

Group Assignment Submission Form

Lecturer:	Dr. Michael Cronin
Module:	ST 6033
Date:	12 th March 2020

We hereby declare that the work submitted is our own and that ideas or extracts from other sources are properly acknowledged and referenced. We are satisfied that this represents a fair contribution in effort on the part of each member. Furthermore, we acknowledge that the penalty for plagiarism incurs serious sanctions.

Student Name	Student Number	Signature
Digant Kumar	119220141	
Shivam Kislay	119220420	

Contents

INTRODUCTION	3
METHODS.....	4
RESULTS	6
CONCLUSION	10
APPENDICES	11
Model Building Details:.....	11
Diagnostics Details:	13
Sparse Data:.....	17
Overdispersion:	17
References:.....	19

INTRODUCTION

The table below shows the data which was collected on a random sample of people to ascertain what factors influenced behavioral change during a Swine Flu outbreak. The data table contains gender of the person, the severity of contracting swine flu, the adequacy of the Information available, the age category of the person, and change in behavior pertaining to hygiene.

In our analysis of these data we will treat the change in behavior as our response variable, and gender of the person, age category, severity and information available as the four discrete predictors or factors.

Research question being addressed: Whether the person changed behavior relating to hygiene.

Purpose of the statistical analyses: The purpose of this analysis is to determine what factors influenced the change in behavior in a person during a Swine Flu outbreak.

Gender	Severe	Information	Age	n	Changed Behaviour
Male	Yes	Adequate	0 - 40	18	11
			40 - 50	23	14
			50 - 60	22	11
			60+	17	5
		Not Adequate	0 - 40	19	4
			40 - 50	35	15
			50 - 60	30	8
			60+	22	8
	No	Adequate	0 - 40	24	10
			40 - 50	37	13
			50 - 60	29	8
			60+	24	6
		Not Adequate	0 - 40	28	11
			40 - 50	42	14
			50 - 60	37	15
			60+	30	9
Female	Yes	Adequate	0 - 40	11	6
			40 - 50	25	13
			50 - 60	12	7
			60+	8	8
		Not Adequate	0 - 40	14	7
			40 - 50	34	15
			50 - 60	22	8
			60+	21	5
	No	Adequate	0 - 40	18	12
			40 - 50	28	15
			50 - 60	24	7
			60+	8	1
		Not Adequate	0 - 40	28	13
			40 - 50	47	21
			50 - 60	45	11
			60+	30	6

METHODS

Statistical Classification:

Statistical classification is the broad supervised learning approach that trains a program to categorize new, unlabeled information based upon its relevance to known, labeled data. The algorithms that sort unlabeled data into labeled classes, or categories of information, are called classifiers.

Generalized linear model:

A generalized linear model is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution.

The following is the structure of the Generalized Linear Model:

Stochastic (random) Component: $Y_i \sim \text{Exponential family density}$

Systematic Component: $g(\mu_i) = \eta_i = B_0 + B_1X_{i1} + B_2X_{i2} + B_3X_{i3} + B_pX_{ip}$

$E(Y_i) = \mu_i$ is linked to a linear combination of the predictors through some function $g()$

Diagnostics Test:

A diagnostics test is used for model checking, detecting outliers, checking high leverage and high influential values in the data. Moreover, it is also used for checking normality and detecting if the residuals have constant or non-constant variance.

Chi-square test:

Chi-squared test or χ^2 test, refers to certain types of statistical hypothesis tests that are valid to perform when the test statistic is chi-squared distributed under the null hypothesis. It is used to determine whether there is a statistically significant difference between two models.

Overdispersion:

Overdispersion is when the variability in the binomial data is greater than what would be expected if the responses were truly binomial. Overdispersion will manifest itself with a large deviance, assuming the fitted model is correct. A value greater than 1 may be indicative of overdispersion while a value of less than 1 may be indicative of under dispersion.

Sparse Data:

A variable with sparse data is one in which a relatively high percentage of the variable's cells do not contain actual data. Such "empty," or NA, values take up storage space in the file. There are two types of sparsity:

1) Controlled sparsity occurs when a range of values of one or more dimensions has no data; for example, a new variable dimensioned by MONTH for which you do not have data for past months. The cells exist because you have past months in the MONTH dimension, but the data is NA.

2) Random sparsity occurs when NA values are scattered throughout the data variable, usually because some combinations of dimension values never have any data. For example, a district might only sell certain products and never have data for other products. Other districts might sell some of those products and other ones, too.

Likelihood Ratio Test:

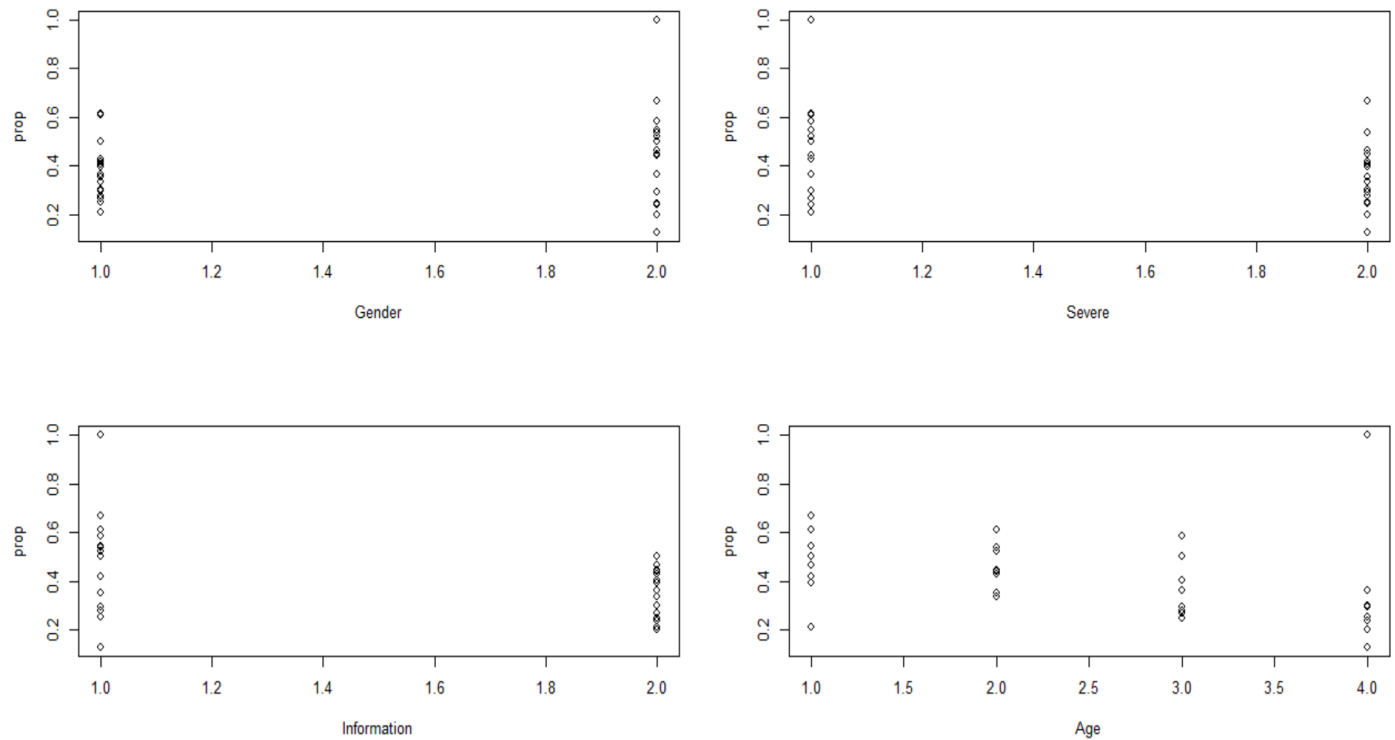
In statistics, the likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods, specifically one found by maximization over the entire parameter space and another found after imposing some constraint.

Odds Ratio:

An odds ratio (OR) is a statistic that quantifies the strength of the association between two events, A and B. The odds ratio is defined as the ratio of the odds of A in the presence of B and the odds of A in the absence of B, or equivalently (due to symmetry), the ratio of the odds of B in the presence of A and the odds of B in the absence of A. Two events are independent if and only if the OR equals 1, i.e., the odds of one event are the same in either the presence or absence of the other event.

RESULTS

Preliminary Analysis:



- **Gender and proportion:** Females are more likely to have had change in behavior than males relating to hygiene.
- **Severe and proportion:** The people who believed the consequences of contracting swine flu were severe changed behavior relating to hygiene almost as equally as much as those who thought otherwise.
- **Information and proportion:** The people who believed the information available about Swine Flu to be adequate had more behavioral change relating to hygiene than those who thought otherwise.
- **Age and proportion:** Change in behavior relating to hygiene decreases with the increase in Age.

R code:

```
prop <- Changed_Behaviour/n
plot(Gender, prop)
plot(Severe, prop)
plot(Information, prop)
plot(Age, prop)
```

Final Model:

R code:

```
y <- cbind(Changed_Behaviour, n - Changed_Behaviour)
glm_full_SI <- glm(y ~ factor(Severe) + factor(Information) +
  factor(Age) + factor(Gender) + factor(Severe):factor(Information),
  family = binomial(link="logit"))
summary(glm_full_SI)
```

Summary of the final model:

Call:

```
glm(formula = y ~ factor(Severe) + factor(Information) + factor(Age) +
  factor(Gender) + factor(Severe):factor(Information), family = binomial(link = "logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7944	-0.5956	0.0378	0.3577	3.4126

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.40976	0.23268	1.761	0.07823	.
factor(Severe)2	-0.73110	0.23047	-3.172	0.00151	**
factor(Information)2	-0.79852	0.23044	-3.465	0.00053	***
factor(Age)2	-0.08784	0.20312	-0.432	0.66543	
factor(Age)3	-0.50742	0.21553	-2.354	0.01856	*
factor(Age)4	-0.68213	0.23745	-2.873	0.00407	**
factor(Gender)2	0.20409	0.14825	1.377	0.16863	
factor(Severe)2:factor(Information)2	0.68915	0.30231	2.280	0.02263	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 66.061 on 31 degrees of freedom

Residual deviance: 32.582 on 24 degrees of freedom

AIC: 157.73

Number of Fisher Scoring iterations: 4

Interpretation of the model parameter estimates:

Intercept -> $\exp(0.40976) = 1.506$

Log odds for males in the age category 0-40, who believed the consequences of contracting swine flu were severe and believed the information available about swine flu to be adequate were 1.506.

Mathematical but no practical meaning.

Factor(Severe)2 -> $\exp(-0.73110) = 0.48$

People who believed the consequences of contracting swine flu were not severe changed behavior relating to hygiene **less likely** than people who believed otherwise by **52%**.
p value = 0.00151. The difference is significant.

Factor(Information)2 -> $\exp(-0.79852) = 0.45$

People who believed the information available about Swine Flu was not adequate changed behavior relating to hygiene **less likely** than people who believed otherwise by **55%**.
p value = 0.00053. The difference is significant.

Factor(Age)2 -> $\exp(-0.08784) = 0.916$

People in the Age category 40-50 changed behavior relating to hygiene **less likely** than people in the age category 0-40 by **8.4%**.
p value = 0.66543. The difference is insignificant.

Factor(Age)3 -> $\exp(-0.50742) = 0.60$

People in the Age category 50-60 changed behavior relating to hygiene **less likely** than people in the age category 0-40 by **40%**.
p value = 0.01856. The difference is significant.

Factor(Age)4 -> $\exp(-0.68213) = 0.50$

People in the Age category 60+ changed behavior relating to hygiene **less likely** than people in the age category 0-40 by **50%**.
p value = 0.00407. The difference is significant.

Factor(Gender)2 -> $\exp(0.20409) = 1.23$

Females changed behavior relating to hygiene **more likely** than males by **23%**.
p value = 0.16863. The difference is insignificant.

Factor(Severe)2 : Factor(Information) 2 -> $\exp(0.68915) = 1.99$

Case 1: When the person thought that the consequences of contracting swine flu was not severe and the information available was adequate: ($\exp(-0.73110) = 0.48$)

People who believed that the consequences of contracting swine flu were not severe changed behavior relating to hygiene **less likely** than people who believed otherwise by **52%**.
p value = 0.00151. The difference is significant.

Case 2: When the person thought that the consequences of contracting swine flu was severe but the information available was not adequate: $(\exp(-0.79852) = 0.45)$

People who believed that the information available about Swine Flu was not adequate changed behavior relating to hygiene **less likely** than people who believed otherwise by **55%**.

p value = 0.00053. The difference is significant.

Case 3: When the person thought that the consequence of contracting swine flu was not severe and the information available was not adequate: $(\exp(-0.73110 - 0.79852 + 0.68916) = 0.431512)$

People who believed that the information available about swine flu was not adequate and the consequences of contracting swine flu were not severe, were **56.8%** less likely to have had change in behavior than people who thought that the information available was adequate and the consequences were severe.

p value = 0.02263. The difference is significant.

Sparse Data:

There is no issue of sparse data on our dataset as the standard error values are small.

Over Dispersion:

There is a case of over dispersion as the mean deviance = 1.35 (>1). A new model is fitted with family as quasibinomial. The new model will have the same coefficients and deviance. Only the coefficient of the standard error will increase by the original standard error times the square root of the dispersion factor which in this case is 1.213141.

CONCLUSION

Factor's response to the change in behavior relating to hygiene in case of contracting swine flu:

- People who believed that the consequences of contracting swine flu were severe had **52%** less changes in their behavior relating to hygiene than people who believed otherwise.
- People who believed that the information available about contracting swine flu was not adequate had **55%** less changes in their behavior relating to hygiene than people who believed otherwise.
- People of the age group of 0 – 40 had the **maximum** change in behavior compared to the other age groups.
 - People in the age category of 40 – 50 years had **8.4% less** change in their behavior relating to hygiene than the people in the age group of 0 – 40.
 - People in the age category of 50 – 60 years had **40% less** change in their behavior relating to hygiene than people in the age group of 0 – 40.
 - People in the age category of 60+ years had **50% less** change in their behavior to hygiene than people in the age group of 0 – 40.
- Females had a **23% more** change in their behavior relating to hygiene than males.

APPENDICES

Model Building Details:

Predictor variables were converted into their respective categorical data i.e. **Gender** was split into two categories, 1: if the person is male and 2: if the person is female. Similarly, **Severity** spread out into two categories, 1: if the person believed the consequences of contracting swine flu were severe, 2: if the person believed otherwise. Likewise, for **Information**, 1: if the person believed the information available about swine flu was adequate, 2: if the person believed otherwise, and finally for **Age**, 4 categories, 1: for Age group 0-40, 2: for age group 40-50, 3: for age group 50-60 and 4: for age group 60+.

Preliminary analysis was conducted to see the various trends in response variable with respect to predictor variables

And finally, generalized linear model was fitted to the data, with change in behavior as the response variable and gender, age, severe and information as the predictor variables.

R code for implementing GLM on the data:

```
glm_full <- glm(y~factor(Gender) + factor(Severe) +  
factor(Information) + factor(Age), family = binomial(link=logit))  
# .....Full Model)
```

Here y is a 2x2 matrix with Changed behavior as the first column and (Total – Changed Behaviour) as the 2nd column. Family is taken as binomial, as the data is categorical and the link function as logit, since logit will give values ranging from 0 to 1 which will link the categories.

During the forward selection process of building the model, full model was observed as not an adequate fit as its residual deviance at 25 degrees of freedom was **37.813** and it exceeded the critical value (**37.65**) at 25 degrees of freedom and hence interactions terms were added to the model. Table below illustrates the forward selection process for best model fitting. (where G stands for Gender, S for Severe, I for Information and A for Age)

Model	Deviance	DOF	Likelihood Ratio Test
Null	66.06	31	First Null model
			Reduction in DOF 6 and reduction in Deviance = $28.25 > 12.59$ (Critical Chi Sq for 6 DOF). Therefore, a significant reduction.
Full	37.81	25	Hence this model is our new base model
Full + GS	37.771	24	Insignificant reduction in deviance for 1 D.O.F
Full + GI	36.495	24	Insignificant reduction in deviance for 1 D.O.F
Full + GA	35.093	22	Insignificant reduction in deviance for 3 D.O.F
			Reduction in DOF 1 and reduction in Deviance = $5.29 > 3.84$ (Critical ChiSq for 1 DOF). Therefore, a significant reduction.
Full + SI	32.582	24	Hence this model is our new base model
Full + SI + SA	29.647	21	Insignificant reduction in deviance for 3 D.O.F
Full + SI + SG	32.414	23	Insignificant reduction in deviance for 1 D.O.F
Full + SI + GA	29.872	21	Insignificant reduction in deviance for 3 D.O.F
Full + SI + GI	31.073	23	Insignificant reduction in deviance for 1 D.O.F
Full + SI + IA	31.731	21	Insignificant reduction in deviance for 1 D.O.F
Full + SI + SA + SG	29.478	20	Insignificant reduction in deviance for 4 D.O.F
Full + SI + SA + GA	27.212	18	Insignificant reduction in deviance for 6 D.O.F
Full + SI + SA + GI	28.157	20	Insignificant reduction in deviance for 4 D.O.F
Full + SI + SA + IA	28.691	18	Insignificant reduction in deviance for 6 D.O.F
Full + SI + SG + GA	29.679	20	Insignificant reduction in deviance for 4 D.O.F
Full + SI + SG + GI	30.919	22	Insignificant reduction in deviance for 2 D.O.F
Full + SI + SG + IA	31.555	20	Insignificant reduction in deviance for 4 D.O.F
Full + SI + SG + IA	31.555	20	Insignificant reduction in deviance for 4 D.O.F
Full + SI + GA + GI	28.471	20	Insignificant reduction in deviance for 4 D.O.F
Full + SI + GA + IA	28.771	18	Insignificant reduction in deviance for 6 D.O.F
Full + SI + GI + IA	30.138	20	Insignificant reduction in deviance for 4 D.O.F
FULL + SI + SA + SG + GA	26.999	17	Insignificant reduction in deviance for 7 D.O.F
FULL + SI + SA + SG + GI	27.99	19	Insignificant reduction in deviance for 5 D.O.F
FULL + SI + SA + SG + IA	28.512	17	Insignificant reduction in deviance for 7 D.O.F
FULL + SI + SG + GA + GI	28.296	19	Insignificant reduction in deviance for 5 D.O.F
FULL + SI + SG + GA + IA	28.558	17	Insignificant reduction in deviance for 7 D.O.F
Full + SI + GA + GI + IA	27.326	17	Insignificant reduction in deviance for 1 D.O.F
FULL + SI + SA + SG + GA + GI	25.657	16	Insignificant reduction in deviance for 8 D.O.F
FULL + SI + SA + SG + GA + IA	25.767	14	Insignificant reduction in deviance for 10 D.O.F
FULL + SI + SG + GA + GI + IA	27.131	16	Insignificant reduction in deviance for 8 D.O.F
Full + SI + GI + GA + SA + IA	24.611	14	Insignificant reduction in deviance for 10 D.O.F
Full + SI + SG + GI + SA + IA	26.959	16	Insignificant reduction in deviance for 8 D.O.F
FULL + SI + SA + SG + GA + GI + IA	24.398	13	Insignificant reduction in deviance for 11 D.O.F

Likelihood Ratio Test was performed on each model, where the reduction in deviance was compared against the critical chi-square value for the reduced degree of freedom.

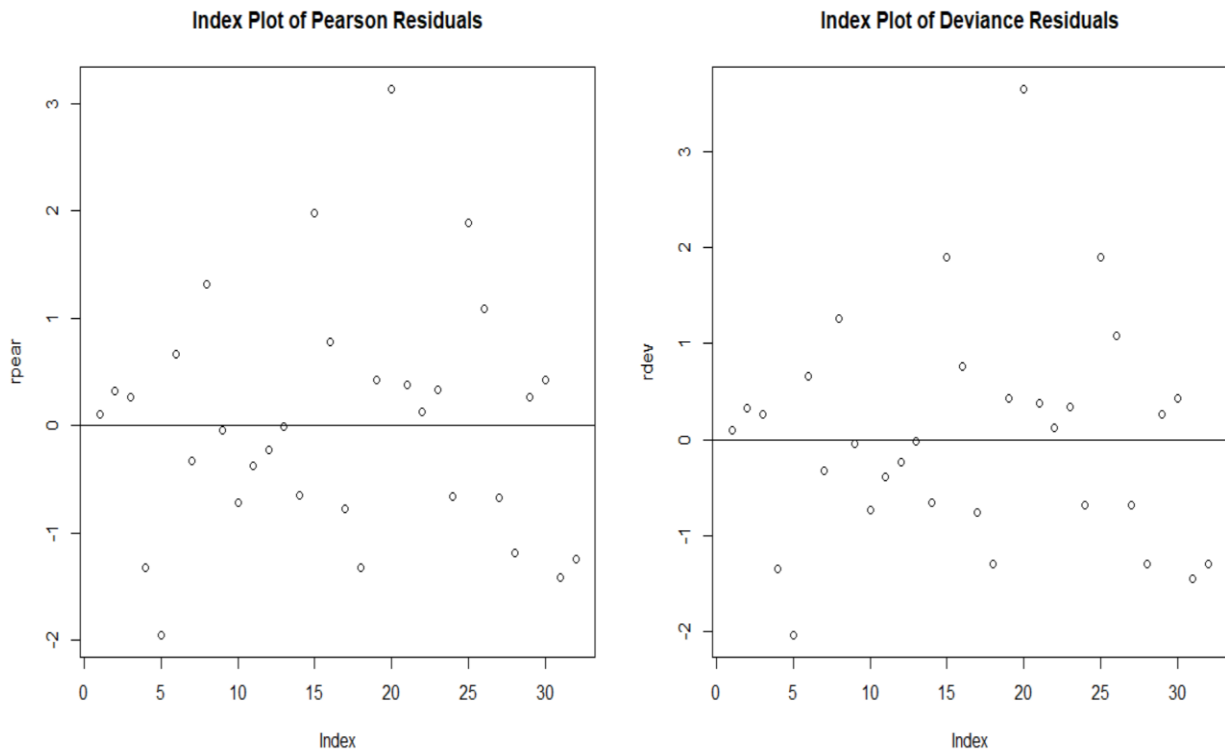
Diagnostics Details:

	Changed_Behaviour	Changed_behaviour_fitted	n	phat	h	rpear	rdev	D
1	11	10.818555	18	0.6010308	0.23368397	0.09976694	0.09991224	1.517625e-03
2	14	13.335256	23	0.5797937	0.25014911	0.32429114	0.32541715	1.754140e-02
3	11	10.463333	22	0.4756060	0.27169604	0.26846331	0.26830338	1.344344e-02
4	5	7.349538	17	0.4323257	0.24103430	-1.32035800	-1.34726920	2.768281e-01
5	4	7.676327	19	0.4040172	0.22407832	-1.95124416	-2.03708911	5.497626e-01
6	15	13.407021	35	0.3830577	0.30991150	0.66675949	0.66231229	9.982546e-02
7	8	8.695147	30	0.2898382	0.26824005	-0.32702081	-0.32972683	1.960089e-02
8	8	5.615167	22	0.2552348	0.21276215	1.31435914	1.26469911	2.334460e-01
9	10	10.088390	24	0.4203496	0.26497399	-0.04263435	-0.04265168	3.276347e-04
10	13	14.767079	37	0.3991102	0.32755657	-0.72340812	-0.72893610	1.274581e-01
11	8	8.813350	29	0.3039086	0.25700088	-0.38096013	-0.38445997	2.510009e-02
12	6	6.438238	24	0.2682599	0.24911480	-0.23300301	-0.23474264	9.005733e-03
13	11	11.030824	28	0.3939580	0.27907339	-0.01404061	-0.01404291	3.815658e-05
14	14	15.674131	42	0.3731936	0.32695109	-0.65104009	-0.65620244	1.029489e-01
15	15	10.407365	37	0.2812801	0.27679369	1.97460654	1.90482535	7.461483e-01
16	9	7.420279	30	0.2473426	0.25575232	0.77484157	0.75824332	1.031568e-01
17	6	7.137003	11	0.6488184	0.14858807	-0.77833722	-0.76444439	5.286276e-02
18	13	15.713825	25	0.6285530	0.27715063	-1.32119972	-1.30112861	3.346378e-01
19	7	6.319001	12	0.5265834	0.15974575	0.42953176	0.43087308	1.753796e-02
20	8	3.863491	8	0.4829364	0.12731524	3.13287702	3.65307278	7.159451e-01
21	7	6.355496	14	0.4539640	0.17717424	0.38140474	0.38059327	1.566156e-02
22	15	14.697958	34	0.4322929	0.31074323	0.12594645	0.12584712	3.575712e-03
23	8	7.338540	22	0.3335700	0.21642698	0.33789512	0.33551575	1.576761e-02
24	5	6.214343	21	0.2959211	0.23524973	-0.66385532	-0.67803097	6.778388e-02
25	12	8.472974	18	0.4707208	0.22346845	1.89003027	1.90028654	5.140018e-01
26	15	12.569406	28	0.4489073	0.27105678	1.08167289	1.07797907	2.175347e-01
27	7	8.369301	24	0.3487209	0.24609464	-0.67547959	-0.68519922	7.446973e-02
28	1	2.481263	8	0.3101578	0.09964092	-1.19319912	-1.29551223	7.878034e-02
29	13	12.420431	28	0.4435868	0.28371303	0.26049225	0.26012040	1.343853e-02
30	21	19.835324	47	0.4220282	0.34719002	0.42573427	0.42470545	4.819785e-02
31	11	14.593964	45	0.3243103	0.34563873	-1.41483117	-1.45076016	5.286693e-01
32	6	8.617682	30	0.2872561	0.28203138	-1.24652829	-1.29218176	3.051871e-01

Checking the Systematic Component:

R code:

```
plot(rpear, main="Index Plot of Pearson Residuals")
abline(h=0)
plot(rdev, main="Index Plot of Deviance Residuals")
abline(h=0)
```

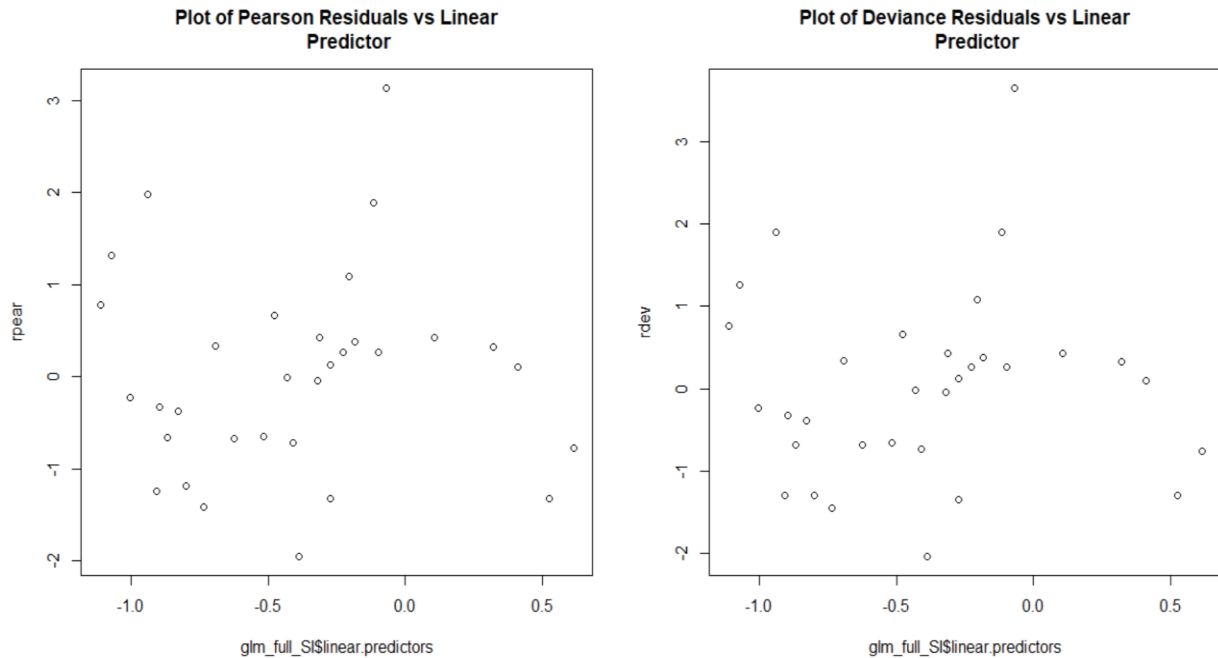


Conclusions from Index Plot:

- Both the plots seem to have a rectangular pattern, therefore, no problem with the systematic component
- Pearson and Deviance residuals plotting randomly about zero with constant range.
- There seems to be a possible case of an outlier which is plotting above the +3 range

R code:

```
plot(glm_full_SI$linear.predictors, rpear, main="Plot of Pearson  
Residuals vs Linear Predictor")  
plot(glm_full_SI$linear.predictors, rdev, main="Plot of Deviance  
Residuals vs Linear Predictor")
```



Conclusions from the Linear Predictor plot:

- Both the plots seem to have a rectangular pattern, therefore, no problem with the systematic component
- Pearson and Deviance residuals plotting randomly about zero with constant range.
- There seems to be a possible case of an outlier which is plotting above the +3 range

Checking for Outliers:

R code:

```
rpear <- residuals(glm_full_SI, "pearson")/sqrt(1-h)
rdev <- residuals(glm_full_SI, "deviance")/sqrt(1-h)
```

On observing the index plot and the linear predictor plots, it is seen that one point appears to be plotting above the +3 range.

Another test for outliers is based on the approximate standard normality of the residuals. Since it is an aggregated data we can assume that the residual deviance plot are approximately normally distributed and hence we can expect 95% of the residuals to lie within ± 1.96 or ± 2 . There is one extreme outlier (case 20) and 2 probable outliers (case 5 and case 15). These 2 cases need to be investigated further.

Checking for High Leverage Cases:

R code:

```
h<-lm.influence(glm_full_SI)$hat
```

A case to be of high leverage is considered if it exceeds $2(p+1)/n$ where p is the number of predictors in the model and n , the number of cases. In this a case, high leverage is considered if it exceeds $2(5)/10 = 10/32 = 0.3125$ value.

By looking at the diagnostics model on page 13, we conclude **case 10, case 14, case 30 and 31** to be case of high leverage.

Checking for High Influence Cases:

R code:

```
D = rpear*rpear*h/(2*(1-h))
```

A case of high influence is considered if its omission results in substantial changes to the fit of the model. By looking at the diagnostics model on page 13, it can be concluded that **case 15** is a case of high influence.

Conclusion from diagnostics test:

There are two possible outliers in the model one being **case 20**, as its residual value is above +3 range and another one being **case 15**, as it's on the border of residual value ± 1.96 , and is of high influence and almost high leverage.

Sparse Data:

There is no issue of sparse data on our dataset as the standard error values are small.

Overdispersion:

From the Model summary of the best model:

Mean Deviance = Residual Deviance/Degree of Freedom = $32.582/24 = 1.35$

Since the Mean Deviance is greater than 1, there is a case of over dispersion.

Over dispersed logistic regression model is fitted by specifying quasibinomial distribution.

R Code:

```
ovd <- glm(y ~ factor(Gender) + factor(Severe) + factor(Information) +  
factor(Age) + factor(Severe):factor(Information), family =  
quasibinomial(link = logit))
```

```
> summary(ovd)
```

Call:

```
glm(formula = y ~ factor(Gender) + factor(Severe) + factor(Information) +  
factor(Age) + factor(Severe):factor(Information), family = quasibinomial(  
link = logit))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7944	-0.5956	0.0378	0.3577	3.4126

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.40976	0.25628	1.599	0.12293
factor(Gender)2	0.20409	0.16329	1.250	0.22340
factor(Severe)2	-0.73110	0.25385	-2.880	0.00824 **
factor(Information)2	-0.79852	0.25382	-3.146	0.00438 **
factor(Age)2	-0.08784	0.22373	-0.393	0.69808
factor(Age)3	-0.50742	0.23739	-2.137	0.04295 *
factor(Age)4	-0.68213	0.26153	-2.608	0.01542 *
factor(Severe)2:factor(Information)2	0.68915	0.33297	2.070	0.04941 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 1.213141)

Null deviance: 66.061 on 31 degrees of freedom
Residual deviance: 32.582 on 24 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

From the summary of the over dispersion model:

- Dispersion Factor = 1.213141
- Coefficient of Standard Error is calculated as:

Coefficient of the binomial model * $\sqrt{\text{dispersion parameter}}$

Therefore, coefficient of the standard error for intercept in the quasibinomial model:

$$0.23268 * \sqrt{1.213141} = 0.25628$$

Similarly, the standard error coefficients are calculated for other parameters as well.

- Coefficient Estimates are same for both the models.
- Residual Deviance is same for both the models.

References:

<https://deepai.org/machine-learning-glossary-and-terms/statistical-classification>
https://en.wikipedia.org/wiki/Generalized_linear_model
https://en.wikipedia.org/wiki/Chi-squared_test
https://docs.oracle.com/cd/A91202_01/901_doc/olap.901/a86720/esdatao6.htm
https://en.wikipedia.org/wiki/Likelihood-ratio_test
https://en.wikipedia.org/wiki/Odds_ratio