Experiment 2

Load packages and data files

```
library(languageR)
library(Matrix)
library(lme4)
library(lattice)
library(MASS)
library(plyr)
library(tables)

rm(list=ls())

datafile = read.csv("/Users/maryamaljassmi/Arabic Predictability Study/Experiment 2.csv", sep = ",", de
colnames(datafile)
```

```
##  [1] "participant"        "item"               "predictability"
##  [4] "target_word_length" "cloze_scores"       "norm_scores"
##  [7] "ortho_frequency"    "word.class"         "TRT"
## [10] "SFD"                "FFD"                "GD"
## [13] "RP"                 "FFC"                "LP"
## [16] "LS"                 "LS_M"               "SA"
## [19] "FC"                 "RI"                 "RO"
## [22] "RPD"                "SKIP"               "SPILLOVER"
## [25] "Blinks"             "Track_loss"         "Long_saccades"
## [28] "Index"              "remove.trial"
```

```
####  Specify which column your participant, stimuli and condition are in ####
col.subject = 1
col.stim = 2
col.condition = 3

# choose dependent variable
measure = "SKIP"

# add to dataframe
datafile$depvar = datafile[,measure]
```

Assign the correct class

```
# Work out which columns the fixed and random factors are in
datafile$pp = datafile[,col.subject]
datafile$condition = datafile[,col.condition]
```

```r
datafile$stim = datafile[,col.stim]

# make sure all the variables are from the correct class
datafile$depvar = as.numeric(datafile$depvar)
datafile$pp = as.factor(datafile$pp)
datafile$stim = as.factor(datafile$stim)
datafile$condition = as.factor(datafile$condition)


####  Inspect and double check ####
str(datafile)
```

```
## 'data.frame':    2880 obs. of  33 variables:
##  $ participant       : chr  "p1" "p1" "p1" "p1" ...
##  $ item              : chr  "i1" "i2" "i3" "i4" ...
##  $ predictability    : chr  "Predictable" "Predictable" "Predictable" "Unpredictable" ...
##  $ target_word_length: int  4 4 3 3 3 3 3 3 3 3 ...
##  $ cloze_scores      : num  95.8 100 91.7 0 95.8 0 4.2 0 95.8 87.5 ...
##  $ norm_scores       : num  4.7 4.8 4.8 3 4.8 3.9 2.9 3.3 4.8 4.8 ...
##  $ ortho_frequency   : num  42.5 73.1 138.4 34 23.9 ...
##  $ word.class        : chr  "noun" "noun" "noun" "noun" ...
##  $ TRT               : int  NA NA 260 186 244 NA NA NA NA 242 ...
##  $ SFD               : int  NA NA 260 186 244 NA NA NA NA 242 ...
##  $ FFD               : int  NA NA 260 186 244 NA NA NA NA 242 ...
##  $ GD                : int  NA NA 260 186 244 NA NA NA NA 242 ...
##  $ RP                : int  NA NA 0 0 0 NA NA NA NA 0 ...
##  $ FFC               : int  NA NA 1 1 1 NA NA NA NA 1 ...
##  $ LP                : num  NA NA 31.5 22.3 3.4 NA NA NA NA 37.6 ...
##  $ LS                : num  2.6 NA 26.3 15.4 46.6 NA NA 8.3 30.3 46.5 ...
##  $ LS_M              : num  2.6 NA 26.3 15.4 46.6 NA NA 8.3 30.3 46.5 ...
##  $ SA                : num  NA NA 0.94 0.32 0.66 NA NA NA NA 1.24 ...
##  $ FC                : int  0 NA 1 1 1 NA NA 0 0 1 ...
##  $ RI                : int  NA NA 0 0 0 NA NA NA NA 0 ...
##  $ RO                : int  NA NA 0 0 1 NA NA NA NA 0 ...
##  $ RPD               : int  NA NA 260 186 710 NA NA NA NA 242 ...
##  $ SKIP              : int  1 NA 0 0 0 NA NA 1 1 0 ...
##  $ SPILLOVER         : int  NA NA 221 438 245 NA NA NA NA 184 ...
##  $ Blinks            : int  0 1 0 0 0 1 1 0 0 0 ...
##  $ Track_loss        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Long_saccades     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Index             : chr  "p1i50" "p1i58" "p1i3" "p1i4" ...
##  $ remove.trial      : int  0 1 0 0 0 1 1 0 0 0 ...
##  $ depvar            : num  1 NA 0 0 0 NA NA 1 1 0 ...
##  $ pp                : Factor w/ 40 levels "p1","p10","p11",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ condition         : Factor w/ 2 levels "Predictable",..: 1 1 1 2 1 2 2 2 1 1 ...
##  $ stim              : Factor w/ 72 levels "i1","i10","i11",..: 1 12 23 34 45 56 67 71 72 2 ...
```

```r
summary(datafile)
```

```
##  participant            item           predictability     target_word_length
##  Length:2880        Length:2880        Length:2880        Min.   :3.0
##  Class :character   Class :character   Class :character   1st Qu.:3.0
##  Mode  :character   Mode  :character   Mode  :character   Median :3.5
```

```
##                                                          Mean   :3.5
##                                                          3rd Qu.:4.0
##                                                          Max.   :4.0
##
##   cloze_scores     norm_scores    ortho_frequency    word.class
##  Min.   :  0.00   Min.   :2.000   Min.   :   0.75   Length:2880
##  1st Qu.:  0.00   1st Qu.:3.500   1st Qu.:  18.05   Class :character
##  Median : 39.60   Median :4.400   Median :  39.34   Mode  :character
##  Mean   : 45.23   Mean   :4.032   Mean   :  81.04
##  3rd Qu.: 91.70   3rd Qu.:4.700   3rd Qu.:  99.89
##  Max.   :100.00   Max.   :4.900   Max.   :1325.74
##
##       TRT              SFD             FFD              GD
##  Min.   :  85.0   Min.   : 85.0   Min.   : 85.0   Min.   :  85.0
##  1st Qu.: 205.0   1st Qu.:191.0   1st Qu.:190.0   1st Qu.: 197.0
##  Median : 263.0   Median :232.0   Median :231.0   Median : 245.0
##  Mean   : 320.5   Mean   :251.9   Mean   :250.3   Mean   : 279.3
##  3rd Qu.: 383.0   3rd Qu.:287.0   3rd Qu.:287.0   3rd Qu.: 326.0
##  Max.   :1997.0   Max.   :968.0   Max.   :968.0   Max.   :1287.0
##  NA's   :769      NA's   :1124    NA's   :863     NA's   :863
##        RP              FFC             LP              LS
##  Min.   :0.0000   Min.   :1.000   Min.   : 0.10   Min.   :  0.10
##  1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:13.00   1st Qu.: 18.05
##  Median :0.0000   Median :1.000   Median :24.80   Median : 33.20
##  Mean   :0.1294   Mean   :1.139   Mean   :25.07   Mean   : 37.51
##  3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:36.20   3rd Qu.: 53.90
##  Max.   :1.0000   Max.   :5.000   Max.   :73.30   Max.   :430.10
##  NA's   :863      NA's   :863     NA's   :863     NA's   :209
##       LS_M             SA              FC              RI
##  Min.   :  0.10   Min.   :0.200   Min.   :0.000   Min.   :0.0000
##  1st Qu.: 18.50   1st Qu.:1.110   1st Qu.:1.000   1st Qu.:0.0000
##  Median : 33.90   Median :1.430   Median :1.000   Median :0.0000
##  Mean   : 38.89   Mean   :1.458   Mean   :1.051   Mean   :0.0873
##  3rd Qu.: 55.10   3rd Qu.:1.790   3rd Qu.:1.000   3rd Qu.:0.0000
##  Max.   :430.10   Max.   :2.760   Max.   :6.000   Max.   :1.0000
##  NA's   :211      NA's   :769     NA's   :209     NA's   :863
##        RO              RPD            SKIP          SPILLOVER
##  Min.   :0.0000   Min.   :  85   Min.   :0.0000   Min.   : 84.0
##  1st Qu.:0.0000   1st Qu.: 204   1st Qu.:0.0000   1st Qu.:183.0
##  Median :0.0000   Median : 255   Median :0.0000   Median :221.0
##  Mean   :0.0823   Mean   : 319   Mean   :0.2449   Mean   :238.6
##  3rd Qu.:0.0000   3rd Qu.: 358   3rd Qu.:0.0000   3rd Qu.:268.0
##  Max.   :1.0000   Max.   :3101   Max.   :1.0000   Max.   :922.0
##  NA's   :863      NA's   :863    NA's   :209      NA's   :1410
##      Blinks          Track_loss        Long_saccades        Index
##  Min.   :0.00000   Min.   :0.0000000   Min.   :0.00000   Length:2880
##  1st Qu.:0.00000   1st Qu.:0.0000000   1st Qu.:0.00000   Class :character
##  Median :0.00000   Median :0.0000000   Median :0.00000   Mode  :character
##  Mean   :0.05833   Mean   :0.0006944   Mean   :0.01528
##  3rd Qu.:0.00000   3rd Qu.:0.0000000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.0000000   Max.   :1.00000
##
##   remove.trial         depvar             pp              condition
##  Min.   :0.00000   Min.   :0.0000   p1     :  72   Predictable  :1440
```

```
##  1st Qu.:0.00000   1st Qu.:0.0000   p10    :  72   Unpredictable:1440
##  Median :0.00000   Median :0.0000   p11    :  72
##  Mean   :0.07257   Mean   :0.2449   p12    :  72
##  3rd Qu.:0.00000   3rd Qu.:0.0000   p13    :  72
##  Max.   :1.00000   Max.   :1.0000   p14    :  72
##                    NA's   :209      (Other):2448
##       stim
##  i1     :  40
##  i10    :  40
##  i11    :  40
##  i12    :  40
##  i13    :  40
##  i14    :  40
##  (Other):2640
```

```
#find means (per participant, per condition)
mean.tt = tapply(datafile$depvar, list(datafile$pp, datafile$condition), mean, na.rm = T)
mean.tt
```

```
##     Predictable Unpredictable
## p1   0.25925926    0.17857143
## p10  0.24242424    0.19354839
## p11  0.34285714    0.19444444
## p12  0.19230769    0.32142857
## p13  0.29411765    0.09090909
## p14  0.21212121    0.20000000
## p15  0.20000000    0.24242424
## p16  0.37142857    0.26470588
## p17  0.36363636    0.27272727
## p18  0.17142857    0.20588235
## p19  0.12903226    0.05714286
## p2   0.15151515    0.17647059
## p20  0.42857143    0.28571429
## p21  0.47222222    0.29411765
## p22  0.31428571    0.22222222
## p23  0.25714286    0.22857143
## p24  0.08571429    0.09090909
## p25  0.39393939    0.36363636
## p26  0.57142857    0.43333333
## p27  0.16666667    0.09090909
## p28  0.26666667    0.12121212
## p29  0.34482759    0.25000000
## p3   0.16666667    0.13888889
## p30  0.17142857    0.28571429
## p31  0.14285714    0.21875000
## p32  0.11764706    0.18750000
## p33  0.39393939    0.35294118
## p34  0.37142857    0.54285714
## p35  0.16666667    0.11764706
## p36  0.25714286    0.34375000
## p37  0.45714286    0.38888889
## p38  0.20588235    0.14705882
## p39  0.30000000    0.25714286
## p4   0.13888889    0.13888889
```

```
## p40  0.39393939    0.17647059
## p5   0.27777778    0.11428571
## p6   0.22222222    0.08333333
## p7   0.20588235    0.17142857
## p8   0.11764706    0.31250000
## p9   0.46666667    0.12000000
```

```
#find sds (per participant, per condition)
sd.tt = tapply(datafile$depvar, list(datafile$pp, datafile$condition), sd, na.rm = T)
sd.tt
```

```
##     Predictable Unpredictable
## p1    0.4465761     0.3900210
## p10   0.4351941     0.4016097
## p11   0.4815940     0.4013865
## p12   0.4019185     0.4755949
## p13   0.4624973     0.2919371
## p14   0.4151488     0.4058397
## p15   0.4058397     0.4351941
## p16   0.4902409     0.4478111
## p17   0.4885042     0.4522670
## p18   0.3823853     0.4104256
## p19   0.3407771     0.2355041
## p2    0.3641095     0.3869530
## p20   0.5020964     0.4583492
## p21   0.5063094     0.4624973
## p22   0.4710082     0.4216370
## p23   0.4434396     0.4260430
## p24   0.2840286     0.2919371
## p25   0.4961977     0.4885042
## p26   0.5039526     0.5040069
## p27   0.3779645     0.2919371
## p28   0.4497764     0.3314340
## p29   0.4837253     0.4409586
## p3    0.3779645     0.3507362
## p30   0.3823853     0.4583492
## p31   0.3550358     0.4200134
## p32   0.3270350     0.3965578
## p33   0.4961977     0.4850713
## p34   0.4902409     0.5054327
## p35   0.3779645     0.3270350
## p36   0.4434396     0.4825587
## p37   0.5054327     0.4944132
## p38   0.4104256     0.3594906
## p39   0.4660916     0.4434396
## p4    0.3507362     0.3507362
## p40   0.4961977     0.3869530
## p5    0.4542568     0.3228029
## p6    0.4216370     0.2803060
## p7    0.4104256     0.3823853
## p8    0.3270350     0.4709291
## p9    0.5074163     0.3316625
```

```
grand.mean=apply(mean.tt, 2, mean, na.rm = T)
grand.sd=apply(sd.tt,2,mean, na.rm = T)
grand.se=grand.sd/sqrt(40)  ## this number is the total number of participants from your data, so it is
summary.ds = rbind(grand.mean, grand.sd, grand.se)
summary.ds
```

Descriptive statistics summary

```
##            Predictable Unpredictable
## grand.mean  0.27013550    0.22192317
## grand.sd    0.43083004    0.40251802
## grand.se    0.06812021    0.06364369
```

```
####  Setting contrasts and table of means ####

contrasts(datafile$condition) <- contr.sdif(2)
(table1 <- ddply(datafile, .(predictability), summarise, M=mean(depvar, na.rm = TRUE), SD=sd(depvar, na
```

```
##   predictability        M       SD    N        SE
## 1    Predictable 0.2681105 0.4431410 1339 0.01211021
## 2  Unpredictable 0.2214715 0.4153931 1332 0.01138170
```

```
# Model 1
depvar.glmeM1 = glmer(depvar ~  predictability + (1 + predictability|participant) + (1 + predictability
print(depvar.glmeM1, corr = FALSE)
```

GLME Models

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 + predictability | participant) +
##     (1 + predictability | item)
##    Data: datafile
##       AIC       BIC    logLik  deviance  df.resid
##  2781.539  2828.661 -1382.770  2765.539      2663
## Random effects:
##  Groups      Name                        Std.Dev. Corr
##  item        (Intercept)                 0.7555
##              predictabilityUnpredictable 0.7493   -0.18
##  participant (Intercept)                 0.5109
##              predictabilityUnpredictable 0.2633   -0.08
## Number of obs: 2671, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                     -1.1706                      -0.3887
```

```r
summary(depvar.glmeM1, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 + predictability | participant) +
##     (1 + predictability | item)
##    Data: datafile
##
##      AIC      BIC   logLik deviance df.resid
##   2781.5   2828.7  -1382.8   2765.5     2663
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5978 -0.5534 -0.3943 -0.2093  4.4876
##
## Random effects:
##  Groups      Name                        Variance Std.Dev. Corr
##  item        (Intercept)                 0.57071  0.7555
##              predictabilityUnpredictable 0.56146  0.7493   -0.18
##  participant (Intercept)                 0.26107  0.5109
##              predictabilityUnpredictable 0.06931  0.2633   -0.08
## Number of obs: 2671, groups:  item, 72; participant, 40
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.1706     0.1400  -8.363  < 2e-16 ***
## predictabilityUnpredictable  -0.3887     0.1472  -2.641  0.00826 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Model 2
depvar.glmeM2 = glmer(depvar ~  predictability + (1 + predictability|participant) + (1 |item), datafile
print(depvar.glmeM2, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 + predictability | participant) +
##     (1 | item)
##    Data: datafile
##       AIC       BIC    logLik  deviance  df.resid
##  2794.562  2829.903 -1391.281  2782.562      2665
## Random effects:
##  Groups      Name                        Std.Dev. Corr
##  item        (Intercept)                 0.7620
##  participant (Intercept)                 0.5157
##              predictabilityUnpredictable 0.2463   -0.21
## Number of obs: 2671, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                     -1.1678                      -0.3038
```

```r
summary(depvar.glmeM2, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 + predictability | participant) +
##     (1 | item)
##    Data: datafile
##
##      AIC      BIC   logLik deviance df.resid
##   2794.6   2829.9  -1391.3   2782.6     2665
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8788 -0.5563 -0.4102 -0.2256  3.6274
##
## Random effects:
##  Groups      Name                        Variance Std.Dev. Corr
##  item        (Intercept)                 0.58060  0.7620
##  participant (Intercept)                 0.26591  0.5157
##              predictabilityUnpredictable 0.06068  0.2463   -0.21
## Number of obs: 2671, groups:  item, 72; participant, 40
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.1678     0.1397  -8.358  < 2e-16 ***
## predictabilityUnpredictable  -0.3038     0.1060  -2.866  0.00415 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Model 3
depvar.glmeM3 = glmer(depvar ~  predictability + (1 |participant) + (1 + predictability|item), datafile
print(depvar.glmeM3, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 | participant) + (1 + predictability |
##     item)
##    Data: datafile
##       AIC       BIC    logLik  deviance  df.resid
##  2778.237  2813.578 -1383.119  2766.237      2665
## Random effects:
##  Groups      Name                        Std.Dev. Corr
##  item        (Intercept)                 0.7556
##              predictabilityUnpredictable 0.7448   -0.19
##  participant (Intercept)                 0.5143
## Number of obs: 2671, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                      -1.171                       -0.375
```

```r
summary(depvar.glmeM3, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 | participant) + (1 + predictability |
##     item)
##    Data: datafile
##
##      AIC      BIC   logLik deviance df.resid
##   2778.2   2813.6  -1383.1   2766.2     2665
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.6899 -0.5517 -0.3942 -0.2147  4.4129
##
## Random effects:
##  Groups      Name                        Variance Std.Dev. Corr
##  item        (Intercept)                 0.5709   0.7556
##              predictabilityUnpredictable 0.5547   0.7448   -0.19
##  participant (Intercept)                 0.2645   0.5143
## Number of obs: 2671, groups:  item, 72; participant, 40
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.1712     0.1398  -8.378  < 2e-16 ***
## predictabilityUnpredictable  -0.3750     0.1376  -2.724  0.00644 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Model 4
depvar.glmeM4 = glmer(depvar ~  predictability + (1 |participant) + (1 |item), datafile, family = binom
print(depvar.glmeM4, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 | participant) + (1 | item)
##    Data: datafile
##       AIC       BIC    logLik  deviance  df.resid
##  2791.052  2814.613 -1391.526  2783.052      2667
## Random effects:
##  Groups      Name        Std.Dev.
##  item        (Intercept) 0.7587
##  participant (Intercept) 0.5034
## Number of obs: 2671, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                     -1.1649                      -0.3008
```

```r
summary(depvar.glmeM4, corr = FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: depvar ~ predictability + (1 | participant) + (1 | item)
##    Data: datafile
##
##      AIC      BIC   logLik deviance df.resid
##   2791.1   2814.6  -1391.5   2783.1     2667
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8861 -0.5590 -0.4112 -0.2277  3.6296
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  item        (Intercept) 0.5757   0.7587
##  participant (Intercept) 0.2534   0.5034
## Number of obs: 2671, groups:  item, 72; participant, 40
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.16486    0.13797  -8.443  < 2e-16 ***
## predictabilityUnpredictable -0.30077    0.09603  -3.132  0.00174 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```