TMA4315: Compulsory exercise 2 Logistic regression and Poisson regression

Group XX: Henrik Syversveen Lie, Mikal Stapnes, Oliver Byhring

12.10.2018

Contents

Part 1: Logistic regression **1**
a) 1

Part 2: Poisson regression - Eliteserien 2018 **1**
a) 1
b) 2
c) 2
d) 4

Part 1: Logistic regression

a)

We let y_i be the number of successfull ascents, and n_i be the total number of attempts (success + fail) of the i'th mountain. We then do binary regression with the logit link to model the probability of success. This gives

1. Model for response: $Y_i \sim \text{Bin}(n_i, \pi_i)$, for $i = 1, \ldots, 113$
2. Linear predictor: $\eta_i = \mathbf{x}_i^T \beta$
3. Link function: $\eta_i = \ln(\frac{\pi_i}{1-\pi_i})$

where x_i is a p dimensional column vector of covariates for observation i, and β is the vector of regression parameters.

Part 2: Poisson regression - Eliteserien 2018

a)

We want to test if the assumption of independence between the goals made by the home and away teams is reasonable. To do this, we first load the data set and make a contingency table of all the results, with the goals of the home team on the rows, and goals of the away team on the columns. We get the following contingency table.

```
##      0  1  2 3 4+
## 0    8 18  3 1  1
## 1   19 26 15 5  3
## 2   10 14 13 4  1
## 3   13 10  7 2  0
## 4+  8  7  3 1  0
```

We then want to test if the number of goals for home and away team are independent. We do this by conducting *Pearson's* χ^2 *test* on the contingency table. The test poses the following hypotheses (SPØR OM DETTE ER RIKTIG HYPOTESE)

H_0 : The sampling distributions are independently chi-squared distributed, H_1 : They are not independently chi-squared d

```
##
##  Pearson's Chi-squared test
##
## data:  contingency
## X-squared = 14.156, df = 16, p-value = 0.5871
```

We get a value of 14.146 for the test statistic, with a corresponding p-value of 0.5871. As this p-value is above any reasonable significance level, we keep the null hypothesis, and confirm that the goals scored by the home and away team are independent.

b)

We now make the current standings in the Eliteserie based on all the results in our data set, and get the following table.

```
##                  Team Played Won Drawn Lost For Against  GD Points
## 1           Rosenborg     24  16     4    4  43      20  23     52
## 2               Brann     24  14     6    4  36      23  13     48
## 3               Molde     24  13     4    7  48      30  18     43
## 4           Haugesund     24  12     5    7  36      28   8     41
## 5          Ranheim_TF     24  11     5    8  38      40  -2     38
## 6          Vaalerenga     24  10     6    8  35      37  -2     36
## 7                 Odd     24   9     7    8  35      29   6     34
## 8             Tromsoe     24  10     3   11  35      33   2     33
## 9         Sarpsborg08     24   9     5   10  39      34   5     32
## 10       Kristiansund     24   8     7    9  32      35  -3     31
## 11         BodoeGlimt     24   6     9    9  28      30  -2     27
## 12      Stroemsgodset     24   6     8   10  38      38   0     26
## 13        Lillestroem     24   6     7   11  26      37 -11     25
## 14            Stabaek     24   5     8   11  29      43 -14     23
## 15              Start     24   6     5   13  24      42 -18     23
## 16 Sandefjord_Fotball     24   2     9   13  24      47 -23     15
```

c)

```
##                  Team        Power
## 1           Rosenborg  0.366945548
## 2               Molde  0.279321007
## 3               Brann  0.225715115
## 4           Haugesund  0.141566217
## 5                 Odd  0.099954079
## 6         Sarpsborg08  0.097625830
## 7             Tromsoe  0.060091773
## 8       Stroemsgodset  0.049639590
## 9          Vaalerenga  0.014445633
## 10       Kristiansund  0.012621369
## 11         Ranheim_TF  0.008439525
## 12         BodoeGlimt  0.000000000
```

```
## 13        Lillestroem -0.132589021
## 14            Stabaek -0.148121316
## 15              Start -0.225876528
## 16 Sandefjord_Fotball -0.291815679
```

```
## [1] "Intercept: "
```

```
## [1] 0.1003129
```

```
## [1] "Home advantage: "
```

```
## [1] 0.4020541
```

```
##                  Power            Ranking
## 1           Rosenborg          Rosenborg
## 2               Molde              Brann
## 3               Brann              Molde
## 4           Haugesund          Haugesund
## 5                 Odd         Ranheim_TF
## 6          Sarpsborg08         Vaalerenga
## 7             Tromsoe                Odd
## 8        Stroemsgodset            Tromsoe
## 9           Vaalerenga        Sarpsborg08
## 10       Kristiansund       Kristiansund
## 11         Ranheim_TF         BodoeGlimt
## 12         BodoeGlimt      Stroemsgodset
## 13        Lillestroem        Lillestroem
## 14            Stabaek            Stabaek
## 15              Start              Start
## 16 Sandefjord_Fotball Sandefjord_Fotball
```

```
##
## Call:
## glm(formula = goals ~ -1 + X, family = "poisson")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0205  -0.8748  -0.2014   0.5761   2.8679
##
## Coefficients: (1 not defined because of singularities)
##                      Estimate Std. Error z value Pr(>|z|)
## XIntercept           0.100304   0.068489   1.465   0.1430
## XHomeAdvantage       0.402068   0.087521   4.594 4.35e-06 ***
## XRosenborg           0.366956   0.168373   2.179   0.0293 *
## XMolde               0.279264   0.168369   1.659   0.0972 .
## XLillestroem        -0.132857   0.168934  -0.786   0.4316
## XOdd                 0.099975   0.166394   0.601   0.5480
## XHaugesund           0.141121   0.166320   0.848   0.3962
## XSandefjord_Fotball -0.291865   0.164767  -1.771   0.0765 .
## XRanheim_TF          0.008343   0.169495   0.049   0.9607
## XBrann               0.225678   0.165557   1.363   0.1728
## XSarpsborg08         0.097553   0.166444   0.586   0.5578
## XStabaek            -0.148047   0.168914  -0.876   0.3808
## XTromsoe             0.060348   0.166332   0.363   0.7167
## XStart              -0.225884   0.165079  -1.368   0.1712
## XVaalerenga          0.014465   0.169280   0.085   0.9319
## XKristiansund        0.012376   0.166170   0.074   0.9406
```

```
## XStroemsgodset       0.049657   0.166211   0.299   0.7651
## XBodoeGlimt                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 499.35  on 384  degrees of freedom
## Residual deviance: 384.12  on 367  degrees of freedom
## AIC: 1135.3
##
## Number of Fisher Scoring iterations: 5
```

DISCUSS THE RESULTS FROM THIS TASK

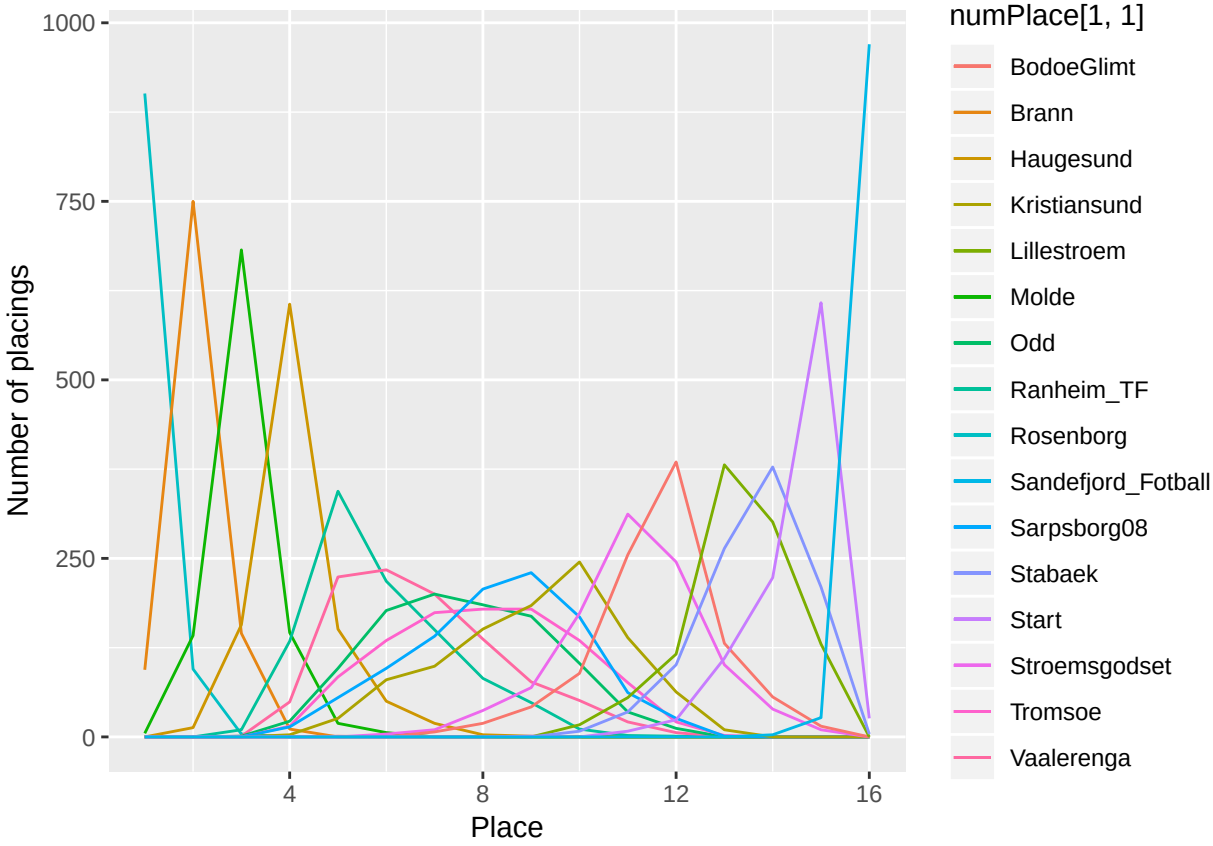
d)

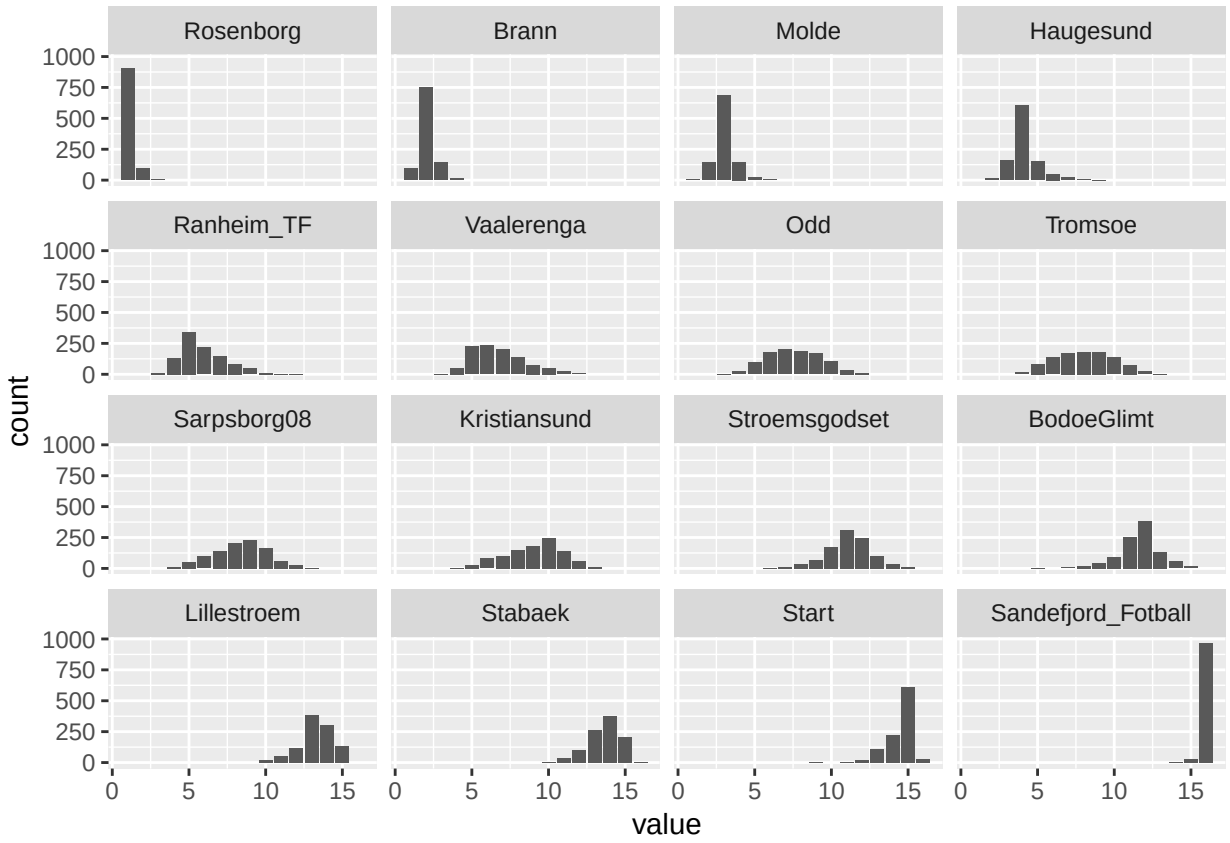
```
##   Rosenborg Brann Molde Haugesund Ranheim_TF Vaalerenga Odd Tromsoe
## 1        62    64    53        53         49         40  44      38
## 2        64    54    57        51         44         43  44      42
## 3        65    61    58        50         48         38  39      43
## 4        66    57    54        52         48         41  40      37
## 5        65    58    54        48         51         44  41      43
## 6        65    62    54        50         45         41  42      42
##   Sarpsborg08 Kristiansund Stroemsgodset BodoeGlimt Lillestroem Stabaek
## 1          40           43            38         40          28      29
## 2          38           44            40         39          26      26
## 3          44           44            32         33          33      28
## 4          39           40            42         40          25      32
## 5          43           44            35         33          29      27
## 6          39           41            39         38          28      25
##   Start Sandefjord_Fotball
## 1    25                17
## 2    31                19
## 3    26                21
## 4    27                21
## 5    30                15
## 6    27                20
```

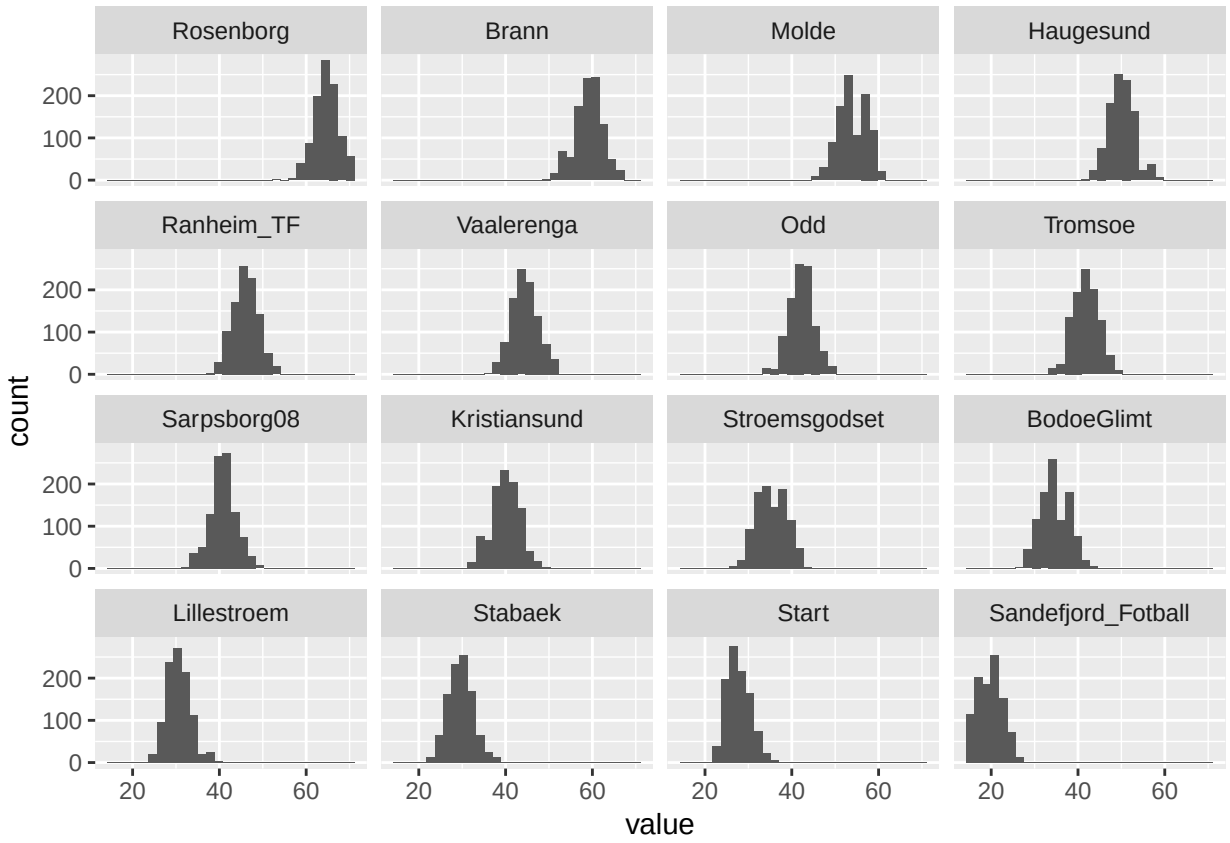
```
##                  Team Played  Won Drawn Lost  For Against    GD Points
## 1           Rosenborg     30 19.8   5.1  5.0 55.9    25.5  30.4   64.7
## 2               Brann     30 17.2   7.3  5.5 46.7    29.5  17.3   59.0
## 3               Molde     30 16.2   5.4  8.4 58.5    36.4  22.2   54.0
## 4           Haugesund     30 14.6   6.3  9.0 45.2    35.8   9.4   50.2
## 5          Ranheim_TF     30 13.2   6.4 10.4 46.2    48.6  -2.4   46.0
## 6          Vaalerenga     30 12.3   7.4 10.3 43.4    45.2  -1.9   44.3
## 7                 Odd     30 11.2   8.3 10.4 43.4    37.8   5.6   42.1
## 8             Tromsoe     30 12.4   4.4 13.2 43.6    41.2   2.4   41.6
## 9          Sarpsborg08     30 11.4   6.3 12.2 47.9    42.1   5.7   40.7
## 10       Kristiansund     30 10.4   8.5 11.1 40.8    42.8  -2.0   39.8
## 11      Stroemsgodset     30  8.6   9.4 12.0 47.2    45.6   1.6   35.3
## 12         BodoeGlimt     30  8.1  10.4 11.5 35.8    39.0  -3.2   34.6
## 13        Lillestroem     30  7.5   8.3 14.2 32.6    47.8 -15.2   30.7
## 14            Stabaek     30  6.8   9.4 13.8 36.2    52.5 -16.4   29.7
```

```
## 15              Start     30  7.2   6.3 16.6 29.8    53.5 -23.7   27.8
## 16 Sandefjord_Fotball     30  3.2  10.1 16.7 29.9    59.7 -29.7   19.7
```

```
##                  Team One Two Three Four Five Six Seven Eight Nine Ten
## 1           Rosenborg 901  95     4    0    0   0     0     0    0   0
## 2               Brann  94 750   145   11    0   0     0     0    0   0
## 3               Molde   5 142   682  146   19   6     0     0    0   0
## 4           Haugesund   0  13   157  606  151  50    19     3    1   0
## 5          Ranheim_TF   0   0    10  134  344 218   150    82   48  11
## 6          Vaalerenga   0   0     1   49  224 234   200   137   77  51
## 7                 Odd   0   0     1   22   96 177   200   185  169 103
## 8             Tromsoe   0   0     0   15   84 135   174   179  179 135
## 9         Sarpsborg08   0   0     0   14   55  96   141   207  230 168
## 10       Kristiansund   0   0     0    3   26  80    99   151  184 245
## 11      Stroemsgodset   0   0     0    0    0   4    10    37   69 173
## 12         BodoeGlimt   0   0     0    0    1   0     7    19   42  89
## 13        Lillestroem   0   0     0    0    0   0     0     0    0  17
## 14            Stabaek   0   0     0    0    0   0     0     0    0   8
## 15              Start   0   0     0    0    0   0     0     0    1   0
## 16 Sandefjord_Fotball   0   0     0    0    0   0     0     0    0   0
##    Eleven Twelve Thirteen Fourteen Fifteen Sixteen
## 1       0      0        0        0       0       0
## 2       0      0        0        0       0       0
## 3       0      0        0        0       0       0
## 4       0      0        0        0       0       0
## 5       2      1        0        0       0       0
## 6      21      6        0        0       0       0
## 7      35     12        0        0       0       0
## 8      76     21        2        0       0       0
## 9      62     26        1        0       0       0
## 10    139     63       10        0       0       0
## 11    312    245      101       39      10       0
## 12    255    385      131       56      15       0
## 13     55    116      381      301     130       0
## 14     35    101      264      378     210       4
## 15      8     24      110      223     608      26
## 16      0      0        0        3      27     970
```


1000
750
500
250
0
Number of placings
4
8
12
16
Place
numPlace[1, 1]
BodoeGlimt
Brann
Haugesund
Kristiansund
Lillestroem
Molde
Odd
Ranheim_TF
Rosenborg
Sandefjord_Fotball
Sarpsborg08
Stabaek
Start
Stroemsgodset
Tromsoe
Vaalerenga


Rosenborg
Brann
Molde
Haugesund
Ranheim_TF
Vaalerenga
Odd
Tromsoe
Sarpsborg08
Kristiansund
Stroemsgodset
BodoeGlimt
Lillestroem
Stabaek
Start
Sandefjord_Fotball
count
value
0
250
500
750
1000
0
5
10
15



By the histogram of points, the amount of points achieved in a season looks to be normal. Seeing as the amount of points is a random variable, we can by the central limit theorem say that the mean of the points is normally distributed. We therefore find the average number of points for each team, as well as the standard deviation, and construct 90 % confidence intervals for the points of each team.

```
##                Teams mean  sd  low high
## 1          Rosenborg 64.7 2.8 60.0 69.3
## 2              Brann 59.0 3.1 53.9 64.1
## 3              Molde 54.0 3.1 49.0 59.1
## 4           Haugesund 50.2 3.0 45.4 55.1
## 5          Ranheim_TF 46.0 3.0 41.1 51.0
## 6          Vaalerenga 44.3 3.2 39.1 49.6
## 7                 Odd 42.1 3.0 37.2 47.0
## 8             Tromsoe 41.6 3.1 36.5 46.6
## 9         Sarpsborg08 40.7 3.0 35.7 45.6
## 10       Kristiansund 39.8 3.2 34.6 45.0
## 11      Stroemsgodset 35.3 3.2 30.1 40.5
## 12         BodoeGlimt 34.6 3.0 29.7 39.6
## 13        Lillestroem 30.7 2.8 26.1 35.3
## 14            Stabaek 29.7 3.0 24.8 34.7
## 15              Start 27.8 2.9 23.1 32.4
## 16 Sandefjord_Fotball 19.7 2.6 15.5 24.0
```