GLAM exam report Kieran Schubert

1 Exploratory Data Analysis

1.1 Dataset Structure

1.1.1 Dataset size

```
##   nrows ncols
## 1    74     5
```

1.1.2 Variable Type

```
## 'data.frame':    74 obs. of  5 variables:
##  $ pH     : num  5.5 5.5 5.5 5.5 5.5 5.5 5.5 5.5 3.5 3.5 ...
##  $ LL     : int  70 70 50 50 30 30 0 0 70 70 ...
##  $ Temp   : int  50 43 43 35 35 25 50 25 43 35 ...
##  $ Brix   : int  11 19 13 15 13 11 19 15 15 11 ...
##  $ CRA8132: int  0 0 1 1 1 0 0 1 0 0 ...
```

1.1.3 Variable range

```
##     CRA8132.range pH.range LL.range Temp.range Brix.range
## min             0      3.5        0         25         11
## max             1      5.5       70         50         19
```

1.1.4 Correlation & Variable Frequency

```
##                 pH          LL       Temp       Brix    CRA8132
## pH      1.00000000  0.09343169 0.04372266  0.1195965  0.3376306
## LL      0.09343169  1.00000000 0.08741539  0.1792907 -0.4583255
## Temp    0.04372266  0.08741539 1.00000000  0.1059341  0.1909273
## Brix    0.11959646  0.17929070 0.10593414  1.0000000 -0.2618827
## CRA8132 0.33763056 -0.45832552 0.19092730 -0.2618827  1.0000000
```

		pH	Freq	LL	Freq	Temp	Freq	Brix	Freq
CRA8132	Freq	3.5	18	0	22	25	18	11	24
0	48	4	20	30	18	35	18	13	16
1	26	5	18	50	16	43	20	15	16
		5.5	18	70	18	50	18	19	18

2 Modelling

2.1 Model 1: Full model

```
##
## Call:
## glm(formula = CRA8132 ~ ., family = binomial, data = orange)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3614  -0.3990  -0.1585   0.6306   1.6200
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.24633    3.21864  -2.251 0.024362 *
## pH           1.88595    0.54123   3.485 0.000493 ***
## LL          -0.06628    0.01905  -3.479 0.000503 ***
## Temp         0.11042    0.04769   2.316 0.020585 *
## Brix        -0.31173    0.14317  -2.177 0.029458 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 95.945  on 73  degrees of freedom
## Residual deviance: 52.331  on 69  degrees of freedom
## AIC: 62.331
##
## Number of Fisher Scoring iterations: 6
```

2.2 Model 2: Removing obs 7 and 44

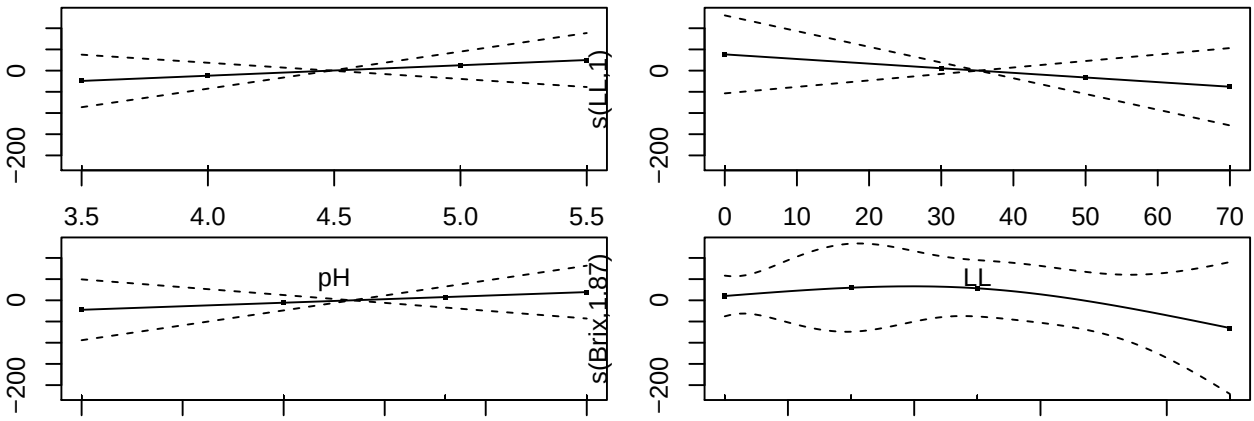
```
##
## Call:
## glm(formula = CRA8132 ~ ., family = binomial, data = orange.omit)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.74072  -0.11295  -0.00597   0.19355   1.98661
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -25.48628    8.30440  -3.069  0.00215 **
## pH            4.52335    1.41007   3.208  0.00134 **
## LL           -0.18304    0.05598  -3.270  0.00108 **
## Temp          0.31870    0.11070   2.879  0.00399 **
## Brix         -0.21455    0.18929  -1.133  0.25702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 94.184  on 71  degrees of freedom
```

```
## Residual deviance: 29.822  on 67  degrees of freedom
## AIC: 39.822
##
## Number of Fisher Scoring iterations: 8
```

2.3 Model 3 : Removing Brix

```
##
## Call:
## glm(formula = CRA8132 ~ pH + LL + Temp, family = binomial, data = orange.omit)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.19290  -0.19117  -0.00962   0.16554   1.81293
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -25.79714    7.75485  -3.327 0.000879 ***
## pH            4.24304    1.28643   3.298 0.000973 ***
## LL           -0.17894    0.05164  -3.465 0.000530 ***
## Temp          0.28512    0.09437   3.021 0.002517 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 94.184  on 71  degrees of freedom
## Residual deviance: 31.190  on 68  degrees of freedom
## AIC: 39.19
##
## Number of Fisher Scoring iterations: 7
```

2.4 Model 4: GAM



2.5 Model 5: Exhaustive Search (with all observations, 1st order interactions, Brix^2)

```
## $`1201`
##
## Call:  glm(formula = x, family = binomial, data = orange)
##
## Coefficients:
## (Intercept)           pH           LL         Temp         Brix
##   -260.7453      43.0019      -1.9031       3.1852      13.4122
##     pH_Brix      LL_Brix    Temp_Brix
##     -2.2149       0.1002      -0.1688
##
## Degrees of Freedom: 73 Total (i.e. Null);  66 Residual
## Null Deviance:       95.95
## Residual Deviance: 21.95     AIC: 37.95
```

```
##
## Call:
## glm(formula = CRA8132 ~ pH + LL + Temp + Brix + pH_Brix + LL_Brix +
##     Temp_Brix, family = binomial, data = orange.aug)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.02710  -0.27399  -0.00002   0.00207   2.44630
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -260.74530  129.81297  -2.009   0.0446 *
## pH            43.00190   21.91533   1.962   0.0497 *
## LL            -1.90315    0.95384  -1.995   0.0460 *
## Temp           3.18523    1.53655   2.073   0.0382 *
## Brix          13.41225    6.82054   1.966   0.0492 *
## pH_Brix       -2.21485    1.14186  -1.940   0.0524 .
## LL_Brix        0.10020    0.05028   1.993   0.0463 *
## Temp_Brix     -0.16884    0.08321  -2.029   0.0425 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 95.945  on 73  degrees of freedom
## Residual deviance: 21.953  on 66  degrees of freedom
## AIC: 37.953
##
## Number of Fisher Scoring iterations: 10
```

2.6 Model 6: bias reduction glm without obs 41

```
##
## Call:
## brglm(formula = CRA8132 ~ pH + LL + Temp + Brix + pH_Brix + LL_Brix +
##     Temp_Brix, family = binomial, data = data)
```

```
##
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -172.00022   70.43454  -2.442   0.0146 *
## pH            28.29872   11.79710   2.399   0.0164 *
## LL            -1.24843    0.52052  -2.398   0.0165 *
## Temp           2.09945    0.86417   2.429   0.0151 *
## Brix           8.89295    3.73629   2.380   0.0173 *
## pH_Brix       -1.48177    0.62561  -2.369   0.0179 *
## LL_Brix        0.06557    0.02759   2.377   0.0175 *
## Temp_Brix     -0.10972    0.04713  -2.328   0.0199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 79.199  on 72  degrees of freedom
## Residual deviance: 18.077  on 65  degrees of freedom
## Penalized deviance: -20.68985
## AIC:  34.077
```

2.7 Model comparison

2.7.1 Model Formulas

```
## [1] "Model 1:  pH + LL + Temp + Brix"
```

```
## [1] "Model 2:  pH + LL + Temp + Brix"
```

```
## [1] "Model 3:  pH + LL + Temp"
```

```
## [1] "Model 4:  pH + LL + Temp + I(Brix^2)"
```

```
## [1] "Model 5:  pH + LL + Temp + Brix + pH_Brix + LL_Brix + Temp_Brix"
```

```
## [1] "Model 6:  pH + LL + Temp + Brix + pH_Brix + LL_Brix + Temp_Brix"
```

2.7.2 Coefficients

```
##   predictors.mod1   beta.mod1 predictors.mod2   beta.mod2 predictors.mod3
## 1     (Intercept) -7.24633384     (Intercept) -25.4862763     (Intercept)
## 2              pH  1.88595099              pH   4.5233522              pH
## 3              LL -0.06627626              LL  -0.1830397              LL
## 4            Temp  0.11042240            Temp   0.3186962            Temp
## 5            Brix -0.31173235            Brix  -0.2145464            <NA>
##     beta.mod3 predictors.mod4   beta.mod4
## 1 -25.7971410     (Intercept) -9.56324192
## 2   4.2430375              pH  1.96762810
## 3  -0.1789428              LL -0.06723706
## 4   0.2851156            Temp  0.11691610
## 5          NA       I(Brix^2) -0.01328526
```

```
##             predictors.mod5    beta.mod5 predictors.mod6     beta.mod6
## (Intercept)     (Intercept) -260.7453040     (Intercept) -172.00022101
```

```
## pH                     pH   43.0019046             pH   28.29872382
## LL                     LL   -1.9031460             LL   -1.24843193
## Temp                 Temp    3.1852345          Temp    2.09944553
## Brix                 Brix   13.4122468          Brix    8.89294795
## pH_Brix           pH_Brix   -2.2148523       pH_Brix   -1.48176529
## LL_Brix           LL_Brix    0.1002039       LL_Brix    0.06556802
## Temp_Brix       Temp_Brix   -0.1688439     Temp_Brix   -0.10971775
```

2.7.3 Standard Errors

```
##   predictors.mod1 stderr.mod1 predictors.mod2 stderr.mod2 predictors.mod3
## 1     (Intercept)  3.21863772     (Intercept)  8.30439822     (Intercept)
## 2              pH  0.54122982              pH  1.41007061              pH
## 3              LL  0.01904888              LL  0.05597954              LL
## 4            Temp  0.04768831            Temp  0.11070371            Temp
## 5            Brix  0.14317366            Brix  0.18928521            <NA>
##   stderr.mod3 predictors.mod4 stderr.mod4
## 1  7.75484506     (Intercept) 3.210336076
## 2  1.28643120              pH 0.555731215
## 3  0.05163926              LL 0.019422448
## 4  0.09437054            Temp 0.049166805
## 5          NA       I(Brix^2) 0.005463129
```

```
##             predictors.mod5  stderr.mod5 predictors.mod6 stderr.mod6
## (Intercept)     (Intercept) 129.81297434     (Intercept) 70.43453512
## pH                      pH  21.91532830              pH 11.79710452
## LL                      LL   0.95384330              LL  0.52051668
## Temp                  Temp   1.53654865            Temp  0.86416989
## Brix                  Brix   6.82053559            Brix  3.73628964
## pH_Brix            pH_Brix   1.14186340         pH_Brix  0.62560662
## LL_Brix            LL_Brix   0.05028277         LL_Brix  0.02758616
## Temp_Brix        Temp_Brix   0.08321463       Temp_Brix  0.04713419
```

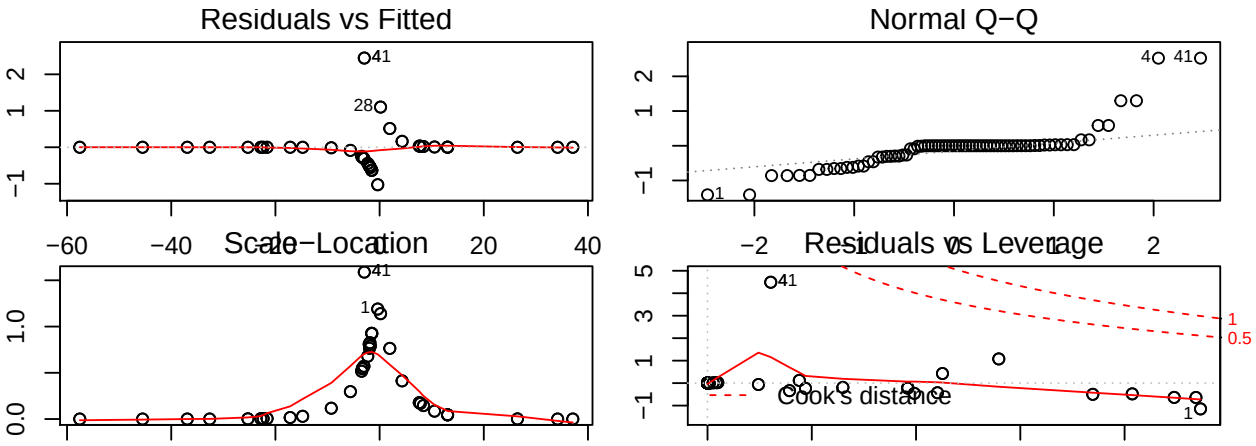
2.7.4 AIC

```
##   mod1.aic mod2.aic mod3.aic mod4.aic mod5.aic mod6.aic
## 1 62.33065 39.82176 39.19047 59.98388 37.95315 34.07701
```

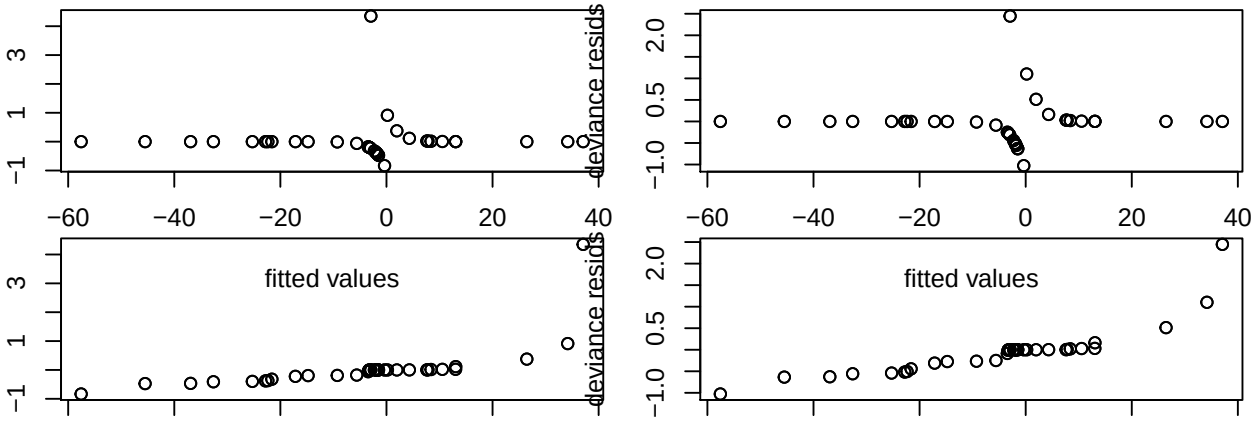
2.7.5 Model p-values

```
##        mod1      mod2     mod3      mod4      mod5 mod6
## 1 0.9324982 0.9999764 0.999963 0.9589803 0.9999999    1
```

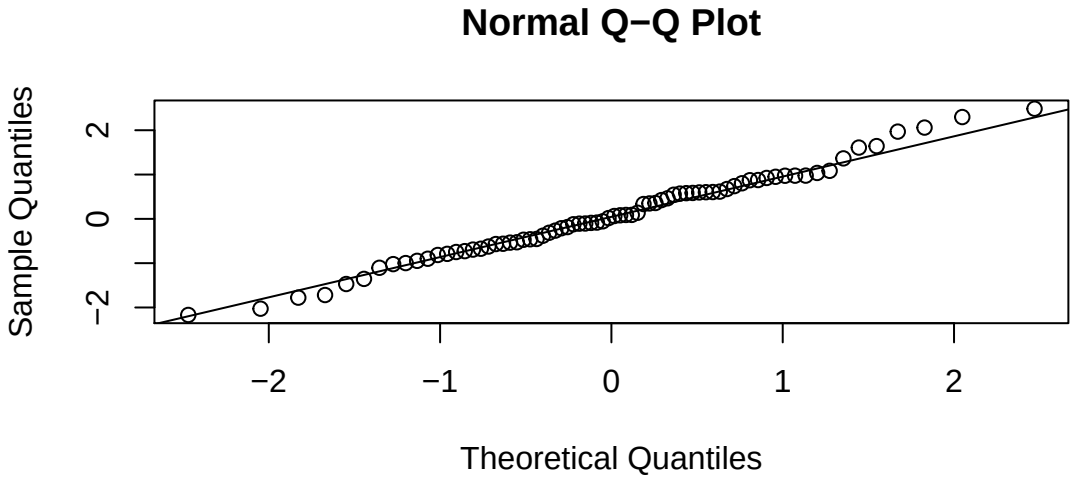
2.8 Residuals and Model Checking of Final Model (Model 5)



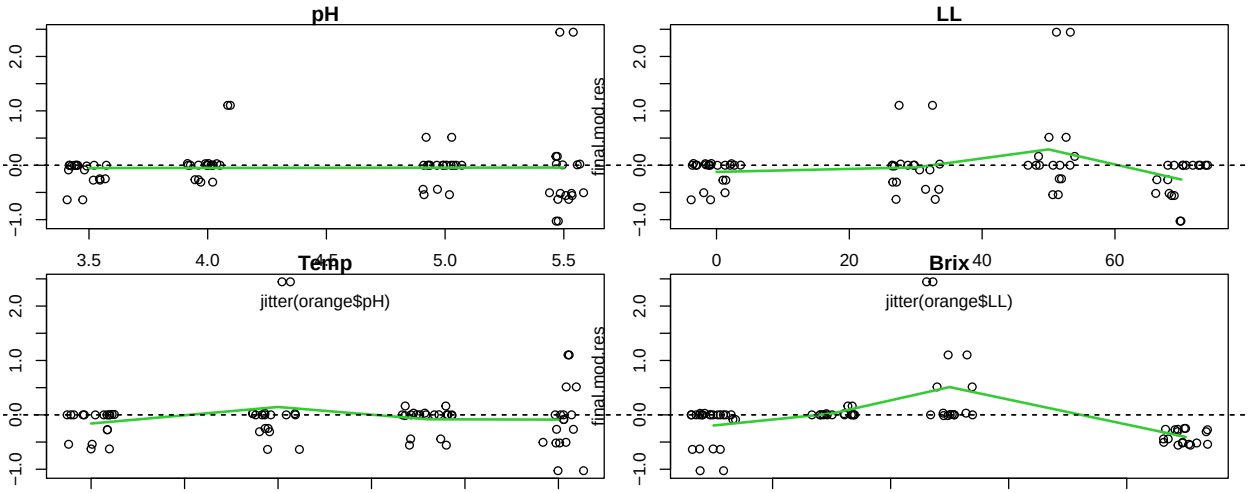
2.8.1 Pearson/Deviance residuals



2.8.2 Randomized quantile residuals



2.8.3 Residuals vs covariates



2.9 Interpretation

Table 1: Temp=25 and Temp=35

	pH 3.5	pH 4	pH 4.5	pH 5	pH 5.5	pH 3.5	pH 4	pH 4.5	pH 5	pH 5.5
LL 0	0.02	0.00	0.00	0.00	0	0.79	0.14	0.01	0.00	0.00
LL 17.5	0.15	0.01	0.00	0.00	0	0.97	0.58	0.06	0.00	0.00
LL 35	0.60	0.06	0.00	0.00	0	1.00	0.92	0.33	0.02	0.00
LL 52.5	0.93	0.36	0.02	0.00	0	1.00	0.99	0.80	0.15	0.01
LL 70	0.99	0.82	0.17	0.01	0	1.00	1.00	0.97	0.60	0.06

Table 2: Temp=43 and Temp=50

	pH 3.5	pH 4	pH 4.5	pH 5	pH 5.5	pH 3.5	pH 4	pH 4.5	pH 5	pH 5.5
LL 0	0.7881	0.1397	0.0070	0.0003	0.0000	0.9647	0.5443	0.0496	0.0023	0.0001
LL 17.5	0.9688	0.5753	0.0558	0.0026	0.0001	0.9956	0.9088	0.3032	0.0186	0.0008
LL 35	0.9962	0.9187	0.3304	0.0211	0.0009	0.9995	0.9881	0.7840	0.1368	0.0069
LL 52.5	0.9995	0.9895	0.8046	0.1523	0.0078	0.9999	0.9986	0.9680	0.5694	0.0546
LL 70	0.9999	0.9987	0.9717	0.5999	0.0614	1.0000	0.9998	0.9961	0.9169	0.3251