Emodash Statistical Analysis

We use R kernel for jupyter (http://jupyter.org/) for this notebook

Experiment Design

We conducted an analysis of variance on the feedback report content. The study was a 2x4 repeated measures design (within-subject), with the following factors and levels:

- *Condition*: With, Without (Emodash)
- *Category* (of the feedback): Affective, Motivational, Summative, Formative

The measured variable was the number of *utterances*.

The analysis was carried out using Generalized Linear Mixed Model (GLMM) with Anova and Bonferroni adjustments for *post hoc* comparisons.

We are going to use each feedback report for each participant as a trial in the model. This gives us a quite good amount of data points to fit the the model.

GLMM deals with winthin-subject experiment, and does not require the three assumptions of the ANOVA (normality, independence, and homogeneity), which is best suited for our case. GLMM deals with missing data as we don't have a full balancing.

The data on which we conduct this analysis, can be found inside the folder *data* associated with this notebook.

R packages

In [4]:
```
## Imports
## -------

# Data Manipulation Packages
library(readxl)
library(dplyr)
library(tidyr)
library(tidyverse)
library(reshape2)

# For plots
library(ggplot2)

# Fit distribution
library(fitdistrplus)
# For GLMM
library(lme4)
# For ANOVA
library(car)
# Post Hoc Pairwise Comparision Test,
# emmeans replaces lsmeans (deprecated)
library(emmeans)
```

Loading Data

In [5]:
```
# Check path
#getwd()

# Back to data folder
setwd('./data')
```

In [6]:
```
# load data

feedbackWith = read_excel('emodash-workbook.v2.xlsx', sheet="feedback-with-e
feedbackWithout = read_excel('emodash-workbook.v2.xlsx', sheet="feedback-wit
```

Check Feedback Data

In [7]:
```
# Check With Emodash head

head(feedbackWith)
```

bloc	session_id	pair	pair_session	unit_id	MA	SF
bloc1	5a0bfe9bb2ee7900015fcd38	P1	P1S1	1	Affective	None
bloc1	NA	P1	P1S1	2	Motivational	Summative
bloc1	NA	P1	P1S1	3	None	Summative
bloc1	NA	P1	P1S1	4	Affective	None
bloc1	NA	P1	P1S1	5	None	Formative
bloc1	NA	P1	P1S1	6	None	Formative

In [8]:
```
# Check Without Emodash head

head(feedbackWithout)
```

bloc	session_id	pair	pair_session	unit_id	MA	SF
1	59eef14ed7ca3d0001ea9f4c	P5	P5S1	1	Affective	None
1	NA	P5	P5S1	2	Motivational	Summative
1	NA	P5	P5S1	3	Motivational	Summative
1	NA	P5	P5S1	4	Motivational	Formative
1	NA	P5	P5S1	5	Motivational	Formative
1	NA	P5	P5S1	6	None	Formative

Data Overview

- bloc: bloc index of learning sessions
- session_id: id of the learning session
- pair: id of the participant (also id of the pair as we have one-to-one relationship b/ tutor and learner)
- pair_session: id of the session of pair
- unit_id: utterance id per pair, per session
- MA: Motivational and Affective feedback content coding
- MA: Summative and Formative feedback content coding

Stacking Feedback with/without emodash data frames

As each line the data frames feedbackWith and feedbackWithout is a participant ('pair' column) utterance.

- first, we stack both frames together
- second, we count the number of utterances by (pair, pair_session, category, condition)

In [9]:
```
# Reshape data
with = melt(feedbackWith, id.vars = c('pair','pair_session'), measure.vars =
without = melt(feedbackWithout, id.vars = c('pair','pair_session'), measure.
```

In [8]:
```
# Rename 'value' col to 'category'
names(with)[names(with) == 'value'] = 'category'
names(without)[names(without) == 'value'] = 'category'
```

In [9]:
```
# Remove None utterances that fall in any category
with = filter(with, with$category != 'None')
without = filter(without, without$category != 'None')
```

In [10]:
```
# Add condition col
without$condition = factor('without')
with$condition = factor('with')
```

In [11]:
```
# Stack with and without data frames
contentAnalysis = rbind(with, without)
```

Convert pair_session into long

Here we convert pair_session into a long factor, by keeping index of the session (1, 2, 3, 4, 5).

For instance: P1S1 => 1

pair_session will be considered as the 'trial' in GLMM.

In [12]:
```
contentAnalysis$pair_session = substr(contentAnalysis$pair_session, start=4,
```

Making sure that columns are factors

In [13]:
```
contentAnalysis$pair = factor(contentAnalysis$pair)
contentAnalysis$pair_session = factor(contentAnalysis$pair_session)
contentAnalysis$condition = factor(contentAnalysis$condition)
contentAnalysis$category = factor(contentAnalysis$category)
```

Count number of utterances by pair, pair_session, condition, category

In [14]:
```
# count by our factors
contentAnalysis = dplyr::count(contentAnalysis, pair, pair_session, conditic
```

Check data

In [15]:
```
head(contentAnalysis)
```

pair	pair_session	condition	category	n
P1	1	with	Affective	3
P1	1	with	Formative	4
P1	1	with	Motivational	2
P1	1	with	Summative	4
P1	1	without	Affective	1
P1	1	without	Formative	6

Rename *n* column to utterance

In [16]:
```
names(contentAnalysis)[names(contentAnalysis) == 'n'] = 'utterance'

# recheck what we have so far
head(contentAnalysis)

print('Nbr of data points:')
print(count(contentAnalysis))
```

pair	pair_session	condition	category	utterance
P1	1	with	Affective	3
P1	1	with	Formative	4
P1	1	with	Motivational	2
P1	1	with	Summative	4
P1	1	without	Affective	1
P1	1	without	Formative	6

```
[1] "Nbr of data points:"
# A tibble: 1 x 1
      n
  <int>
1   157
```

Summarising data

For instance we can see that on average, we have 4.261 utterances per participant per trial (feedback categories combined).

In [17]:
```
# summarise data
summary(contentAnalysis)
```

```
 pair    pair_session   condition         category     utterance
 P1:38   1:34         with   :75   Affective   :28   Min.   : 1.000
 P2:32   2:30         without:82   Formative   :43   1st Qu.: 2.000
 P3:30   3:33                      Motivational:41   Median : 4.000
 P4:26   4:29                      Summative   :45   Mean   : 4.261
 P5:31   5:31                                        3rd Qu.: 6.000
                                                     Max.   :16.000
```

Summarising data by both factors *condition* and *category*

In [18]:
```
# summarise by ~ condition + catagory
plyr::ddply(contentAnalysis, ~ condition  + category , function (data) { sur
```

condition	category	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
with	Affective	2	3.00	3.0	4.307692	6.00	9
with	Formative	1	3.00	5.0	6.300000	9.25	16
with	Motivational	1	1.75	2.5	3.350000	5.00	9
with	Summative	1	2.00	4.0	3.681818	4.00	10
without	Affective	1	1.00	2.0	1.933333	2.50	4
without	Formative	1	3.00	4.0	4.608696	6.00	9
without	Motivational	1	1.00	4.0	4.095238	6.00	10
without	Summative	1	3.50	5.0	5.130435	7.00	11

Distribution of the data

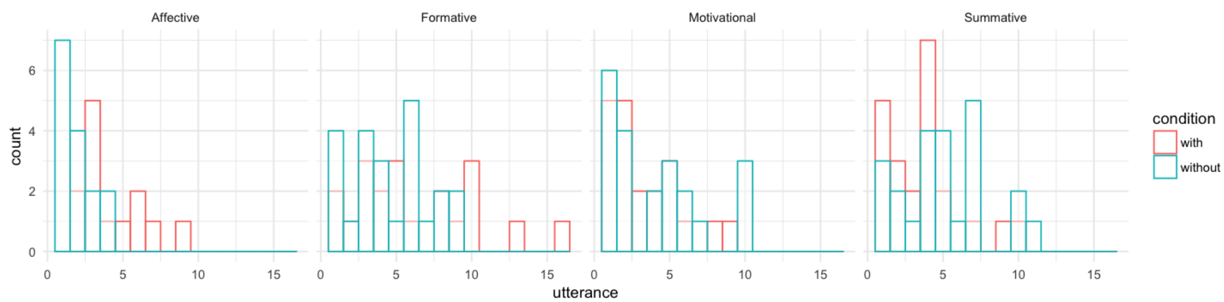
Here we plot the hist of each *catagory* under each *condition* to see how the distribution is.

As we can see below is not a normal distibution, but rather it tends towards a *poisson* one which is usually the case with count data (like, error rate etc.). We will verify this below.

In [19]:
```
# Hist plots

options(repr.plot.width=12, repr.plot.height=3)

ggplot(contentAnalysis) +
geom_histogram(aes (x = utterance, color = condition),
               binwidth = 1,
               fill = "white",
               alpha = 0.5,
               position = "identity") +
facet_grid(~ category) +
theme_minimal()
```

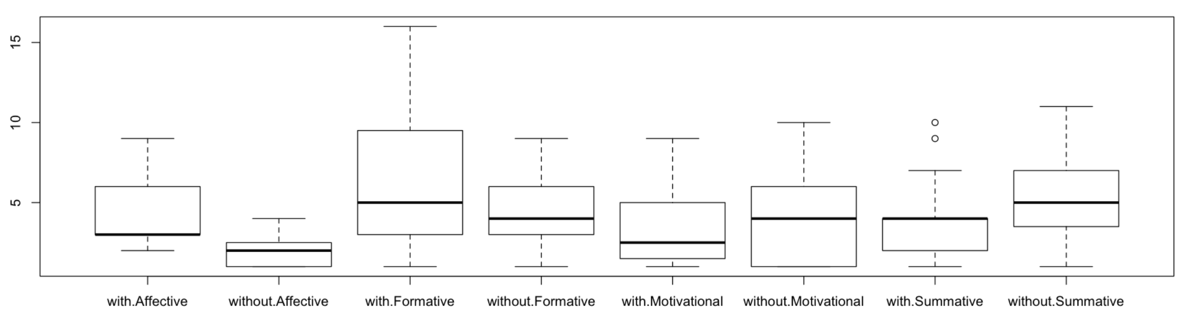


Boxplot of ~ condition + category

In [20]:
```
# Boxplots

options(repr.plot.width=15, repr.plot.height=5)

boxplot( utterance ~ condition + category, data = contentAnalysis)
```



As we can see in the boxplots above:

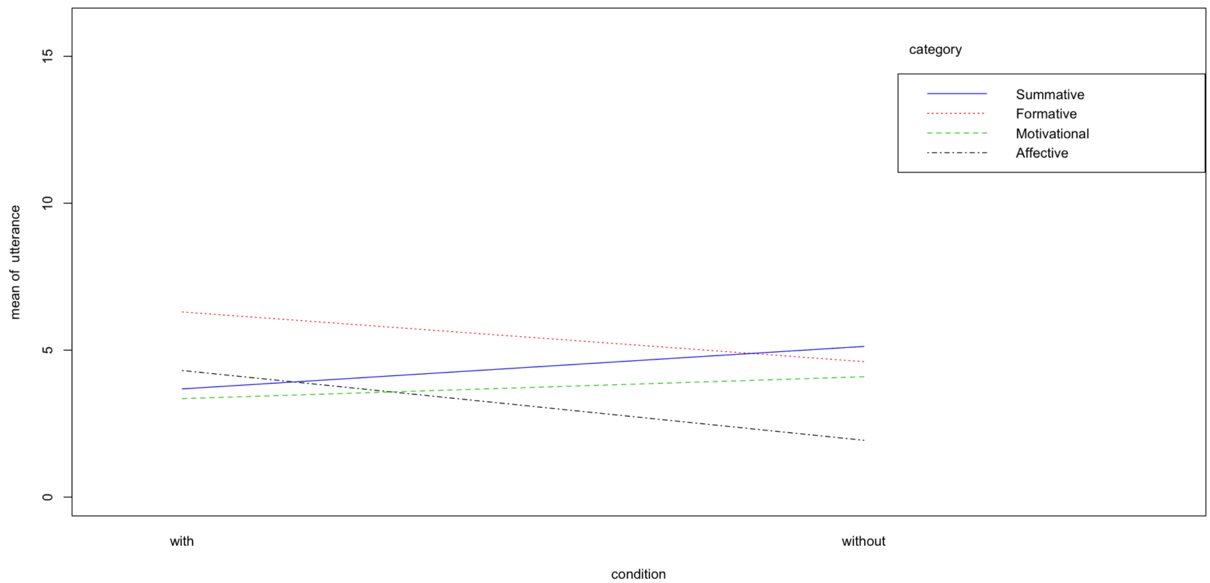
- We can see that there may be a difference in the Affective level which is increased with Emodash, and same thing with Formaive level.
- Also, there may be a difference in the Summative level which is decreased with Emodash.
- As we can see for the Motivational level, probably there is no difference as there is too much overlap in the box plot.

Interaction plot

In [21]:
```
# interaction between categories

options(repr.plot.width=15, repr.plot.height=8)

with(contentAnalysis, interaction.plot( col = 1:4, leg.bty = "o", condition,
```



Interaction plot shows that there is an interaction between our variables (levels).

What is interesting is that:

- Without Emodash Formative and Summative are almost the same, but, they get different With Emodash where Formative increased and Summative decreased.
- Same thing with Affective and Motivational, where Affective level decreased without Emodash and Motivational increased.

Checking the distribution of the data

In [22]:
```
# check if the data fit a poisson distribution as illustrated by histograms

fit = fitdist(contentAnalysis[contentAnalysis$condition == 'with' & contentA
gofstat(fit)
```

```
Chi-squared statistic:  1.719688
Degree of freedom of the Chi-squared distribution:  1
Chi-squared p-value:  0.1897332
   the p-value may be wrong with some theoretical counts < 5
Chi-squared table:
     obscounts theocounts
<= 3  7.000000   4.885042
<= 6  4.000000   6.228297
> 6   2.000000   1.886661

Goodness-of-fit criteria
                               1-mle-pois
Akaike's Information Criterion   56.02720
Bayesian Information Criterion   56.59215
```

In [23]:
```
fit = fitdist(contentAnalysis[contentAnalysis$condition == 'with' & contentA
gofstat(fit)
```

```
Chi-squared statistic:  3.820542
Degree of freedom of the Chi-squared distribution:  3
Chi-squared p-value:  0.2815057
   the p-value may be wrong with some theoretical counts < 5
Chi-squared table:
     obscounts theocounts
<= 1  5.000000   3.052339
<= 2  5.000000   3.937342
<= 4  4.000000   8.078933
<= 5  3.000000   2.467097
> 5   3.000000   2.464289

Goodness-of-fit criteria
                               1-mle-pois
Akaike's Information Criterion   89.50635
Bayesian Information Criterion   90.50208
```

In [24]:
```
fit = fitdist(contentAnalysis[contentAnalysis$condition == 'with' & contentA
gofstat(fit)
```

```
Chi-squared statistic:  10.33606
Degree of freedom of the Chi-squared distribution:  4
Chi-squared p-value:  0.03513167
   the p-value may be wrong with some theoretical counts < 5
Chi-squared table:
      obscounts theocounts
<= 2  3.0000000  0.9969298
<= 3  3.0000000  1.5305416
<= 5  5.0000000  5.4479630
<= 8  3.0000000  8.3198984
<= 10 4.0000000  2.5790769
> 10  2.0000000  1.1255902

Goodness-of-fit criteria
                               1-mle-pois
Akaike's Information Criterion   122.0106
Bayesian Information Criterion   123.0063
```

In [25]:
```
fit = fitdist(contentAnalysis[contentAnalysis$condition == 'with' & contentA
gofstat(fit)
```

```
Chi-squared statistic:  5.858433
Degree of freedom of the Chi-squared distribution:  2
Chi-squared p-value:  0.0534389
   the p-value may be wrong with some theoretical counts < 5
Chi-squared table:
     obscounts theocounts
<= 1  5.000000   2.593247
<= 3  5.000000   8.361755
<= 4  7.000000   4.240992
> 4   5.000000   6.804006

Goodness-of-fit criteria
                               1-mle-pois
Akaike's Information Criterion   100.6796
Bayesian Information Criterion   101.7706
```

Same for without..

As we can see in result above, the p-values are not significant, except for Formative level but with high Chi-square, so, we dont have a significant departure from poisson distribution.

Fit a GLMM

In [26]:
```
# Needed as we are using Anova from 'car' package not anova from default R
contrasts(contentAnalysis$condition) = 'contr.sum'
contrasts(contentAnalysis$category) = 'contr.sum'
contrasts(contentAnalysis$pair_session) = 'contr.sum'
```

In [27]:
```
# Fit a generalized linear mixed-effects model (GLMM)
m = glmer(utterance ~ (condition + category)/pair_session + (1|pair), data=c
```

```
fixed-effect model matrix is rank deficient so dropping 5 columns / coeff
icients
```

Overall ANOVA Test

In [28]:
```
Anova(m, type=3)
```

	Chisq	Df	Pr(>Chisq)
(Intercept)	9.583846	1	0.001962968
condition	4.443061	1	0.035043354
category	9.381812	3	0.024622504
condition:category:pair_session	57.006976	35	0.010800070

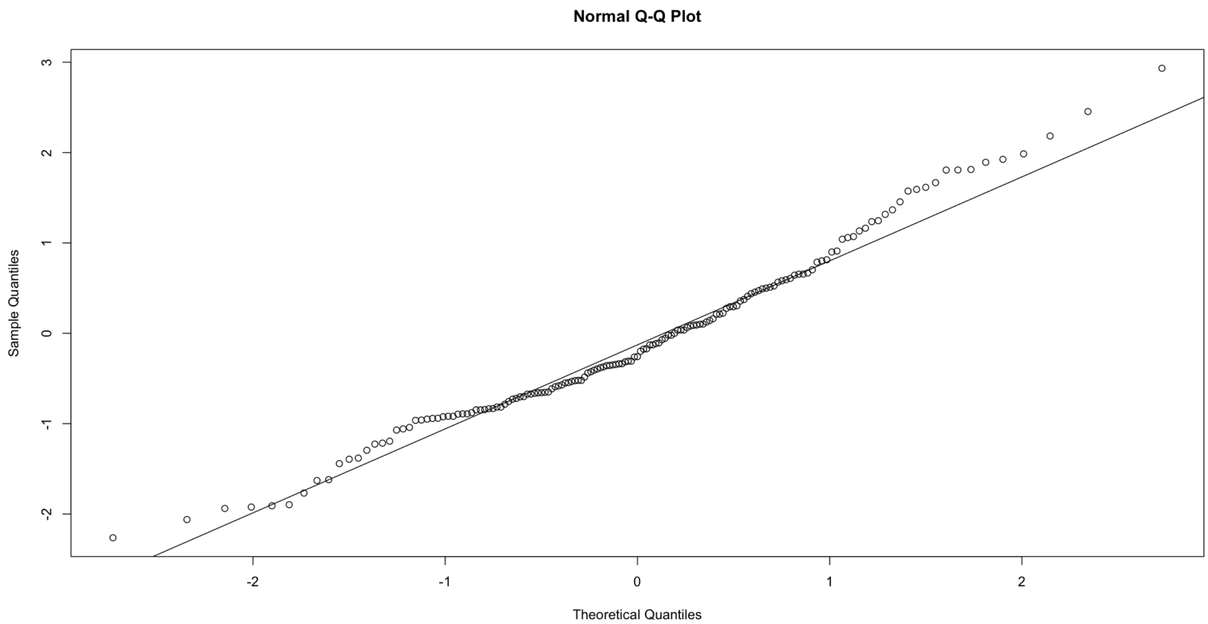
The overall ANOVA (omnibus test) tells us that:

- There is a significant difference among the levels (with and without Emodash) of *condition* factor
$$(\chi^2 = (1, N = 157) = 4.443, p < 0.05)$$

- There is a significant difference among the levels (Affective, Motivational, Summative, Formative) of *category* factor
$$(\chi^2 = (1, N = 157) = 9.381, p < 0.05)$$

QQNorm with residuals of the built model

In [29]:
```
qqnorm(residuals(m));qqline(residuals(m))
```



As we can see in the result above, *qqnorm* with residuals shows that there is no departure from the normal.

Post Hoc Test

Now we run a pairwise comparison using Holm-Bonferroni for adjustments.

In [30]:
```
# Get pairwise summary

emmeans(m, list(pairwise ~ condition | category), adjust = "holm")
```

```
NOTE: Results may be misleading due to involvement in interactions

$`emmeans of condition | category`
category = Affective:
 condition    emmean        SE  df  asymp.LCL asymp.UCL
 with      1.3911005 0.2188517 Inf  0.9621590  1.820042
 without   0.4564178 0.3054943 Inf -0.1423400  1.055176

category = Formative:
 condition    emmean        SE  df  asymp.LCL asymp.UCL
 with      1.7950457 0.1902091 Inf  1.4222428  2.167849
 without   1.4423502 0.1937194 Inf  1.0626671  1.822033

category = Motivational:
 condition    emmean        SE  df  asymp.LCL asymp.UCL
 with      1.1265018 0.2112598 Inf  0.7124403  1.540563
 without   1.2749335 0.2007223 Inf  0.8815250  1.668342

category = Summative:
 condition    emmean        SE  df  asymp.LCL asymp.UCL
 with      1.1604161 0.2064785 Inf  0.7557257  1.565107
 without   1.5017805 0.1921657 Inf  1.1251427  1.878418

Results are averaged over the levels of: pair_session
Results are given on the log (not the response) scale.
Confidence level used: 0.95

$`pairwise differences of condition | category`
category = Affective:
 contrast         estimate        SE  df z.ratio p.value
 with - without  0.9346827 0.2901955 Inf   3.221  0.0013

category = Formative:
 contrast         estimate        SE  df z.ratio p.value
 with - without  0.3526955 0.1344426 Inf   2.623  0.0087

category = Motivational:
 contrast         estimate        SE  df z.ratio p.value
 with - without -0.1484317 0.1712002 Inf  -0.867  0.3859

category = Summative:
 contrast         estimate        SE  df z.ratio p.value
 with - without -0.3413643 0.1547195 Inf  -2.206  0.0274

Results are averaged over the levels of: pair_session
Results are given on the log (not the response) scale.
```

Test results

As we can see from the results above.

- There was a significant effect of the *condition* on the Affective level as it is increased with Emodash ($z=3.221$, $p=0.0013$).
- Similarly, there was a significant effect on the Formative as it is increased with Emodash ($z=2.623$, $p=0.0087$).
- Regarding the Summative level, we notice that there was a significant effect but, the number of utterances is rather decreased With Emodash ($z=-2.206$, $p=0.0274$).
- There was no significant effect on the Motivational ($Z = -0.867$, p=n.s).

As illustrated in boxplot figure above there is too much overlap in the boxplot of the Motivational under both *conditions* compared to the others levels.

In [ ]: