Society of Actuaries in Ireland

How Powerful are your Rating Factors?
Chris Reynolds and Andrei Halasz

2012

Disclaimer

The following presentation is for general information, education and discussion purposes only.

Views or opinions expressed, whether oral or in writing do not constitute legal or professional advice.

PartnerRe

How do factors interact?

Gender/ Age

Distribution

BANK

Smoker

Units

Mortgage

BMI/WHR

Postcode

PartnerRe

Source: istockphoto

Linear Regression

Random Structure

Responses vary even for constant values of the predictor

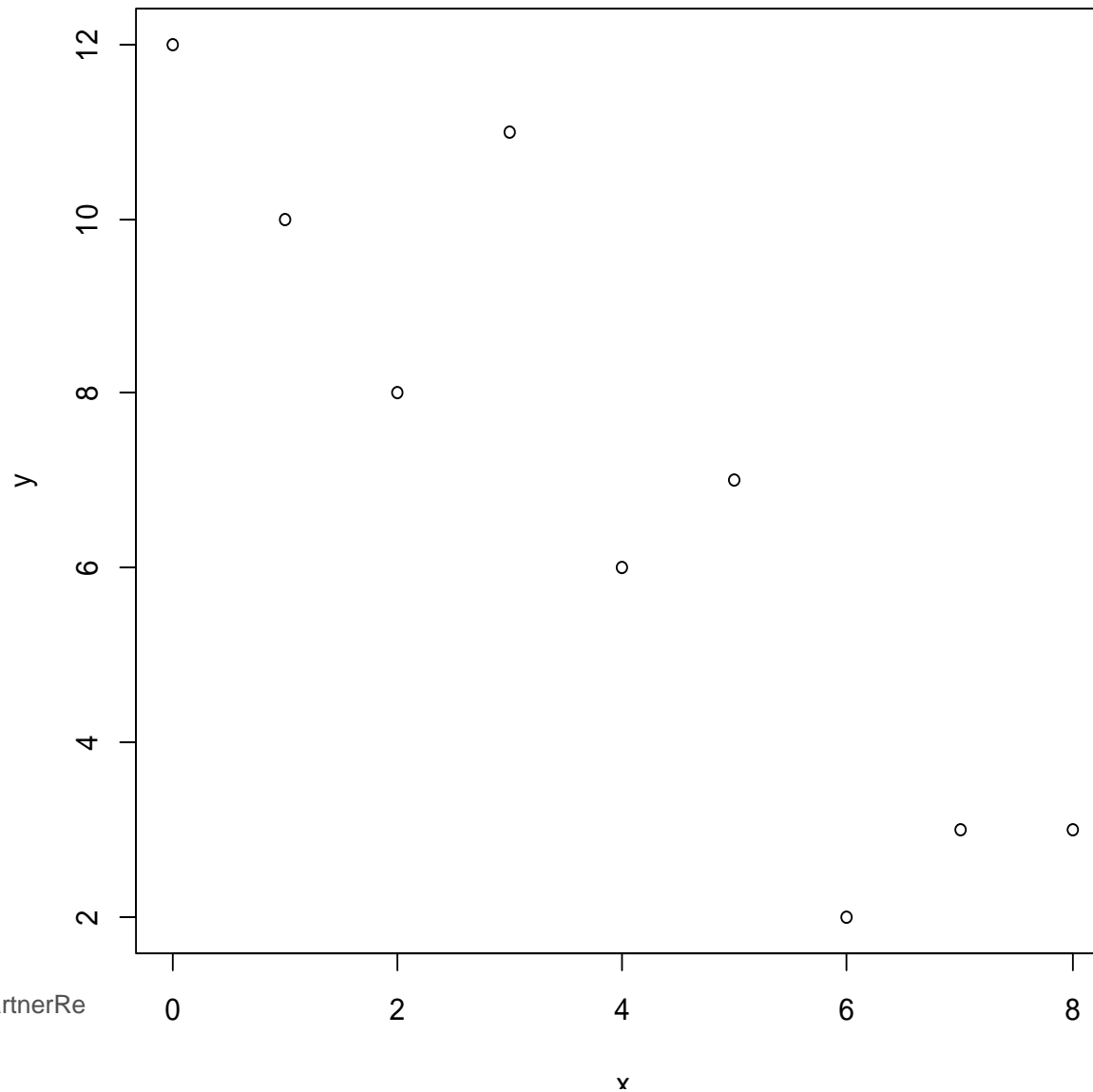$$Y_i \sim N(\mu_i, \sigma^2)$$

Systematic Structure

The simplest way to express the dependence of the response μ_i on the predictor x_i is to assume a linear function
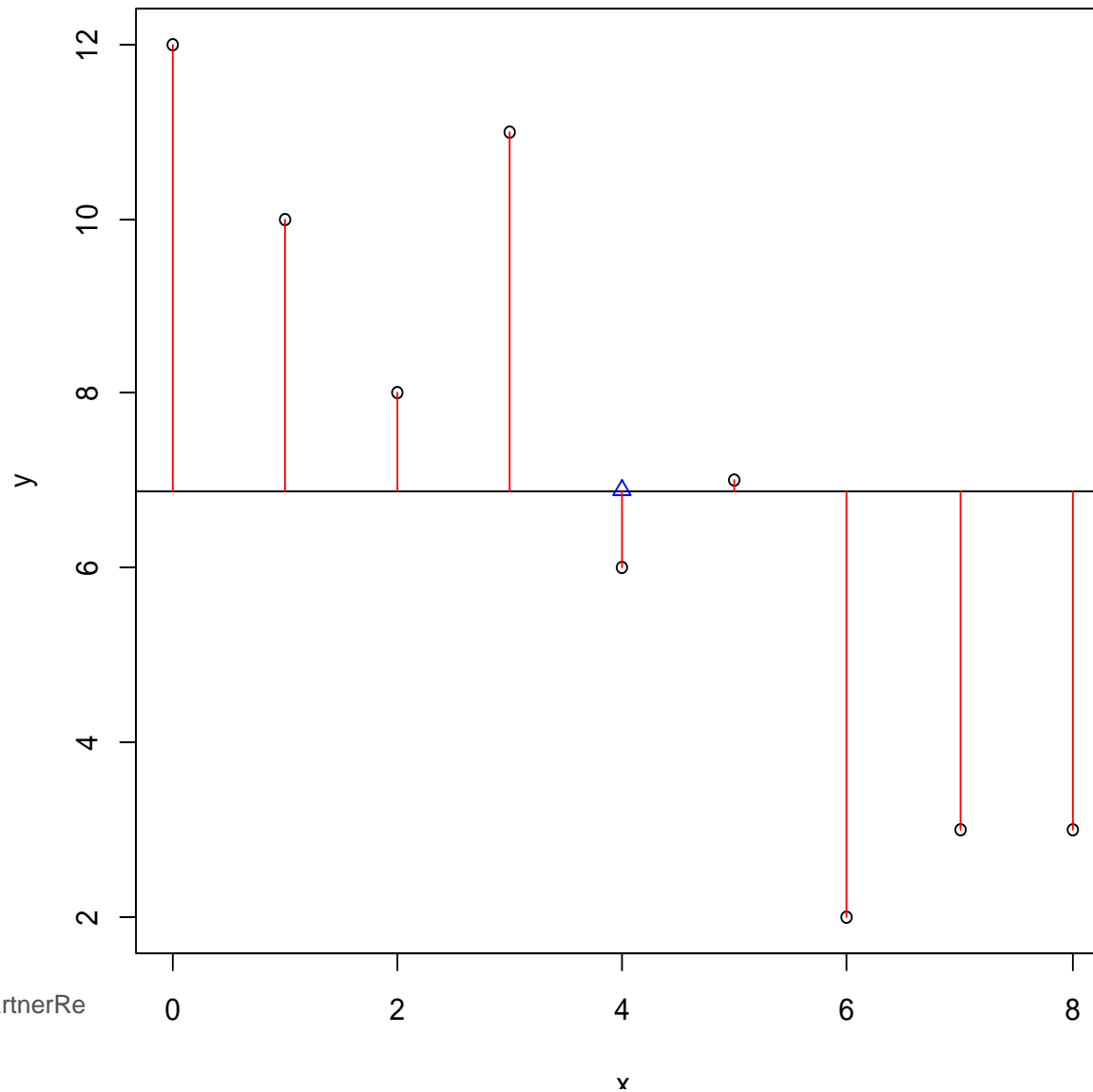
$$\mu_i = \alpha + \beta x_i$$

PartnerRe

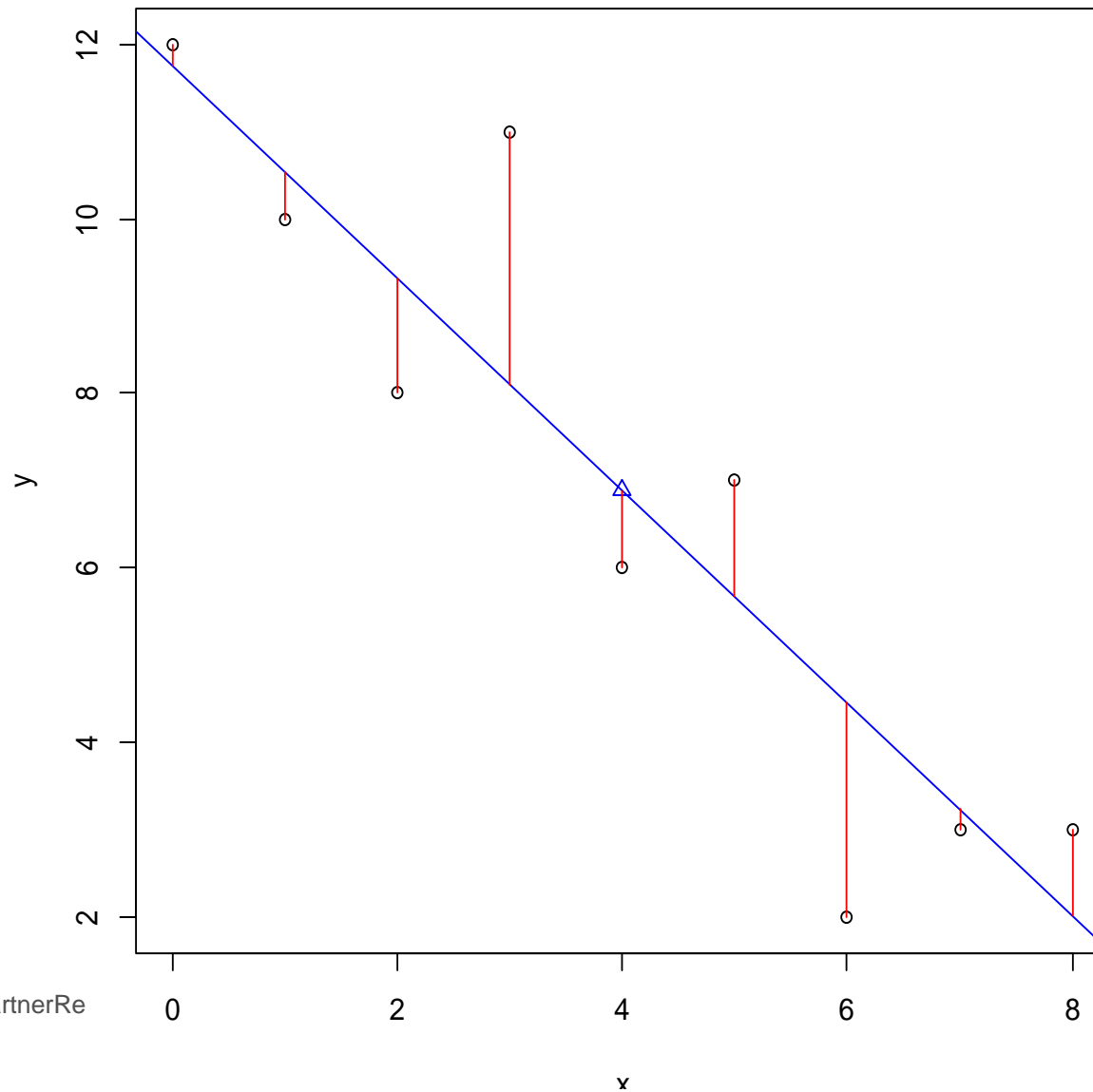# Linear Regression

Source: PartnerRe

Linear Regression

Source: PartnerRe

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

7

Linear Regression

Source: PartnerRe

PartnerRe

Linear Regression
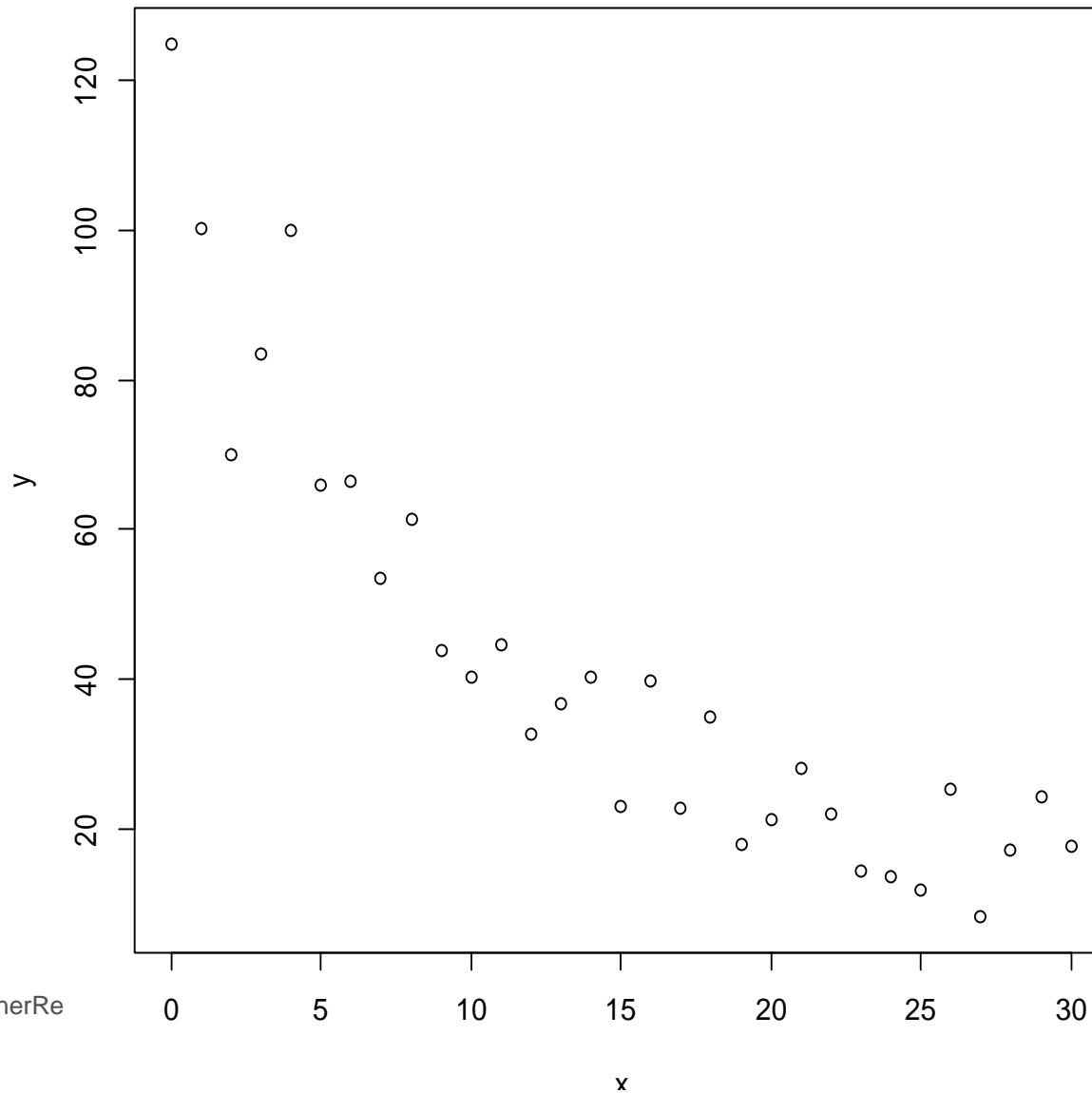
Source: PartnerRe

Linear Regression - Limitations

Source: PartnerRe

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

10

Linear Regression - Limitations

Source: PartnerRe

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

11
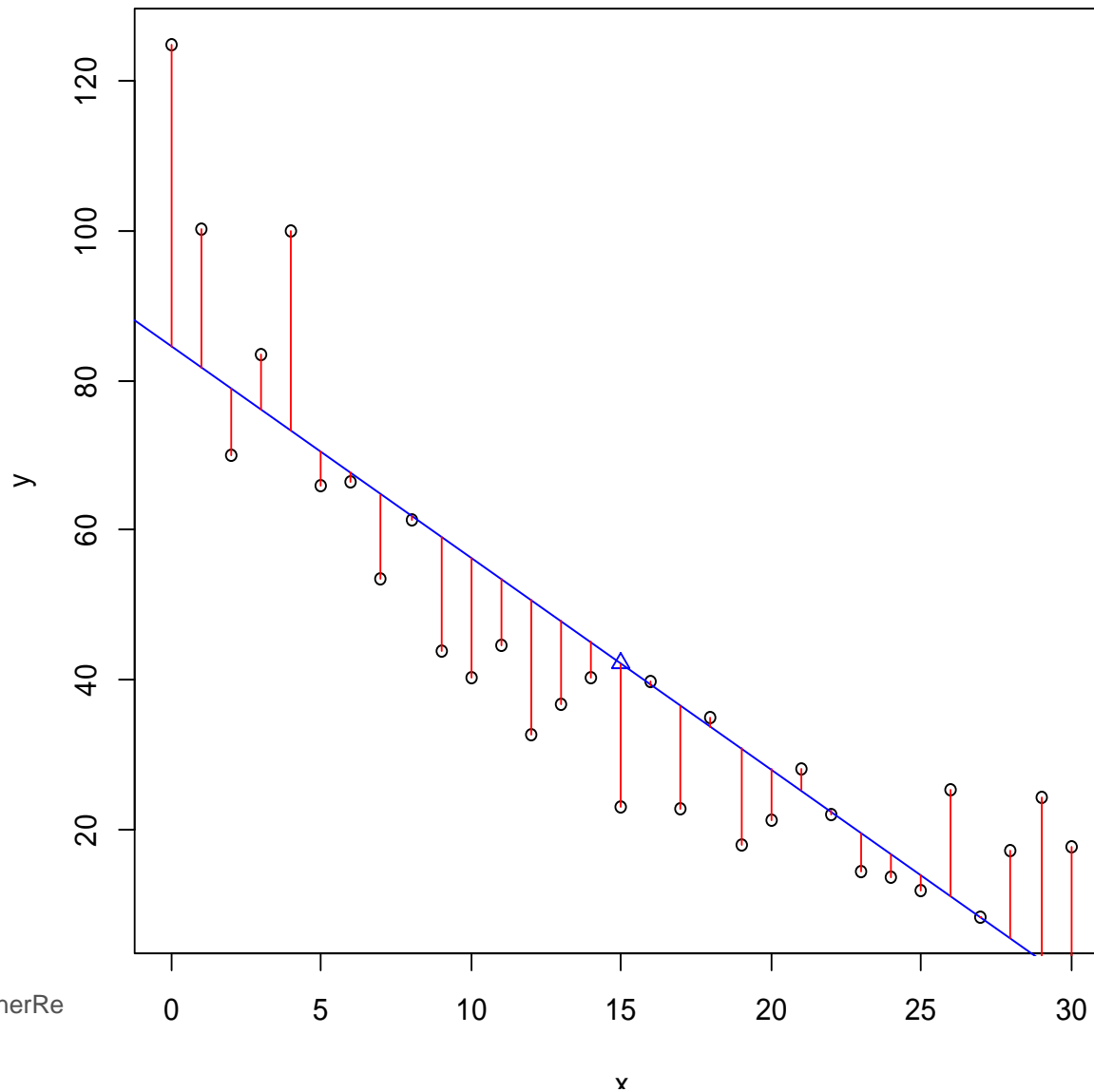
Linear Regression - Limitations

Residuals vs Fitted
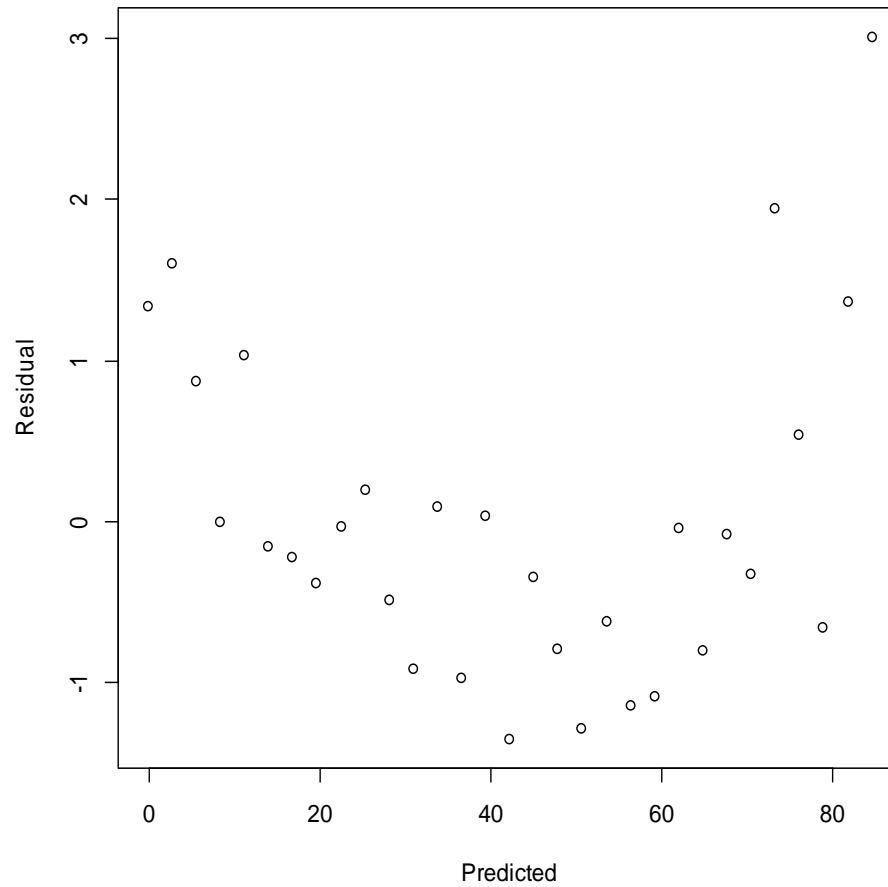
Normal Q-Q

Source: PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

12
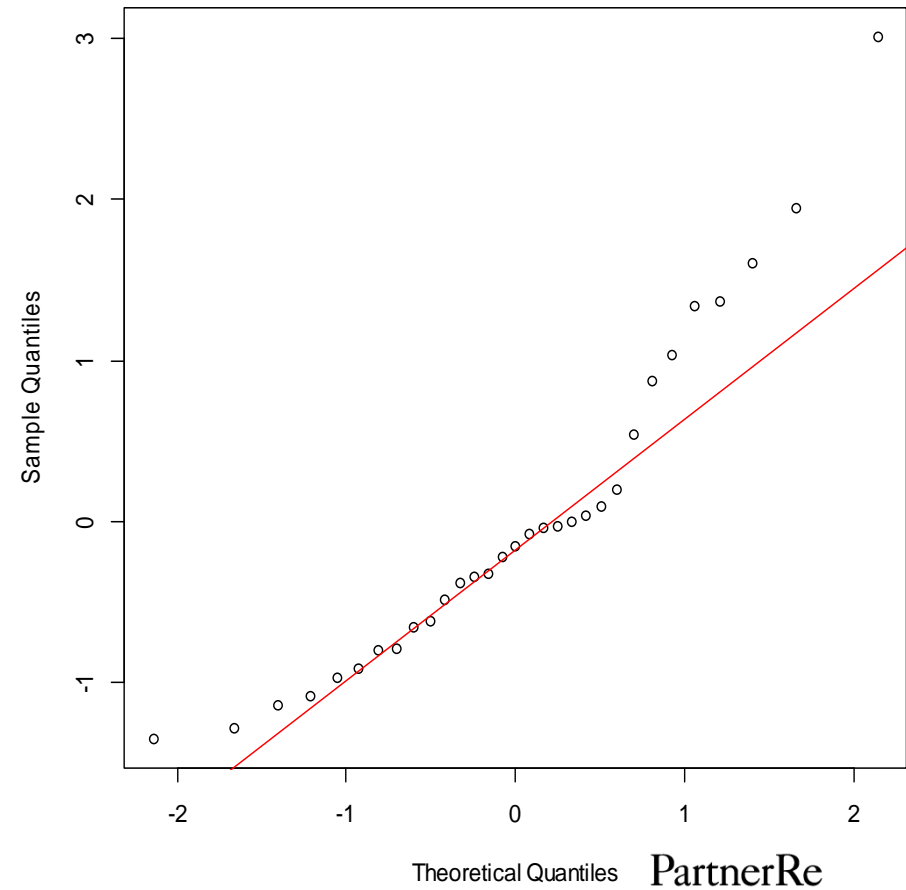
Linear Regression - Limitations

1 The relationship between the response and the predictor may not be linear;

2 A normal distribution for the response may be inappropriate;

3 The variance will often increase linearly with the mean, so a constant variance assumption may be inappropriate.

What do we do???

We **generalise** the model framework.

PartnerRe

The 3 part GLM Recipe

1 Random Component
Identify the response variable Y and assume a
probability distribution for it

2 Systematic Component
Specify what the explanatory variables X are. This
gives the linear component $\alpha + \beta X_i$

3 Link
Specify the relationship g between the mean E(Y) and
the systematic component X:

$$g\left(\mathrm{E}[Y_i]\right) = \alpha + \beta X_i$$

PartnerRe

Poisson Regression

Number of Deaths: $Y_i \sim Po(n_i \mu_i) \Rightarrow E(Y_i) = n_i \mu_i$

Gompertz: $\mu_i = e^{x_i \beta}$

GLM: $\log[E(Y_i)] = \log n_i + x_i \beta$

This is called the "offset"

PartnerRe

R Software – What is it?

"R is an open-source, object-oriented statistical programming language. In the past decade, it has become the global lingua franca of statistics"

www.r-project.org

PartnerRe

R – Should a company trust free software?
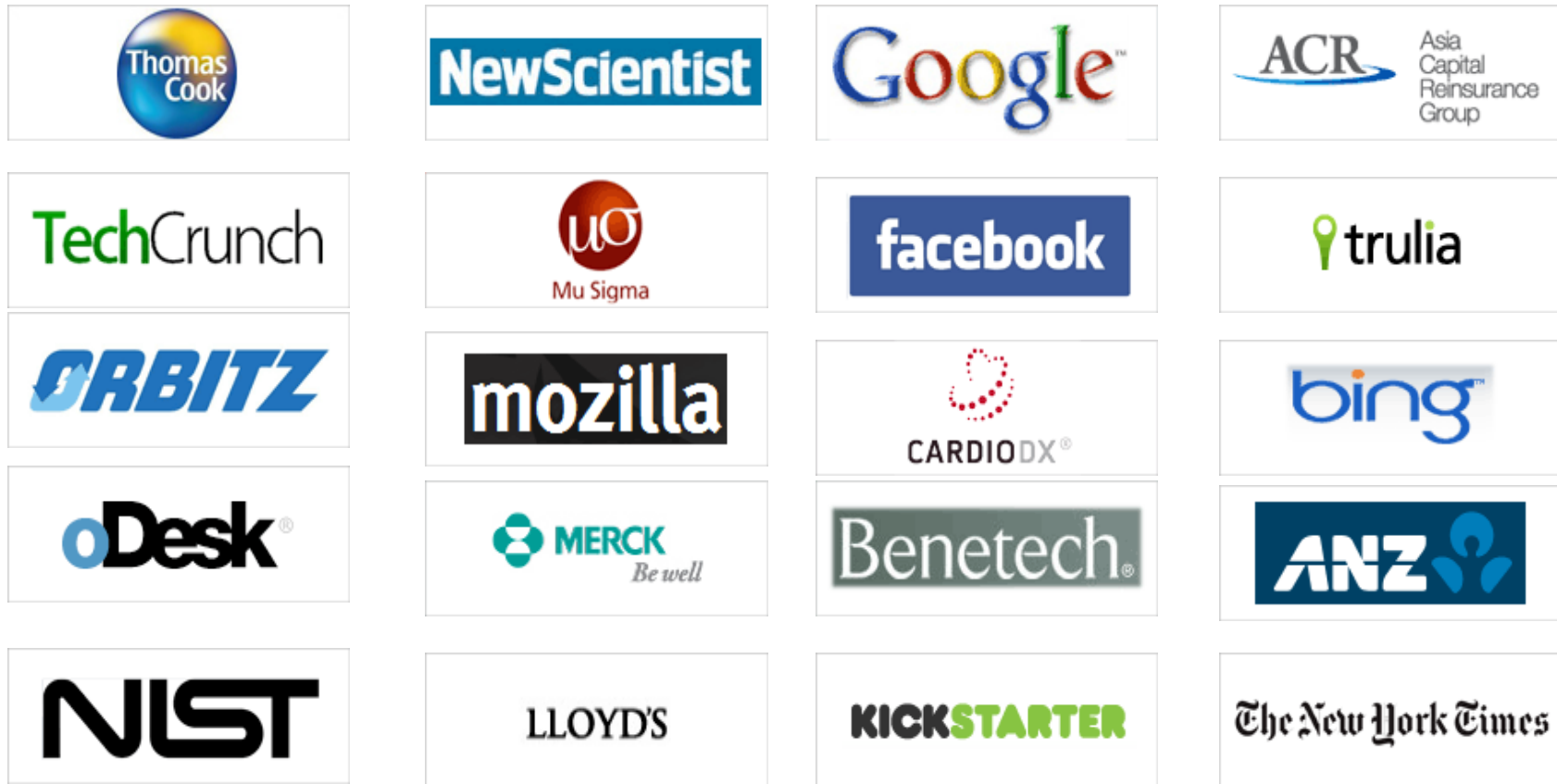
These companies do*

PartnerRe

*Source : http://www.revolutionanalytics.com/what-is-open-source-r/companies-using-r.php

How Powerful are your Rating Factors?
December 3, 2012

17

Real time demonstration

Using HMD* mortality data

Years 1960 onwards

Ages 20 – 65

UK, Ireland, Poland & Russia

* Human Mortality Database (www.mortality.org)

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

18

Creating a GLM model - start simple & build vs start with saturated and strip down

- **One approach might be to start simple and build up**

- **But it's often easier in the GLM world to start with all available rating factors**
  - strip out factors (or combinations of factors) that don't add much value

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

19

Term Assurance data - Our dataset

- **Several companies**

- **Many thousands of claims**

- **All UK, "normal" term assurance**

- **Chosen datasets with rich rating factors**
  - e.g. rated status, joint v single, distribution channel, underwriting year, sum assured size, etc.

PartnerRe

Model – Investigation 1

Take a grouped dataset split by

- Age Bands of 5
- Sex
- Smoker
- Duration

PartnerRe

Model – Investigation 1

Fit an AGE * SEX * SMK * DUR model

Use Exposure as the "offset"

Analyse the significance of each factor

The principle of Parsimony (Occam's Razor) – simplify the model to the "simplest acceptable model"

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

22

Model – Investigation 1 : Age * Sex * Smk * Dur

```
Console ~/

Call:
glm(formula = MG1$DTHS ~ offset(log(MG1$EXPO)) + MG1$AGE * MG1$SEX *
    MG1$SMOKER * MG1$DUR, family = quasipoisson)

Deviance Residuals:
    Min       1Q    Median       3Q       Max
-3.2464   -0.6364   0.0055    0.7708    2.5557

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                           -1.210e+01  2.846e-01 -42.498  < 2e-16 ***
MG1$AGE                                9.095e-02  5.582e-03  16.295  < 2e-16 ***
MG1$SEXM                               9.848e-01  3.464e-01   2.843 0.004884 **
MG1$SMOKERS                           -1.034e+00  5.251e-01  -1.969 0.050243 .
MG1$DUR                                1.405e-01  7.600e-02   1.849 0.065830 .
MG1$AGE:MG1$SEXM                      -8.488e-03  6.699e-03  -1.267 0.206497
MG1$AGE:MG1$SMOKERS                    3.756e-02  1.008e-02   3.726 0.000246 ***
MG1$SEXM:MG1$SMOKERS                   1.458e+00  6.194e-01   2.354 0.019457 *
MG1$AGE:MG1$DUR                       -2.373e-04  1.476e-03  -0.161 0.872363
MG1$SEXM:MG1$DUR                      -2.842e-01  9.357e-02  -3.038 0.002668 **
MG1$SMOKERS:MG1$DUR                    1.531e-01  1.447e-01   1.058 0.291215
MG1$AGE:MG1$SEXM:MG1$SMOKERS          -2.955e-02  1.189e-02  -2.486 0.013663 *
MG1$AGE:MG1$SEXM:MG1$DUR               4.330e-03  1.780e-03   2.432 0.015808 *
MG1$AGE:MG1$SMOKERS:MG1$DUR           -3.618e-03  2.746e-03  -1.317 0.189065
MG1$SEXM:MG1$SMOKERS:MG1$DUR          -1.875e-01  1.738e-01  -1.079 0.281919
MG1$AGE:MG1$SEXM:MG1$SMOKERS:MG1$DUR   4.276e-03  3.284e-03   1.302 0.194135
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.271398)

    Null deviance: 10548.47  on 239  degrees of freedom
Residual deviance:   272.92  on 224  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

>
```

PartnerRe

Model – Investigation 1 : Age * Sex * Smk * Dur

```
Console ~/

Call:
glm(formula = MG1$DTHS ~ MG1$AGE + MG1$SEX + MG1$SMOKER + MG1$DUR +
    MG1$AGE:MG1$SEX + MG1$AGE:MG1$SMOKER + MG1$SEX:MG1$SMOKER +
    MG1$AGE:MG1$DUR + MG1$SEX:MG1$DUR + MG1$AGE:MG1$SEX:MG1$SMOKER +
    MG1$AGE:MG1$SEX:MG1$DUR + offset(log(MG1$EXPO)), family = quasipoisson)

Deviance Residuals:
    Min      1Q    Median      3Q      Max
 -3.4709  -0.6370  -0.0157   0.7362   2.5731

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -12.261799   0.250419 -48.965  < 2e-16 ***
MG1$AGE                       0.094960   0.004853  19.569  < 2e-16 ***
MG1$SEXM                      1.184994   0.303860   3.900 0.000127 ***
MG1$SMOKERS                  -0.513008   0.253163  -2.026 0.043889 *
MG1$DUR                       0.189013   0.064526   2.929 0.003742 **
MG1$AGE:MG1$SEXM             -0.013148   0.005837  -2.253 0.025237 *
MG1$AGE:MG1$SMOKERS           0.025282   0.004787   5.282 2.98e-07 ***
MG1$SEXM:MG1$SMOKERS          0.835884   0.307339   2.720 0.007036 **
MG1$AGE:MG1$DUR              -0.001400   0.001240  -1.129 0.259894
MG1$SEXM:MG1$DUR             -0.343461   0.078550  -4.372 1.87e-05 ***
MG1$AGE:MG1$SEXM:MG1$SMOKERS -0.015298   0.005778  -2.648 0.008669 **
MG1$AGE:MG1$SEXM:MG1$DUR      0.005693   0.001490   3.820 0.000172 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.272148)

    Null deviance: 10548.47  on 239  degrees of freedom
Residual deviance:  277.45  on 228  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

>
```

PartnerRe

Model – Investigation 1 : Age * Sex * Smk * Dur

```
Console ~/

Call:
glm(formula = MG1$DTHS ~ MG1$AGE + MG1$SEX + MG1$SMOKER + MG1$DUR +
    MG1$AGE:MG1$SEX + MG1$AGE:MG1$SMOKER + MG1$SEX:MG1$SMOKER +
    MG1$SEX:MG1$DUR + MG1$AGE:MG1$SEX:MG1$SMOKER + offset(log(MG1$EXPO)),
    family = quasipoisson)

Deviance Residuals:
    Min      1Q   Median       3Q      Max
-3.4289  -0.6567  -0.0245   0.7375   3.4940

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -12.025798   0.144692 -83.113  < 2e-16 ***
MG1$AGE                         0.090308   0.002721  33.194  < 2e-16 ***
MG1$SEXM                        0.210538   0.179932   1.170  0.24317
MG1$SMOKERS                    -0.519027   0.268181  -1.935  0.05417 .
MG1$DUR                         0.117522   0.012862   9.137  < 2e-16 ***
MG1$AGE:MG1$SEXM                0.005571   0.003324   1.676  0.09514 .
MG1$AGE:MG1$SMOKERS             0.025376   0.005071   5.005 1.11e-06 ***
MG1$SEXM:MG1$SMOKERS            0.934566   0.326615   2.861  0.00461 **
MG1$SEXM:MG1$DUR               -0.042665   0.015503  -2.752  0.00639 **
MG1$AGE:MG1$SEXM:MG1$SMOKERS   -0.017194   0.006139  -2.801  0.00553 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.437646)

    Null deviance: 10548.47  on 239  degrees of freedom
Residual deviance:   313.08  on 230  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

>
```

PartnerRe

Model – Investigation 1 : Age * Sex * Smk * Dur

```
Console ~/

Call:
glm(formula = MG1$DTHS ~ MG1$AGE + MG1$SMOKER + MG1$DUR + MG1$AGE:MG1$SEX +
    MG1$AGE:MG1$SMOKER + MG1$SEX:MG1$SMOKER + MG1$SEX:MG1$DUR +
    MG1$AGE:MG1$SEX:MG1$SMOKER + offset(log(MG1$EXPO)), family = quasipoisson)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-3.4806  -0.6930   -0.0417   0.7281   3.2266

Coefficients:
                               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)                   -11.890175   0.086314 -137.756  < 2e-16 ***
MG1$AGE                         0.087918   0.001805   48.704  < 2e-16 ***
MG1$SMOKERS                    -0.649344   0.244500   -2.656  0.00846 **
MG1$DUR                         0.114614   0.012615    9.086  < 2e-16 ***
MG1$AGE:MG1$SEXM                0.009228   0.001145    8.063 4.02e-14 ***
MG1$AGE:MG1$SMOKERS             0.027839   0.004628    6.016 6.96e-09 ***
MG1$SMOKERS:MG1$SEXM            1.137887   0.277364    4.102 5.67e-05 ***
MG1$DUR:MG1$SEXM               -0.038779   0.015160   -2.558  0.01117 *
MG1$AGE:MG1$SMOKERS:MG1$SEXM   -0.020943   0.005254   -3.986 9.02e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.445737)

    Null deviance: 10548.47  on 239  degrees of freedom
Residual deviance:   315.05  on 231  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

>
```

PartnerRe

Model – Investigation 1 – Model Comparison

Variable removed	Residual Deviance	Degrees of Freedom	Anova (F test) Pr(> F)
FULL MODEL	272.92	224	
Age:Sex:Smoker:Dur	Δ 2.16	Δ 1	0.19
Sex:Smoker:Dur	Δ 1.56	Δ 1	0.27
Age:Smoker:Dur	Δ 0.18	Δ 1	0.71
Smoker:Dur	Δ 0.60	Δ 1	0.48
Age:Dur	Δ 35.63	Δ 2	***
Sex	Δ 1.98	Δ 1	0.24

PartnerRe

Model – Investigation 1 – Final model

So the simplified model is:

$$\mu(x, \mathrm{dur}, \mathrm{Sex}, \mathrm{Smoker}) =$$

$$\exp\begin{pmatrix} -12.262 + 0.095x + 1.185\chi_{\mathrm{MALE}} \\ -0.513\chi_{\mathrm{SMOKER}} + 0.189\,dur \\ -0.013x\chi_{\mathrm{MALE}} + 0.025x\chi_{\mathrm{SMOKER}} \\ +0.836\chi_{\mathrm{MALE}}\chi_{\mathrm{SMOKER}} - 0.001xdur \\ -0.343dur\chi_{\mathrm{MALE}} - 0.015x\chi_{\mathrm{MALE}}\chi_{\mathrm{SMOKER}} \\ +0.006xdur\chi_{\mathrm{MALE}} \end{pmatrix}$$

E.g. $\mu(37.5, 5, \mathrm{M}, \mathrm{NS}) = 0.0003958$

Observed = 0.0003601

PartnerRe

Model – Investigation 1 – Diagnostic plots

PartnerRe

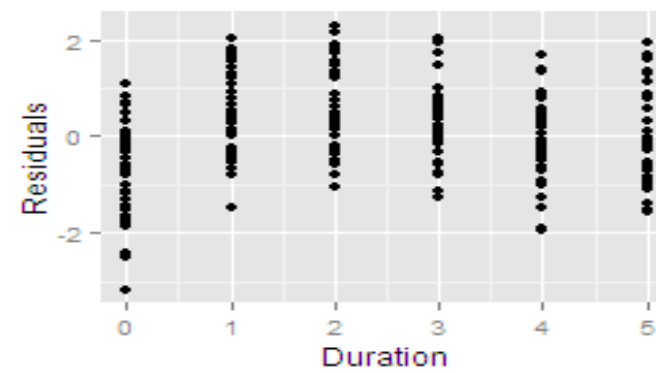# Model – Investigation 1 – Diagnostic plots

PartnerRe
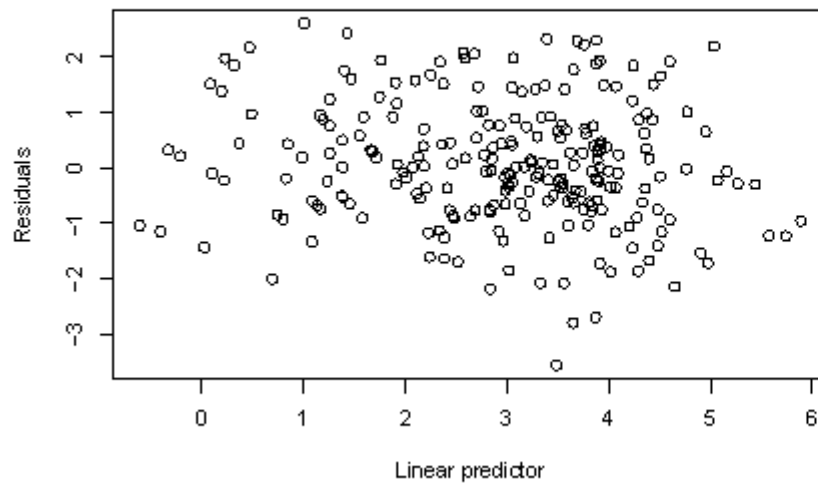
Model – Investigation 1 – Diagnostic plots

Model – Investigation 1 – Overdispersion

```
Console ~/

Call:
glm(formula = MG1$DTHS ~ MG1$AGE + MG1$SEX + MG1$SMOKER + MG1$DUR +
    MG1$AGE:MG1$SEX + MG1$AGE:MG1$SMOKER + MG1$SEX:MG1$SMOKER +
    MG1$AGE:MG1$DUR + MG1$SEX:MG1$DUR + MG1$AGE:MG1$SEX:MG1$SMOKER +
    MG1$AGE:MG1$SEX:MG1$DUR + offset(log(MG1$EXPO)), family = quasipoisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4709  -0.6370  -0.0157   0.7362   2.5731

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -12.261799   0.250419 -48.965  < 2e-16 ***
MG1$AGE                         0.094960   0.004853  19.569  < 2e-16 ***
MG1$SEXM                        1.184994   0.303860   3.900 0.000127 ***
MG1$SMOKERS                    -0.513008   0.253163  -2.026 0.043889 *
MG1$DUR                         0.189013   0.064526   2.929 0.003742 **
MG1$AGE:MG1$SEXM               -0.013148   0.005837  -2.253 0.025237 *
MG1$AGE:MG1$SMOKERS             0.025282   0.004787   5.282 2.98e-07 ***
MG1$SEXM:MG1$SMOKERS            0.835884   0.307339   2.720 0.007036 **
MG1$AGE:MG1$DUR                -0.001400   0.001240  -1.129 0.259894
MG1$SEXM:MG1$DUR               -0.343461   0.078550  -4.372 1.87e-05 ***
MG1$AGE:MG1$SEXM:MG1$SMOKERS   -0.015298   0.005778  -2.648 0.008669 **
MG1$AGE:MG1$SEXM:MG1$DUR        0.005693   0.001490   3.820 0.000172 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.272148)

    Null deviance: 10548.47  on 239  degrees of freedom
Residual deviance:   277.45  on 228  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

>
```

```
    Null deviance: 9457.61  on 239  degrees of freedom
Residual deviance:  254.71  on 228  degrees of freedom
AIC: 1456.5
```

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

32

Model – Investigation 2

Take a grouped dataset split by

- Age Bands of 5

- Sex

- Smoker

- Duration

and SA Band, Rated, JL Status, Distribution Channel

PartnerRe

Model – Investigation 2

Fit the following model :
AGE * SEX * (SMK + DUR) + SA + R + JS + CHANNEL

Use Exposure as the "offset"

Analyse the significance of each factor and simplify

PartnerRe

Model – Investigation 2 - Results


```
R Console                                                    _ | □ | X
File  Edit  Misc  Packages  Windows  Help


Call:
glm(formula = MG2$DTHS ~ offset(log(MG2$EXPO)) + MG2$AGE * MG2$SEX *
    MG2$SMOKER + MG2$AGE * MG2$SEX * MG2$DUR + MG2$R + MG2$SJ +
    MG2$SA + MG2$CHANNEL, family = "quasipoisson")

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-3.5530  -0.6371  -0.2856   0.1106   4.2731

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -1.055e+01  1.675e-01 -62.957  < 2e-16 ***
MG2$AGE                          7.348e-02  3.067e-03  23.955  < 2e-16 ***
MG2$SEXF                        -1.211e+00  2.832e-01  -4.278 1.92e-05 ***
MG2$SMOKERS                      2.317e-01  1.628e-01   1.425 0.154699
MG2$DUR                         -1.282e-01  4.188e-02  -3.061 0.002213 **
MG2$RR                           5.521e-01  2.912e-02  18.958  < 2e-16 ***
MG2$SJJ                         -9.762e-02  2.332e-02  -4.186 2.89e-05 ***
MG2$SASA_H                      -2.857e-01  2.977e-02  -9.597  < 2e-16 ***
MG2$SASA_M                      -9.944e-02  2.953e-02  -3.367 0.000765 ***
MG2$CHANNELDirect sales force    7.972e-02  3.593e-02   2.219 0.026560 *
MG2$AGE:MG2$SEXF                 1.347e-02  5.437e-03   2.478 0.013249 *
MG2$AGE:MG2$SMOKERS              1.102e-02  3.023e-03   3.647 0.000268 ***
MG2$SEXF:MG2$SMOKERS            -8.466e-01  2.872e-01  -2.948 0.003211 **
MG2$AGE:MG2$DUR                  3.965e-03  7.739e-04   5.123 3.11e-07 ***
MG2$SEXF:MG2$DUR                 3.387e-01  7.322e-02   4.626 3.81e-06 ***
MG2$AGE:MG2$SEXF:MG2$SMOKERS     1.510e-02  5.398e-03   2.796 0.005185 **
MG2$AGE:MG2$SEXF:MG2$DUR        -5.616e-03  1.388e-03  -4.046 5.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.124377)

    Null deviance: 15026.8  on 5504  degrees of freedom
Residual deviance:  4251.4  on 5488  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5

> |
```
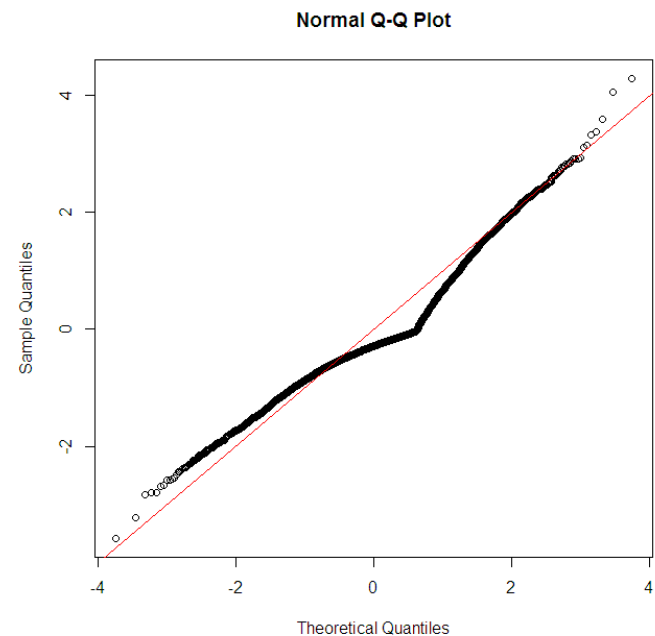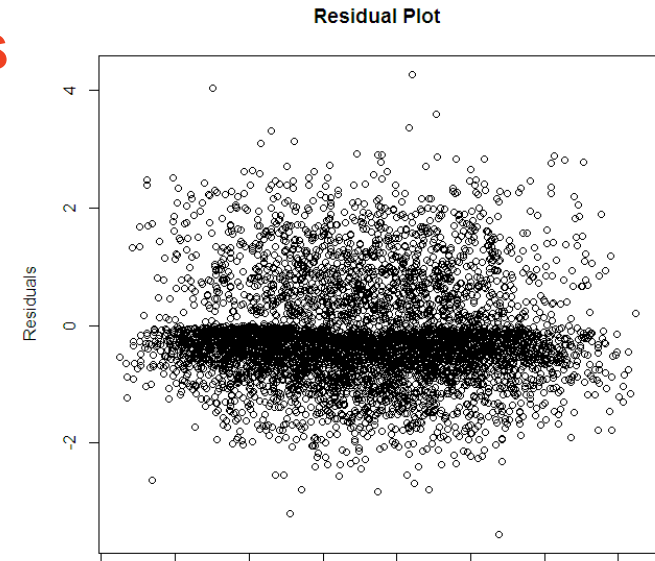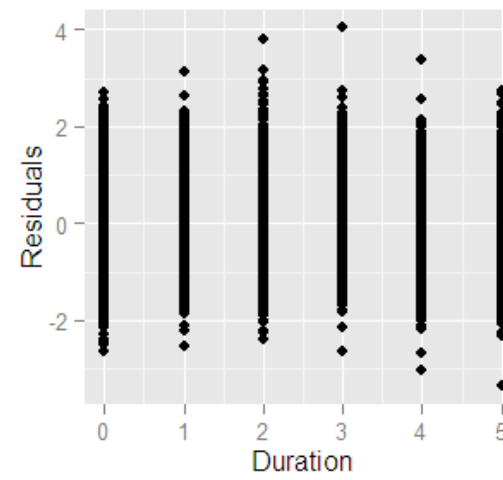
Residual Plot

Normal Q-Q Plot

Model – Investigation 2 - Results

PartnerRe
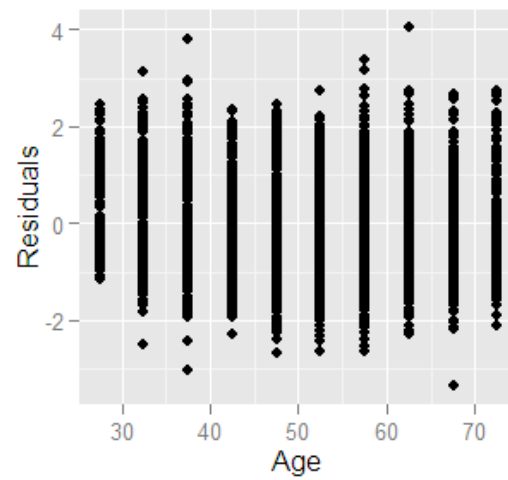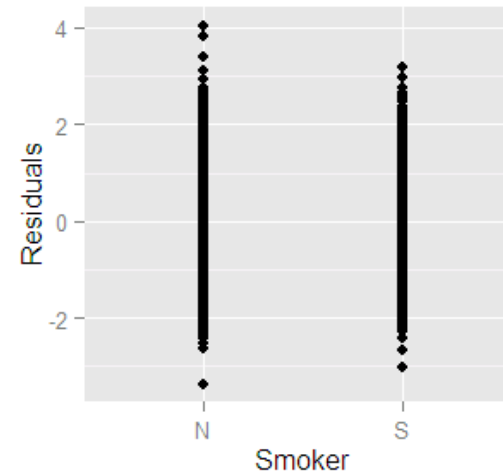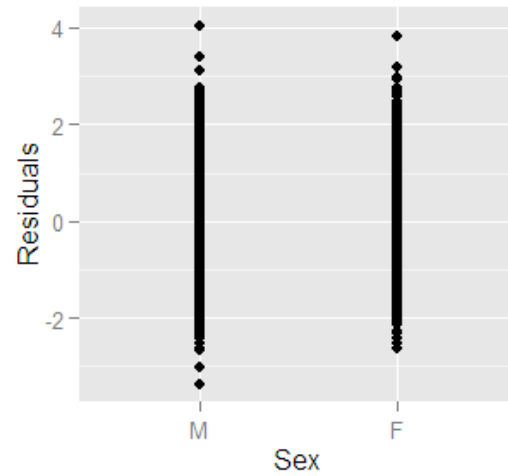
Model – Investigation 2 - Results
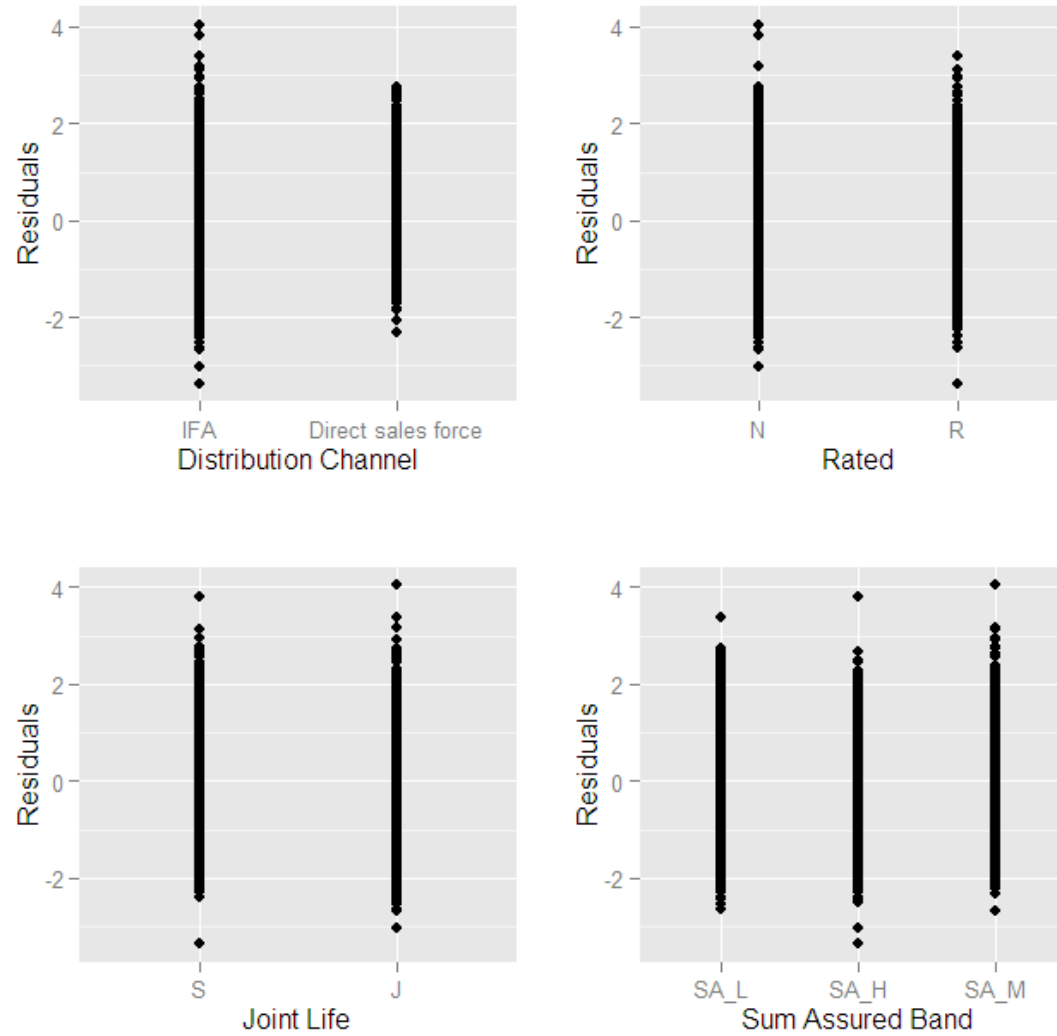
Typical model outputs from Model 2

Rating Factor	Loading to Intercept
Rated	174%
Joint Life	91%
Medium SA	91%
High SA	75%
Direct Sales Force	108%

Intercept:

Male – Non Smoker – Non Rated – IFA – Single Life

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

38

Model – Investigation 3

Take a grouped dataset split by

- 5 Year Age Band

- Sex

- Smoker

- Duration

- and SA Band, Rated, JL Status, Channel.

Fit the following model :
AGE * [SEX * (SMK + DUR) + SA + R + JS + CHANNEL]

Use Expected as the "offset"

PartnerRe

Model – Investigation 3 - Results

```
R Console                                                          _ □ x
File  Edit  Misc  Packages  Windows  Help

Deviance Residuals:
    Min      1Q    Median       3Q      Max
-3.4957  -0.6593  -0.3260   0.0712   3.9661

Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                          1.0066208  0.1977162   5.091 3.56e-07 ***
MG3$AGE                             -0.0137769  0.0035806  -3.848 0.000119 ***
MG3$SEXF                            -0.2004565  0.2578580  -0.777 0.436928
MG3$SMOKERS                          0.1483953  0.1500640   0.989 0.322722
MG3$DUR                             -0.0890091  0.0389286  -2.286 0.022227 *
MG3$RR                               1.1375597  0.1429907   7.955 1.78e-15 ***
MG3$SJJ                             -0.6422588  0.1138074  -5.643 1.67e-08 ***
MG3$SASA_H                          -0.6269160  0.1486198  -4.218 2.46e-05 ***
MG3$SASA_M                          -0.1881701  0.1623397  -1.159 0.246410
MG3$CHANNELDirect sales force        0.2371696  0.1693607   1.400 0.161399
MG3$AGE:MG3$SEXF                     0.0042572  0.0049667   0.857 0.391355
MG3$AGE:MG3$SMOKERS                 -0.0020921  0.0027911  -0.750 0.453520
MG3$SEXF:MG3$SMOKERS                -0.5187904  0.2663103  -1.948 0.051407 .
MG3$AGE:MG3$DUR                      0.0008031  0.0007195   1.116 0.264330
MG3$SEXF:MG3$DUR                     0.1255568  0.0676442   1.856 0.063434 .
MG3$AGE:MG3$RR                      -0.0100303  0.0025711  -3.901 9.57e-05 ***
MG3$AGE:MG3$SJJ                      0.0107826  0.0021192   5.088 3.62e-07 ***
MG3$AGE:MG3$SASA_H                   0.0073186  0.0027044   2.706 0.006807 **
MG3$AGE:MG3$SASA_M                   0.0020528  0.0029125   0.705 0.480908
MG3$AGE:MG3$CHANNELDirect sales force -0.0022516 0.0030134  -0.747 0.454960
MG3$AGE:MG3$SEXF:MG3$SMOKERS         0.0102123  0.0050058   2.040 0.041339 *
MG3$AGE:MG3$SEXF:MG3$DUR            -0.0026716  0.0012864  -2.077 0.037815 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 4907.4  on 5504  degrees of freedom
Residual deviance: 4219.2  on 5483  degrees of freedom
AIC: 10512

Number of Fisher Scoring iterations: 5

> |
```

PartnerRe

Model – Investigation 3 - Results

Residual Plot

Normal Q-Q Plot


```
R Console                                                    _ □ ×
File  Edit  Misc  Packages  Windows  Help

Call:
glm(formula = MG3$DTHS ~ MG3$AGE + MG3$DUR + MG3$R + MG3$SJ +
    MG3$SA + MG3$CHANNEL + MG3$AGE:MG3$R + MG3$AGE:MG3$SJ + MG3$AGE:MG3$
    offset(log(MG3$EXPE)), family = poisson)

Deviance Residuals:
    Min       1Q    Median        3Q       Max
-3.4249  -0.7128   -0.3382    0.2258    3.9327

Coefficients:
                                Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)                     0.812147    0.102096    7.955  1.79e-15 ***
MG3$AGE                        -0.010465    0.001801   -5.810  6.26e-09 ***
MG3$DUR                        -0.051076    0.006180   -8.265  < 2e-16 ***
MG3$RR                          1.136706    0.142338    7.986  1.39e-15 ***
MG3$SJJ                        -0.643878    0.112556   -5.721  1.06e-08 ***
MG3$SASA_H                     -0.474477    0.116498   -4.073  4.64e-05 ***
MG3$CHANNELDirect sales force   0.109819    0.033996    3.230   0.00124 **
MG3$AGE:MG3$RR                 -0.009979    0.002558   -3.901  9.58e-05 ***
MG3$AGE:MG3$SJJ                 0.010811    0.002090    5.172  2.31e-07 ***
MG3$AGE:MG3$SASA_H             0.005190    0.002227    2.330   0.01979 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3752.2  on 3695  degrees of freedom
Residual deviance: 3087.4  on 3686  degrees of freedom
AIC: 8237.9

Number of Fisher Scoring iterations: 5

> |
```
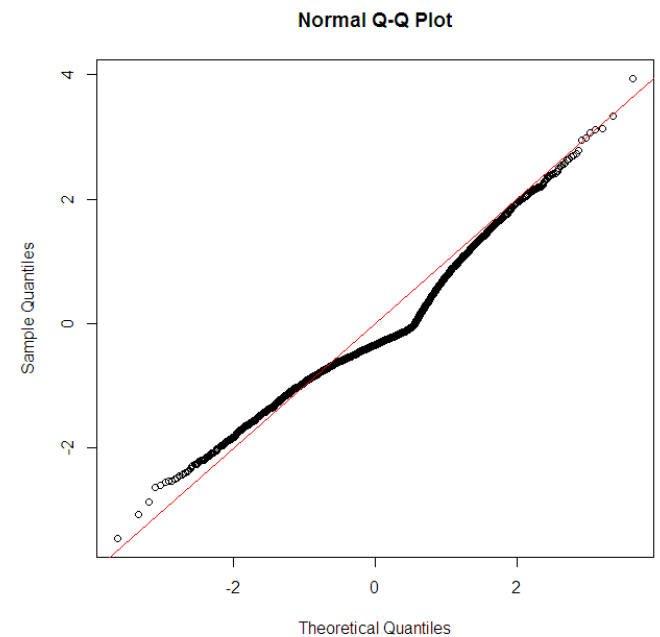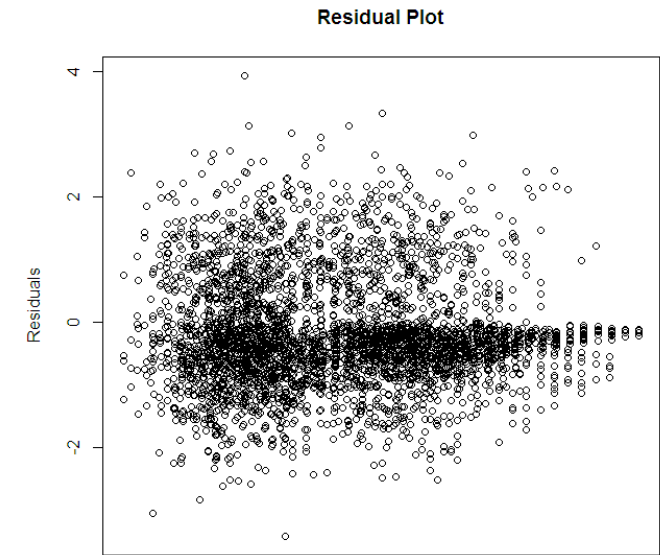
Model – Investigation 3 – Model Comparison

Variable removed	Residual Dev	DOF	AIC	Anova (χ^2 test)
FULL MODEL	3071.6	3676	8242	
Age:Channel	Δ 0.43	Δ 1	Δ -1.6	0.51
Age:Duration	Δ 0.00	Δ 1	Δ -2.0	0.93
Sex:Duration	Δ 0.81	Δ 1	Δ -1.2	0.37
Sex:Smoker	Δ -0.17	Δ 1	Δ -0.8	0.69
Age:Smoker	Δ -0.25	Δ 1	Δ -1.7	0.62
Sex	Δ -0.88	Δ 1	Δ -1.2	0.35
Smoker	Δ 3.98	Δ 1	Δ 2.0	*

PartnerRe

What happens if we remove gender from Model 2?

Variable removed	Residual Dev	AIC	Anova (χ^2 test)
FINAL MODEL	4251.4	10535	
Gender	Δ 346.6	Δ 334	***
Age	Δ 5925	Δ 5919	***

PartnerRe

How Powerful are your Rating Factors?
December 3, 2012

43

Removing gender & age from Model 2

qx	27.5 MNS	37.5 MNS	47.5 MNS	27.5 FNS	37.5 FNS	47.5 FNS
Age & Sex + RF	0.000167	0.000426	0.001089	0.000181	0.000399	0.000876
Age + RF	0.000151	0.000380	0.000956	0.000151	0.000380	0.000956
Just RF	0.001467	0.001467	0.001467	0.001467	0.001467	0.001467
(Age+RF) vs (Age+Sex+RF)	-10%	-11%	-12%	-17%	-5%	9%
(Just RF) vs (Age+Sex+RF)	776%	244%	36%	710%	268%	67%

PartnerRe

Summary

- Generalised Linear Models can be easily fitted with R;

- Data should be analysed and understood before fitting a model ;

- There are dangers of over-fitting or under-fitting a model;

- Over-dispersion may be a common feature of mortality data;

- Simplifying a model involves subjective decisions. No substitute for experience and common sense;

- Checking the fit of your model is essential.

PartnerRe

Further Reading

CMI – Working Paper 58 (2011)

An Introduction to Generalized Linear Models, Dobson & Barnett (2008)

Statistics : An Introduction using R, Crawley (2005)

Generalized Linear Models for Insurance Data, Jong & Heller (2008)

Demystifying GLMs (Sessional Meeting - Australia), Henwood et al (1991)

Risk classification in life insurance: methodology and case study, Gschlössl, Schoenmaekers and Denuit (2011)

Actuarial Graduation Practice and Generalised Linear and Non-Linear Models, Renshaw (1991)

PartnerRe