Forecasting Long-Term Conditions

Technical Report

Contact details:

Bespoke Services – Consultancy Services
ISD - Public Health & Intelligence SBU
NHS National Services Scotland
rosalia.munoz-arroyo@nhs.net
Tel: 0131 2756619

Date: 1st June 2016
1. Introduction

This Forecasting Long Term Conditions project was designed to provide evidence on the potential future demand for healthcare contacts to NHS Scotland by examining existing information on a number of long-term conditions and forecasting the likely future changes in prevalence of these conditions.

Five long-term conditions were examined: Chronic Obstructive Pulmonary Disease (COPD), Heart failure, Hypertension and Type 1 and Type 2 Diabetes. We used existing national data to derive observed trends in prevalence over the past 15 years in relation to age, time period and cohort as a basis to forecast future prevalence.

The aim of this document is to provide technical information to aid interpretation of the results presented in the data explorer and projection output files. It also includes definitions, data sources and caveats related to data quality. Further details on the statistical methods used in forecasting and an example of more specific results from the modelling process for the COPD condition can be also found in this report.

2. Data sources and data quality

2.1 Data sources

Population estimates and projections (National Records of Scotland, NRS)

NRS produces mid-year population estimates for a fixed point in time (30th June). The methodology used to produce the population projections involves assumptions about migration, fertility and mortality and are based on analysis of past trends (see NRS website for further information\(^1\)).

Mortality Information (National Records of Scotland, NRS)

The Statutory Register of Deaths contains detailed information about each person who has died in Scotland since 1st January 1855. Registering deaths with the Statutory Register of Deaths is compulsory, so the information is complete\(^2\).

Diabetes (Scottish Diabetes Survey)

The Scottish Diabetes Survey is the most complete routine data source for diabetes in Scotland. Based on annual returns from Scottish Health Boards, it combines information from both primary and secondary care settings. The reports present a wide range of information on patients registered as having diabetes in Scotland, at both national and health board levels. This includes the number of registered patients, information on the management and health status of registered patients, and an

---


\(^2\) [http://www.nrscotland.gov.uk/research/guides/statutory-registers/deaths](http://www.nrscotland.gov.uk/research/guides/statutory-registers/deaths)
assessment of the completeness of the data. The most recent published survey is the 2013 Scottish Diabetes Survey.

COPD, Heart Failure and Hypertension (Secondary Care Data, SMR01 Linked Catalog)

ISD Scotland’s linked database combines the Scottish Cancer Registry and National Records of Scotland Death Records with Acute (SMR01) and Psychiatric (SMR04) hospital admissions. This linked dataset matches all individual patient-level hospital episodes from 1981 to the present day. It is an effective data source on which to derive the prevalence of conditions where hospital admission is likely to result. Notice that this may be an appropriate assumption for heart failure, however not all cases of COPD or Hypertension will result in hospitalisation.

2.2 Data quality

Prevalence Predictions

The prediction results are trend based and assume that past trends in disease prevalence will continue into the future. They take no account of the possible effect of future health interventions or changes to socioeconomic factors that may influence disease risk.

Populations

Information about the quality of the population estimates can be found on the NRS website.

NRS draw attention to the limitations of their population projections stating ‘a projection is a calculation showing what happens if particular assumptions are made. The population projections are trend-based. Many social and economic factors influence population change, including policies adopted by both central and local government. The relationships between various factors are complex and largely unknown. While future policy changes are not taken into account projections will reflect the impact of past policy and economic changes’. For further information on the quality of the population projections please see the NRS website.

Diabetes

The diabetes register is not considered complete (>95% coverage) until 2004, so statistics from earlier years possibly underestimate the incidence of the disease and as such should be treated with caution.

---

As it was necessary to use five-year time periods in the projections, we used earlier data (1999-2003) in the estimation models as a minimum of three time periods were required for estimation. This probably resulted in an exaggeration of the apparent rate of increase in prevalence in time as some of the increases in earlier years would be due to the register being brought up to date. However, as only the last 10 years of data were used in projecting the trend in time into the future, the impact of this on the projections should be minimised. There is no evidence to support that the incompleteness of the data was biased towards any particular age group, so age specific prevalence estimates should not be affected.

**Diabetes Type 1**

The number of cases of Type 1 Diabetes was relatively small compared to other conditions. There is a lot of inter-annual variation in incidence even when all ages are considered, this is not surprising as only approximately 350-450 cases are diagnosed annually in females and 500-650 cases in males in the last 10 years. Prevalence rates in 5-year age groups and time periods also show a lot of variation between age groups rather than the smooth curve with age which might be expected in a larger population.

**COPD**

COPD is thought to be under-diagnosed, even in a primary care setting, which would be expected to record more cases than hospitalisation records as it will also capture less severe cases that do not ever result in a hospital admission ([ISD Quality and Outcomes framework](http://www.isdscotland.org/Health-Topics/General-Practice/Publications/2015-10-13/2015-10-13-QOF-Report.pdf)). Therefore SMR01 as a source will underestimate the prevalent cases.

The estimated prevalence of COPD recorded from a Primary care setting in 2013-14 was 2.2% ([ISD Quality and outcomes framework](http://www.isdscotland.org/Health-Topics/General-Practice/Publications/2015-10-13/2015-10-13-QOF-Report.pdf)). This compares to estimates of 1.7% in 2013 and 1.8% in 2014 in the total population based on hospitalisation data. Secondary care data are consistently 20% lower than estimates from general practice records over the last 10 years.

**Heart Failure**

SMR01 is likely to be a good source of data for heart failure, as it is very likely to result in hospital admission.

**Hypertension**

We considered only primary hypertension in this analysis (i.e. not caused by another condition such as heart or kidney disease). Prevalence of Hypertension will almost certainly be under estimated using SMR01, as it is unlikely to directly result in a

---


hospital admission unless it becomes severe (hypertensive crisis). It may, however, be recorded on admissions primarily related to another condition, either as a co-morbidity or in cases where it is a complication of the main condition.

Estimates from primary health care data put the prevalence of Hypertension at 13.9% in 2013/14 (ISD Quality and outcomes framework⁹). Although the QoF register includes both primary and secondary hypertension (i.e. the higher prevalence rate is partly due to the inclusion of secondary hypertension), this is more than double the prevalence derived from SMR01 for the most closely comparable time period (6.5% prevalence in the total population in 2014).

2.3 ICD 9 and ICD10 codes used to select SMR01 records

Table 1: ICD 9 and ICD 10 codes used to select conditions (COPD, Hypertension and Heart Failure)

<table>
<thead>
<tr>
<th>ICD-10 codes</th>
<th>ICD-9 code</th>
<th>ICD short description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>490</td>
<td>Bronchitis, not specified as acute or chronic</td>
<td></td>
</tr>
<tr>
<td>491</td>
<td>Chronic bronchitis</td>
<td></td>
</tr>
<tr>
<td>492.9</td>
<td>Emphysema</td>
<td></td>
</tr>
<tr>
<td>494</td>
<td>Bronchiectasis</td>
<td></td>
</tr>
<tr>
<td>496</td>
<td>Chronic airways obstruction, not elsewhere classified</td>
<td></td>
</tr>
<tr>
<td>J40</td>
<td>Chronic lower respiratory diseases</td>
<td></td>
</tr>
<tr>
<td>J41</td>
<td>Simple and mucopurulent chronic bronchitis</td>
<td></td>
</tr>
<tr>
<td>J42</td>
<td>Unspecified chronic bronchitis</td>
<td></td>
</tr>
<tr>
<td>J43</td>
<td>Emphysema</td>
<td></td>
</tr>
<tr>
<td>J44</td>
<td>Other chronic obstructive pulmonary disease</td>
<td></td>
</tr>
<tr>
<td>J47</td>
<td>Bronchiectasis</td>
<td></td>
</tr>
<tr>
<td>Heart Failure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I50</td>
<td>428</td>
<td>Heart failure</td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I10</td>
<td>401</td>
<td>Hypertension</td>
</tr>
</tbody>
</table>

3. Definitions and measures

The following definitions and measures are used in this study. Refer to Section 2 for further details on Data Sources.

**Incidence**
Incidence refers to the number of new cases of a specific disease occurring during a certain period in a population at risk. In this study, incidence of COPD, Hypertension and Heart Failure is determined if a relevant ICD code is present in any diagnostic position and is the first recorded diagnosis of this condition for the individual (since 1981, from when SMR01 information is available).

For Diabetes Type 1 and 2 the incidence was determined using the date of diagnosis from the Scottish Diabetes Survey.

**Prevalence**
Prevalence of a condition is the estimated number of cases of a condition in the population at a given point in time. For COPD, hypertension and heart failure, prevalence was calculated using a 20-year look-back in SMR01. As the records are only available from 1981 reliable prevalence estimates were only available from 2000 onwards. In this study we have calculated annual end-of-year prevalence by taking the previous year’s prevalence, adding new incidences in that year and subtracting the mortality of individuals known to have the disease. As the conditions in this study are classified as long term conditions it is assumed that once an individual has the condition that it is life-long.

For Diabetes Type 1 and 2 the prevalence was extracted from the Scottish Diabetes Survey.

**Mortality**
Mortality records were selected if any of the LTC’s were recorded as a cause of death in any of the cause positions on the death record (underlying and secondary causes).

**Populations**
Populations and projections are all mid-year estimates.

**Projected counts**
The projected prevalence rates are for five-year periods, therefore projected counts are the projected average annual number of expected prevalent cases in the five year period. The five-year projected prevalence rate is multiplied by the average of the population projections over the five-year period.

**Crude rates**
The crude rates are calculated by dividing the total number of incident and prevalent cases, and deaths in a given time period by the total number of persons in the population at that time. The crude rates in this study are expressed per 10,000 population. As there were clear differences between male and female incidence, mortality and prevalence rates for each LTC the crude rates were calculated separately for males and females.
**Age-specific rates**

The age-specific rates are calculated by dividing the total number of incident and prevalent cases, and death in a given age group in a given time period by the total number of relevant persons in the population at that time. The age-specific rates in this study are expressed per 10,000 population. As there were clear differences between male and female rates, the age-specific rates were calculated separately for males and females.

**European Age Standardised Rates (EASR)**

Directly age standardised rates have been calculated to adjust the overall rate, which may vary with the age structure of the populations. The direct standardisation method was used, with the age-sex specific rates applied to the age structure of a standard population (in this case the European standard population for 2013). This gives the overall rate that would have occurred in the estimated and projected Scottish population (for males and females) if it had the same age profile as the standard population. It allows valid comparisons to be made between time periods with differing population age structures. In this study, age standardised rates are expressed per 10,000 population per year. Male and female age standardised rates are calculated separately.
4. Method used to calculate projections

Producing the projections of prevalence of the four long-term conditions included three main stages:

1. Data exploration
2. Fitting a model to the observed data and evaluating the best fit model
3. Generating the projections based on the fitted model

4.1. Data exploration

The first step in producing the projections was creating data exploration spreadsheets for each of the long-term conditions of interest. The aim of the data exploration stage is to assess how much variation in the incidence, prevalence and mortality rates are affected by age, period and cohort.

The data was aggregated into five-year time periods. This is a common practice for the predicted counts or rates of the disease in order to smooth out random year to year variation in the numbers. Similarly, rates are often grouped by age group rather than age in years. We grouped the data by 5-year age bands and 5-year periods. Cohort is also defined by 5-years intervals, derived by subtracting age from period.

Producing descriptive analyses in a graphical format has allowed us to identify trends, patterns, outliers or discontinuities in the data in relation to age, period and cohort. The following graphs were produced for incidence, prevalence and mortality rates:

1. Rates for the whole population by year (split by males and females)
2. Rates for each age group by period (observations within each period connected)
3. Rates versus age (observations within each birth cohort connected)
4. Rates versus period (observations within each age-class connected)
5. Rates versus cohort (observations within each age-class connected)

The purpose of producing the graphs is to assess the following effects:

- **Age effects** are variations in the rate of disease across age groups. This is linked to the biological and social processes of ageing specific to individuals. They are unrelated to the time period or birth cohort to which an individual belongs.
- **Period effects** result from external factors that equally affect all age groups at a particular calendar time. If there were some similar change in disease risk for all individuals alive at a particular point of time or changes in treatment regardless of age this would result in an increase or decrease in the prevalence rate across all age groups that could be assigned to a period effect.
- **Cohort effects** are variation related to changes across groups of individuals who were born in the same year or years. Different birth cohorts may have different exposure to risk factors which may produce a change in disease incidence for individuals born at particular time and this would be described as a cohort effect.

Appendix A details the questions and criteria relevant to the APC model components that were checked and recorded for the COPD condition as an example.
4.2 Fitting the APC model and model selection

Age Period Cohort models (APC) provide a useful method for modelling incidence, prevalence and mortality rates. The aim of the analysis is to describe and estimate the independent effects of age, period and cohort on the health outcome of interest. However, APC models suffer from an ‘identifiability problem’, which is that age, period, and cohort are all exactly related. Given the calendar year and age of a subject, one can determine the cohort (birth year):

\[ \text{Cohort} = \text{Period} - \text{Time} \]

If fitted directly in a generalized linear model (GLM) this leads to overparameterization and, consequently, the exclusion of one of the terms. It is therefore necessary to fit constraints to the model to extract identifiable answers for each of the parameters and this is what the NORPRED approach in R does (see below for further details). The consequence of the identification problem for the modelling used in this study is that one category of cohort (the youngest cohort) and one category of period (the last period) cannot be estimated. However, this does not have an impact on the overall results but limits our ability to determine how much variation in rates is due to each of the different components (age, period and cohort).

4.2.1 Age Period Cohort estimation method: NORDPRED

There are several means of estimating age, period and cohort effects. The approach taken in this project was based on using the NORDPRED package in the statistical program R. The package was written by Harald Fekjær and Bjørn Møller at the Cancer Registry of Norway\(^\text{10}\) as part of a larger cancer prediction project. NORDPRED was developed for studying cancer incidence, but (as it is a version of APC modelling) can be used to model the incidence or prevalence of other conditions provided that sufficiently high quality records and population data are available. NORDPRED uses observed incidence or prevalence numbers and population data to estimate and predict the incidence or prevalence rates for individual age groups or the whole population in a particular time period (Møller et al. 2002, Møller et al. 2003). NORDPRED fits a Poisson regression APC model with age, period and cohort as factors and a linear trend term for period.

The standard NORDPRED program can only fit a full APC model. We altered the program to allow it to fit alternative models (age only, age-period, and age-cohort), as fitting only one or two of these terms might be adequate to estimate the prevalence of a disease (Dyba 2000). R codes for the modified version used in this study are available on request. The original version of NORDPRED is available to download \(\text{here}\)\(^\text{11}\). Additional details of the use of the software can be found in R documentation for nordpred\(^\text{12}\).

---

\(^{10}\) https://www.kreftregisteret.no/en/

\(^{11}\) https://www.kreftregisteret.no/en/Research/Projects/Nordpred/Nordpred-software/#Download the Nordpred package

\(^{12}\) https://www.kreftregisteret.no/globalassets/gammelt/dokumenter/nordpred/nordpred-docpages.pdf
4.2.2 Selecting the model type
We compared the APC model with age-cohort (AC) and age-period (AP) models. The fit of the three models were compared using the following criteria:

- Akaike Information Criteria (AIC) where models with lower scores indicate a better fit
- The residual deviance where models with lower values indicate a better fit
- Consideration of the residuals plots checking for points for which the model is a poor fit and points that exert undue influence on the model (see Appendix B)

Poisson regression normally uses a log-link (a data transformation) to connect the response and explanatory variables, where the coefficient needs to be exponentiated in order to obtain the estimated parameter. However, NORDPRED also has an option to use a power link function, where the estimated parameters must be raised to the power of 5 to find the estimated response value. In this study, the choice of whether to use a log or power 5 link functions was made based on comparing the estimated model fit.

4.2.3 Selecting age groups to model
As some prevalence rates were very low and fairly constant in time in some age groups (such as the very young and very old) not all models included all age groups. For each condition we chose a minimum starting age to be included in the estimation model based on decisions made after the data exploration stage. For most conditions prevalence remained low in the youngest age groups and at some point began to rise more rapidly. We selected the age at which this more rapid increase began as the starting age. Type 1 Diabetes is an exception as it has a different prevalence pattern from the other LTCs in this study. The prevalence is very low in the youngest and oldest age groups, therefore both the youngest and oldest age groups were excluded from the estimation model.

4.3 Generating projections
NORDPRED uses the estimated model parameters to project future trends. For age groups not included in estimates, the future prevalence rate is determined by the average of the rate in the previous two periods.

NORDPRED has two options that need to be considered for how the linear period component is used to project:

4.3.1 Length of prediction base
If the trend in prevalence rates has changed in recent observed periods (as evident in the data exploration stage), it is possible to only use the most recent trend, as this is more likely to be relevant to future projections.

A decision on whether to use the recent or whole observed trend in projecting the prevalence rate of each condition was made based on testing whether the model fitted better if the linear period trend was replaced with a curved trend (the square of the period). If the curved trend was a better fit, then this indicates that there has been a significant change in the prevalence rates over the observed time period, and in this case the recent trend was used in projections.
4.3.2 Cutting the trend over time
The second option is whether to project the entire trend into the future indefinitely or whether to reduce the amount of the trend that is projected in later periods. It is not normally realistic to expect a linear trend (whether increasing or decreasing) to continue indefinitely into the future, in reality it will level off at some point. However, how much to reduce the trend into the future and deciding which prediction scenario is most likely requires expert advice. For all models where period trend was predicted, we have presented two options:

1. The first ‘Full trend’ projects the whole trend.
2. The second (cut trend) projects the full trend for the first projected period, 75% of the trend in the second projected period and 50% of the trend in the third projected period. The effect of the cut trend option is to reduce the steepness of the increasing (or decreasing) trend in time.
References


Harald Fekjær and Bjørn Møller (Cancer Registry of Norway), 2006, nordpred package, https://www.kreftregisteret.no


<table>
<thead>
<tr>
<th>Assumption / issue</th>
<th>Charts / checks</th>
<th>Results and comments</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Should males and females be modelled separately?</td>
<td>Compare all rates and all effect components by gender (all charts)</td>
<td>Female incidence, prevalence and mortality rates are consistently lower than males. Period effects are slightly different; decrease in rates for females and an increase for males.</td>
<td>Males and females should be modelled separately</td>
</tr>
<tr>
<td><strong>Age effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Should all age groups be included in the model? Are incidence rates very similar to the adjacent age groups (0-4, 5-9 etc)?</td>
<td>Check period by age charts</td>
<td>All rates very low until age 45 for males and females. Linear trend from age 45+</td>
<td>Check the residuals in the model and consider using constant rate for younger age groups</td>
</tr>
<tr>
<td>Are the period trends similar across the age groups? Should different slopes be fitted for different age groups?</td>
<td>Check age by period charts</td>
<td>The prevalence of heart failure fell in females in all age groups over 60 between the period 2000-2004 and 2005-2009, but increased in all age groups under 60. Prevalence rates rose in most age groups in males between the period 2000-2004 and 2005-2009, although declined in age groups 55-59, 60-64 and 85+.</td>
<td>Check the residuals in the model and consider whether including the cohort effect alleviates this issue.</td>
</tr>
<tr>
<td><strong>Period effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How long should the prediction base be/how many observed years should be included?</td>
<td>Fit GLM models and check the overall fit of the model. Fit the model using more recent observed years and consider if the model is a better fit.</td>
<td>Trends relatively stable over time. As many data points as possible should be included.</td>
<td>Use all available data</td>
</tr>
<tr>
<td><strong>Cohort</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is there a cohort effect?</td>
<td>Produce the Epi plots to check for cohort effects</td>
<td></td>
<td>Cohort effect should be included in the model</td>
</tr>
</tbody>
</table>
Appendix B
Example of R Outputs, Males COPD – Estimation Models

This appendix gives an example of the R outputs that were produced for the chosen estimation model for males with COPD. The model estimated parameters, AIC score and diagnostic plots are shown.

Males COPD model outputs

An age-cohort (AC) model was fitted to age groups 40 years and over. Only age and cohort effects were fitted in the model. Period was found to be non-significant. The model was fitted with a power 5 link function.

Table A1: Estimation model for Males with COPD

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 40-44</td>
<td>0.267</td>
<td>0.0022</td>
<td>121.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 45-49</td>
<td>0.310</td>
<td>0.0020</td>
<td>153.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 50-54</td>
<td>0.360</td>
<td>0.0019</td>
<td>190.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 55-59</td>
<td>0.413</td>
<td>0.0018</td>
<td>231.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 60-64</td>
<td>0.468</td>
<td>0.0017</td>
<td>276.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 65-69</td>
<td>0.520</td>
<td>0.0016</td>
<td>320.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 70-74</td>
<td>0.563</td>
<td>0.0016</td>
<td>362.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 75-79</td>
<td>0.600</td>
<td>0.0015</td>
<td>409.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 80-84</td>
<td>0.628</td>
<td>0.0014</td>
<td>444.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age 85 and over</td>
<td>0.662</td>
<td>0.0011</td>
<td>598.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1915-1919  (Baseline)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Cohort 1920-1924</td>
<td>-0.008</td>
<td>0.0014</td>
<td>-5.90</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1925-1929</td>
<td>-0.004</td>
<td>0.0014</td>
<td>-2.92</td>
<td>0.004</td>
</tr>
<tr>
<td>Cohort 1930-1934</td>
<td>0.002</td>
<td>0.0015</td>
<td>1.55</td>
<td>0.120</td>
</tr>
<tr>
<td>Cohort 1935-1939</td>
<td>0.006</td>
<td>0.0015</td>
<td>4.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1940-1944</td>
<td>0.017</td>
<td>0.0016</td>
<td>10.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1945-1949</td>
<td>0.015</td>
<td>0.0017</td>
<td>8.65</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1950-1954</td>
<td>0.018</td>
<td>0.0018</td>
<td>10.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1955-1959</td>
<td>0.027</td>
<td>0.0019</td>
<td>14.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1960-1964</td>
<td>0.034</td>
<td>0.0020</td>
<td>17.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1965-1969</td>
<td>0.042</td>
<td>0.0022</td>
<td>19.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cohort 1970-1974</td>
<td>0.050</td>
<td>0.0025</td>
<td>20.2</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Null deviance: 4.57x 10^7 on 30 degrees of freedom
Residual deviance: 30.5 on 9 degrees of freedom
AIC: 412.42
Diagnostic plots

Residuals vs Fitted

Normal Q-Q

Scale-Location

Residuals vs Leverage