1 Checking Proportions

Here we will look at how our data is distributed between conditions after removing incorrect responses and outliers.

Use the 'xtabs' function to get a table of counts by conditions

```r
> acc.words.xtab <- xtabs(~Frequency + block, acc.words)
> acc.words.xtab

block
Frequency  2  3
 HF 1264 1281
 LF  971 1056

> acc.nonwords.xtab <- xtabs(~PSHstatus + block, acc.nonwords)
> acc.nonwords.xtab

block
PSHstatus  2  3
 PSH 1165 1214
 nonPSH 1167 1200
```
The nice thing about the tables from xtabs is that you can plot them out easily using ‘mosaicplot’.

> mosaicplot(acc.words.xtab, main = "Accurate Words")

![Accurate Words mosaic plot]

Figure 1: Proportions of correct word responses
> mosaicplot(acc.nonwords.xtab, main = "Accurate Nonwords")

**Accurate Nonwords**

![Mosaic plot showing proportions of correct word responses](image)

Figure 2: Proportions of correct word responses
**Figure 3:** RT Distributions for Correct Word Responses for all subjects

<table>
<thead>
<tr>
<th>shapiro.test.pvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
</tr>
<tr>
<td>logRT</td>
</tr>
<tr>
<td>trimmed.rt</td>
</tr>
<tr>
<td>log.trimmed.rt</td>
</tr>
<tr>
<td>invRT</td>
</tr>
<tr>
<td>inv.trimmed.rt</td>
</tr>
<tr>
<td>ormRT</td>
</tr>
<tr>
<td>invormRT</td>
</tr>
</tbody>
</table>

Table 1: shapiro test of normality for group RTs
Here we will check the normality of each subject’s RTs using the 8 different ‘versions’ of RTs from above. We do this using the “not_normal” function which you can see by typing ‘not_normal’.

```r
> subjpvals <- not_normal(subj = acc.words$Subj, RT = RT)
> subjpvals

Subjects pvals
S1   S1   0
S2   S2   0
S3   S3   0
S4   S4   0
S5   S5   0
S6   S6   0
S7   S7   0
S8   S8   0
S9   S9   0
S10  S10  0
S11  S11  0
S12  S12  0
S13  S13  0
S14  S14  0
S15  S15  0
S16  S16  0

> subjpvals <- not_normal(subj = acc.words$Subj, RT = logRT)
> subjpvals

Subjects pvals
S1   S1  0.000
S2   S2  0.000
S3   S3  0.000
S4   S4  0.000
S5   S5  0.000
S6   S6  0.000
S7   S7  0.000
S8   S8  0.000
S10  S10 0.000
S11  S11 0.000
S12  S12 0.000
S13  S13 0.022
S14  S14 0.000
S15  S15 0.000
S16  S16 0.028

> subjpvals <- not_normal(subj = acc.words$Subj, RT = trimmed.rt)
> subjpvals

Subjects pvals
S1   S1   0
S2   S2   0
S3   S3   0
S4   S4   0
S5   S5   0
S6   S6   0
S7   S7   0
```
> subjpvals <- not_normal(subj = acc.words$Subj, RT = log.trimmed.rt)
> subjpvals
  Subjects  pvals
S1       S1  0.000
S2       S2  0.000
S3       S3  0.000
S4       S4  0.000
S5       S5  0.000
S6       S6  0.000
S7       S7  0.000
S8       S8  0.035
S9       S9  0.000
S10      S10 0.000
S11      S11 0.000
S12      S12 0.000
S13      S13 0.011
S14      S14 0.003
S15      S15 0.000
S16      S16 0.006

> subjpvals <- not_normal(subj = acc.words$Subj, RT = invRT)
> subjpvals
  Subjects  pvals
S2       S2  0.026
S5       S5  0.000
S7       S7  0.000
S9       S9  0.000
S12      S12 0.000
S16      S16 0.003

> subjpvals <- not_normal(subj = acc.words$Subj, RT = inv.trimmed.rt)
> subjpvals
  Subjects  pvals
S2       S2  0.006
S3       S3  0.021
S5       S5  0.000
S6       S6  0.002
S7       S7  0.000
S9       S9  0.000
S10      S10 0.026
S11      S11 0.043
S12      S12 0.000
S13      S13 0.011
S15      S15 0.006
S16      S16 0.000
> subjpvals <- not_normal(subj = acc.words$Subj[acc.words$outliers == 0], RT = ormRT)
> subjpvals

<table>
<thead>
<tr>
<th>Subjects</th>
<th>pvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.000</td>
</tr>
<tr>
<td>S2</td>
<td>0.000</td>
</tr>
<tr>
<td>S3</td>
<td>0.000</td>
</tr>
<tr>
<td>S4</td>
<td>0.000</td>
</tr>
<tr>
<td>S5</td>
<td>0.000</td>
</tr>
<tr>
<td>S6</td>
<td>0.000</td>
</tr>
<tr>
<td>S7</td>
<td>0.000</td>
</tr>
<tr>
<td>S8</td>
<td>0.000</td>
</tr>
<tr>
<td>S9</td>
<td>0.001</td>
</tr>
<tr>
<td>S10</td>
<td>0.000</td>
</tr>
<tr>
<td>S11</td>
<td>0.000</td>
</tr>
<tr>
<td>S12</td>
<td>0.000</td>
</tr>
<tr>
<td>S13</td>
<td>0.000</td>
</tr>
<tr>
<td>S14</td>
<td>0.000</td>
</tr>
<tr>
<td>S15</td>
<td>0.000</td>
</tr>
<tr>
<td>S16</td>
<td>0.000</td>
</tr>
</tbody>
</table>

> subjpvals <- not_normal(subj = acc.words$Subj[acc.words$outliers == 0], RT = invormRT)
> subjpvals

<table>
<thead>
<tr>
<th>Subjects</th>
<th>pvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3</td>
<td>0.019</td>
</tr>
<tr>
<td>S5</td>
<td>0.000</td>
</tr>
<tr>
<td>S6</td>
<td>0.004</td>
</tr>
<tr>
<td>S7</td>
<td>0.002</td>
</tr>
<tr>
<td>S9</td>
<td>0.000</td>
</tr>
<tr>
<td>S10</td>
<td>0.049</td>
</tr>
<tr>
<td>S12</td>
<td>0.000</td>
</tr>
<tr>
<td>S13</td>
<td>0.005</td>
</tr>
<tr>
<td>S16</td>
<td>0.000</td>
</tr>
</tbody>
</table>

It appears that using ‘invRT’ is the best as it yields the fewest subjects with non-normal distributions. So let’s use invRT and then isolate those subjects that are still flagged by the Shapiro test.

> subjpvals <- not_normal(subj = acc.words$Subj, RT = invRT)
> subjpvals

<table>
<thead>
<tr>
<th>Subjects</th>
<th>pvals</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2</td>
<td>0.026</td>
</tr>
<tr>
<td>S5</td>
<td>0.000</td>
</tr>
<tr>
<td>S7</td>
<td>0.000</td>
</tr>
<tr>
<td>S9</td>
<td>0.000</td>
</tr>
<tr>
<td>S12</td>
<td>0.000</td>
</tr>
<tr>
<td>S16</td>
<td>0.003</td>
</tr>
</tbody>
</table>

> newsubjs <- acc.words[acc.words$Subj %in% subjpvals$Subjects, ]
> l <- split(newsubjs, newsubjs$Subj, drop = TRUE)
> lmmodels <- lapply(l, function(x) lmModel <- lm(-1000/RT ~ Frequency, x))
Now we plot out the residuals

```r
> par(mfrow = c(3, 2), mar = c(4, 3, 3, 1) + 0.1)
> for (i in 1:length(lmmodels)) {
+   x <- lmmodels[[i]]
+   r <- resid(x)
+   upper <- mean(r) + 2.5 * sd(r)
+   lower <- mean(r) - 2.5 * sd(r)
+   plot(r, ylim = c(upper + 0.5, lower - 0.5), main = names(lmmodels[[i]]))
+   abline(h = upper, lty = 2)
+   abline(h = lower, lty = 2)
+   bound <- abs(scale(resid(x)))
+   points(which(bound > 2.5), r[which(bound > 2.5)], pch = 16, col = "red")
+ }
```
OK, now we can remove the RTs whose residuals in the simple lm exceed 2.5SDs.

```r
> l <- split(newsubjs, newsubjs$Subj, drop = TRUE)
> trimmed <- lapply(l, function(x) {
+   lmModel <- lm(-1000/RT ~ Frequency, x)
+   d <- x[abs(scale(resid(lmModel))) < 2.5, ]
+   return(d)
+ })
> trimmed.df <- ldply(trimmed)
> subjvals <- not_normal(subj = trimmed.df$Subj, RT = (-1000/trimmed.df$RT))
> subjvals

Subjects pvals
S2     S2  0.007
S5     S5  0.000
S7     S7  0.000
S12    S12 0.000
```
Finally, let’s compare the qqplots for pre- and post-trimmed RTs for these 6 subjects:

```r
> par(mfcol = c(6, 2), mar = c(2, 4, 3, 1) + 0.1)
> for (i in 1:length(l)) {
+ df <- l[[i]]
+ invRT = -1000/df$RT
+ qqnorm(invRT, main = paste(names(l)[i], " no trimming", sep = ""))
+ qqline(invRT, col = "red")
+ }
> for (i in 1:length(trimmed)) {
+ df <- trimmed[[i]]
+ invRT = -1000/df$RT
+ qqnorm(invRT, main = paste(names(trimmed)[i], " after trimming", sep = ""))
+ qqline(invRT, col = "red")
+ }
```