Experiment 1

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 1.csv", sep = ",", de
colnames(datafile)
```

```
##  [1] "participant"             "item"
##  [3] "predictability"          "target_word_length"
##  [5] "ortho_frequency"         "cloze_scores"
##  [7] "norm_scores"             "word.class"
##  [9] "TRT"                     "SFD"
## [11] "FFD"                     "GD"
## [13] "RP"                      "FFC"
## [15] "LP"                      "LS"
## [17] "LS_including_skipped_TW" "SA"
## [19] "FC"                      "RI"
## [21] "RO"                      "RPD"
## [23] "SKIP"                    "SPILLOVER"
## [25] "Blinks"                  "Track_loss"
## [27] "Long_saccades"           "Index"
## [29] "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]
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" "Unpredictable" "Predictable" "Unpredictable" ...
##  $ target_word_length     : int  7 6 5 6 7 8 8 6 6 6 ...
##  $ ortho_frequency        : num  109.1 101.6 30.7 10.8 17.4 ...
##  $ cloze_scores           : num  100 0 100 0 91.7 91.7 79.2 100 66.7 0 ...
##  $ norm_scores            : num  4.8 3.5 4.9 3.3 4.7 4.5 4.9 4.6 4.8 3.5 ...
##  $ word.class             : chr  "noun" "noun" "noun" "noun" ...
##  $ TRT                    : int  228 NA 392 NA 411 225 NA 187 207 499 ...
##  $ SFD                    : int  228 NA 392 NA 411 225 NA 187 207 NA ...
##  $ FFD                    : int  228 NA 392 NA 411 225 NA 187 207 320 ...
##  $ GD                     : int  228 NA 392 NA 411 225 NA 187 207 499 ...
##  $ RP                     : int  0 NA 0 NA 0 0 NA 0 0 1 ...
##  $ FFC                    : int  1 NA 1 NA 1 1 NA 1 1 2 ...
##  $ LP                     : num  43.6 NA 11.2 NA 53.4 23.8 NA 37.6 24.5 10 ...
##  $ LS                     : num  28.1 NA 66.3 NA 28.5 62.8 NA 57.7 39.9 52.7 ...
##  $ LS_including_skipped_TW: num  28.1 NA 66.3 NA 28.5 62.8 NA 57.7 39.9 52.7 ...
##  $ SA                     : num  1.56 NA 1.63 NA 1.4 1.78 NA 1.67 1.12 1.34 ...
##  $ FC                     : int  1 NA 1 NA 1 1 NA 1 1 2 ...
##  $ RI                     : int  0 NA 0 NA 0 0 NA 0 0 0 ...
##  $ RO                     : int  0 NA 0 NA 0 1 NA 1 0 0 ...
##  $ RPD                    : int  228 NA 392 NA 411 427 NA 582 207 499 ...
##  $ SKIP                   : int  0 NA 0 NA 0 0 NA 0 0 0 ...
##  $ SPILLOVER              : int  197 NA 421 NA 410 291 NA 379 NA NA ...
##  $ Blinks                 : int  0 1 0 1 0 0 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  "p1i7" "p1i6" "p1i9" "p1i11" ...
##  $ remove.trial           : int  0 1 0 1 0 0 1 0 0 0 ...
##  $ depvar                 : num  0 NA 0 NA 0 0 NA 0 0 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 2 1 2 1 1 1 1 1 2 ...
##  $ stim                   : Factor w/ 72 levels "i1","i10","i11",..: 1 12 23 34 45 56 67 71 72 2 ...
```

```
summary(datafile)
```

```
##  participant            item           predictability     target_word_length
##  Length:2880        Length:2880        Length:2880        Min.   :4.000
##  Class :character   Class :character   Class :character   1st Qu.:5.000
##  Mode  :character   Mode  :character   Mode  :character   Median :6.000
##                                                           Mean   :5.958
##                                                           3rd Qu.:7.000
##                                                           Max.   :8.000
##
##  ortho_frequency     cloze_scores     norm_scores     word.class
##  Min.   :   0.030   Min.   :  0.00   Min.   :2.400   Length:2880
##  1st Qu.:   5.195   1st Qu.:  0.00   1st Qu.:3.600   Class :character
##  Median :  23.640   Median : 37.50   Median :4.400   Mode  :character
##  Mean   :  68.271   Mean   : 44.47   Mean   :4.117
##  3rd Qu.:  57.653   3rd Qu.: 91.70   3rd Qu.:4.700
##  Max.   :1388.180   Max.   :100.00   Max.   :5.000
##
##       TRT              SFD              FFD              GD
##  Min.   :  82.0   Min.   :  82.0   Min.   :  82.0   Min.   :  82.0
##  1st Qu.: 223.5   1st Qu.: 207.0   1st Qu.: 201.0   1st Qu.: 215.0
##  Median : 290.0   Median : 246.0   Median : 241.0   Median : 267.0
##  Mean   : 349.7   Mean   : 267.1   Mean   : 262.2   Mean   : 306.9
##  3rd Qu.: 422.0   3rd Qu.: 300.0   3rd Qu.: 295.0   3rd Qu.: 363.0
##  Max.   :1462.0   Max.   :1033.0   Max.   :1033.0   Max.   :1420.0
##  NA's   :501      NA's   :984      NA's   :529      NA's   :529
##        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.:16.75   1st Qu.: 17.60
##  Median :0.0000   Median :1.000   Median :29.50   Median : 34.55
##  Mean   :0.1935   Mean   :1.205   Mean   :29.88   Mean   : 39.19
##  3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:42.30   3rd Qu.: 56.58
##  Max.   :1.0000   Max.   :4.000   Max.   :74.90   Max.   :327.00
##  NA's   :529      NA's   :529     NA's   :529     NA's   :338
##  LS_including_skipped_TW       SA              FC              RI
##  Min.   :  0.10          Min.   :0.200   Min.   :0.000   Min.   :0.0000
##  1st Qu.: 17.60          1st Qu.:1.120   1st Qu.:1.000   1st Qu.:0.0000
##  Median : 34.55          Median :1.430   Median :1.000   Median :0.0000
##  Mean   : 39.19          Mean   :1.455   Mean   :1.297   Mean   :0.0621
##  3rd Qu.: 56.58          3rd Qu.:1.770   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :327.00          Max.   :2.760   Max.   :6.000   Max.   :1.0000
##  NA's   :338             NA's   :501     NA's   :338     NA's   :529
##        RO            RPD              SKIP          SPILLOVER
##  Min.   :0.00   Min.   :  82.0   Min.   :0.0000   Min.   :  82.0
##  1st Qu.:0.00   1st Qu.: 221.0   1st Qu.:0.0000   1st Qu.: 199.0
##  Median :0.00   Median : 277.0   Median :0.0000   Median : 239.0
##  Mean   :0.06   Mean   : 336.5   Mean   :0.0751   Mean   : 260.1
##  3rd Qu.:0.00   3rd Qu.: 388.5   3rd Qu.:0.0000   3rd Qu.: 297.0
##  Max.   :1.00   Max.   :2714.0   Max.   :1.0000   Max.   :1036.0
##  NA's   :529    NA's   :529      NA's   :338      NA's   :1313
##      Blinks         Track_loss        Long_saccades       Index
##  Min.   :0.0000   Min.   :0.0000000   Min.   :0.0000   Length:2880
##  1st Qu.:0.0000   1st Qu.:0.0000000   1st Qu.:0.0000   Class :character
```

```
##  Median :0.0000   Median :0.0000000   Median :0.0000   Mode  :character
##  Mean   :0.1052   Mean   :0.0003472   Mean   :0.0125
##  3rd Qu.:0.0000   3rd Qu.:0.0000000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000000   Max.   :1.0000
##
##   remove.trial        depvar              pp               condition
##  Min.   :0.0000   Min.   :0.0000   p1     :  72   Predictable  :1440
##  1st Qu.:0.0000   1st Qu.:0.0000   p10    :  72   Unpredictable:1440
##  Median :0.0000   Median :0.0000   p11    :  72
##  Mean   :0.1174   Mean   :0.0751   p12    :  72
##  3rd Qu.:0.0000   3rd Qu.:0.0000   p13    :  72
##  Max.   :1.0000   Max.   :1.0000   p14    :  72
##                   NA's   :338      (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   0.16666667    0.21739130
## p10  0.00000000    0.03448276
## p11  0.13888889    0.11111111
## p12  0.00000000    0.03225806
## p13  0.00000000    0.03225806
## p14  0.11764706    0.05714286
## p15  0.08571429    0.03030303
## p16  0.02857143    0.02857143
## p17  0.03225806    0.11111111
## p18  0.00000000    0.03846154
## p19  0.11111111    0.05555556
## p2   0.05714286    0.08823529
## p20  0.00000000    0.07692308
## p21  0.18750000    0.12121212
## p22  0.00000000    0.03448276
## p23  0.12500000    0.29032258
## p24  0.09375000    0.03030303
## p25  0.02857143    0.00000000
## p26  0.09090909    0.06060606
## p27  0.05882353    0.17142857
## p28  0.00000000    0.04347826
## p29  0.22857143    0.19444444
## p3   0.02941176    0.09375000
## p30  0.03030303    0.00000000
## p31  0.08571429    0.03333333
## p32  0.06451613    0.03225806
## p33  0.08571429    0.14285714
```

```
## p34  0.08333333    0.03125000
## p35  0.12500000    0.05882353
## p36  0.13888889    0.05714286
## p37  0.03030303    0.08571429
## p38  0.17241379    0.15625000
## p39  0.18518519    0.19047619
## p4   0.00000000    0.00000000
## p40  0.06060606    0.12121212
## p5   0.03030303    0.12500000
## p6   0.02857143    0.06060606
## p7   0.00000000    0.00000000
## p8   0.06896552    0.06451613
## p9   0.05714286    0.02777778
```

```
#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    0.3790490     0.4217412
## p10   0.0000000     0.1856953
## p11   0.3507362     0.3187276
## p12   0.0000000     0.1796053
## p13   0.0000000     0.1796053
## p14   0.3270350     0.2355041
## p15   0.2840286     0.1740777
## p16   0.1690309     0.1690309
## p17   0.1796053     0.3202563
## p18   0.0000000     0.1961161
## p19   0.3187276     0.2323107
## p2    0.2355041     0.2879022
## p20   0.0000000     0.2717465
## p21   0.3965578     0.3314340
## p22   0.0000000     0.1856953
## p23   0.3360108     0.4614144
## p24   0.2961446     0.1740777
## p25   0.1690309     0.0000000
## p26   0.2919371     0.2423058
## p27   0.2388326     0.3823853
## p28   0.0000000     0.2085144
## p29   0.4260430     0.4013865
## p3    0.1714986     0.2961446
## p30   0.1740777     0.0000000
## p31   0.2840286     0.1825742
## p32   0.2497310     0.1796053
## p33   0.2840286     0.3550358
## p34   0.2803060     0.1767767
## p35   0.3360108     0.2388326
## p36   0.3507362     0.2355041
## p37   0.1740777     0.2840286
## p38   0.3844259     0.3689020
## p39   0.3958474     0.4023739
## p4    0.0000000     0.0000000
## p40   0.2423058     0.3314340
```

```
## p5    0.1740777     0.3360108
## p6    0.1690309     0.2423058
## p7    0.0000000     0.0000000
## p8    0.2578807     0.2497310
## p9    0.2355041     0.1666667
```

```
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  0.07068746    0.07852626
## grand.sd    0.21404602    0.24013647
## grand.se    0.03384365    0.03796891
```

```
####  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 0.07248636 0.2593926 1283 0.007241763
## 2  Unpredictable 0.07783956 0.2680254 1259 0.007553759
```

```
# 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
## 1285.8643 1332.5900 -634.9322 1269.8643      2534
## Random effects:
##  Groups      Name                        Std.Dev. Corr
##  item        (Intercept)                 0.93138
##              predictabilityUnpredictable 0.65543  -0.34
##  participant (Intercept)                 0.73311
```

```
##              predictabilityUnpredictable 0.08086  -1.00
## Number of obs: 2542, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                     -3.1151                       0.1078
```

```
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
##   1285.9   1332.6   -634.9   1269.9     2534
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0592 -0.2817 -0.2113 -0.1581  6.0381
##
## Random effects:
##  Groups      Name                        Variance Std.Dev. Corr
##  item        (Intercept)                 0.867464 0.93138
##              predictabilityUnpredictable 0.429585 0.65543  -0.34
##  participant (Intercept)                 0.537452 0.73311
##              predictabilityUnpredictable 0.006539 0.08086  -1.00
## Number of obs: 2542, groups:  item, 72; participant, 40
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -3.1151     0.2429 -12.825   <2e-16 ***
## predictabilityUnpredictable   0.1078     0.2619   0.411    0.681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## boundary (singular) fit: see ?isSingular
```

```
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
## 1284.1642 1319.2085 -636.0821 1272.1642      2536
```

```
## Random effects:
##  Groups      Name                        Std.Dev. Corr
##  item        (Intercept)                 0.87847
##  participant (Intercept)                 0.74519
##              predictabilityUnpredictable 0.08207  -1.00
## Number of obs: 2542, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                     -3.0824                       0.1177
## optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
```

```
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
##   1284.2   1319.2   -636.1   1272.2     2536
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0944 -0.2877 -0.2131 -0.1577  6.3978
##
## Random effects:
##  Groups      Name                        Variance Std.Dev. Corr
##  item        (Intercept)                 0.771709 0.87847
##  participant (Intercept)                 0.555306 0.74519
##              predictabilityUnpredictable 0.006735 0.08207  -1.00
## Number of obs: 2542, groups:  item, 72; participant, 40
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -3.0824     0.2235 -13.794   <2e-16 ***
## predictabilityUnpredictable   0.1177     0.1905   0.618    0.537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
# 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
## 1282.0088 1317.0531 -635.0044 1270.0088      2536
## Random effects:
##  Groups      Name                        Std.Dev. Corr
##  item        (Intercept)                 0.9250
##              predictabilityUnpredictable 0.6553   -0.32
##  participant (Intercept)                 0.6890
## Number of obs: 2542, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                    -3.08857                      0.06209
```

```
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
##   1282.0   1317.1   -635.0   1270.0     2536
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0246 -0.2806 -0.2110 -0.1585  6.0954
##
## Random effects:
##  Groups      Name                        Variance Std.Dev. Corr
##  item        (Intercept)                 0.8556   0.9250
##              predictabilityUnpredictable 0.4294   0.6553   -0.32
##  participant (Intercept)                 0.4747   0.6890
## Number of obs: 2542, groups:  item, 72; participant, 40
##
## Fixed effects:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -3.08857    0.22688 -13.613   <2e-16 ***
## predictabilityUnpredictable  0.06209    0.23167   0.268    0.789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 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
## 1280.3157 1303.6786 -636.1579 1272.3157      2538
```

```
## Random effects:
##  Groups      Name        Std.Dev.
##  item        (Intercept) 0.8768
##  participant (Intercept) 0.7005
## Number of obs: 2542, groups:  item, 72; participant, 40
## Fixed Effects:
##                 (Intercept)  predictabilityUnpredictable
##                    -3.05881                      0.07808
```

```
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
##   1280.3   1303.7   -636.2   1272.3     2538
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0620 -0.2880 -0.2145 -0.1594  6.4579
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  item        (Intercept) 0.7688   0.8768
##  participant (Intercept) 0.4906   0.7005
## Number of obs: 2542, groups:  item, 72; participant, 40
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -3.05881    0.20962 -14.592   <2e-16 ***
## predictabilityUnpredictable  0.07808    0.15996   0.488    0.625
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```