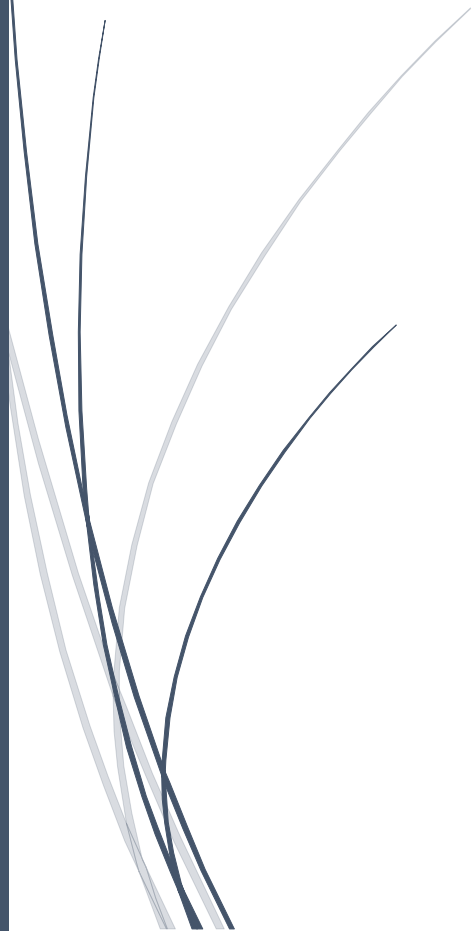
4/6/2020

Clinical Biostatistics

(BCA_CLB – Assignment 1)



Student ID: 6707

(Unikey-skau6707)

Report 1 (Statistical Process control analysis for Bloodstream infections)

Introduction

Bacteraemia is a blood stream infection which can be life threatening if not treated on time, although such infections cannot be completely eliminated but is definitely preventable. In order to prevent these infections to ensure patient's safety, the hospital can monitor bacteraemia to flag higher than expected infection rates and implement changes at the hospital, which could bring the system in control. This study examines the bacteraemia dataset "weeklybacts" collected over a period of 96 consecutive weeks through statistical process control charts, to figure out the special or common cause of variation in the process which could be a sign of increasing rate of infections.

Methodology

Statistical process control methods such as Shewhart C chart and EWMA (*Exponentially weighted moving average*) control charts were used to detect unacceptable rate of infections as these are best suited with count data. Shewhart chart detects large and more abrupt changes in the number of infections which arise due to special causes whereas EWMA charts helps in detecting sustained rise in the mean infection rate, which is the result of common cause of variation.

To begin with the analysis, it was assumed that the data of weekly counts of infections follow Poisson distribution and number of infections per week are independent and stable. Thus, the infections of the hospital were examined visually through Shewhart and EWMA control charts. In addition, the central line (sample mean) and control limits (at 2 standard deviation from the mean) were defined for each graphical method, which were used to indicate statistically significant increase in the infection rate. The significant increase detected in the Shewhart chart (i.e. when exceeds the upper control limit) suggests that increase is more likely to be a result of the process that has not been implemented appropriately. On the other hand, if the upper control limit is exceeded in EWMA chart it indicates statistically significant increase in the mean value, which is due to common cause of variation.

Descriptive analysis

The number of infections recorded from 2 May,2018 to 26 Feb,2020 over a time span of 96 weeks had the minimum and maximum number of infections as 2 and 14 per week respectively, with the average of around 7 infections per week. Also, the median number of infections were found to be 7 per week means 50% of the time number of infections are 7. Since, the mean is similar to variance, which indicates Poisson distribution is appropriate for further statistical analysis. (as mean = 7.3 & variance = 6.59).

	Minimum	Maximum	Mean	Variance	Median
Number of infections	2	14	7.3	6.6	7

Table1: Descriptive statistics of number of infections per week.

Statistical analysis

To determine any special cause of variation in the process, the number of infections were plotted against time (weeks) with control limits set to 2 standard deviation from the mean where sample mean is used for central line. The control limits are calculated using square root of the mean as the standard deviation.
i.e.

Control limit = mean $\pm$ 2* standard deviation
(& if the lower limit is negative then it is set to zero)

Thus, the lower and upper control limits for Shewhart chart are: (1.89, 12.71) infections per week.
The Shewhart chart is shown in figure 1 below.

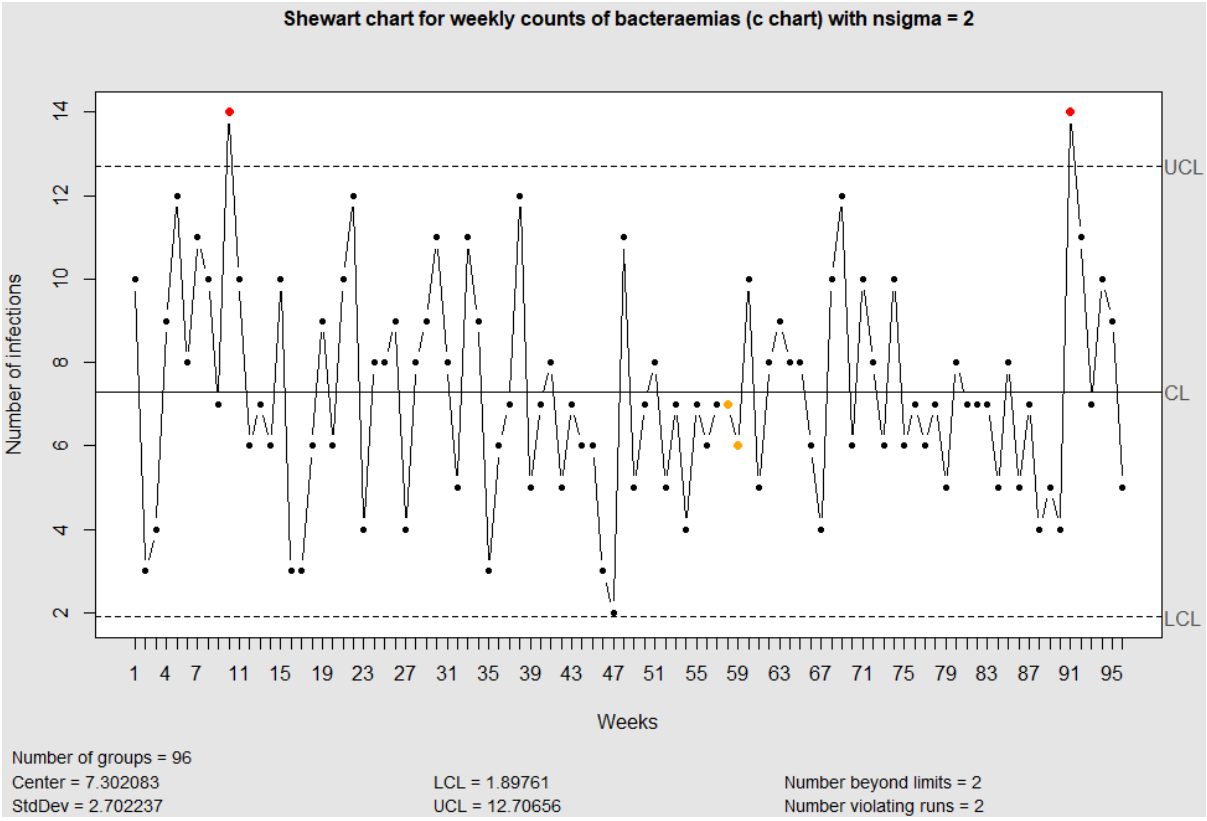


Figure 1: Shewhart chart for weekly counts of bacteraemia.

From above chart it is evident that there exists a positive trend in rate of infections as two observations are outside the upper limit indicating the variation in the process is not random but is due to some special cause. And two of the observations are below the central tendency which means there is a decrease in the infections rate within a time frame of 52nd to 59th week (i.e. 24th April 2019 to 12th June 2019) which is another important time frame to consider to figure out what leads to decrease in number of infections than expected during that time as for 8 continuous weeks the infections are low than average.
Now to determine any small persistent shift in the system mean EWMA chart was plotted. To construct the plot EWMA values and confidence limits were computed using weight w = 0.2 and the calculation for first 10 weeks is illustrated in Table 2. The EWMA

Value for the week is 0.2 times the count for the week added to 0.8 times the EWMA Value for the previous week and for calculating the control limits 2 standard deviations from the mean - the formula for standard deviation is

$SD_E = k \times SD_s \times \sqrt{\frac{w}{2-w}}$ where, SD_E and SD_s are the EWMA and Shewhart standard deviation respectively. For first six weeks the $k = 0.6, 0.77, 0.86, 0.91, 0.94$, and 0.97 respectively and for remaining time $k \approx 1$.

Week	Count	Mean	EWMA	UCL	LCL
1	10	7.3	(7.3*0.8) +(10*0.2) = 7.84	6.27	8.33
2	3	7.3	(7.84*0.8) +(3*0.2) = 6.87	5.99	8.61
3	4	7.3	(6.87*0.8) +(4*0.2) = 6.30	5.83	8.77
4	9	7.3	(6.30*0.8) +(9*0.2) = 6.84	5.74	8.86
5	12	7.3	(6.84*0.8) +(12*0.2) = 7.87	5.68	8.92
6	8	7.3	(7.87*0.8) +(8*0.2) = 7.90	5.64	8.95
7	11	7.3	(7.90*0.8) +(11*0.2) = 8.52	5.63	8.97
8	10	7.3	(8.52*0.8) +(10*0.2) = 8.82	5.61	8.98
9	7	7.3	(8.82*0.8) +(7*0.2) = 8.46	5.60	8.99
10	14	7.3	(8.46*0.8) +(14*0.2) = 9.57	5.60	9.00

Table 2: Calculation for EWMA and control limits.

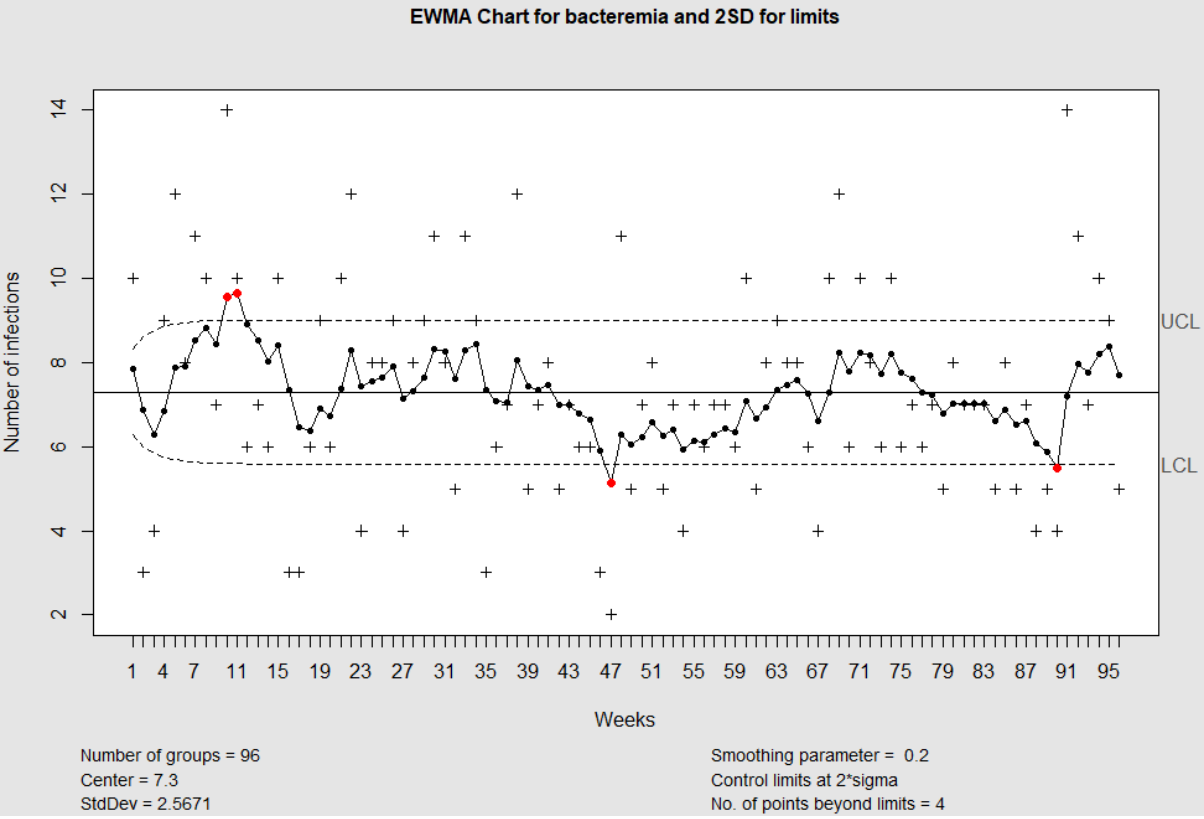


Figure 2: EWMA chart for bacteraemia.

The EWMA chart plotted for the data in figure 2 indicates statistically significant result as EWMA line exceeds the upper control limit at two points and this is more likely due to common cause of variation means the system itself is suboptimal.

Discussion

Statistical control charts detect the unnatural variation that helps to bring the system in control through rapid assessment and clinical intervention. The Shewhart control chart indicating presence of non-random variation in the process, means the hospital staff members such as nurse or physician is either not following measures of hygiene to the accepted standard or the cleaners are not cleaning the hospital as usual. In contrast, the common cause of variation detected by EWMA chart suggests that the system is itself not optimal which could be because of the failures in the hygiene or antibiotic control programmes, so to run an optimal system, hospital management should ensure the proper use of personal protective equipment such as gloves, masks and gowns and implement environment infection control measures, hand hygiene and aseptic technique. However, for the system to be in control with only random variation, it is advised to first target common cause of variation, and then figure out the special causes. Moreover, it would be good to have a detail of what specific measures were adopted during weeks 52-59 as this time period have lower number of infections than expected. This could be because of some better hygiene standards practiced during that time by staff members.

Apart from the uses of methods of statistical process control, there are few limitations −the control charts just indicate for an out of norm rather than indicating the system is working fine. Further these methods just measure variation in the system rather than finding underlying cause of the infection in each patient. On the other hand, the target value used to define threshold to detect increase in rate of infections is the sample mean rather than value obtained from the historical data in control, so the results obtained through analysis does not suggest that system is completely out of control.

Report 2 (Assessing the agreement between two raters for % of stuttered syllables)

Introduction

In a clinical trial, for stuttering treatment, the standard outcome measure is the percentage of stuttered syllables (i.e. %SS) which is observed by two independent raters for 59 subjects. This study is thus conducted to determine the extent of agreement between two observers which would be effective in further treatment to minimise the stuttering among adults. Although the primary outcome is percentage of stuttered syllables, however ordinal rating procedures have advantage over that method, so during the study, two analysis were performed based on numerical data and ordinal data derived from the numerical data.

Methodology

To assess inter-rater reliability, (for primary outcome), Bland-Altman plot is used in which differences are plotted against the mean for each subject, in addition to the percentage of agreement (computed in R software) assuming the differences are approximately normally distributed. Also, the limits of agreement are defined at 2 standard deviations from the mean. (A similar analysis was performed on transformed data which also leads to similar results but for better interpretation we follow the results obtained using raw data (see Appendix 2))
For second analysis, the values were first dichotomised into "minimal stuttering" (0%SS to <1%SS) or "mild to severe stuttering" (1%SS+). Then to assess the agreement between two raters, Cohen's Kappa was used.

i.e. $\kappa = \frac{observed\ agreement - chance\ agreement}{1 - chance\ agreement}$

The kappa coefficient ranges between -1 and 1 where, 0 means no agreement and 1 indicates perfect agreement. To incorporate the ordered nature of the outcomes weighted kappa coefficient was used to decide on agreement of measurements taking into account agreement that could occur just by chance. Thus, Kappa statistic was computed in R software. (unweighted and weighted kappa gave the same result as there are only two variables)

Descriptive analysis

The minimum %SS observed by two raters is 2% whereas maximum %SS observed are 11%, 7.2% by rater 1 and rater 2 respectively. On average, the %SS recorded by rater 1 is 2.84% and by rater 2 is 1.96%.

Variable	Minimum (%)	Maximum (%)	Mean (%)	Median (%)	Variance (%)
Rater 1	0	11	2.84	1.4	8.45
Rater 2	0	7.2	1.96	1.1	3.69

Table 1: Summary of percentage of stuttered syllables observed by two raters.

The plot between the reading of two raters (in figure 1) indicates that there is weak agreement between the two observers as the points are quite scattered from the equity line. (i.e. line of agreement) but statistically superior results were obtained using Bland-Altman method.

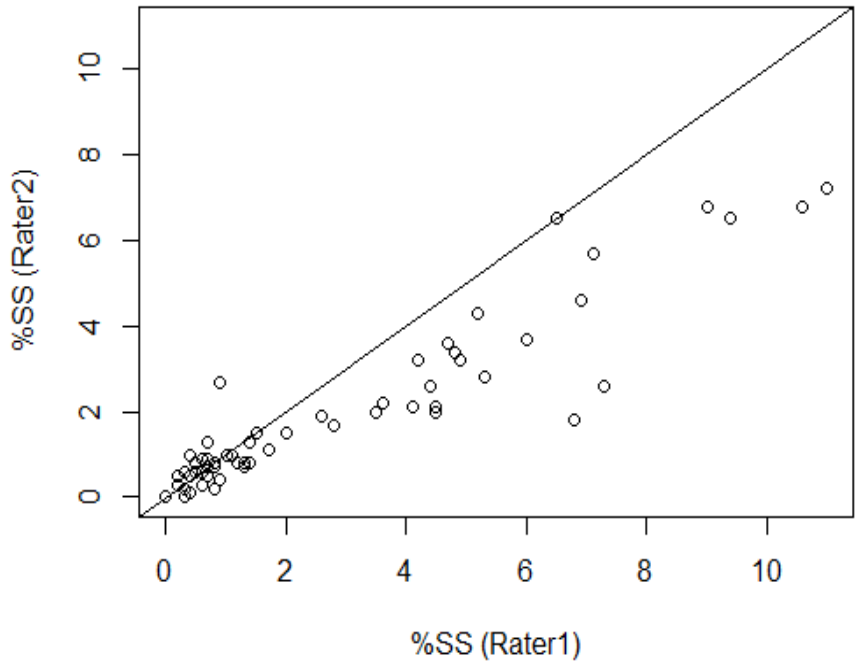


Figure 1: Scatter plot of the measurements of two raters.

Statistical analysis

The Bland-Altman plot constructed for the primary outcome of the trial is shown in figure 2.

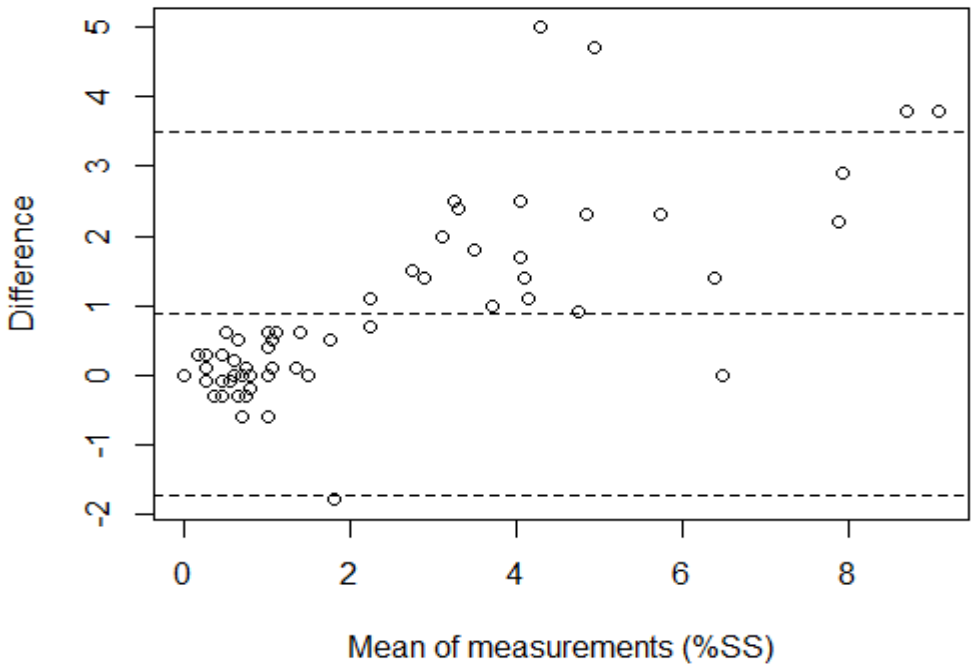


Figure 2: Bland-Altman plot for %SS rated by two raters.

It is evident from the plot (in figure 2) that mean difference is quite different from zero and 5 out of 59 observations are outside the limits of agreement (-1.72, 3.49) and the observations within the limit are unevenly scattered around the difference which indicates systematic bias and heterogeneous variance in the measurements, therefore, the estimates are less reliable to treat the stuttering adults. Also, the percentage of agreement obtained for continuous measurements is 11.9% which is poor agreement percentage and concludes similar result as obtained from the plot.

Further, to decide inter-rater reliability of two raters for categorical outcomes, Kappa coefficient computed through R was obtained as 0.72. As kappa coefficient is > 0 and close to 1 means there is good agreement, hence the measurements are reliable. The agreement table gives a more detail about the agreement of the measurements as shown in table 2 below:

		Rater 2	
		Minimal stuttering	Mild-Severe stuttering
Rater 1	Minimal stuttering	21	3
	Mild-Severe stuttering	5	30

Table 2: Agreement table for percentage of stuttered syllables.

So, both raters agree that 21 subjects have minimal stuttering and 30 adults have mild to severe stuttering and for remaining 8 measurements the observers disagree. As, raters agree for 51 out of 59 measurements thus, the percentage of agreement is 86.4 which is much better than that obtained for continuous measurements. Hence, measurements from the dichotomised data are reliable indicating the two raters are performing well.

Discussion

In conclusion, the results suggest that measurements observed by rater 1 and rater 2 can be used interchangeably for dichotomised outcomes whereas the measurements are not reliable for the primary outcomes of a trial (i.e. raw data) which indicates either the raters need to retrain or the raters could observe the measurement again. Moreover, similar results were obtained by the percent agreement statistic, which is easily calculated and directly interpretable but its key limitation is that it does not take account of the possibility that raters guessed the measurement; hence it should be used cautiously. While interpreting Kappa coefficient it is hard to determine what exact value should be used as threshold to assess agreement, as 0 indicates no agreement and 1 indicates perfect agreement and any value between 0 and 1 could define some extent of agreement. Perhaps the best advice for researchers is to use both percent agreement and kappa to reach a best possible conclusion for stuttering treatment when dealing with dichotomised data.

Appendix 1

```
library(tidyverse)

bact<-read_csv("weeklybacts.csv")

#bact

summary(bact)

(mean_count<-mean(bact$count))

(variance <-var(bact$count))



library(qcc)



## Shewhart chart with limits at +/- 2 SDs

Title1 <- 'Shewart chart for weekly counts of bacteraemias (c chart) with 2SD limits'

Count1 <- with(bact, qcc(count, type = "c", nsigmas = 2, title = Title1,ylab="Number of infections", restore.par = FALSE,xlab="Weeks"))




## EWMA chart

Title_e <- "EWMA Chart for bacteremia with 2SD limits"

ewma_bact <- ewma(bact$count, nsigmas = 2, title = Title_e,xlab="Weeks", ylab="Number of infections",center = 7.30, std.dev = sqrt(6.59), restore.par = FALSE)



## we get a summary of calculation involved in EWMA chart including LCL and UCL

summary(ewma_bact)
```

Appendix 2

The data was transformed using log transformation on measurements of outcome (%SS) after incrementing the measurements by 1 to avoid numerical issues. The differences on transformation appeared to be normally distributed but the results for agreement were quite similar so we rely on results obtained using raw data. The percentage of agreement is 11.9% and 3 out of 59 observation fall outside limits of agreement and there is a systematic bias as well as heterogenous variance among the readings as observations are not scattered evenly around the bias. Hence the measurements are not reliable.

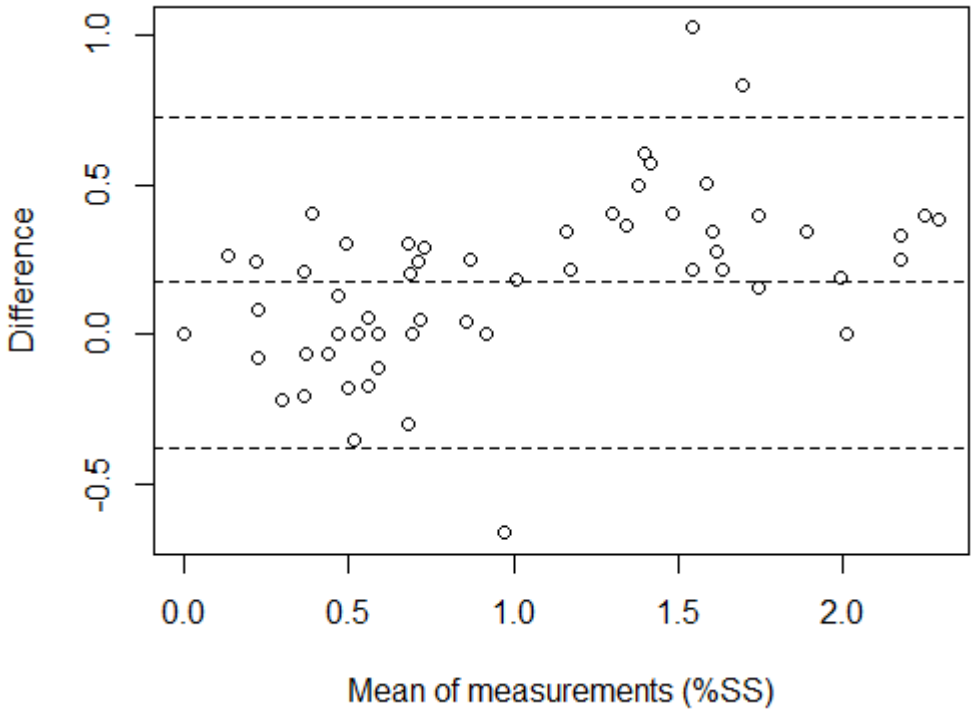


Figure A: Bland-Altman plot for transformed data.

Appendix 3

R code for question 2

```
library(tidyverse)

ss <- read_csv("stuttering_ratings.csv",col_types = cols())

attach(ss)

summary(ss)

var(ss$Rater1);var(ss$Rater2)
```

```
plot(Rater1,Rater2,xlim = c(0,11),ylim = c(0,11),xlab="%SS (Rater1)",ylab="%SS (Rater2)");abline(0,1)

cor(Rater1,Rater2)


## first check the assumption whether differences follow normal distribution

ss$d = Rater1-Rater2

qqnorm(ss$d);qqline(ss$d) ## difference is not normally distributed

shapiro.test(ss$d)


## since readings contain 0 values so to for log transformation we add 1 to each value

ss$rplus1=Rater1+1

ss$rplus2=Rater2+1


## now transforming the data and finding differnce of transformed reading

ss$r1_log=log(ss$rplus1)

ss$r2_log=log(ss$rplus2)

ss$diff=ss$r1_log-ss$r2_log

qqnorm(ss$diff);qqline(ss$diff) ## now the assumption seems verified as difference is normal

shapiro.test(ss$diff)


library(BlandAltmanLeh)

with(ss,bland.altman.plot(Rater1,Rater2,xlab="Mean of measurements (%SS)",ylab="Difference"))

with(ss,bland.altman.stats(Rater1,Rater2))
```

```
with(ss,bland.altman.plot(r1_log,r2_log,xlab="Mean of measurements
(%SS)",ylab="Difference"))

with(ss,bland.altman.stats(r1_log,r2_log))


## alternatively we can get the plot of differnces vs average as follows:

library(blandr)

with(ss,blandr.draw(Rater1,Rater2))

with(ss,blandr.output.text(Rater1,Rater2))


with(ss,blandr.draw(r1_log,r2_log))

with(ss,blandr.output.text(r1_log,r2_log))




## we can test the correlation as follows

suppressMessages(library(PairedData))

r<- with(ss, paired(Rater1, Rater2))

grambsch.Var.test(r)


##alternatively, correlation can be tested as:

with(ss, cor.test(Rater1-Rater2, (Rater1+Rater2)/2))


with(ss, cor.test(r1_log-r2_log, (r1_log+r2_log)/2))


## check for the agreement  for raw data
```

```
library(irr)

agree(ss[,c(1,2)])

kappa2(ss[,c(1,2)]) ##computes cohen kappa between two raters


## check agreement for transformed data

agree(ss[,c(6,7)])

kappa2(ss[,c(6,7)])


############ Analysis on dichotomise variable #############

ss$R1<- ss$Rater1;ss$R2<-ss$Rater2

ss$R1[ss$R1<1]=0;ss$R1[ss$R1>1]=1

ss$R2[ss$R2<1]=0;ss$R2[ss$R2>1]=1


agree(ss[,c(9,10)])

kappa2(ss[,c(9,10)]) ## computes unweighted cohen kappa


## computes weighted cohen kappa

kappa2(ss[,c(9,10)],"squared") ## Compare raters 1 and 2 with squared weights

kappa2(ss[,c(9,10)],"equal") ## Compare raters 1 and 2 with linear weights
```