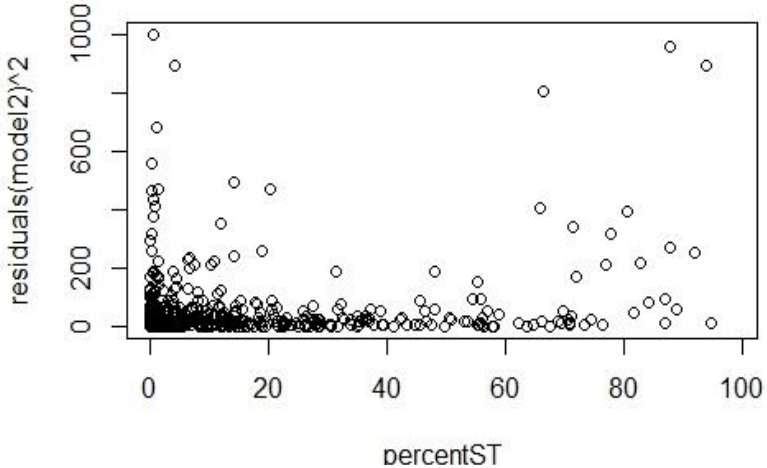
```
> vif(model1)
percent_urban      sexratio     percentSC     percentST     literacyF     literacyM
     1.356198      1.068021      1.457285      1.640592      4.776771      4.392062

> vif(model2)
percent_urban      sexratio     percentSC     percentST     literacyF
     1.347384      1.067684      1.454877      1.607605      1.459960

> vif(model3)
percent_urban      sexratio     percentSC     percentST
     1.065158      1.011600      1.451684      1.520337

> vif(model4)
percent_urban      sexratio     percentSC     literacyF
     1.332039      1.054988      1.017855      1.380707

> vif(model5)
literacyM literacyF
 4.191542  4.191542
```

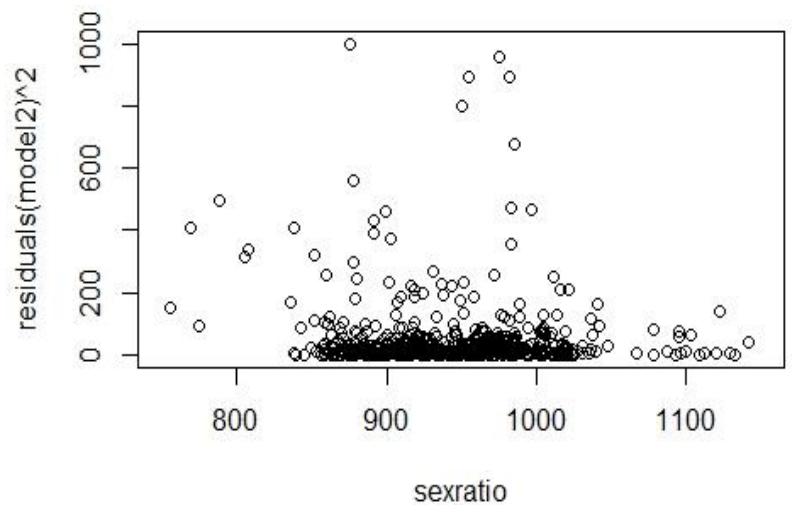
All VIF < 10 → No Issue regarding Multicollinearity.

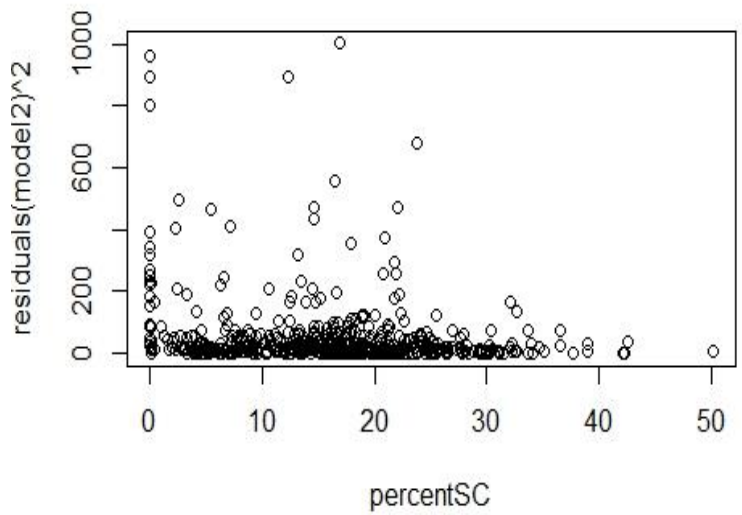
HETEROSCEDASTICITY

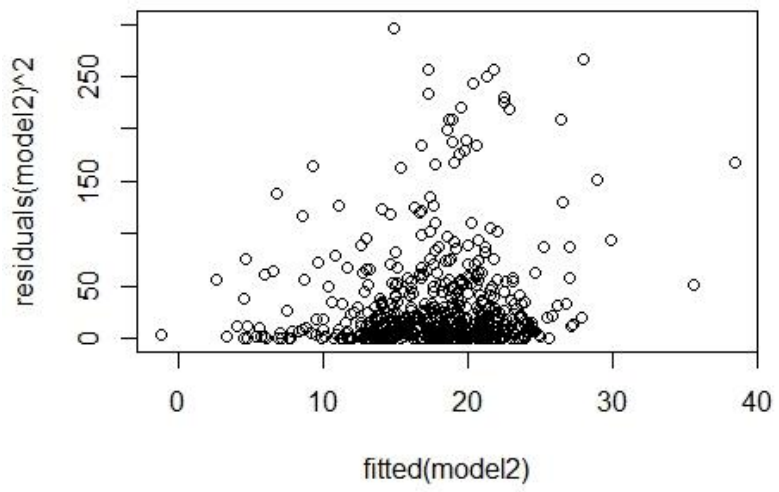












TEST FOR HETEROSCEDASTICITY

- Breusch Pegan Test –
  - p-value < 0.05
  - Reject Null Hypothesis that –
    - Variance is not changing with residual.
  - Heteroscedasticity present.

```
> bptest(model2)

        studentized Breusch-Pagan test

data:  model2
BP = 29.405, df = 5, p-value = 1.93e-05
```

- White Test –
  - p-value < 0.05
  - Reject Null Hypothesis that –
    - Variance is not changing with residual.
  - Heteroscedasticity present.

```
> bptest(model2, ~ literacyF*sexratio + literacyF*percent_urban + literacyF*percentSC + literacyF*percentS
T+ sexratio*percent_urban + sexratio*percentSC + sexratio*percentST + percent_urban*percentSC + percent_ur
ban*percentST + percentSC*percentST + I(literacyF^2) + I(sexratio^2) + I(percent_urban^2)  + I(percentSC^
2) + I(percentST^2),data = new_data[,c(2,3,4,5,7,9)])

        studentized Breusch-Pagan test

data:  model2
BP = 99.359, df = 20, p-value = 1.64e-12
```

CONCLUSION

- Population growth depends upon more factors than what are assumed in the current model. However, the current model illustrates that the relation is substantial.
  - Parameters like – Lack of Healthcare infrastructure, Minority feelings & culture.
- The relation is dynamic meaning the presence of more variables affects the dependent variable differently.
- Male literacy rate is mostly insignificant in the whole analysis. Hence, policies aimed at men aren't going to curb population growth much.
- Female literacy and sex ratio are very significant in determining the population growth in all cases. Hence, policies aimed at women are important to reduce population and educating them is even more crucial.
- Population growth is a complex phenomenon hence complete explanation is impossible and error are bound to come in.
- The current model is based on Indian population hence its not universal.

REFERENCES

- E. Wesley F. Peterson – "The Role of Population in Economic Growth," SAGE Journals, Volume: 7 issue: 4, 2017, 1-15.
- J. Bongaarts – "A Framework for Analyzing the Proximate Determinants of Fertility," *Population and Development Review* Vol. 4, No. 1 (1978), 105-132.
- Findley, Sally E. 1980. "A Suggested Framework for Analysis of Urban-Rural Fertility Differentials with an Illustration of the Tanzanian Case," *Population and Environment* 3 (3/4): 237–261.
- M. Nag. "Modernization and its Impact on Fertility: The Indian Scene," India International Centre Quarterly, Vol. 8, No. 3/4, India's Population: Problems & Prospects (1981), 235-247.
- M. M. Gandotra, R. D. Retherford, A. Pandey, N. Y. Luther, and V. K. Mishra. "Fertility in India," National Family Health Survey Subject Reports Number 9 • May 1998.

THANK YOU!