

Statistical Methods – COSC 6323 - Home-Work-10

By Dinesh Narlakanti (2083649)

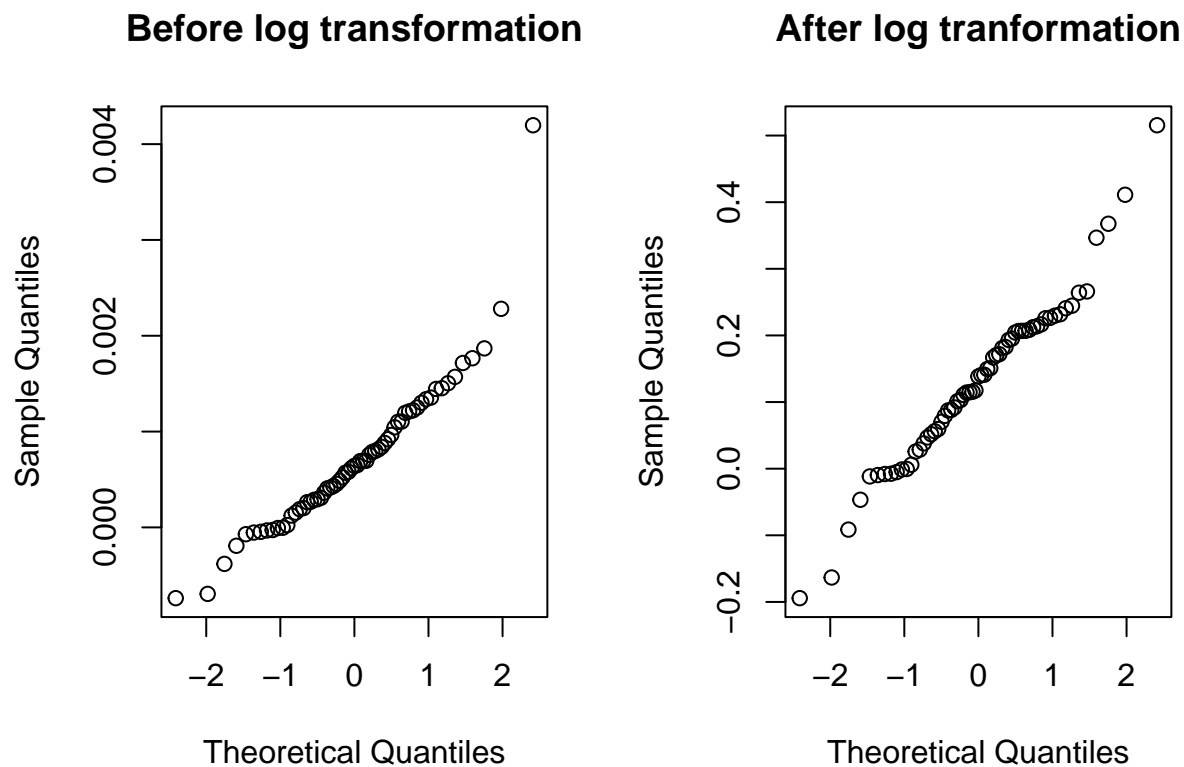
INTRODUCTION

63 participants were randomly assigned to four groups for a writing an essay task. Mainly divided into two groups (Email mode: Batch and Continual) and each group is further divided into two other groups (Anticipatory stress: High and Low). High stressor was implemented by informing the participant about an upcoming presentation after the writing task.

This report mainly concentrated on:

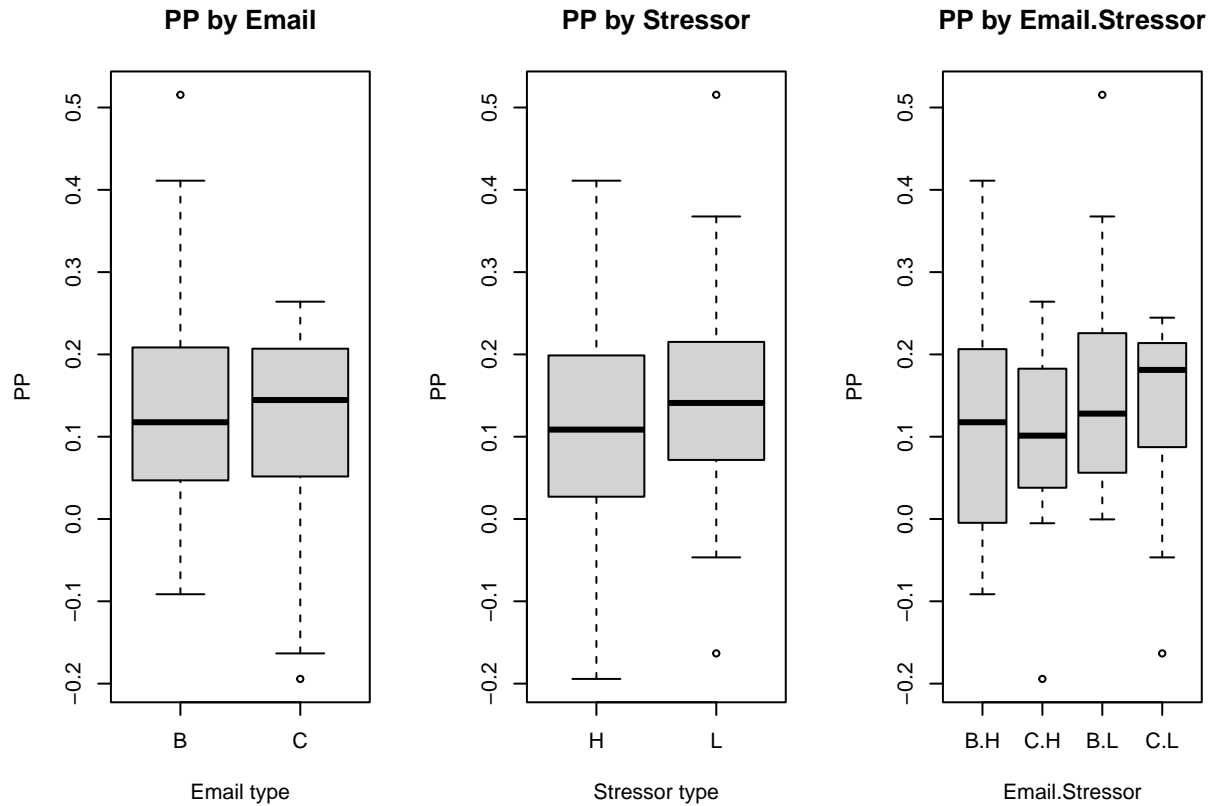
- Whether the manipulation (B/c and H/L) affects stress of the participants.
- Whether or not to log transform the pp(perinasal perspiration) values.
- Performing anova and generalized linear model. Also comparison of the two models.
- Graphs and plots for visualization.

Checking whether to log tranform the delta_pp values:



As log-corrected values are more normalized, so, considering log corrected mean difference as response variable.

Box-plots:



Inference from above boxplots:

- 1) By looking at the box plots, we can say that there is no significance difference in the stress levels by manipulating participants into B/C and H/L.
- 2) From 'PP by Email', we can conclude that participants undergo little extra stress when they receive continual emails.
- 3) Also, the participants who doesn't know about the upcoming presentation had more stress arousal than the participants who already know the the upcoming presentation.
- 4) Although, there is no significance difference in the stress levels but from the data we have, we can say there is slight extra stress in the participants who received continual emails and doesn't know about the presentation(C.L). And has slight less stress in the participants who received continual emails and know about the presentation(C.H).

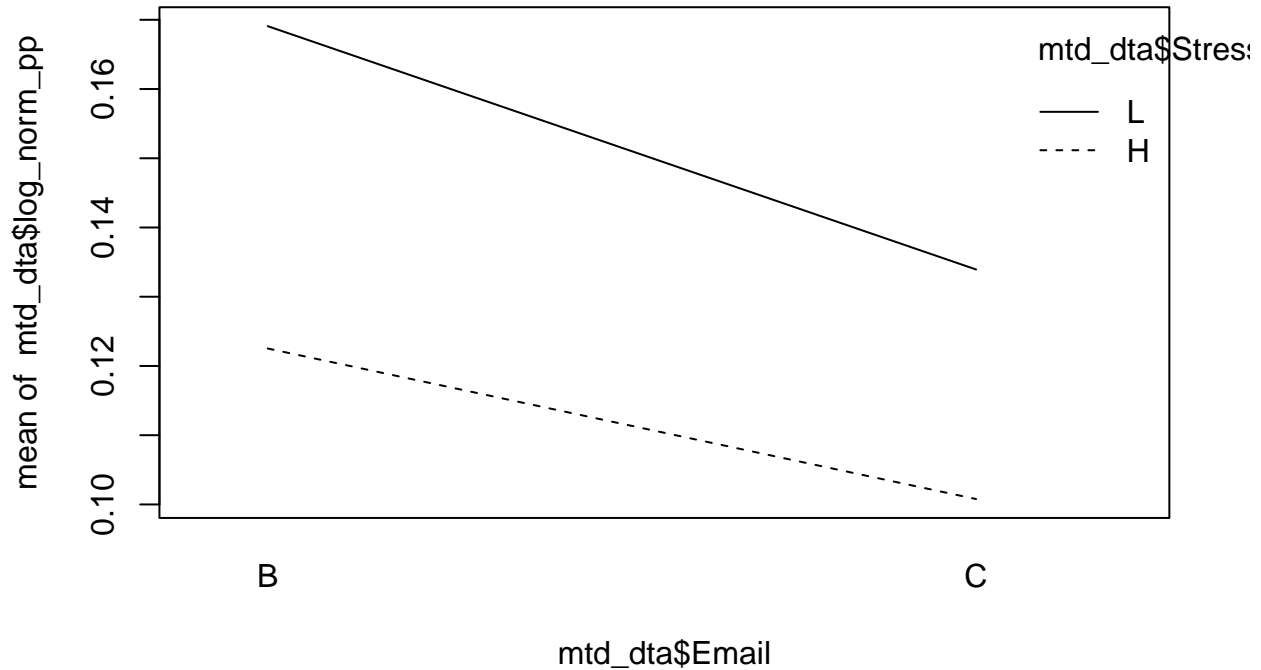
ANOVA test results

##	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## Email	1	0.0120	0.011957	0.744	0.392
## Stressor	1	0.0243	0.024338	1.514	0.223
## Email:Stressor	1	0.0007	0.000703	0.044	0.835
## Residuals	59	0.9486	0.016078		

Inference from Anova results.

- 1) As p-value of Email mode(0.392) is greater than 0.05, we can conclude that there is no significant effects of it in participants' stress arousal.

- 2) As p-value of Anticipatory stress(0.223), which is again greater than 0.05, we can conclude that stressor also doesn't have any significant effect on participants' stress arousal.
- 3) p-value of the interaction is also greater than 0.05 (0.835), this reveals that there is no statistically significant interaction between Stressor and Email.
- 4) To support the above statement, below is the interaction plot of the stressor and email features. We can clearly observe, there is no interaction between the features.

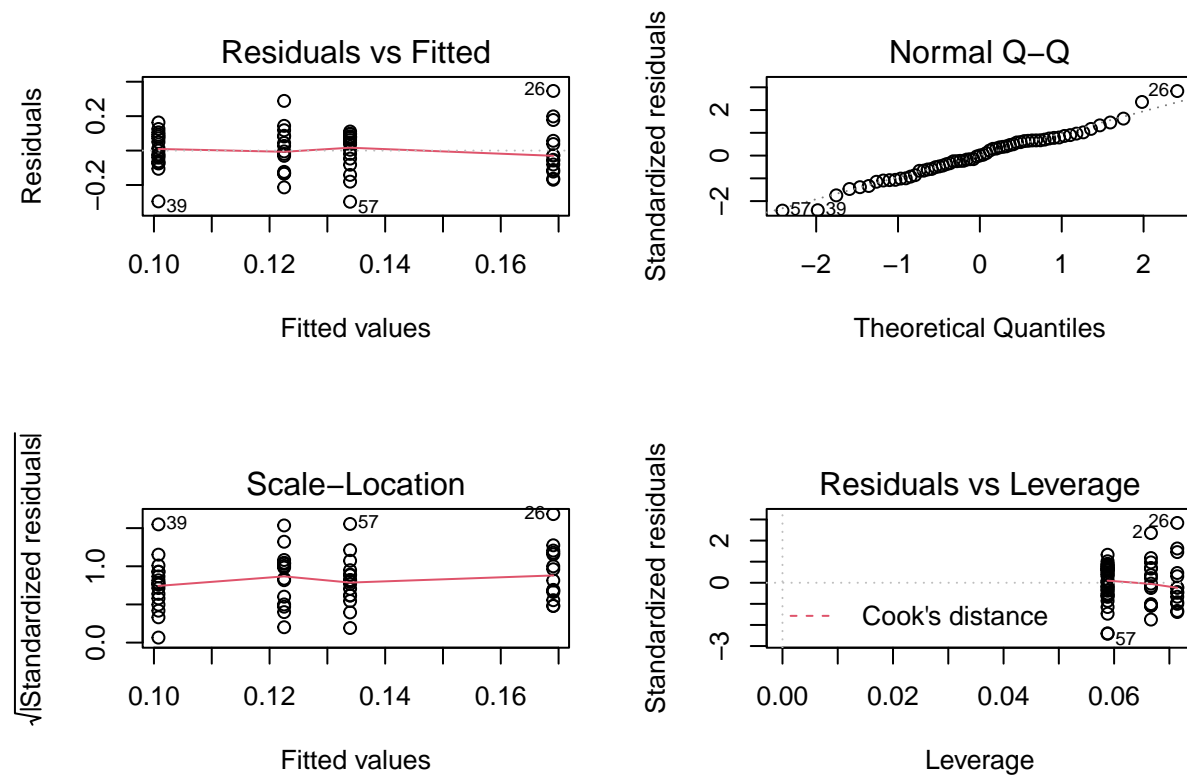


5) We can also confirm that there are no significant interactions from post-hoc test also. From the below post-hoc test, as there are no p-values less than 0.05, we can again conclude that there are no statistically significant interactions.

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = log_norm_pp ~ Email * Stressor, data = mtd_dta)
##
## $Email
##          diff          lwr          upr      p adj
## C-B -0.02764014 -0.09177442 0.03649414 0.391973
##
## $Stressor
##          diff          lwr          upr      p adj
## L-H 0.03930889 -0.02463115 0.1032489 0.2235172
##
## $`Email:Stressor`
##          diff          lwr          upr      p adj
## C:H-B:H -0.02173941 -0.14049243 0.09701361 0.9623612
## B:L-B:H 0.04655998 -0.07801451 0.17113447 0.7567333
```

```
## C:L-B:H  0.01141359 -0.10733943 0.13016661 0.9941821
## B:L-C:H  0.06829939 -0.05268578 0.18928456 0.4484439
## C:L-C:H  0.03315300 -0.08182912 0.14813512 0.8710175
## C:L-B:L -0.03514639 -0.15613156 0.08583878 0.8685130
```

Diagnostic plots:



Interpretations from diagnostic plots:

- 1) Residuals vs Fitted: AS we have a horizontal line without any distinct pattern. We can conclude that we have linear relationship.
- 2) Normal Q-Q: All the residual points are following the dashed line, so, we can conclude that residuals are normally distributed.
- 3) Scale-Location: We have a horizontal line with equally spread points, this is great indication of homoscedasticity.
- 4) Residuals vs Leverage: When checked, there are 2 points(outliers) can be influential against a regression line.

GLM test results

```
##
## Call:
## glm(formula = log_norm_pp ~ Email * Stressor, data = mtd_dta)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29723  -0.07802   0.00053   0.08187   0.34636
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.12252    0.03274   3.742 0.000415 ***
## EmailC         -0.02174    0.04492  -0.484 0.630189
## StressorL       0.04656    0.04712   0.988 0.327127
## EmailC:StressorL -0.01341    0.06412  -0.209 0.835104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.01607773)
##
##      Null deviance: 0.98558  on 62  degrees of freedom
## Residual deviance: 0.94859  on 59  degrees of freedom
## AIC: -75.557
##
## Number of Fisher Scoring iterations: 2
```

Inferences from GLM test results.

1) One unit increase in the predictor variable 'EMailC' (Continual email mode) is associated with an average change on -0.02174 in the log odds of the response variable 'log_norm_pp' (difference of log corrected pp values) taking on a value of 1. So, higher the value of 'EmailC' is associated with lower likelihood of 'log_norm_pp'.

2) As none of the features or interaction between the features have p-values less than 0.05, we can conclude that there are no statistically significant predictor variables in the model.

3) When calculated, P-value of the model is less than 0.05, we can conclude that model is highly useful.

```
## [1] 7.43271e-06
```

4) When compared the 2 models(two way anova and glm) using anova(), we got Res.Df and RSS as same values and 0 df. So, we can conclude that two models are same.

```
## Analysis of Variance Table
##
## Model 1: log_norm_pp ~ Email * Stressor
## Model 2: log_norm_pp ~ Email * Stressor
##   Res.Df    RSS Df Sum of Sq F Pr(>F)
## 1      59 0.94859
## 2      59 0.94859  0 1.1102e-16
```