Analysis of binary repeated measures data with R

Right-handed basketball players take right and left-handed shots from 3 locations in a different random order for each player. Hit or miss is recorded. This is a 2x3 factorial design with repeated measures on both factors: Hand they are shooting with and spot on the court.

```R
> rm(list=ls()); options(scipen=999) # To avoid scientific notation
> # Install packages if necessary. Only need to do this once.
> # install.packages("lme4")
> # install.packages("car")
> # Library packages -- do this every time
> library(lme4) # For lmer function
> library(car)  # For F-tests, likelihood ratio and Wald chi-squared tests
> # Read data into a data frame
> bball = read.table("http://www.utstat.toronto.edu/~brunner/data/legal/Bball1.data.txt")
> head(bball,12); attach(bball)

Subject Hand      Spot   Hit
 1     1 Lhand LeftBaseline      0
 2     1 Lhand     Middle       0
 3     1 Lhand RightBaseline    1
 4     1 Rhand LeftBaseline     1
 5     1 Rhand     Middle       1
 6     1 Rhand RightBaseline    1
 7     2 Lhand LeftBaseline     0
 8     2 Lhand     Middle       1
 9     2 Lhand RightBaseline    0
10     2 Rhand LeftBaseline     1
11     2 Rhand     Middle       1
12     2 Rhand RightBaseline    1

> # Sample sizes
> table(Hand,Spot)

Spot
Hand       LeftBaseline Middle RightBaseline
Lhand      58      58          58
Rhand      58      58          58

> # Frequency table
> net = table(Hit,Spot,Hand); net
,, Hand = Lhand

Spot
  Hit LeftBaseline Middle RightBaseline
0     39       27      47
1     19       31      11

,, Hand = Rhand

Spot
  Hit LeftBaseline Middle RightBaseline
0     31       17      24
1     27       41      34
```
> prophitleft = prop.table(net[,1],2); prophitleft # Col % on first sub-table
  Spot Hit LeftBaseline Middle RightBaseline
  0   0.6724138 0.4655172 0.8103448
  1   0.3275862 0.5344828 0.1896552
> prophitright = prop.table(net[,2],2); prophitright # Col % on second sub-table
  Spot Hit LeftBaseline Middle RightBaseline
  0   0.5344828 0.2931034 0.4137931
  1   0.4655172 0.7068966 0.5862069

> # 2 x 3 table of percentage hits
> # Labels
> handval = as.character(sort(unique(Hand))); handval
[1] "Lhand" "Rhand"
> spotval = as.character(sort(unique(Spot))); spotval
[1] "LeftBaseline" "Middle" "RightBaseline"
> PercentHits = rbind(prophitleft[,2],prophitright[,2]) * 100
> PercentHits = round(PercentHits,2)
> rownames(PercentHits) = handval; colnames(PercentHits) = spotval
> PercentHits

  LeftBaseline Middle RightBaseline
 Lhand        32.76  53.45         18.97
 Rhand        46.55  70.69         58.62

> # Set contrasts of the factors to contr.sum (effect coding)
> contrasts(Hand) = contr.sum(2); contrasts(Spot) = contr.sum(3)
>
> # Naive logistic regression, ignoring repeated measures
> naiveglm = glm(Hit ~ Spot*Hand, family=binomial)
> # summary(naiveglm) # Optional
> # Likelihood ratio tests
> anova(naiveglm, test="Chisq") # Terms added sequentially first to last
Analysis of Deviance Table
Model: binomial, link: logit

Response: Hit

Terms added sequentially (first to last)

        Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL     347     481.04
Spot 2 16.3131       345     464.73   0.0002869 ***
Hand 1 20.5967       344     444.13 0.000005669 ***
Spot:Hand 2 5.3207       342     438.81 0.0699225 .
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Anova(naiveglm, type="III") # Each term is last
Analysis of Deviance Table (Type III tests)

Response: Hit

    LR Chisq Df Pr(>Chisq)
Spot 17.6956 2 0.0001437 ***
Hand 21.0853 1 0.000004393 ***
Spot:Hand 5.3207 2 0.0699225 .
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
# Fit a mixed model
hoops = glmer(Hit ~ Spot*Hand + (1 | Subject), family=binomial)

# summary(hoops)
Anova(hoops, type="III") # Each effect controlled for all others

Analysis of Deviance Table (Type III Wald chisquare tests)

Response: Hit

<table>
<thead>
<tr>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.0290</td>
<td>1 0.31041</td>
</tr>
<tr>
<td>Spot</td>
<td>19.4143</td>
<td>2 0.000060847 ***</td>
</tr>
<tr>
<td>Hand</td>
<td>22.7883</td>
<td>1 0.000001809 ***</td>
</tr>
<tr>
<td>Spot:Hand</td>
<td>6.0659</td>
<td>2 0.04817 *</td>
</tr>
</tbody>
</table>

---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Because the interaction is significant but close, try a likelihood ratio test
# Fit a model with main effects but no interaction, and compare
restrictedmodel = glmer(Hit ~ Spot+Hand + (1 | Subject), family=binomial)
anova(restrictedmodel,hoops)

Data: NULL
Models:
  restrictedmodel: Hit ~ Spot + Hand + (1 | Subject)
  hoops: Hit ~ Spot * Hand + (1 | Subject)

<table>
<thead>
<tr>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>restrictedmodel</td>
<td>5</td>
<td>439.20</td>
<td>458.46</td>
<td>-214.60</td>
<td>429.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hoops</td>
<td>7</td>
<td>436.86</td>
<td>463.83</td>
<td>-211.43</td>
<td>422.86</td>
<td>6.3371</td>
<td>2 0.04206 *</td>
</tr>
</tbody>
</table>

---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Want mean estimated log odds and pairwise comparisons.
X = model.matrix(hoops) # The X matrix!
sumh = summary(hoops)
betahat = cbind(sumh$coef[,1]) # Estimated fixed effects as a column vector.
estlogodds = X %*% betahat # Estimated pop mean log odds for each observation.

# Display estimated pop mean log odds by Hand and Spot
aggmeans = aggregate(estlogodds, by = list(Hand,Spot), FUN = mean)
aggmeans

Group.1 Group.2  V1
1  Lhand LeftBaseline -0.8721978
2  Rhand LeftBaseline -0.1669893
3   Lhand Middle  0.1708591
4  Rhand Middle  1.0680078
5  Lhand RightBaseline -1.7338464
6  Rhand RightBaseline  0.4272533

# aggregate(estlogodds, by = list(Hand,Spot), FUN = sd) # Check, all zeros
treatmeans = aggmeans[,3]
dim(treatmeans) = c(2,3)
rownames(treatmeans) = rownames(PercentHits)
colnames(treatmeans) = colnames(PercentHits)
> addmargins(treatmeans,FUN=mean) # With marginal means
Margins computed over dimensions
in the following order:
1:
   LeftBaseline    Middle RightBaseline       mean
Lhand   -0.8721978 0.1708591    -1.7338464 -0.8117284
Rhand   -0.1669893 1.0680078     0.4272533  0.4427573
mean   -0.5195935 0.6194335    -0.6532965 -0.1844855
> addmargins(PercentHits,FUN=mean) # With marginal means
Margins computed over dimensions
in the following order:
1:
   LeftBaseline Middle RightBaseline  mean
Lhand       32.760  53.45        18.970 35.06
Rhand       46.550  70.69        58.620 58.62
mean        39.655  62.07        38.795 46.84
>
>
> DiffHand = treatmeans[2,] - treatmeans[1,] # Advantage of Right hand
(interaction)
> rbind(treatmeans,DiffHand)
LeftBaseline    Middle RightBaseline
Lhand      -0.8721978 0.1708591    -1.7338464
Rhand      -0.1669893 1.0680078     0.4272533
DiffHand    0.7052085 0.8971486     2.1610997
> # Advantage of the right hand appears to be greatest on the right baseline
> # Compare percentage hits
> Radvantage = PercentHits[2,] - PercentHits[1,]
> rbind(PercentHits,Radvantage)
LeftBaseline Middle RightBaseline
Lhand             32.76  53.45         18.97
Rhand             46.55  70.69         58.62
Radvantage        13.79  17.24         39.65
> # Same general story. This is reassuring.
> # Testing pairwise differences.
> # First make a combination variable. The combination variable HandSpot
> # will have 6 values
>
> n = length(Hit); n
[1] 348
> HandSpot = character(n) # A character-valued variable of length n
> for(j in 1:n) HandSpot[j] = paste(Hand[j],Spot[j],sep='')
> freq = table(HandSpot); cbind(freq)
freq
LhandLeftBaseline 58
LhandMiddle 58
LhandRightBaseline 58
RhandLeftBaseline 58
RhandMiddle 58
RhandRightBaseline 58
>
> # Look at a summary of this one-way
> summary( glmer(Hit ~ HandSpot + (1 | Subject), family=binomial) )

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
  Family: binomial ( logit )
  Formula: Hit ~ HandSpot + (1 | Subject)

AIC      BIC   logLik deviance df.resid
436.9    463.8   -211.4    422.9      341

Scaled residuals:
     Min      1Q  Median      3Q     Max
-1.7842 -0.7273 -0.3379  0.6988  3.5638

Random effects:
  Groups   Name        Variance Std.Dev.
    Subject (Intercept) 0.9844   0.9922
Number of obs: 348, groups:  Subject, 58

Fixed effects:
                         Estimate Std. Error   z value  Pr(>|z|)
(Intercept)                -0.8722     0.3362  -2.595   0.00947 **
HandSpotLhandMiddle        1.0431     0.4258   2.449   0.01431 *
HandSpotLhandRightBaseline -0.8616     0.4735  -1.820   0.06877 .
HandSpotRhandLeftBaseline  0.7052     0.4234   1.666   0.09576 .
HandSpotRhandMiddle        1.9402     0.4527   4.286 0.0000182 ***
HandSpotRhandRightBaseline 1.2994     0.4305   3.018   0.00254 **

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
HndSptLhndM           -0.674
HndSptLhndR           -0.573  0.449
HndSptRhnL            -0.671  0.532  0.456
HndSptRhnM            -0.652  0.521  0.412  0.512
HndSptRhnRB           -0.672  0.535  0.441  0.530  0.524

> # Note how the intercept matches with estimated mean log odds for
> # Left hand, left baseline. Good.

> # Now make a matrix, number of rows = number of columns = number of treatments
> # Upper triangle will have z statistics for pairwise tests
> # Lower triangle will have unadjusted p-values
> handspot = factor(HandSpot) # Dummy variable coding will change in the loop
> ntreat = length(unique(HandSpot))
> PairWize = diag(ntreat) # Ones on main diagonal
> rownames(PairWize) = sort(unique(HandSpot))
> colnames(PairWize) = c("LL","LM","LR","RL","RM","RR")
> for(i in 1:(ntreat-1))
>   { 
>     contrasts(handspot) = contr.treatment(ntreat, base = i) # i is reference category
>     model = glmer(Hit ~ handspot + (1 | Subject), family=binomial)
>     ztable = summary(model)$coef
>     PairWize[i,(i+1):ntreat] = ztable[(i+1):ntreat,3] # z-values
>     PairWize[(i+1):ntreat,i] = ztable[(i+1):ntreat,4] # p-values
>   } # Next i (row)

Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.0285943 (tol = 0.001, component 1)
> # Convergence problems on one of these equivalent models -- the last.
> # Look at the full summary.
> summary( glmer(Hit ~ handspot + (1 | Subject), family=binomial) )
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
'[ 'glmerMod' ']
Family: binomial  ( logit )
Formula: Hit ~ handspot + (1 | Subject)

AIC      BIC   logLik deviance df.resid
436.9    463.8   -211.4    422.9      341

Scaled residuals:
          Min       1Q  Median       3Q      Max
-1.7918  -0.7238  -0.3321   0.6982   3.6153

Random effects:
Groups   Name        Variance Std.Dev.
Subject (Intercept) 1.014    1.007
Number of obs: 348, groups:  Subject, 58

Fixed effects:  
(Intercept)  1.072483   0.002468   434.6 <0.0000000000000002 ***
handspot1   -1.959860   0.002469  -793.9 <0.0000000000000002 ***
handspot2   -0.901053   0.002468  -365.1 <0.0000000000000002 ***
handspot3   -2.825073   0.002468 -1144.5 <0.0000000000000002 ***
handspot4   -1.247175   0.002468  -505.3 <0.0000000000000002 ***
handspot6   -0.644438   0.002468  -261.1 <0.0000000000000002 ***
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr) hndsp1 hndsp2 hndsp3 hndsp4
handspot1 0.000
handspot2 0.000  0.000
handspot3 0.000  0.000  0.000
handspot4 0.000  0.000  0.000  0.000
handspot6 0.000  0.000  0.000  0.000  0.000

convergence code: 0
Model failed to converge with max|grad| = 0.0285943 (tol = 0.001, component 1)

Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  Model failed to converge with max|grad| = 0.0285943 (tol = 0.001, component 1)
>
> # Do the test with a no-intercept model
> noint = glmer(Hit ~ 0 + HandSpot + (1 | Subject), family=binomial)
> L = rbind(c(0,0,0,0,0,1,-1))
> linearHypothesis(noint,L) # Testing H0: L beta = 0
Linear hypothesis test

Hypothesis:
HandSpotRhandMiddle - HandSpotRhandRightBaseline = 0

Model 1: restricted model
Model 2: Hit ~ 0 + HandSpot + (1 | Subject)

Df  Chisq Pr(>Chisq)
1
2 1 2.2065  0.1374
>

Page 6 of 7
> PairWize[5,6] = -sqrt(2.2065); PairWize[6,5] = 0.1374
> # Bonferroni: Compare p-values to alpha/k
> 0.05/15
> [1] 0.003333333
> >
> > round(PairWize,6)

<table>
<thead>
<tr>
<th></th>
<th>LL</th>
<th>LM</th>
<th>LR</th>
<th>RL</th>
<th>RM</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LhandLeftBaseline</td>
<td>1.000000</td>
<td>2.449387</td>
<td>-1.819906</td>
<td>1.665746</td>
<td>4.285583</td>
<td>3.018448</td>
</tr>
<tr>
<td>LhandMiddle</td>
<td>0.014310</td>
<td>1.000000</td>
<td>-4.020645</td>
<td>-0.822777</td>
<td>2.082820</td>
<td>0.621173</td>
</tr>
<tr>
<td>LhandRightBaseline</td>
<td>0.068773</td>
<td>0.000058</td>
<td>1.000000</td>
<td>3.336967</td>
<td>5.575011</td>
<td>4.508563</td>
</tr>
<tr>
<td>RhandLeftBaseline</td>
<td>0.095764</td>
<td>0.410635</td>
<td>0.000847</td>
<td>1.000000</td>
<td>2.848380</td>
<td>1.435711</td>
</tr>
<tr>
<td>RhandMiddle</td>
<td>0.000018</td>
<td>0.037268</td>
<td>0.000000</td>
<td>0.004394</td>
<td>1.000000</td>
<td>-1.485429</td>
</tr>
<tr>
<td>RhandRightBaseline</td>
<td>0.002541</td>
<td>0.534486</td>
<td>0.000007</td>
<td>0.151085</td>
<td>0.137400</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

This document was prepared by Jerry Brunner, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License: [http://creativecommons.org/licenses/by-sa/3.0/deed.en_US](http://creativecommons.org/licenses/by-sa/3.0/deed.en_US). Use any part of it as you like and share the result freely. It is available in OpenOffice.org from the workshop website: [http://www.utstat.toronto.edu/~brunner/workshops/mixed](http://www.utstat.toronto.edu/~brunner/workshops/mixed)