Binomial glm

Greg Deakin

The Date

Load required libraries

```
library(data.table)
library(tidyverse)
library(statfuncs)
library(emmeans)
```

Custom functions

Example data

```
dat <- data.table(
  block=as.factor(1:25),
  treatment = as.factor(c("hand","mangle","spinner")),
  survived = sapply(1:75,function(i)rbinom(1,20,0.9/(i%%3+1))),
  total=20
)
```

Load data

Explorartory analysis

Check for balance

```
dat[,(.N),by=treatment]
```

```
##    treatment V1
## 1:      hand 25
## 2:    mangle 25
## 3:   spinner 25
```

```
dat[,(.N),by=block]
```

```
##    block V1
## 1:     1  3
## 2:     2  3
## 3:     3  3
## 4:     4  3
## 5:     5  3
## 6:     6  3
## 7:     7  3
## 8:     8  3
## 9:     9  3
```

```
## 10:     10  3
## 11:     11  3
## 12:     12  3
## 13:     13  3
## 14:     14  3
## 15:     15  3
## 16:     16  3
## 17:     17  3
## 18:     18  3
## 19:     19  3
## 20:     20  3
## 21:     21  3
## 22:     22  3
## 23:     23  3
## 24:     24  3
## 25:     25  3
##     block V1
```
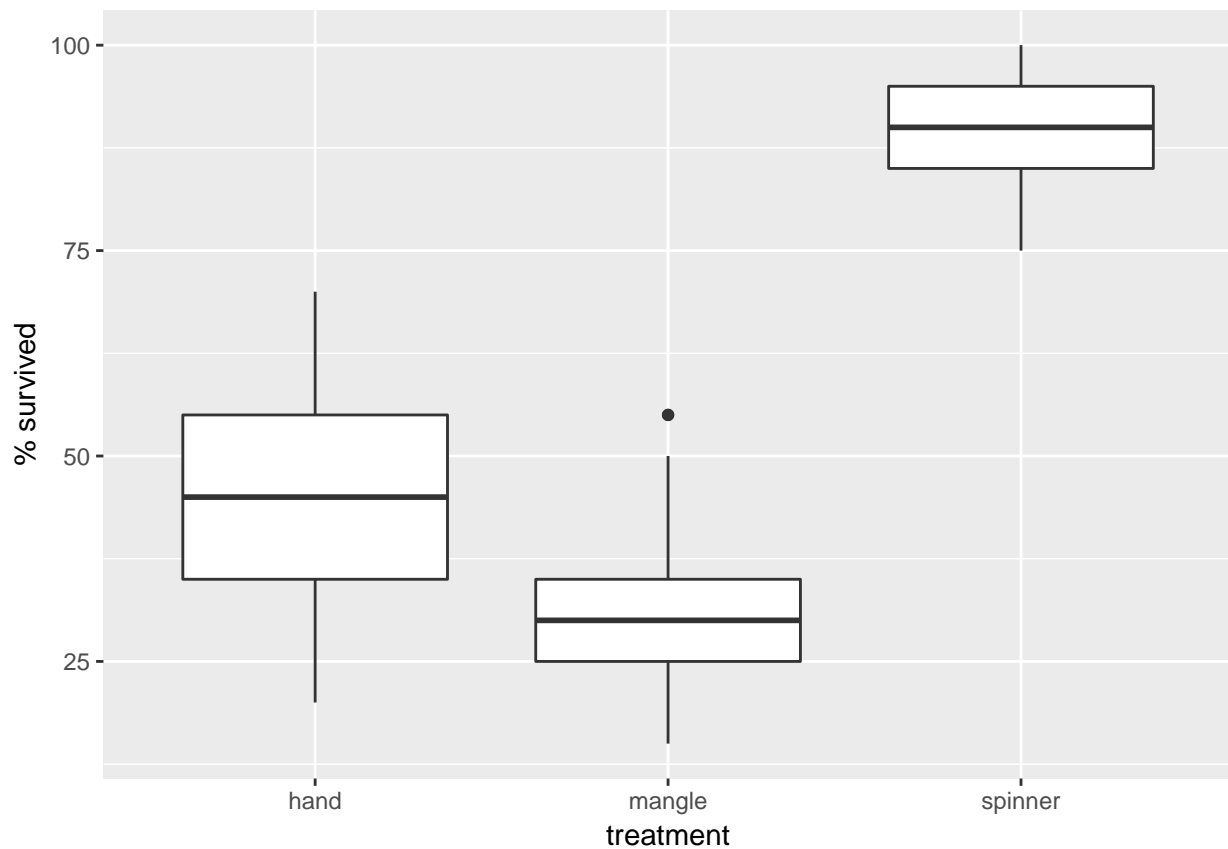
Summary statistics

```
dat[,as.list(summary(survived/total*100)),by=treatment]
```

```
##    treatment Min. 1st Qu. Median Mean 3rd Qu. Max.
## 1:      hand   20      35     45 46.0      55   70
## 2:    mangle   15      25     30 31.8      35   55
## 3:   spinner   75      85     90 88.0      95  100
```

Plots

```
g <- ggplot(dat,aes(x=treatment, y=(survived/total*100)))+ylab("% survived")
g + geom_boxplot()
```

Analysis

```r
model <- glm(survived/total~block+treatment,
        data=dat,
        family=binomial(link="logit"),
        weights=total)
```

model summary

```r
glm.sum(model)
```

```
##              d.f  deviance mean.deviance deviance.ratio     Chisq.p
## Regression    26 411.94423     15.844009       13.49728 4.557682e-71
## Residual      48  56.34561      1.173867             NA           NA
## Total         74 468.28983      6.328241             NA           NA
##                      F.p
## Regression 2.437295e-14
## Residual            NA
## Total               NA
```

model coefficients

```
summary(model)
```

```
## 
## Call:
## glm(formula = survived/total ~ block + treatment, family = binomial(link = "logit"),
##     data = dat, weights = total)
## 
## Deviance Residuals:
##      Min       1Q    Median       3Q       Max
## -1.83356  -0.61679  -0.00016   0.64872   1.96892
## 
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.62577    0.31170  -2.008  0.04468 *
## block2             0.36091    0.42541   0.848  0.39623
## block3             0.71400    0.42504   1.680  0.09299 .
## block4             0.62584    0.42479   1.473  0.14067
## block5             0.62584    0.42479   1.473  0.14067
## block6             1.07159    0.42882   2.499  0.01246 *
## block7             0.89133    0.42634   2.091  0.03656 *
## block8             0.18209    0.42687   0.427  0.66969
## block9             0.18209    0.42687   0.427  0.66969
## block10            0.53773    0.42477   1.266  0.20553
## block11            0.44948    0.42498   1.058  0.29021
## block12            0.09153    0.42789   0.214  0.83062
## block13            0.18209    0.42687   0.427  0.66969
## block14            0.27183    0.42604   0.638  0.52345
## block15            0.53773    0.42477   1.266  0.20553
## block16            0.27183    0.42604   0.638  0.52345
## block17            0.98097    0.42742   2.295  0.02173 *
## block18           -0.09264    0.43049  -0.215  0.82961
## block19            0.18209    0.42687   0.427  0.66969
## block20            0.27183    0.42604   0.638  0.52345
## block21            0.18209    0.42687   0.427  0.66969
## block22            0.71400    0.42504   1.680  0.09299 .
## block23            1.25703    0.43273   2.905  0.00367 **
## block24            0.53773    0.42477   1.266  0.20553
## block25            0.53773    0.42477   1.266  0.20553
## treatmentmangle   -0.61906    0.13331  -4.644 3.42e-06 ***
## treatmentspinner   2.19683    0.16630  13.210  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 468.290  on 74  degrees of freedom
## Residual deviance:  56.346  on 48  degrees of freedom
## AIC: 338.33
## 
## Number of Fisher Scoring iterations: 4
```

Correction for overdispersion

```
# model <- update(model,family=quasibinomial)
```

Anova

```
anova(model,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: survived/total
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                        74     468.29
## block     24    24.26       50     444.03   0.4467
## treatment  2   387.68       48      56.35   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

post hoc tests

```
emmeans(model,pairwise~treatment)
```

```
## $emmeans
##  treatment     emmean         SE  df  asymp.LCL  asymp.UCL
##  hand       -0.1630997 0.09099798 Inf -0.3414525  0.0152530
##  mangle     -0.7821610 0.09749801 Inf -0.9732536 -0.5910684
##  spinner     2.0337279 0.13910788 Inf  1.7610814  2.3063743
##
## Results are averaged over the levels of: block
## Results are given on the logit (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##  contrast           estimate        SE  df z.ratio p.value
##  hand - mangle     0.6190612 0.1333125 Inf   4.644  <.0001
##  hand - spinner   -2.1968276 0.1663035 Inf -13.210  <.0001
##  mangle - spinner -2.8158889 0.1704535 Inf -16.520  <.0001
##
## Results are averaged over the levels of: block
## Results are given on the log odds ratio (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Alternative analysis with brglm2

This is sometimes required if there is a factror equal to 0 or 1 There is an alternative FB/XX method for calculating p values fom the deviance - this to be added here #### Load required libraries

```r
library(brglm2)
```

update model

```r
model <- update(model,method="brglmFit")
```

post hoc tests

```r
emmeans(model,pairwise~treatment)
```

```
## $emmeans
##  treatment     emmean         SE  df  asymp.LCL  asymp.UCL
##  hand       -0.1594409 0.09093346 Inf -0.3376672  0.0187854
##  mangle     -0.7651285 0.09715451 Inf -0.9555479 -0.5747092
##  spinner     1.9932752 0.13706838 Inf  1.7246261  2.2619243
##
## Results are averaged over the levels of: block
## Results are given on the logit (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##  contrast          estimate        SE  df z.ratio p.value
##  hand - mangle    0.6056876 0.1330219 Inf   4.553  <.0001
##  hand - spinner  -2.1527162 0.1645588 Inf -13.082  <.0001
##  mangle - spinner -2.7584038 0.1685504 Inf -16.365  <.0001
##
## Results are averaged over the levels of: block
## Results are given on the log odds ratio (not the response) scale.
## P value adjustment: tukey method for comparing a family of 3 estimates
```