

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: Shooting hand and Spot on the court.

```
> 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")
> # Load packages -- do this every time
> library(lme4) # For lmer function
Loading required package: Matrix
> 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
```

```

>
> # 2 x 3 table of percentage hits
> > phit = tapply(Hit, INDEX=list(Hand,Spot),FUN=mean)
> PercentHits = round(100*phit,2); PercentHits
> PercentHits

      LeftBaseline Middle RightBaseline
Lhand      32.76   53.45         18.97
Rhand      46.55   70.69         58.62

>
> # Set contrasts to contr.sum (effect coding)
> Hand = factor(Hand); contrasts(Hand) = contr.sum(2)
> Spot = factor(Spot); 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)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
Family: binomial ( logit )
Formula: Hit ~ Spot * Hand + (1 | Subject)

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

Scaled residuals:
   Min       1Q   Median       3Q      Max
-1.7841 -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)  -0.1845     0.1819  -1.014   0.3104
Spot1         -0.3351     0.1766  -1.898   0.0578 .
Spot2         0.8039     0.1826   4.403 0.00001068 ***
Hand1        -0.6272     0.1314  -4.774 0.00000181 ***
Spot1:Hand1   0.2746     0.1760   1.560   0.1187
Spot2:Hand1   0.1787     0.1770   1.009   0.3128
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) Spot1 Spot2 Hand1 Sp1:H1
Spot1         -0.030
Spot2         -0.035 -0.465
Hand1         0.056  0.020 -0.163
Spot1:Hand1  -0.012  0.048  0.068 -0.075
Spot2:Hand1  -0.071  0.041  0.009 -0.048 -0.446

> # coefficients(hoops) # Don't go there
> Anova(hoops, type="III") # Each effect controlled for all others

Analysis of Deviance Table (Type III Wald chisquare tests)

Response: Hit
              Chisq Df Pr(>Chisq)
(Intercept)  6.7321  1  0.009469 **
Spot         16.6783  2  0.000239 ***
Hand         2.7745  1  0.095775 .
Spot:Hand    6.0658  2  0.048177 *
---
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) # Compare Wald p-value = .048177
Data: NULL
Models:
restrictedmodel: Hit ~ Spot + Hand + (1 | Subject)
hoops: Hit ~ Spot * Hand + (1 | Subject)

```

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
restrictedmodel	5	439.20	458.46	-214.60	429.20				
hoops	7	436.86	463.83	-211.43	422.86	6.3371		2	0.04206 *

```

---
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!
> head(X)
  (Intercept) Spot1 Spot2 Hand1 Spot1:Hand1 Spot2:Hand1
1            1     1     0     1            1            0
2            1     0     1     1            0            1
3            1    -1    -1     1           -1           -1
4            1     1     0    -1           -1            0
5            1     0     1    -1            0           -1
6            1    -1    -1    -1            1           1

```

```
> sumh = summary(hoops); sumh
```

Showing just part of the output. This table is called `sumh$coef`

```
Fixed effects:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.1845    0.1819  -1.014   0.3104
Spot1       -0.3351    0.1766  -1.898   0.0578 .
Spot2        0.8039    0.1826   4.403 0.00001068 ***
Hand1       -0.6272    0.1314  -4.774 0.00000181 ***
Spot1:Hand1  0.2746    0.1760   1.560   0.1187
Spot2:Hand1  0.1787    0.1770   1.009   0.3128
```

```
> betahat = cbind(sumh$coef[,1]) # Estimated fixed effects as a column vector.
> betahat
```

```
      [,1]
(Intercept) -0.1844855
Spot1       -0.3351080
Spot2        0.8039190
Hand1       -0.6272428
Spot1:Hand1  0.2746385
Spot2:Hand1  0.1786685
```

```
> estlogodds = X %*% betahat # Estimated pop mean log odds for each observation.
```

```

> # Display estimated pop mean log odds by Hand and Spot
> meanlodds = tapply(estlogodds, INDEX=list(Hand,Spot),FUN=mean)
> # round(meanlodds,3)
> round(addmargins(meanlodds,FUN=mean, quiet=TRUE),3)
  LeftBaseline Middle RightBaseline mean
Lhand      -0.872  0.171          -1.734 -0.812
Rhand      -0.167  1.068           0.427  0.443
mean        -0.520  0.619          -0.653 -0.184

> # Corresponding probabilities
> round(exp(meanlodds)/(1+exp(meanlodds)),3)
  LeftBaseline Middle RightBaseline
Lhand      0.295  0.543           0.150
Rhand      0.458  0.744           0.605

> # Compare raw proportion of hits
> PercentHits

  LeftBaseline Middle RightBaseline
Lhand      32.76  53.45           18.97
Rhand      46.55  70.69           58.62

> DiffHand = meanlodds[2,] - meanlodds[1,] # Advantage of Right hand (interaction)
> round(rbind(meanlodds,DiffHand),3)

  LeftBaseline Middle RightBaseline
Lhand      -0.872  0.171          -1.734
Rhand      -0.167  1.068           0.427
DiffHand    0.705  0.897           2.161
> # 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
> HandSpot = paste(Hand,Spot,sep='')
> cbind(table(HandSpot)) # View table in one column instead of one row
      [,1]
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  1.82e-05 ***
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:
      (Intr) HndSLM HnSLRB HnSRLB HndSRM
HndSptLhndM -0.674
HndSptLhnrB -0.573  0.449
HndSptRhnLB -0.671  0.532  0.456
HndSptRhndM -0.652  0.521  0.412  0.512
HndSptRhnrB -0.672  0.535  0.441  0.530  0.524
>
> meanlodds
      LeftBaseline      Middle RightBaseline
Lhand  -0.8721978  0.1708591  -1.7338464
Rhand  -0.1669893  1.0680078   0.4272533
> # Note how the intercept matches with estimated mean log odds for
> # Left hand, left baseline (alphabetically first). 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 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 = 0.002, 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:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.072483   0.002468   434.6   <2e-16 ***
handspot1    -1.959860   0.002469  -793.9   <2e-16 ***
handspot2    -0.901053   0.002468  -365.1   <2e-16 ***
handspot3    -2.825073   0.002468 -1144.5   <2e-16 ***
handspot4    -1.247175   0.002468  -505.3   <2e-16 ***
handspot6    -0.644438   0.002468  -261.1   <2e-16 ***
---
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
optimizer (Nelder_Mead) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.0285943 (tol = 0.002, component 1)

```

```
Warning message:
In checkConv(attr("opt", "derivs"), opt$par, ctrl = control$checkConv, :
Model failed to converge with max|grad| = 0.0285943 (tol = 0.002, component 1)
> contrasts(handspot)
```

```
      1 2 3 4 6
LhandLeftBaseline 1 0 0 0 0
LhandMiddle       0 1 0 0 0
LhandRightBaseline 0 0 1 0 0
RhandLeftBaseline 0 0 0 1 0
RhandMiddle       0 0 0 0 0
RhandRightBaseline 0 0 0 0 1
```

```
> tryagain = glmer(Hit ~ handspot + (1 | Subject), family=binomial, nAGQ=2)
> summary(tryagain) # Reference category is 5 = RhandMiddle
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite
Quadrature,
nAGQ = 2) [glmerMod]
Family: binomial (logit )
Formula: Hit ~ handspot + (1 | Subject)
```

```
      AIC      BIC  logLik deviance df.resid
436.5    463.5   -211.3   422.5     341
```

```
Scaled residuals:
      Min       1Q   Median       3Q      Max
-1.7857 -0.7241 -0.3351  0.6998  3.5850
```

```
Random effects:
Groups Name          Variance Std.Dev.
Subject (Intercept) 1.007    1.004
Number of obs: 348, groups: Subject, 58
```

```
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.0687     0.3473   3.078 0.00209 **
handspot1    -1.9417     0.4548  -4.270 1.96e-05 ***
handspot2    -0.8977     0.4319  -2.078 0.03769 *
handspot3    -2.8056     0.5051  -5.554 2.79e-08 ***
handspot4    -1.2357     0.4351  -2.840 0.00451 **
handspot6    -0.6412     0.4324  -1.483 0.13810
```

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

```
> round(PairWize,3)
```

```
      LL      LM      LR      RL      RM      RR
LhandLeftBaseline 1.000 2.449 -1.820 1.666 4.286 3.018
LhandMiddle       0.014 1.000 -4.021 -0.823 2.083 0.621
LhandRightBaseline 0.069 0.000 1.000 3.337 5.575 4.509
RhandLeftBaseline 0.096 0.411 0.001 1.000 2.848 1.436
RhandMiddle       0.000 0.037 0.000 0.004 1.000 -261.084
RhandRightBaseline 0.003 0.534 0.000 0.151 0.000 1.000
```

```
> PairWize[5,6] = -1.483; PairWize[6,5] = 0.1381
```

```
> round(PairWize,3) # Positive z numbers mean the columns have an advantage
```

```
      LL      LM      LR      RL      RM      RR
LhandLeftBaseline 1.000 2.449 -1.820 1.666 4.286 3.018
LhandMiddle       0.014 1.000 -4.021 -0.823 2.083 0.621
LhandRightBaseline 0.069 0.000 1.000 3.337 5.575 4.509
RhandLeftBaseline 0.096 0.411 0.001 1.000 2.848 1.436
RhandMiddle       0.000 0.037 0.000 0.004 1.000 -1.483
RhandRightBaseline 0.003 0.534 0.000 0.151 0.138 1.000
```


In lmer and glmer, random effects are normally distributed. They are specified by terms of the form (A|B). There can be more than one such expression in the model statement. The random variables from different expressions are independent. Within expressions, they are multivariate normal, with a covariance matrix that needs to be estimated. In (A|B), A is a linear model, specified by the usual model syntax. For example x1+x2 means $b_0 + b_1*x_1 + b_2*x_2 = x'b$, where the $b = (b_0, b_1, b_2)'$ is multivariate normal with mean vector zero. The B part of (A|B) is a "grouping factor." The model has a different independent b vector for each level of B. One could say that B is the way one put subscripts on b.

Separate variances

```
adult = child = numeric(length(Group))
adult[Group=='Adult']=1; child[Group=='Child']=1
```

```
# H0 is that the variance of random shocks is the same for kids and adults
mod1H0 = glmer(Target ~ Group + Language + Group:Language +
               (1|Participant), family=binomial, data=adults5)
```

```
mod1 = glmer(Target ~ Group + Language + Group:Language +
              (0+adult|Participant) + (0+child|Participant),
              family=binomial, data=adults5)
```

```
summary(mod1)
```

```
Anova(mod1, type="III")
```

Testing difference with a LR test. Anova and anova don't work.

```
Gsq = 2 * (logLik(mod1)-logLik(mod1H0)); dfree = 1
```

```
pval = 1-pchisq(Gsq,df=dfree)
```

```
options(scipen=999)
```

```
c(Gsq,dfree,pval) # 30.44195390769801 1.000000000000000 0.00000003440072
```

```
model44 = glmer( AllTargets ~ Group*Item + (1 | Participant), family=binomial,
                nAGQ=1,
```

```
                control = glmerControl(optimizer = "Nelder_Mead", optCtrl =
                list(maxfun=100)) )
```

```
glmerControl(optimizer = c("bobyqa", "Nelder_Mead"))
```

For glmer, if length(optimizer)==2, the first element will be used for the preliminary (random effects parameters only) optimization, while the second will be used for the final (random effects plus fixed effect parameters) phase. See modular for more information on these two phases.

For glmer, the default is optimizer = c("bobyqa", "Nelder_Mead"), so it uses bobyqa and then Nelder_Mead.

Sometimes, control = glmerControl(optimizer = "bobyqa") has worked when the default did not converge. My choices (in quotes) include

```
bobyqa
```

```
Nelder_Mead
```

```
nlminbwrap
```

```
nloptwrap
```

```
mod1 = glmer(Target ~ Group + Language + Condition + Language:Condition +
              (0+adult|Participant) + (0+child|Participant),
              family=binomial, data=a5,
              control = glmerControl(optimizer = "bobyqa")) # Converged!
```

```
# Try a different optimizer. The default is "bobyqa"
model44 = glmer(AllTargets ~ Group*Item + (1 | Participant), family=binomial, nAGQ=1,
               control = glmerControl(optimizer = "Nelder_Mead") )

# Specify maximum number of iterations as well as Nelder_Mead
model44 = glmer(AllTargets ~ Group*Item + (1 | Participant), family=binomial, nAGQ=1,
               control = glmerControl(optimizer = "Nelder_Mead", optCtrl = list(maxfun=100) ))
# Worked!

# Try 100,000 iterations instead of 10,000.
model44 = glmer(AllTargets ~ Group*Item + (1 | Participant), family=binomial, nAGQ=1,
               control = glmerControl( optCtrl = list(maxfun=100000) ))
```

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. Use any part of it as you like and share the result freely. It is available in OpenOffice.org format from the course website: <http://www.utstat.toronto.edu/brunner/oldclass/312f22>