TMA4315: Project 3

jototlan@stud.ntnu.no (10018), martigtu@stud.ntnu.no (10037)

```
long <- read.csv("https://www.math.ntnu.no/emner/TMA4315/2020h/eliteserie.csv", colClasses = c("factor"
head(long)
```

```
##               attack            defence home goals
## 1              Molde Sandefjord_Fotball  yes     5
## 2 Sandefjord_Fotball              Molde   no     0
## 3      Stroemsgodset            Stabaek  yes     2
## 4            Stabaek      Stroemsgodset   no     2
## 5                Odd          Haugesund  yes     1
## 6          Haugesund                Odd   no     2
```

a)

We consider the model

```
library(glmmTMB)
mod <- glmmTMB(goals ~ home + (1|attack) + (1|defence),  poisson, data=long, REML=TRUE)
```

Model and Notation

We use a triple index to denote the response, y_{ijk}, where i is the attacking team, j is the defending team, and $k \in \{0, 1\}$ where $k = 1$ indicates that the attacking team, i, is playing home, while $k = 0$ indicates that the attacking team is playing away. Clearly, $i, j \in \{1, 2, \dots 16\}$ and $i \neq j$. The distributional assumption on the response is $y_{ijk}|\gamma_i^a, \gamma_j^d \sim \text{Poisson}(\lambda_{ijk})$. Here, γ_i^a is the effect of team i attacking an γ_j^d is the effect of team j defending. The conditional mean is connected to the covariates by the canonical link function,

$$\lambda_{ijk} = \exp\left(\beta_0 + \beta_h k + \gamma_i^a + \gamma_j^d\right) = \exp(\eta_{ijk}),$$

where β_0 is the intercept and β_h is the effect of playing home. The random effects are independent and identically distributed, such that

$$\gamma_i^a \overset{iid}{\sim} \mathcal{N}(0, \tau_{\text{a}}) \quad \text{and} \quad \gamma_j^d \overset{iid}{\sim} \mathcal{N}(0, \tau_d).$$

We also assume that γ_i^a and γ_j^d are independent $\forall i, j$.

Distributional Assumption on Response

Assuming that the response follows a Poisson distribution amounts to making the following assumptions:

1. Goals are scored independently, i.e. the number of goals scored within disjoint time intervals is independent.
2. The number of goals scored in a time interval is proportional to the length of the interval.
3. Two (or more) goals cannot be scored at exactly the same instance.

The last two assumptions seem very reasonable; two goals can obviously not occur at the same time, and more time gives more oppurtunities for goal scoring. The first one, however, is more questionable. For example, a team might which has just conceded a goal close to the end of the game might play more aggressively to salvage a draw, hence increasing the likelihood of more goals being scored. Despite this, the Poisson distribution seems like a reasonable choice to model this process. Trenger flere antagelser? Diskutere REML?

b)

```
sum <- summary(mod)
sum
```

```
##  Family: poisson  ( log )
## Formula:          goals ~ home + (1 | attack) + (1 | defence)
## Data: long
##
##      AIC      BIC   logLik deviance df.resid
##   1147.2   1163.1   -569.6   1139.2      382
##
## Random effects:
##
## Conditional model:
##  Groups  Name        Variance Std.Dev.
##  attack  (Intercept) 0.007478 0.08647
##  defence (Intercept) 0.016383 0.12800
## Number of obs: 384, groups:  attack, 16; defence, 16
##
## Conditional model:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.12421    0.07809   1.591    0.112
## homeyes      0.40716    0.08745   4.656 3.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
beta.0 <- sum$coefficients$cond[1]
beta.h <- sum$coefficients$cond[2]

rf <- ranef(mod)
rf
```

```
## $attack
##                     (Intercept)
## BodoeGlimt          -0.036781062
## Brann                0.012026209
## Haugesund            0.011223106
## Kristiansund        -0.011367328
## Lillestroem         -0.049915996
## Molde                0.078390643
## Odd                  0.003654179
## Ranheim_TF           0.023375599
## Rosenborg            0.050622609
## Sandefjord_Fotball -0.058333079
## Sarpsborg08          0.026946364
## Stabaek             -0.026801293
## Start               -0.060500163
```

```
## Stroemsgodset       0.024556017
## Tromsoe             0.005756700
## Vaalerenga          0.007147494
##
## $defence
##                      (Intercept)
## BodoeGlimt         -0.042616090
## Brann              -0.123934761
## Haugesund          -0.061931278
## Kristiansund        0.008112432
## Lillestroem         0.030699257
## Molde              -0.036630979
## Odd                -0.052013600
## Ranheim_TF          0.062209734
## Rosenborg          -0.152631173
## Sandefjord_Fotball  0.133164228
## Sarpsborg08         0.006574064
## Stabaek             0.085376126
## Start               0.081958112
## Stroemsgodset       0.040486666
## Tromsoe            -0.009852817
## Vaalerenga          0.031030079
```

The effect of playing home is positive and statistically significant. According to the output, the team playing home scores $\exp(0.40716) = 1.5025445$ more goals on average, i.e. nearly 50% more goals on average. This seems reasonable from an intuitive perspective. Looking at the estimated random effects, we can e.g. consider $\gamma^{\text{defence}}_{\text{Rosenborg}} \approx -0.153$. This is the lowest value among all teams, which indicates that Rosenborg is the best defending team. To check this, we calculate the average number of goals conceded by every team:

```
no.NA = long[is.na(long$goals) == 0, c("defence", "goals")]
agg = aggregate(no.NA$goals, by = list(no.NA$defence), FUN = mean)
colnames(agg) <- c("Team", "Avg. # of conceded goals")
knitr::kable(agg)
```

Team	Avg. # of conceded goals
BodoeGlimt	1.2500000
Brann	0.9583333
Haugesund	1.1666667
Kristiansund	1.4583333
Lillestroem	1.5416667
Molde	1.2500000
Odd	1.2083333
Ranheim_TF	1.6666667
Rosenborg	0.8333333
Sandefjord_Fotball	1.9583333
Sarpsborg08	1.4166667
Stabaek	1.7916667
Start	1.7500000
Stroemsgodset	1.5833333
Tromsoe	1.3750000
Vaalerenga	1.5416667

As expected, Rosenborg has the lowest average number of conceded goals.

From the model assumptions, namely $\gamma_i^a \sim \mathcal{N}(0, \tau_a)$ and $\gamma_j^d \sim \mathcal{N}(0, \tau_d)$, the average attack strength and the average defense strength are equal to 0. An estimate of the expected number of goals which the average attacking team scores is thus given as

$$\exp(\beta_0 + \beta_h) \approx 1.701265.$$

Since the the number of goals scored is assumed to follow a Poisson distribution, an estimate of the variance is given by the same value. We now consider the the situation when the average defense team is attacking (here we assume that they also have average attck strength and that the average attack team has average defense) spør om dette!!. Then, an estimate of the number of goals scored is given as

$$\exp(\beta_0) \approx 1.132258.$$

Again, since we consider a Poisson-distributed variable, an estimate of the variance i given by the same value.

c)

The expected value and variance of goals scored by two randomly selected teams can be found by the laws of total expectation and total variance conditioned on the random effect given by the two teams. For this, we use that the distribution Lognormal$(0, \tau^2)$ has mean $\exp(\tau^2/2)$ and variance $(\exp(\tau^2) - 1)\exp(\tau^2)$.

The law of total expectation gives

$$\begin{aligned}
\mathrm{E}[y_{ijk}] &= \mathrm{E}\left[\mathrm{E}\left[y_{ijk} \mid \gamma_{ai}, \gamma_{dj}\right]\right] \\
&= \mathrm{E}\left[\exp\left(\beta_0 + \beta_h x_{ijk} + \gamma_{ai} + \gamma_{dj}\right)\right] \\
&= \exp\left(\beta_0 + \beta_h x_{ijk}\right) \mathrm{E}\left[\exp\left(\gamma_{ai} + \gamma_{dj}\right)\right] \\
&= \exp\left(\beta_0 + \beta_h x_{ijk}\right) \exp\left(\frac{\tau_a^2 + \tau_d^2}{2}\right).
\end{aligned}$$

The law of total variance gives

$$\begin{aligned}
\mathrm{Var}[y_{ijk}] &= \mathrm{E}\left[\mathrm{Var}\left[y_{ijk} \mid \gamma_{ai}, \gamma_{dj}\right]\right] + \mathrm{Var}\left[\mathrm{E}\left[y_{ijk} \mid \gamma_{ai}, \gamma_{dj}\right]\right] \\
&= \mathrm{E}\left[\exp\left(\beta_0 + \beta_h x_{ijk} + \gamma_{ai} + \gamma_{dj}\right)\right] + \mathrm{Var}\left[\exp\left(\beta_0 + \beta_h x_{ijk} + \gamma_{ai} + \gamma_{dj}\right)\right] \\
&= \mathrm{E}[y_{ijk}] + \exp\left(2(\beta_0 + \beta_h x_{ijk})\right) \mathrm{Var}\left[\exp\left(\gamma_{ai} + \gamma_{dj}\right)\right] \\
&= \exp\left(\beta_0 + \beta_h x_{ijk}\right) \exp\left(\frac{\tau_a^2 + \tau_d^2}{2}\right) + \exp\left(2\left(\beta_0 + \beta_h x_{ijk}\right)\right)\left(\exp\left(\tau_a^2 + \tau_d^2\right) - 1\right) \exp\left(\tau_a^2 + \tau_d^2\right).
\end{aligned}$$

Note that the total marginal variance consists of two terms, given as

$$\mathrm{Var}[y_{ijk}] = \underbrace{\mathrm{E}\left[\mathrm{Var}\left[y_{ijk} \mid \gamma_{ai}, \gamma_{dj}\right]\right]}_{\text{Variance of the game}} + \underbrace{\mathrm{Var}\left[\mathrm{E}\left[y_{ijk} \mid \gamma_{ai}, \gamma_{dj}\right]\right]}_{\text{Varaince of team strengths}}.$$

The first term is the expected variance of goals, which is a measure of the inherent randomness of the football game given the two teams. The second term is the variance of the expected goals, which is attributed to the variance in the strength of the two teams, since the expected goals are constant given two teams, so the randomness comes from the difference in strengths among the teams. We now calculate the proportion of variance explained by the two terms without and with home filed advantage.

Estimating the proportions with no home field advantage:

```
# Parameters from model
beta <- summary(mod)$coefficients$cond[,1]
beta0 = beta[1]
```

```r
beta.h = beta[2]

# Variance of random effects
tau2.attack <- summary(mod)$var$cond$attack[1]
tau2.defence <- summary(mod)$var$cond$defence[1]
tau2 <- tau2.attack + tau2.defence

# Marginal variance
var0.game <- exp(beta0 + tau2/2)
var0.strength <- exp(2*beta0)*(exp(tau2)-1)*exp(tau2)
var0 <- var0.game + var0.strength

# Proportion of variance
prop0.game = var0.game/var0
prop0.strength = var0.strength/var0
```

With no home field advantage for the attacking team, the total marginal variance of the number of goals is 1.1775525, where the inherent randomness of the game accounts for 0.9731 and the variation in team strengths account for 0.0269.

Estimating the proportions with home field advantage:

```r
# Marginal variance
var1.game <- exp(beta0 + beta.h + tau2/2)
var1.strength <- exp(2*(beta0 + beta.h))*(exp(tau2)-1)*exp(tau2)
var1 <- var1.game + var1.strength

# Proportions of variance
prop1.game = var1.game / var1
prop1.strength = var1.strength / var1
```

If the attacking team has a home field advantage, the total marginal variance of the number of goals is 1.7932623, where the inherent randomness of the game accounts for 0.9601 and the variation in team strengths account for 0.0399. We note that also here the majority of variance can be explained by the game itself.

An interesting observation is that when the home field advantage is present, the proportion of variance explained by the strengths of the teams is higher. This follows directly from the equation of the total marginal variance when $\beta_h > 0$, as the second term is multiplied by an extra factor of $\exp(\beta_h) > 1$.

d)

We want to test if the two random effects in the model are significant. Since $\gamma_a \sim N(0, \tau_a^2)$ and $\gamma_d \sim N(0, \tau_d^2)$, this is equivalent to testing whether the variance of each random effect is positive. The hypothesis test for determining if the effect of attacking is significant can be formulated as

$$\mathrm{H}_0 : \tau_a^2 = 0 \quad \text{vs} \quad \mathrm{H}_1 : \tau_a^2 > 0,$$

and a similar test can be constructed for the effect of defending.

This can be carried out using a likelihood-ratio test. The test statistic is $\lambda_{LR} := -2(l_0 - l_1)$, where l_0 and l_1 are the log-likelihoods of the two models. We cannot apply Wilks' theorem directly here, as standard asymptotic theory is violated. The reason for this is that there is a 50% chance that under H_1, the maximum likelihood estimator $\hat{\tau}_a^2$ falls on the boundary of the parameter space. This is due to the expected value of the score function $s(\tau_a^2) := \frac{\partial l_1}{\partial \tau_a^2}$ evaluated at zero is zero, $E\left[s(0)\right] = 0$, so whenever the slope of $s(0)$ is negative,

the maximum likelihood estimator returns zero. This results in a mixture of two distributions

$$\lambda_{RT} \sim \frac{1}{2}\chi_0^2 : \frac{1}{2}\chi_1^2$$

where χ_0^2 is simply the point mass distribution located at zero. The critical value C of the likelihood-ratio test is given as

$$P(\lambda_{RT} \geq C) = \frac{1}{2}P(\chi_0^2 \geq C) + \frac{1}{2}P(\chi_1^2 \geq C) = \alpha.$$

Since $C > 0$ for $\alpha < \frac{1}{2}$ the first term vanishes, so C is the upper 2α quantile of a chi-squared distribution with one degree of freedom. Using $\alpha = 0.05$ we get the critical value

```
alpha <- 0.05
qchisq(1-2*alpha, df=1)
```

```
## [1] 2.705543
```

Testing significance of effect of attack

```
mod.no_attack <- glmmTMB(goals ~ home + (1|defence), poisson, data=long, REML=TRUE)
LRT.no_attack <- -2*as.numeric(logLik(mod.no_attack) - logLik(mod))

p.value <- 0.5 * pchisq(LRT.no_attack, df=0, lower.tail = FALSE) +
           0.5 * pchisq(LRT.no_attack, df=1, lower.tail = FALSE)
p.value
```

```
## [1] 0.2587558
```

The p-value suggest that we can not reject H_0, meaning that the effect of attack is not significant. The test statistic has value 0.4188496, which is lower than the critical value.

Testing significance of effect of defence

```
mod.no_defence <- glmmTMB(goals ~ home + (1|attack), poisson, data=long, REML=TRUE)
LRT.no_defence <- -2*as.numeric(logLik(mod.no_defence) - logLik(mod))

p.value <- 0.5 * pchisq(LRT.no_defence, df=0, lower.tail = FALSE) +
           0.5 * pchisq(LRT.no_defence, df=1, lower.tail = FALSE)
p.value
```

```
## [1] 0.09843786
```

The p-value suggest that the effect of defence is not significant. The test statistic has value has value 1.6654006, which is again lower than the critical value.

Testing the effect of home field advantage

Our hypothesis can be formulated as

$$H_0 : \beta_h = 0 \quad \text{vs} \quad H_1 : \beta_h \neq 0,$$

where under H_1, the home field advantage could be either negative or positive.

The restricted maximum likelihood (REML) is often preferred when fitting generalized mixed models as the calculations are cheaper than using normal maximum likelihood estimation. However, this is not possible here, as REML uses a likelihood function on a transformed set of data, so that nuisance parameters have no effect. Under H_0, this is simply the intercept β_0. But under H_1, we also have the parameter β_h for the home

field advantage. This means that the two models lead to different means using REML, making the likelihood non-comparable. Setting REML = FALSE avoids this issue and fits the models using normal maximum likelihoods.

```
mod.ML <- glmmTMB(goals ~ home + (1|attack) + (1|defence), poisson, data=long, REML=FALSE)
mod.no_home <- glmmTMB(goals ~ (1|attack) + (1|defence), poisson, data=long, REML=FALSE)

LRT.no_home <- -2*as.numeric(logLik(mod.no_home) - logLik(mod.ML))

p.value <- pchisq(LRT.no_home, df=1, lower.tail = FALSE)
p.value
```

```
## [1] 2.507328e-06
```

The low p-value is smaller than the significance level, indicating that the effect of home field is significant.

e)

```
frankv <- data.table::frankv

ranking <- function(df){
  n.teams <- length(unique(df$attack))
  stats <- data.frame(row.names = unique(df$attack), points = rep(0, n.teams),
                      goal.diff = rep(0, n.teams), goals.scored = rep(0, n.teams))

  for(n in seq(1, dim(df)[1], 2)){
    team1 <- as.character(df$attack[n])
    team2 <- as.character(df$defence[n])
    goals1 = df$goals[n]
    goals2 = df$goals[n+1]

    # skip missing values
    if (is.na(goals1) | is.na(goals2)){
      next
    }
    else if(goals1 > goals2){
      stats[team1,"points"] = stats[team1,"points"] + 3
    }
    else if(goals2 > goals1){
      stats[team2,"points"] = stats[team2,"points"] + 3
    }
    else {
      stats[team1,"points"] = stats[team1,"points"] + 1
      stats[team2,"points"] = stats[team2,"points"] + 1
    }
    stats[team1,"goal.diff"] = stats[team1,"goal.diff"] + goals1 - goals2
    stats[team2,"goal.diff"] = stats[team2,"goal.diff"] + goals2 - goals1
    stats[team1,"goals.scored"] = stats[team1,"goals.scored"] + goals1
    stats[team2,"goals.scored"] = stats[team2,"goals.scored"] + goals2
  }
  stats$rank <- frankv(stats, cols=c("points","goal.diff","goals.scored"),
                       order=-1, ties.method="random")
  stats
}
```

```
stats <- ranking(long)
knitr::kable(stats[order(stats$rank),])
```

	points	goal.diff	goals.scored	rank
Rosenborg	52	23	43	1
Brann	48	13	36	2
Molde	43	18	48	3
Haugesund	41	8	36	4
Ranheim_TF	38	-2	38	5
Vaalerenga	36	-2	35	6
Odd	34	6	35	7
Tromsoe	33	2	35	8
Sarpsborg08	32	5	39	9
Kristiansund	31	-3	32	10
BodoeGlimt	27	-2	28	11
Stroemsgodset	26	0	38	12
Lillestroem	25	-11	26	13
Stabaek	23	-14	29	14
Start	23	-18	24	15
Sandefjord_Fotball	15	-23	24	16

f)

Simulating 1000 complete seasons using the predicted mean $\hat{\mu}$ from the model created in a).

```
set.seed(0)
# number of matches
n = dim(long)[1]

# predicted mean goals scored in each match
mu.hat <- predict(mod, newdata=long, type="response")

# matrix to store all simulated rankings
rank.matrix <- matrix(0, nrow = 1000, ncol = 16)
colnames(rank.matrix) = unique(long$attack)

# doing the simulations
temp <- long$goals
for(i in 1:1000){
  long$goals <- rpois(n, mu.hat)
  rank.matrix[i,] <- ranking(long)$rank
}
long$goals <- temp
```

```
# probabilities of each observation
probabilities <- apply(rank.matrix, 2, table)/1000

# summary with mean ranking and probabilities
rank.summary <- data.frame(row.names = unique(long$attack),
                           mean = colMeans(rank.matrix),
                           P = t(probabilities))
```

```r
# print the summary
output <- rank.summary[order(rank.summary$mean),]
library(kableExtra)
knitr::kable(output) %>% kableExtra::landscape()
```

	mean	P.1	P.2	P.3	P.4	P.5	P.6	P.7	P.8	P.9	P.10	P.11	P.12	P.13	P.14	P.15	P.16
Rosenborg	5.239	0.184	0.130	0.125	0.097	0.076	0.064	0.071	0.053	0.035	0.034	0.034	0.027	0.026	0.016	0.016	0.012
Brann	6.276	0.140	0.102	0.098	0.087	0.070	0.074	0.066	0.056	0.050	0.064	0.058	0.027	0.040	0.025	0.024	0.019
Molde	6.422	0.120	0.107	0.090	0.086	0.081	0.082	0.066	0.057	0.053	0.060	0.044	0.038	0.040	0.033	0.028	0.015
Haugesund	7.221	0.094	0.098	0.074	0.074	0.079	0.076	0.056	0.058	0.070	0.047	0.057	0.056	0.052	0.038	0.037	0.034
Odd	7.651	0.060	0.079	0.087	0.078	0.077	0.068	0.060	0.071	0.055	0.065	0.063	0.061	0.051	0.057	0.040	0.028
Sarpsborg08	8.133	0.061	0.062	0.068	0.056	0.082	0.075	0.066	0.071	0.068	0.060	0.071	0.046	0.062	0.050	0.053	0.049
Tromsoe	8.278	0.045	0.061	0.074	0.074	0.069	0.059	0.061	0.074	0.069	0.065	0.066	0.071	0.065	0.056	0.059	0.032
BodoeGlimt	8.422	0.053	0.064	0.048	0.066	0.067	0.080	0.058	0.061	0.081	0.061	0.062	0.075	0.074	0.056	0.047	0.047
Stroemsgodset	8.825	0.048	0.056	0.067	0.061	0.051	0.055	0.067	0.053	0.074	0.072	0.065	0.060	0.061	0.082	0.068	0.060
Kristiansund	8.872	0.047	0.051	0.051	0.062	0.057	0.058	0.065	0.072	0.063	0.065	0.074	0.084	0.072	0.056	0.069	0.054
Ranheim_TF	9.035	0.038	0.050	0.055	0.055	0.062	0.070	0.070	0.059	0.066	0.064	0.062	0.061	0.066	0.079	0.080	0.063
Vaalerenga	9.079	0.036	0.044	0.051	0.063	0.051	0.062	0.079	0.069	0.068	0.064	0.071	0.058	0.077	0.079	0.066	0.062
Lillestroem	9.962	0.019	0.036	0.032	0.049	0.053	0.057	0.051	0.066	0.071	0.082	0.064	0.073	0.079	0.094	0.079	0.095
Stabaek	10.201	0.027	0.030	0.036	0.043	0.047	0.039	0.056	0.065	0.057	0.067	0.061	0.099	0.094	0.086	0.103	0.090
Start	10.819	0.016	0.023	0.024	0.021	0.046	0.042	0.059	0.067	0.059	0.070	0.078	0.080	0.073	0.094	0.112	0.136
Sandefjord_Fotball	11.565	0.012	0.007	0.020	0.028	0.032	0.039	0.049	0.048	0.061	0.060	0.070	0.084	0.068	0.099	0.119	0.204

TODO: Add some comment here.

g)

We start by creating a data frame with the rankings from f) and the corresponding random effects, as well as the difference between attack strength and defense strength.

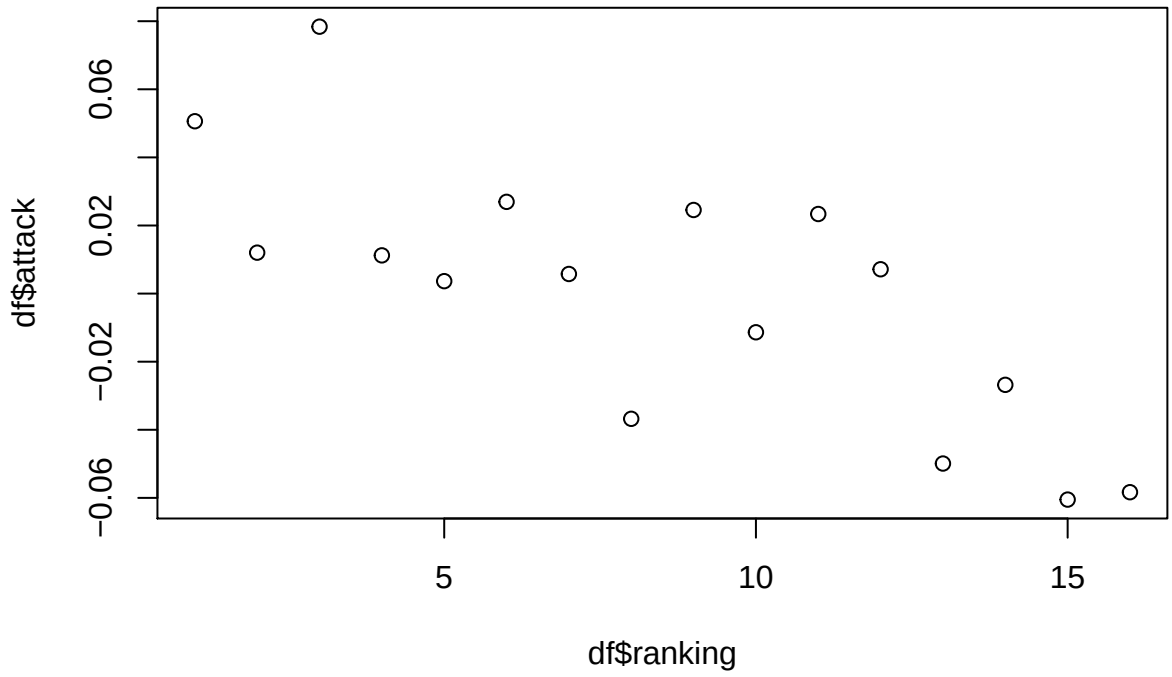
```
re <- ranef(mod)

# Create new data frame for rankings and random effects.
df <- data.frame("ranking" = 1:16)
row.names(df) <- row.names(output)
df$attack <- re$cond$attack[row.names(df), ]
df$defence <- re$cond$defence[row.names(df), ]
df$difference = df$attack - df$defence
df
```

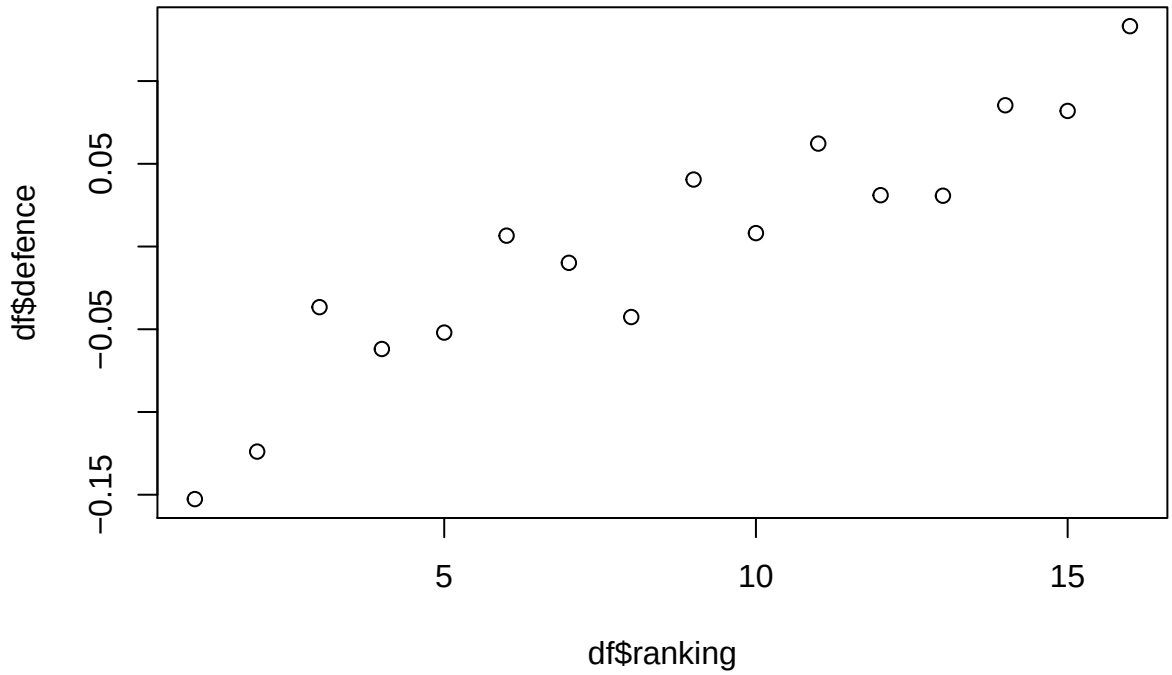
```
##                    ranking       attack      defence   difference
## Rosenborg                1  0.050622609 -0.152631173  0.203253783
## Brann                    2  0.012026209 -0.123934761  0.135960970
## Molde                    3  0.078390643 -0.036630979  0.115021622
## Haugesund                4  0.011223106 -0.061931278  0.073154385
## Odd                      5  0.003654179 -0.052013600  0.055667779
## Sarpsborg08              6  0.026946364  0.006574064  0.020372301
## Tromsoe                  7  0.005756700 -0.009852817  0.015609517
## BodoeGlimt               8 -0.036781062 -0.042616090  0.005835029
## Stroemsgodset            9  0.024556017  0.040486666 -0.015930650
## Kristiansund            10 -0.011367328  0.008112432 -0.019479760
## Ranheim_TF              11  0.023375599  0.062209734 -0.038834135
## Vaalerenga              12  0.007147494  0.031030079 -0.023882585
## Lillestroem             13 -0.049915996  0.030699257 -0.080615253
## Stabaek                 14 -0.026801293  0.085376126 -0.112177420
## Start                   15 -0.060500163  0.081958112 -0.142458276
## Sandefjord_Fotball      16 -0.058333079  0.133164228 -0.191497307
```

Next, we do some plotting:

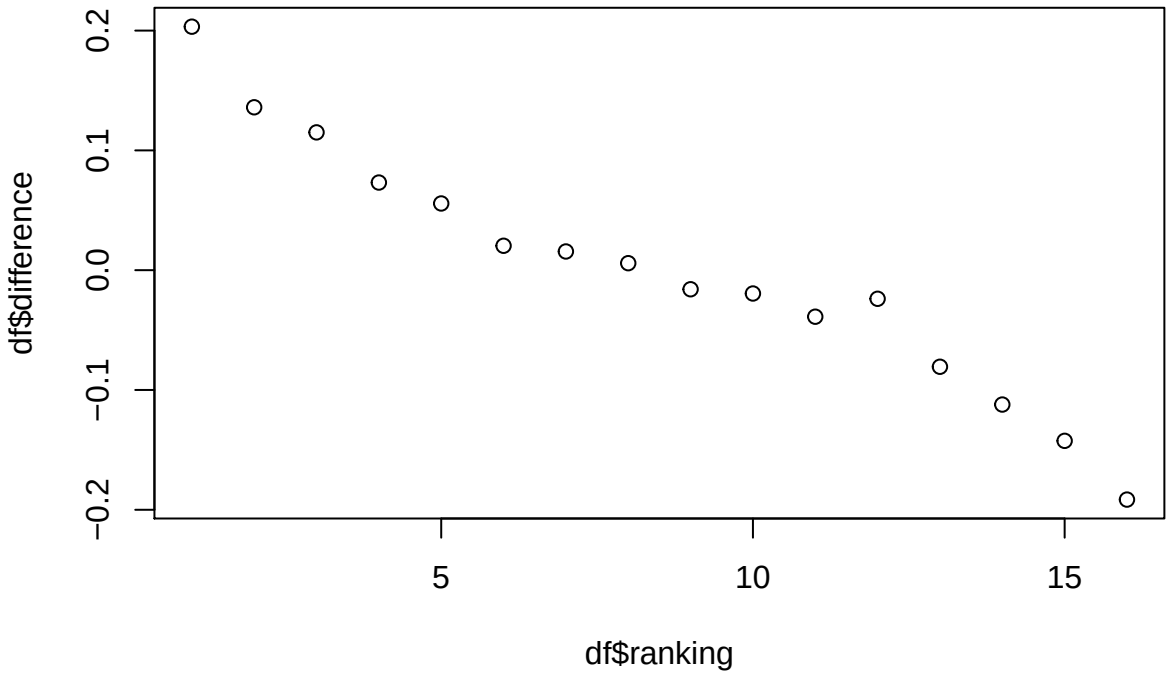
```
plot(df$ranking, df$attack)
```




```
plot(df$ranking, df$defence)
```




```
plot(df$ranking, df$difference)
```



From the plots, there seems to be a linear relationship between both the random effects and the simulated ranking. The linear relationship is clearly visible when we plot the difference against the ranking.