

useful resources

<https://stats.oarc.ucla.edu/r/dae/poisson-regression/>

Read in the data file

```
df <- read_excel("ErinAveryPrimaryParenthesisClean.xlsx")
```

Read in the required packages to run a poisson regression analysis

```
library(sandwich) # this package is to compute the robust standard errors
library(msm)
library(ggeffects) # this package is to calculate marginal effects and adjusted predictions, plus to create interaction plots
library(apaTables)
```

explore the data

```
names(df) # Get the column names
```

```
FALSE [1] "experiment_id"
FALSE [2] "student_id"
FALSE [3] "release_date"
FALSE [4] "due_date"
FALSE [5] "start_time"
FALSE [6] "end_time"
FALSE [7] "assignment_session_count"
FALSE [8] "pretest_problem_count"
FALSE [9] "pretest3_correct"
FALSE [10] "pretest_correct"
FALSE [11] "pretest_time_on_task"
FALSE [12] "pretest_average_first_response_time"
FALSE [13] "pretest_session_count"
FALSE [14] "assigned_condition"
FALSE [15] "condition_time_on_task"
FALSE [16] "condition_average_first_response_or_request_time"
FALSE [17] "condition_problem_count"
FALSE [18] "condition_total_correct"
FALSE [19] "condition_total_attempt_count"
FALSE [20] "assistments_reference_assignment_log_id"
FALSE [21] "Mastery"
FALSE [22] "Parentheses"
FALSE [23] "Position"
FALSE [24] "PrimaryLast"
FALSE [25] "class_id"
FALSE [26] "teacher_id"
```

```
FALSE [27] "Zpretest_average_first_response_time"
FALSE [28] "Zcondition_problem_count"
```

```
summary(df) # run the summary statistics (e.g., mean, median, min, max)
```

```
FALSE experiment_id      student_id      release_date      due_date
FALSE Length:538         Min.   : 129245    Length:538         Length:538
FALSE Class :character    1st Qu.: 885469    Class :character    Class :character
FALSE Mode  :character     Median :1054069    Mode  :character    Mode  :character
FALSE                                     Mean   : 945452
FALSE                                     3rd Qu.:1079301
FALSE                                     Max.   :1152928
FALSE
FALSE start_time          end_time          assignment_session_count
FALSE Length:538          Length:538          Min.   :1.000
FALSE Class :character     Class :character    1st Qu.:1.000
FALSE Mode  :character     Mode  :character    Median :1.000
FALSE                                     Mean   :1.195
FALSE                                     3rd Qu.:1.000
FALSE                                     Max.   :6.000
FALSE
FALSE pretest_problem_count pretest3_correct pretest_correct pretest_time_on
_task
FALSE Min.   :4           Min.   :0.000     Min.   :1.000     Min.   : 30.
3
FALSE 1st Qu.:4           1st Qu.:1.000     1st Qu.:2.000     1st Qu.: 89.
1
FALSE Median :4           Median :2.000     Median :3.000     Median : 144.
8
FALSE Mean   :4           Mean   :2.145     Mean   :3.145     Mean   : 7637.
1
FALSE 3rd Qu.:4           3rd Qu.:3.000     3rd Qu.:4.000     3rd Qu.: 281.
8
FALSE Max.   :4           Max.   :3.000     Max.   :4.000     Max.   :855802.
8
FALSE
FALSE pretest_average_first_response_time pretest_session_count assigned_con
dition
FALSE Min.   : 7.58           Min.   :1.000           Length:538
FALSE 1st Qu.: 22.28           1st Qu.:1.000           Class :character
FALSE Median : 36.19           Median :1.000           Mode  :character
FALSE Mean   : 1909.28           Mean   :1.123
FALSE 3rd Qu.: 70.43           3rd Qu.:1.000
FALSE Max.   :213950.70         Max.   :3.000
FALSE
FALSE condition_time_on_task condition_average_first_response_or_request_tim
```

```

e
FALSE Length:538 Length:538
FALSE Class :character Class :character
FALSE Mode :character Mode :character
FALSE
FALSE
FALSE
FALSE
FALSE condition_problem_count condition_total_correct condition_total_attempt_count
FALSE Min. :3.000 Length:538 Min. : 3.000
FALSE 1st Qu.:3.000 Class :character 1st Qu.: 3.000
FALSE Median :3.000 Mode :character Median : 3.000
FALSE Mean :3.812 Mean : 4.532
FALSE 3rd Qu.:4.000 3rd Qu.: 5.000
FALSE Max. :9.000 Max. :23.000
FALSE
FALSE assistments_reference_assignment_log_id Mastery Parentheses
FALSE Min. :16770661 Min. :1 Min. :0.0000
FALSE 1st Qu.:18014622 1st Qu.:1 1st Qu.:0.0000
FALSE Median :18506915 Median :1 Median :1.0000
FALSE Mean :18733431 Mean :1 Mean :0.5483
FALSE 3rd Qu.:19747412 3rd Qu.:1 3rd Qu.:1.0000
FALSE Max. :19888117 Max. :1 Max. :1.0000
FALSE
FALSE Position PrimaryLast class_id teacher_id
FALSE Min. :1.000 Min. :1 Min. :60296 Min. : 4416
FALSE 1st Qu.:1.000 1st Qu.:1 1st Qu.:86251 1st Qu.: 9658
FALSE Median :2.000 Median :1 Median :87742 Median : 121864
FALSE Mean :1.948 Mean :1 Mean :88249 Mean : 344150
FALSE 3rd Qu.:3.000 3rd Qu.:1 3rd Qu.:91864 3rd Qu.: 947900
FALSE Max. :3.000 Max. :1 Max. :96812 Max. :1144011
FALSE NA's :217 NA's :3
FALSE Zpretest_average_first_response_time Zcondition_problem_count
FALSE Min. :-0.12291 Min. :-0.51876
FALSE 1st Qu.: -0.12272 1st Qu.: -0.51876
FALSE Median : -0.12254 Median : -0.51876
FALSE Mean : -0.09858 Mean : -0.08140
FALSE 3rd Qu.: -0.12210 3rd Qu.: 0.01969
FALSE Max. : 2.61356 Max. : 2.71190
FALSE

```

creating dummy variables

```

df$middle <- ifelse(df$Position == '2', 1, 0) # Yes = 1, No = 0
df$left <- ifelse(df$Position == '1', 1, 0)
df$right <- ifelse(df$Position == '3', 1, 0)

# change the type of the variables

```

```
df$left <- as.factor(df$left)
df$right <- as.factor(df$right)
df$middle <- as.factor(df$middle)
```

build a prediction model (left as a reference group)

testing the main effects

```
model1 <- glm(condition_total_attempt_count ~ pretest3_correct + Parentheses
+ middle + right, family= "poisson", data= df) # Note: "glm" function is used
to fit generalized linear models. If you put "gaussian" instead of "poisson"
you can run a linear regression model for the normal distribution. (or use "l
m" instead of "glm" for linear regression models)
```

```
summary(model1) # obtain coefficients, SE, p-values
```

```
FALSE
FALSE Call:
FALSE glm(formula = condition_total_attempt_count ~ pretest3_correct +
FALSE Parentheses + middle + right, family = "poisson", data = df)
FALSE
FALSE Deviance Residuals:
FALSE      Min       1Q   Median       3Q      Max
FALSE -1.7704  -0.7188  -0.4081   0.3501   5.6151
FALSE
FALSE Coefficients:
FALSE              Estimate Std. Error z value Pr(>|z|)
FALSE (Intercept)      1.62158    0.06286  25.795 < 2e-16 ***
FALSE pretest3_correct -0.09605    0.02117  -4.536 5.72e-06 ***
FALSE Parentheses      -0.12164    0.04060  -2.996 0.00274 **
FALSE middle1          0.11354    0.05112   2.221 0.02636 *
FALSE right1           0.35097    0.04997   7.024 2.16e-12 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE (Dispersion parameter for poisson family taken to be 1)
FALSE
FALSE      Null deviance: 679.51  on 537  degrees of freedom
FALSE Residual deviance: 593.77  on 533  degrees of freedom
FALSE AIC: 2364.9
FALSE
FALSE Number of Fisher Scoring iterations: 4
```

testing the interaction effects

```
model2 <- glm(condition_total_attempt_count ~ pretest3_correct + Parentheses
+ middle + right + right*Parentheses + middle*Parentheses, family= "poisson",
data= df)
```

```
summary(model2)
```

```

FALSE
FALSE Call:
FALSE glm(formula = condition_total_attempt_count ~ pretest3_correct +
FALSE     Parentheses + middle + right + right * Parentheses + middle *
FALSE     Parentheses, family = "poisson", data = df)
FALSE
FALSE Deviance Residuals:
FALSE      Min       1Q   Median       3Q      Max
FALSE -1.6909  -0.6980  -0.3506   0.2554   5.6728
FALSE
FALSE Coefficients:
FALSE              Estimate Std. Error z value Pr(>|z|)
FALSE (Intercept)      1.50296    0.07490  20.067 < 2e-16 ***
FALSE pretest3_correct -0.09093    0.02125  -4.279 1.87e-05 ***
FALSE Parentheses      0.06423    0.07571   0.848 0.396232
FALSE middle1          0.32440    0.07560   4.291 1.78e-05 ***
FALSE right1           0.43576    0.07546   5.775 7.70e-09 ***
FALSE Parentheses:right1 -0.13865    0.10088  -1.374 0.169317
FALSE Parentheses:middle1 -0.40106    0.10364  -3.870 0.000109 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE (Dispersion parameter for poisson family taken to be 1)
FALSE
FALSE      Null deviance: 679.51  on 537  degrees of freedom
FALSE Residual deviance: 577.89  on 531  degrees of freedom
FALSE AIC: 2353
FALSE
FALSE Number of Fisher Scoring iterations: 4

```

build a prediction model (middle as a reference group)

testing the main effects

```

model3 <- glm(condition_total_attempt_count ~ pretest3_correct + Parentheses
+ left + right, family= "poisson", data= df)

```

```
summary(model3)
```

```

FALSE
FALSE Call:
FALSE glm(formula = condition_total_attempt_count ~ pretest3_correct +
FALSE     Parentheses + left + right, family = "poisson", data = df)
FALSE
FALSE Deviance Residuals:
FALSE      Min       1Q   Median       3Q      Max
FALSE -1.7704  -0.7188  -0.4081   0.3501   5.6151
FALSE
FALSE Coefficients:
FALSE              Estimate Std. Error z value Pr(>|z|)
FALSE (Intercept)      1.73512    0.05979  29.022 < 2e-16 ***

```

```

FALSE pretest3_correct -0.09605    0.02117   -4.536 5.72e-06 ***
FALSE Parentheses      -0.12164    0.04060   -2.996 0.00274 **
FALSE left1            -0.11354    0.05112   -2.221 0.02636 *
FALSE right1           0.23743    0.04845    4.900 9.57e-07 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE (Dispersion parameter for poisson family taken to be 1)
FALSE
FALSE      Null deviance: 679.51  on 537  degrees of freedom
FALSE Residual deviance: 593.77  on 533  degrees of freedom
FALSE AIC: 2364.9
FALSE
FALSE Number of Fisher Scoring iterations: 4

m3_results <- apa_print(model3) # make the table in APA format
apa_table(m3_results,
          caption = "Table for model 3", , escape = TRUE) # print the table i
n APA format

```

(#tab:unnamed-chunk-7)

Table for model 3

Predictor	<i>b</i>	95% CI	<i>z</i>	<i>p</i>
Intercept	1.74	[1.62, 1.85]	29.02	<.001
Pretest3 correct	-0.10	[-0.14, -0.05]	-4.54	<.001
Parentheses	-0.12	[-0.20, -0.04]	-3.00	.003
Left1	-0.11	[-0.21, -0.01]	-2.22	.026
Right1	0.24	[0.14, 0.33]	4.90	<.001

testing the interaction effects

```

model4 <- glm(condition_total_attempt_count ~ pretest3_correct + Parentheses
+ left + right + right*Parentheses + left*Parentheses, family= "poisson", dat
a= df)

```

```
summary(model4)
```

```

FALSE
FALSE Call:
FALSE glm(formula = condition_total_attempt_count ~ pretest3_correct +
FALSE      Parentheses + left + right + right * Parentheses + left *
FALSE      Parentheses, family = "poisson", data = df)
FALSE
FALSE Deviance Residuals:
FALSE      Min        1Q      Median        3Q        Max
FALSE -1.6909   -0.6980   -0.3506    0.2554    5.6728
FALSE
FALSE Coefficients:
FALSE              Estimate Std. Error z value Pr(>|z|)

```

```

FALSE (Intercept)      1.82736    0.06339   28.827 < 2e-16 ***
FALSE pretest3_correct -0.09093    0.02125   -4.279 1.87e-05 ***
FALSE Parentheses      -0.33683    0.07068   -4.766 1.88e-06 ***
FALSE left1            -0.32440    0.07560   -4.291 1.78e-05 ***
FALSE right1           0.11136    0.06705    1.661 0.096730 .
FALSE Parentheses:right1 0.26241    0.09735    2.695 0.007031 **
FALSE Parentheses:left1 0.40106    0.10364    3.870 0.000109 ***
FALSE ---
FALSE Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
FALSE
FALSE (Dispersion parameter for poisson family taken to be 1)
FALSE
FALSE      Null deviance: 679.51  on 537  degrees of freedom
FALSE Residual deviance: 577.89  on 531  degrees of freedom
FALSE AIC: 2353
FALSE
FALSE Number of Fisher Scoring iterations: 4

m4_results <- apa_print(model4) # make the table in APA format
apa_table(m4_results,
          caption = "Table for model 4", , escape = TRUE) # print the table i
n APA format

```

(#tab:unnamed-chunk-7)

Table for model 4

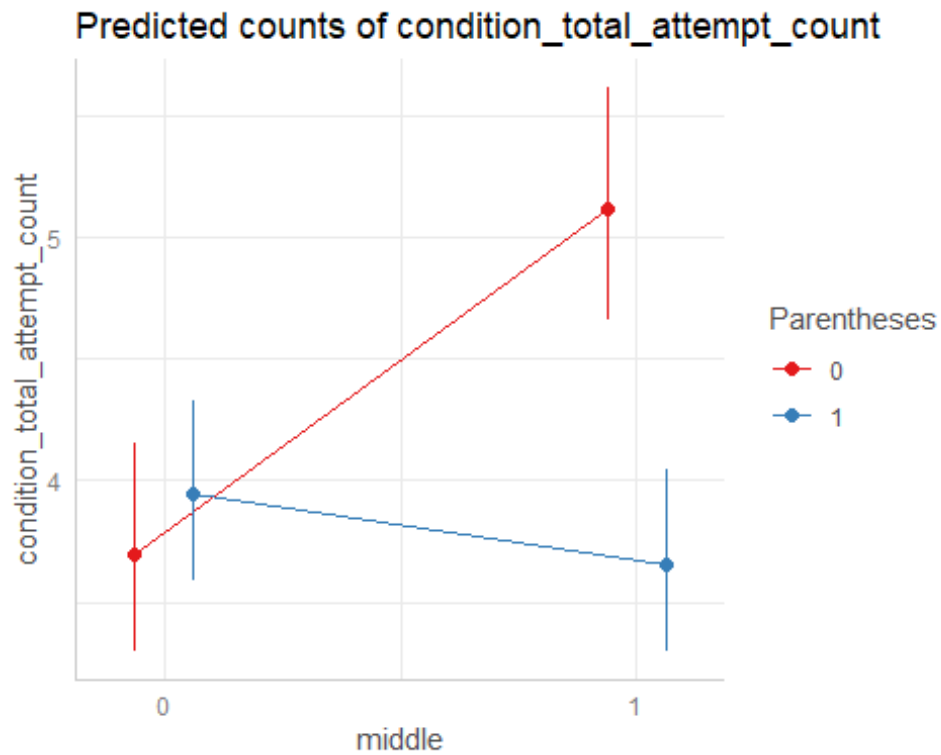
Predictor	<i>b</i>	95% CI	<i>z</i>	<i>p</i>
Intercept	1.83	[1.70, 1.95]	28.83	< .001
Pretest3 correct	-0.09	[-0.13, -0.05]	-4.28	< .001
Parentheses	-0.34	[-0.48, -0.20]	-4.77	< .001
Left1	-0.32	[-0.47, -0.18]	-4.29	< .001
Right1	0.11	[-0.02, 0.24]	1.66	.097
Parentheses × Right1	0.26	[0.07, 0.45]	2.70	.007
Parentheses × Left1	0.40	[0.20, 0.60]	3.87	< .001

Generating the interaction plots

```

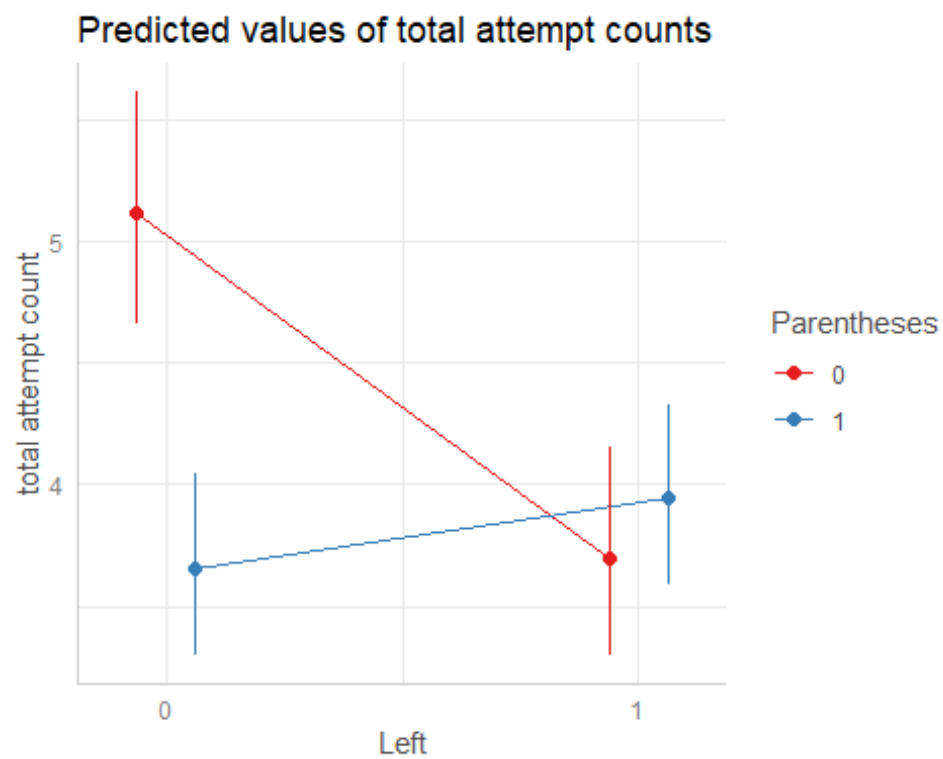
# interaction plot for model 2
interaction1 <- ggpredict(model2, terms = c("middle", "Parentheses"))
plot(interaction1, connect.lines = TRUE)

```

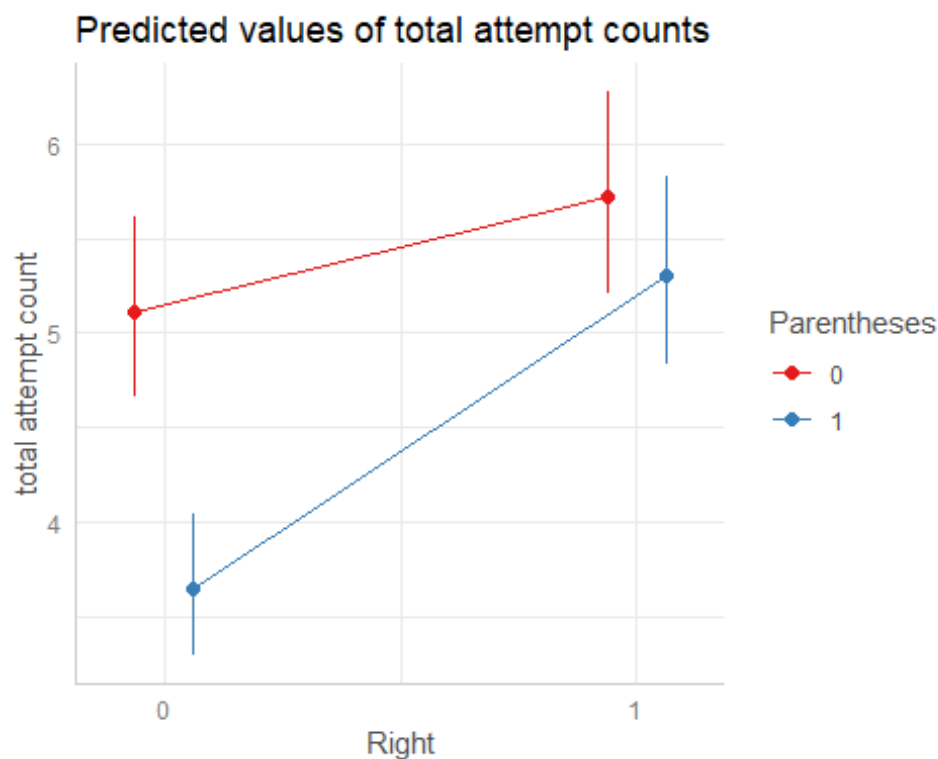


```
# interaction plot for model 4
interaction2 <- ggpredict(model4, terms = c("left", "Parentheses"))
interaction3 <- ggpredict(model4, terms = c("right", "Parentheses"))

plot(interaction2, connect.lines = TRUE) + labs(
  x = "Left",
  y = "total attempt count",
  title = "Predicted values of total attempt counts")
```

```
plot(interaction3, connect.lines = TRUE) + labs(  
  x = "Right",  
  y = "total attempt count",  
  title = "Predicted values of total attempt counts")
```



building a model predicting mastery speed

```
model5 <- lm(condition_average_first_response_or_request_time ~ pretest3_correct + Paratheses + left + right + Paratheses*left + Paratheses*right, data = df)
```

```
summary(model5)
```

FALSE

FALSE Call:

```
FALSE lm(formula = condition_average_first_response_or_request_time ~
FALSE   pretest3_correct + Paratheses + left + right + Paratheses *
FALSE   left + Paratheses * right, data = df)
```

FALSE

FALSE Residuals:

```
FALSE   Min      1Q  Median      3Q      Max
FALSE -1337  -766   -425   -90  100659
```

FALSE

FALSE Coefficients:

```
FALSE               Estimate Std. Error t value Pr(>|t|)
FALSE (Intercept)      601.5      860.4   0.699    0.485
FALSE pretest3_correct    -46.2      277.1  -0.167    0.868
FALSE Paratheses       -387.8      879.4  -0.441    0.659
FALSE left1            -116.0      932.9  -0.124    0.901
FALSE right1           -440.1      934.9  -0.471    0.638
FALSE Paratheses:left1    818.0     1245.0   0.657    0.511
```

```
FALSE Parentheses:right1 1590.4 1289.2 1.234 0.218
FALSE
FALSE Residual standard error: 5975 on 531 degrees of freedom
FALSE Multiple R-squared: 0.004865, Adjusted R-squared: -0.006379
FALSE F-statistic: 0.4327 on 6 and 531 DF, p-value: 0.8572
```