MANOVA outlier assessment

2022-08-18

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(readxl)  
library(officer)

##   
## Attaching package: 'officer'  
##   
## The following object is masked from 'package:readxl':  
##   
## read\_xlsx

library(flextable)

##   
## Attaching package: 'flextable'  
##   
## The following object is masked from 'package:purrr':  
##   
## compose

library(extrafont)

## Registering fonts with R

library(writexl)  
library(rstatix)

##   
## Attaching package: 'rstatix'  
##   
## The following object is masked from 'package:stats':  
##   
## filter

FOR EACH MEASURE:

To check normality: QQ plot (A Q-Q plot or Quantile-Quantile plot is a common graphical way to check data for non-normality. Quantile is another term for percentile. A Q-Q plot is a plot of the quantiles of a data set versus the quantiles of a reference theoretical distribution. This plot is used to assess normality of residuals. Curvature and outliers indicate problems. Note: The QQ plot is not useful until variance is approximately equal.)

shapiro.test with p value greater than 0.05 (Histograms and QQ plots are usually more informative than the tests, because small sample sizes generally “pass” the test (high p-value, no evidence against normality), and large sample sizes generally “fail” (small p-value, evidence against normality).)

To check for outliers:

Standardized (or studentized) residuals are residuals that have been “standardized” by dividing each residual by it’s SE. • They have approximately a t-distribution, which is approximately normal for moderate sample sizes, so we expect that about 95% of the standardized residuals will be between -2 and +2. Values greater in absolute value than 3.5 are usually considered outliers. • In R, standardized residuals can be found using rstandard(ModelObject).

To check homogeneity of covariances/equal variance:

Box’s test (homegeneity of variance-covariance matrices)

Levene’s test of equality of variances (homogeneity of variances)

Plot of residuals vs fitted values (equal scatter; This is the primary diagnostic plot for assessing linearity and constant variance. Curvature, unequal variance (megaphone) or outliers indicate problems.)

To check linear response:

Plot of resids vs fitted values (should not show a trend)

Look to pages 20-end 512 MultReg1

For y variables: You should have an adequate sample size. Although the larger your sample size, the better, at a bare minimum, there needs to be as many cases (e.g., particpants) in each cell of the design as there are number of dependent variables. 12 dependent variables

For MANOVA, the dependent variables should be somewhat correlated.

STEPS: Power analysis to find sample size - 278 NAs shouldn’t be over 20% Find outliers & remove Double check correlations check assumptions run manova run anovas

#times new roman tables  
my\_ft\_theme <- function(ft, ...) {  
 # Remove vertical cell padding  
 ft <- padding(ft, padding.top = 0, padding.bottom = 0, part = "all")  
   
 # Change font to TNR 11  
 ft <- font(ft, fontname = "Times New Roman", part = "all")  
 ft <- fontsize(ft, part = "all", size = 12)  
 ft  
}

#remove identifiers and demographic data  
  
headscan\_full<-read\_excel("C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\headscan\_full.xlsx")  
  
headscan\_full$AA\_C <- headscan\_full$AA\_C \*10  
headscan\_full$BGl\_C <- headscan\_full$BGl\_C \* 10  
headscan\_full$BiW\_C <- headscan\_full$BiW\_C \*10  
headscan\_full$BiW\_L <- headscan\_full$BiW\_L \*10  
headscan\_full$ChCh\_C <- headscan\_full$ChCh\_C \*10  
headscan\_full$GoSub\_C <- headscan\_full$GoSub\_C \*10  
headscan\_full$NRB\_L <- headscan\_full$NRB\_L \*10  
headscan\_full$ProA\_L <- headscan\_full$ProA\_L \*10  
headscan\_full$ProA\_C <- headscan\_full$ProA\_C \*10  
headscan\_full$ProS\_C <- headscan\_full$ProS\_C \*10  
headscan\_full$ProS\_L <- headscan\_full$ProS\_L \*10  
headscan\_full$SelP\_C <- headscan\_full$SelP\_C \*10  
headscan\_full$SelP\_L <- headscan\_full$SelP\_L \*10  
headscan\_full$SelDH\_C <- headscan\_full$SelDH\_C \*10  
headscan\_full$SelM\_L <- headscan\_full$SelM\_L \*10  
headscan\_full$SnasM\_C <- headscan\_full$SnasM\_C \*10  
headscan\_full$SmanM\_C <- headscan\_full$SmanM\_C \*10  
headscan\_full$SmanM\_L <- headscan\_full$SmanM\_L \*10  
headscan\_full$SnasM\_L <- headscan\_full$SnasM\_L \*10  
headscan\_full$TrHO\_C <- headscan\_full$TrHO\_C \*10  
headscan\_full$TrEJ\_C <- headscan\_full$TrEJ\_C \*10  
headscan\_full$TrGo\_C <- headscan\_full$TrGo\_C \*10  
headscan\_full$TrSel\_C <- headscan\_full$TrSel\_C \*10  
headscan\_full$TrSman\_C <- headscan\_full$TrSman\_C \*10  
headscan\_full$TrSnas\_C <- headscan\_full$TrSnas\_C \*10  
headscan\_full$TrTr\_C <- headscan\_full$TrTr\_C \*10  
headscan\_full$TrTr\_L <- headscan\_full$TrTr\_L \*10  
  
headscan\_full$gender <- as.factor(headscan\_full$gender)  
headscan\_full$race\_eth <- as.factor(headscan\_full$race\_eth)  
headscan\_full$age\_group <- as.factor(headscan\_full$age\_group)  
  
str(headscan\_full)

## tibble [2,016 × 33] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:2016] "400-20201012-002" "400-20201012-003" "400-20201012-004" "400-20201012-005" ...  
## $ AA\_C : num [1:2016] 65 55 70 58 67 60 59 59 65 65 ...  
## $ BGl\_C : num [1:2016] 315 289 293 313 288 306 320 NA 300 277 ...  
## $ BiW\_C : num [1:2016] 130 127 143 140 137 130 141 138 143 150 ...  
## $ BiW\_L : num [1:2016] 115 108 121 109 104 106 109 111 113 116 ...  
## $ ChCh\_C : num [1:2016] 62 64 68 70 70 70 67 69 67 63 ...  
## $ GoSub\_C : num [1:2016] 93 93 115 93 103 100 79 106 85 102 ...  
## $ NRB\_L : num [1:2016] 17 18 19 21 19 14 17 18 16 17 ...  
## $ ProA\_L : num [1:2016] 28 25 31 23 28 28 26 27 32 28 ...  
## $ ProA\_C : num [1:2016] 31 27 33 27 31 29 27 29 34 31 ...  
## $ ProS\_C : num [1:2016] 18 20 14 13 22 22 19 14 26 24 ...  
## $ ProS\_L : num [1:2016] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_C : num [1:2016] 42 41 51 45 47 48 46 42 47 44 ...  
## $ SelP\_L : num [1:2016] 42 41 51 44 47 48 46 41 46 44 ...  
## $ SelDH\_C : num [1:2016] 15 9 9 11 13 15 9 9 12 14 ...  
## $ SelM\_L : num [1:2016] 122 99 130 115 119 126 117 112 117 117 ...  
## $ SnasM\_C : num [1:2016] 82 55 84 74 73 80 78 76 64 75 ...  
## $ SmanM\_C : num [1:2016] 59 51 45 43 33 34 55 37 61 41 ...  
## $ SmanM\_L : num [1:2016] 55 50 45 42 33 34 50 36 59 40 ...  
## $ SnasM\_L : num [1:2016] 75 53 78 69 67 76 69 71 62 69 ...  
## $ TrHO\_C : num [1:2016] 179 163 169 166 159 162 169 NA 167 166 ...  
## $ TrEJ\_C : num [1:2016] 40 32 39 29 46 42 29 32 29 33 ...  
## $ TrGo\_C : num [1:2016] 84 57 70 61 68 70 75 61 67 64 ...  
## $ TrSel\_C : num [1:2016] 149 138 150 133 140 151 140 138 156 143 ...  
## $ TrSman\_C : num [1:2016] 177 145 178 147 157 164 149 159 151 160 ...  
## $ TrSnas\_C : num [1:2016] 163 142 167 145 152 157 148 149 157 NA ...  
## $ TrTr\_C : num [1:2016] 296 276 292 273 279 300 283 275 307 286 ...  
## $ TrTr\_L : num [1:2016] 155 141 156 149 146 146 147 151 157 144 ...  
## $ coder : chr [1:2016] "Kayna" "Kayna" "Kayna" "Kayna" ...  
## $ age : num [1:2016] 31 49 49 34 49 55 26 18 25 27 ...  
## $ gender : Factor w/ 4 levels "Female","Male",..: 2 1 2 2 2 2 2 2 2 2 ...  
## $ race\_eth : Factor w/ 8 levels "AIAN","Asian",..: 3 8 8 8 8 8 3 3 8 8 ...  
## $ age\_group: Factor w/ 3 levels "18-36","37-54",..: 1 2 2 1 2 3 1 1 1 1 ...

SA3\_data <- headscan\_full[c(1,2,4,5,7,8,12,14,16,17,25,27,28,31,32,33)]

SA3\_NAsums <- colSums(is.na(SA3\_data))  
  
SA3\_NAprops <- colMeans(is.na(SA3\_data))  
  
SA3\_NAprops1 <- as.data.frame(SA3\_NAprops)  
SA3\_NAprops1 <- rownames\_to\_column(SA3\_NAprops1, "var\_name")  
SA3\_NAprops1 <- SA3\_NAprops1 %>% slice(-c(1, 29:33))  
  
SA3\_NAsums1 <- as.data.frame(SA3\_NAsums)  
SA3\_NAsums1 <- rownames\_to\_column(SA3\_NAsums1, "var\_name")  
SA3\_NAsums1 <- SA3\_NAsums1 %>% slice(-c(1, 29:33))  
  
SA3\_NAs <- inner\_join(SA3\_NAprops1, SA3\_NAsums1, by = "var\_name", desc)  
  
SA3\_NAs$SA3\_NAprops<- round(SA3\_NAs$SA3\_NAprops, digits=4)  
  
SA3\_NAs$var\_name <- fct\_reorder(SA3\_NAs$var\_name, SA3\_NAs$SA3\_NAsums, .desc=TRUE)  
  
str(SA3\_NAs$SA3\_name)

## NULL

#Size 12 Table TNR  
flextable(SA3\_NAs) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("NA values for each MANOVA Variable") %>%   
 autofit() %>%   
 set\_header\_labels(values = list(var\_name = "Variable",  
 SA3\_NAprops = "Proportion of NA values",  
 SA3\_NAsums = "Count of NA values"))

**Table** : NA values for each MANOVA Variable

| **Variable** | **Proportion of NA values** | **Count of NA values** |
| --- | --- | --- |
| AA\_C | 0.0084 | 17 |
| BiW\_C | 0.0084 | 17 |
| BiW\_L | 0.0084 | 17 |
| GoSub\_C | 0.0630 | 127 |
| NRB\_L | 0.0074 | 15 |
| ProS\_L | 0.0094 | 19 |
| SelP\_L | 0.0074 | 15 |
| SelM\_L | 0.1111 | 224 |
| SnasM\_C | 0.1171 | 236 |
| TrSman\_C | 0.0655 | 132 |
| TrTr\_C | 0.0188 | 38 |
| TrTr\_L | 0.0169 | 34 |
| gender | 0.0040 | 8 |
| race\_eth | 0.0000 | 0 |
| age\_group | 0.0005 | 1 |

SA3\_data1 <- SA3\_data  
  
#race/eth  
SA3\_data1$race\_eth <-   
 recode\_factor(SA3\_data1$race\_eth, 'AIAN'= "Other",  
 'NHOPI' = "Other",  
 'PTNS' = "Other")  
  
#gender  
SA3\_data1$gender <-   
 recode\_factor(SA3\_data1$gender, 'Non-binary or Other'= "Other",  
 'Prefer not to say' = "Other")  
  
SA3\_data1$gender[is.na(SA3\_data1$gender)]="Other"  
  
  
summary(SA3\_data1)

## ID AA\_C BiW\_C BiW\_L   
## Length:2016 Min. :44.00 Min. :101.0 Min. : 82.0   
## Class :character 1st Qu.:57.00 1st Qu.:124.0 1st Qu.:104.0   
## Mode :character Median :61.00 Median :133.0 Median :111.0   
## Mean :61.25 Mean :133.4 Mean :111.2   
## 3rd Qu.:65.00 3rd Qu.:141.0 3rd Qu.:118.0   
## Max. :87.00 Max. :188.0 Max. :152.0   
## NA's :17 NA's :17 NA's :17   
## GoSub\_C NRB\_L ProS\_L SelP\_L   
## Min. : 45.00 Min. : 3.00 Min. :12.00 Min. :16.00   
## 1st Qu.: 88.00 1st Qu.:15.00 1st Qu.:17.00 1st Qu.:42.00   
## Median : 99.00 Median :18.00 Median :19.00 Median :44.00   
## Mean : 99.05 Mean :17.98 Mean :19.16 Mean :44.53   
## 3rd Qu.:108.00 3rd Qu.:21.00 3rd Qu.:21.00 3rd Qu.:47.00   
## Max. :424.00 Max. :72.00 Max. :42.00 Max. :65.00   
## NA's :127 NA's :15 NA's :19 NA's :15   
## SelM\_L SnasM\_C TrSman\_C TrTr\_C   
## Min. : 69.0 Min. : 44.0 Min. : 64.0 Min. :241.0   
## 1st Qu.:110.0 1st Qu.: 68.0 1st Qu.:143.0 1st Qu.:272.2   
## Median :116.0 Median : 75.0 Median :152.0 Median :282.0   
## Mean :116.2 Mean : 75.1 Mean :153.4 Mean :282.7   
## 3rd Qu.:123.0 3rd Qu.: 82.0 3rd Qu.:162.0 3rd Qu.:293.0   
## Max. :145.0 Max. :163.0 Max. :298.0 Max. :332.0   
## NA's :224 NA's :236 NA's :132 NA's :38   
## TrTr\_L gender race\_eth age\_group   
## Min. :127.0 Other : 14 Other : 38 18-36:991   
## 1st Qu.:141.0 Female:1063 Asian : 91 37-54:940   
## Median :146.0 Male : 939 Black : 548 55-72: 84   
## Mean :146.5 LatinX: 100 NA's : 1   
## 3rd Qu.:152.0 white :1239   
## Max. :173.0   
## NA's :34

#Dr. Erin Buchanan from Missouri State If there were missing data categories over 20%, I could replace the data? According to manova-9 notes

Otherwise, it seems like I have to delete the missing values. I did this for PCA as well.

SA3\_data2 <- SA3\_data1 %>% drop\_na()  
  
sum(is.na(SA3\_data1))

## [1] 892

2016-1710

## [1] 306

#306 observations were deleted when we dropped NAs. That is not bad considering there are 892 missing values

str(SA3\_data2)

## tibble [1,710 × 16] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:1710] "400-20201012-002" "400-20201012-003" "400-20201012-004" "400-20201012-005" ...  
## $ AA\_C : num [1:1710] 65 55 70 58 67 60 59 59 65 65 ...  
## $ BiW\_C : num [1:1710] 130 127 143 140 137 130 141 138 143 150 ...  
## $ BiW\_L : num [1:1710] 115 108 121 109 104 106 109 111 113 116 ...  
## $ GoSub\_C : num [1:1710] 93 93 115 93 103 100 79 106 85 102 ...  
## $ NRB\_L : num [1:1710] 17 18 19 21 19 14 17 18 16 17 ...  
## $ ProS\_L : num [1:1710] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_L : num [1:1710] 42 41 51 44 47 48 46 41 46 44 ...  
## $ SelM\_L : num [1:1710] 122 99 130 115 119 126 117 112 117 117 ...  
## $ SnasM\_C : num [1:1710] 82 55 84 74 73 80 78 76 64 75 ...  
## $ TrSman\_C : num [1:1710] 177 145 178 147 157 164 149 159 151 160 ...  
## $ TrTr\_C : num [1:1710] 296 276 292 273 279 300 283 275 307 286 ...  
## $ TrTr\_L : num [1:1710] 155 141 156 149 146 146 147 151 157 144 ...  
## $ gender : Factor w/ 3 levels "Other","Female",..: 3 2 3 3 3 3 3 3 3 3 ...  
## $ race\_eth : Factor w/ 5 levels "Other","Asian",..: 3 5 5 5 5 5 3 3 5 5 ...  
## $ age\_group: Factor w/ 3 levels "18-36","37-54",..: 1 2 2 1 2 3 1 1 1 1 ...

Next, identify univariate outliers

#AA\_C  
AA\_C\_out <- SA3\_data2 %>%   
 identify\_outliers(AA\_C)  
  
AA\_C\_out$extreme\_overall <- AA\_C\_out$is.extreme  
  
AA\_C\_out <- AA\_C\_out[c(1,2,19)]  
  
  
  
AA\_C\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(AA\_C)  
  
AA\_C\_out1$extreme\_race\_eth <- AA\_C\_out1$is.extreme  
  
AA\_C\_out1 <- AA\_C\_out1[c(2,3,19)]  
  
  
  
  
AA\_C\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(AA\_C)  
  
AA\_C\_out2$extreme\_gender <- AA\_C\_out2$is.extreme  
  
AA\_C\_out2 <- AA\_C\_out2[c(2,3,19)]  
  
  
  
AA\_C\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(AA\_C)  
  
AA\_C\_out3$extreme\_age <- AA\_C\_out3$is.extreme  
  
AA\_C\_out3 <- AA\_C\_out3[c(2,3,19)]  
  
  
AA\_C\_extreme <- full\_join(AA\_C\_out, AA\_C\_out1, by=c('ID', 'AA\_C'))  
AA\_C\_extreme <- full\_join(AA\_C\_extreme, AA\_C\_out2, by=c('ID', 'AA\_C'))  
AA\_C\_extreme <- full\_join(AA\_C\_extreme, AA\_C\_out3, by=c('ID', 'AA\_C'))  
  
AA\_C\_extreme <-   
 filter(AA\_C\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
  
AA\_C\_extreme$extreme\_overall <- as.character(AA\_C\_extreme$extreme\_overall)  
AA\_C\_extreme$extreme\_race\_eth <- as.character(AA\_C\_extreme$extreme\_race\_eth)  
AA\_C\_extreme$extreme\_gender <- as.character(AA\_C\_extreme$extreme\_gender)  
AA\_C\_extreme$extreme\_age <- as.character(AA\_C\_extreme$extreme\_age)  
   
AA\_C\_extreme[is.na(AA\_C\_extreme)]="Not Outlier"  
  
#Size 12 Table TNR  
flextable(AA\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("AA\_C Extreme Outliers (mean=61.25mm)")

**Table** : AA\_C Extreme Outliers (mean=61.25mm)

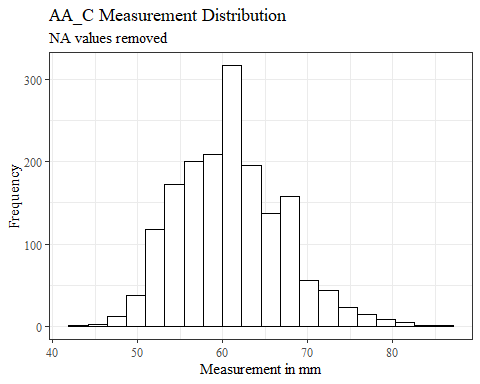
| **ID** | **AA\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201012-012 | 45 | Not Outlier | TRUE | Not Outlier | Not Outlier |

#%>% set\_header\_labels(values = list(AA\_C = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(AA\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("AA\_C Extreme Outliers (mean=61.25mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : AA\_C Extreme Outliers (mean=61.25mm)

| **ID** | **AA\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201012-012 | 45 | Not Outlier | TRUE | Not Outlier | Not Outlier |

#%>% set\_header\_labels(values = list(AA\_C = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=AA\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="AA\_C Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#BiW\_C  
BiW\_C\_out <- SA3\_data2 %>%   
 identify\_outliers(BiW\_C)  
  
BiW\_C\_out$extreme\_overall <- BiW\_C\_out$is.extreme  
  
BiW\_C\_out <- BiW\_C\_out[c(1,3,19)]  
  
  
  
BiW\_C\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(BiW\_C)  
  
BiW\_C\_out1$extreme\_race\_eth <- BiW\_C\_out1$is.extreme  
  
BiW\_C\_out1 <- BiW\_C\_out1[c(2,4,19)]  
  
  
  
  
BiW\_C\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(BiW\_C)  
  
BiW\_C\_out2$extreme\_gender <- BiW\_C\_out2$is.extreme  
  
BiW\_C\_out2 <- BiW\_C\_out2[c(2,4,19)]  
  
  
  
BiW\_C\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(BiW\_C)  
  
BiW\_C\_out3$extreme\_age <- BiW\_C\_out3$is.extreme  
  
BiW\_C\_out3 <- BiW\_C\_out3[c(2,4,19)]  
  
  
BiW\_C\_extreme <- full\_join(BiW\_C\_out, BiW\_C\_out1, by=c('ID', 'BiW\_C'))  
BiW\_C\_extreme <- full\_join(BiW\_C\_extreme, BiW\_C\_out2, by=c('ID', 'BiW\_C'))  
BiW\_C\_extreme <- full\_join(BiW\_C\_extreme, BiW\_C\_out3, by=c('ID', 'BiW\_C'))  
  
BiW\_C\_extreme <-   
 filter(BiW\_C\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
  
BiW\_C\_extreme$extreme\_overall <- as.character(BiW\_C\_extreme$extreme\_overall)  
BiW\_C\_extreme$extreme\_race\_eth <- as.character(BiW\_C\_extreme$extreme\_race\_eth)  
BiW\_C\_extreme$extreme\_gender <- as.character(BiW\_C\_extreme$extreme\_gender)  
BiW\_C\_extreme$extreme\_age <- as.character(BiW\_C\_extreme$extreme\_age)  
   
BiW\_C\_extreme[is.na(BiW\_C\_extreme)]="Not Outlier"  
  
  
#Size 12 Table TNR  
flextable(BiW\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("BiW\_C Extreme Outliers (mean=133.4mm)")

**Table** : BiW\_C Extreme Outliers (mean=133.4mm)

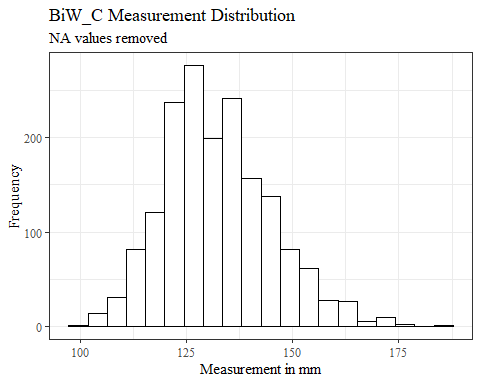
| **ID** | **BiW\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20210108-006 | 187 | FALSE | FALSE | TRUE | FALSE |

#%>% set\_header\_labels(values = list(BiW\_C = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(BiW\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("BiW\_C Extreme Outliers (mean=133.4mm)") %>%   
 fit\_to\_width(7.5)

**Table** : BiW\_C Extreme Outliers (mean=133.4mm)

| **ID** | **BiW\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20210108-006 | 187 | FALSE | FALSE | TRUE | FALSE |

#%>% set\_header\_labels(values = list(BiW\_C = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=BiW\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="BiW\_C Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#BiW\_L  
BiW\_L\_out <- SA3\_data2 %>%   
 identify\_outliers(BiW\_L)  
  
BiW\_L\_out$extreme\_overall <- BiW\_L\_out$is.extreme  
  
BiW\_L\_out <- BiW\_L\_out[c(1,4,19)]  
  
  
  
BiW\_L\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(BiW\_L)  
  
BiW\_L\_out1$extreme\_race\_eth <- BiW\_L\_out1$is.extreme  
  
BiW\_L\_out1 <- BiW\_L\_out1[c(2,5,19)]  
  
  
  
  
BiW\_L\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(BiW\_L)  
  
BiW\_L\_out2$extreme\_gender <- BiW\_L\_out2$is.extreme  
  
BiW\_L\_out2 <- BiW\_L\_out2[c(2,5,19)]  
  
  
  
BiW\_L\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(BiW\_L)  
  
BiW\_L\_out3$extreme\_age <- BiW\_L\_out3$is.extreme  
  
BiW\_L\_out3 <- BiW\_L\_out3[c(2,5,19)]  
  
  
BiW\_L\_extreme <- full\_join(BiW\_L\_out, BiW\_L\_out1, by=c('ID', 'BiW\_L'))  
BiW\_L\_extreme <- full\_join(BiW\_L\_extreme, BiW\_L\_out2, by=c('ID', 'BiW\_L'))  
BiW\_L\_extreme <- full\_join(BiW\_L\_extreme, BiW\_L\_out3, by=c('ID', 'BiW\_L'))  
  
BiW\_L\_extreme <-   
 filter(BiW\_L\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
  
BiW\_L\_extreme$extreme\_overall <- as.character(BiW\_L\_extreme$extreme\_overall)  
BiW\_L\_extreme$extreme\_race\_eth <- as.character(BiW\_L\_extreme$extreme\_race\_eth)  
BiW\_L\_extreme$extreme\_gender <- as.character(BiW\_L\_extreme$extreme\_gender)  
BiW\_L\_extreme$extreme\_age <- as.character(BiW\_L\_extreme$extreme\_age)  
   
BiW\_L\_extreme[is.na(BiW\_L\_extreme)]="Not Outlier"  
  
  
#Size 12 Table TNR  
flextable(BiW\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("BiW\_L Extreme Outliers (mean=111.2mm)")

**Table** : BiW\_L Extreme Outliers (mean=111.2mm)

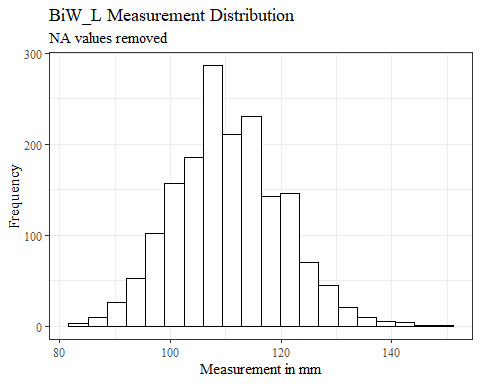
| **ID** | **BiW\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |

#%>% set\_header\_labels(values = list(BiW\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(BiW\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("BiW\_L Extreme Outliers (mean=111.2mm)") %>%   
 fit\_to\_width(7.5)

**Table** : BiW\_L Extreme Outliers (mean=111.2mm)

| **ID** | **BiW\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |

#%>% set\_header\_labels(values = list(BiW\_L = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=BiW\_L))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="BiW\_L Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#GoSub\_C  
GoSub\_C\_out <- SA3\_data2 %>%   
 identify\_outliers(GoSub\_C)  
  
GoSub\_C\_out$extreme\_overall <- GoSub\_C\_out$is.extreme  
  
GoSub\_C\_out <- GoSub\_C\_out[c(1,5,19)]  
  
  
  
GoSub\_C\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(GoSub\_C)  
  
GoSub\_C\_out1$extreme\_race\_eth <- GoSub\_C\_out1$is.extreme  
  
GoSub\_C\_out1 <- GoSub\_C\_out1[c(2,6,19)]  
  
  
  
  
GoSub\_C\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(GoSub\_C)  
  
GoSub\_C\_out2$extreme\_gender <- GoSub\_C\_out2$is.extreme  
  
GoSub\_C\_out2 <- GoSub\_C\_out2[c(2,6,19)]  
  
  
  
GoSub\_C\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(GoSub\_C)  
  
GoSub\_C\_out3$extreme\_age <- GoSub\_C\_out3$is.extreme  
  
GoSub\_C\_out3 <- GoSub\_C\_out3[c(2,6,19)]  
  
  
GoSub\_C\_extreme <- full\_join(GoSub\_C\_out, GoSub\_C\_out1, by=c('ID', 'GoSub\_C'))  
GoSub\_C\_extreme <- full\_join(GoSub\_C\_extreme, GoSub\_C\_out2, by=c('ID', 'GoSub\_C'))  
GoSub\_C\_extreme <- full\_join(GoSub\_C\_extreme, GoSub\_C\_out3, by=c('ID', 'GoSub\_C'))  
  
GoSub\_C\_extreme <-   
 filter(GoSub\_C\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
GoSub\_C\_extreme$extreme\_overall <- as.character(GoSub\_C\_extreme$extreme\_overall)  
GoSub\_C\_extreme$extreme\_race\_eth <- as.character(GoSub\_C\_extreme$extreme\_race\_eth)  
GoSub\_C\_extreme$extreme\_gender <- as.character(GoSub\_C\_extreme$extreme\_gender)  
GoSub\_C\_extreme$extreme\_age <- as.character(GoSub\_C\_extreme$extreme\_age)  
  
GoSub\_C\_extreme[is.na(GoSub\_C\_extreme)]="Not Outlier"   
  
#Size 12 Table TNR  
flextable(GoSub\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("GoSub\_C Extreme Outliers (mean=99.05mm)")

**Table** : GoSub\_C Extreme Outliers (mean=99.05mm)

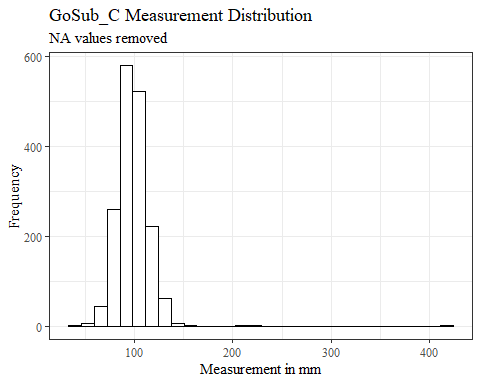
| **ID** | **GoSub\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201203-004 | 208 | TRUE | TRUE | TRUE | TRUE |
| 400-20210216-006 | 217 | TRUE | TRUE | TRUE | TRUE |
| 400-20210129-009 | 424 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(GoSub\_C = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(GoSub\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("GoSub\_C Extreme Outliers (mean=99.05mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : GoSub\_C Extreme Outliers (mean=99.05mm)

| **ID** | **GoSub\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201203-004 | 208 | TRUE | TRUE | TRUE | TRUE |
| 400-20210216-006 | 217 | TRUE | TRUE | TRUE | TRUE |
| 400-20210129-009 | 424 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(GoSub\_C = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=GoSub\_C))+  
 geom\_bar(stat="bin", bins=30, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="GoSub\_C Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#NRB\_L  
NRB\_L\_out <- SA3\_data2 %>%   
 identify\_outliers(NRB\_L)  
  
NRB\_L\_out$extreme\_overall <- NRB\_L\_out$is.extreme  
  
NRB\_L\_out <- NRB\_L\_out[c(1,6,19)]  
  
  
  
NRB\_L\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(NRB\_L)  
  
NRB\_L\_out1$extreme\_race\_eth <- NRB\_L\_out1$is.extreme  
  
NRB\_L\_out1 <- NRB\_L\_out1[c(2,7,19)]  
  
  
  
  
NRB\_L\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(NRB\_L)  
  
NRB\_L\_out2$extreme\_gender <- NRB\_L\_out2$is.extreme  
  
NRB\_L\_out2 <- NRB\_L\_out2[c(2,7,19)]  
  
  
  
NRB\_L\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(NRB\_L)  
  
NRB\_L\_out3$extreme\_age <- NRB\_L\_out3$is.extreme  
  
NRB\_L\_out3 <- NRB\_L\_out3[c(2,7,19)]  
  
  
NRB\_L\_extreme <- full\_join(NRB\_L\_out, NRB\_L\_out1, by=c('ID', 'NRB\_L'))  
NRB\_L\_extreme <- full\_join(NRB\_L\_extreme, NRB\_L\_out2, by=c('ID', 'NRB\_L'))  
NRB\_L\_extreme <- full\_join(NRB\_L\_extreme, NRB\_L\_out3, by=c('ID', 'NRB\_L'))  
  
NRB\_L\_extreme <-   
 filter(NRB\_L\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
NRB\_L\_extreme$extreme\_overall <- as.character(NRB\_L\_extreme$extreme\_overall)  
NRB\_L\_extreme$extreme\_race\_eth <- as.character(NRB\_L\_extreme$extreme\_race\_eth)  
NRB\_L\_extreme$extreme\_gender <- as.character(NRB\_L\_extreme$extreme\_gender)  
NRB\_L\_extreme$extreme\_age <- as.character(NRB\_L\_extreme$extreme\_age)  
  
  
NRB\_L\_extreme[is.na(NRB\_L\_extreme)]="Not Outlier"   
  
#Size 12 Table TNR  
flextable(NRB\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("NRB\_L Extreme Outliers (mean=17.98mm)")

**Table** : NRB\_L Extreme Outliers (mean=17.98mm)

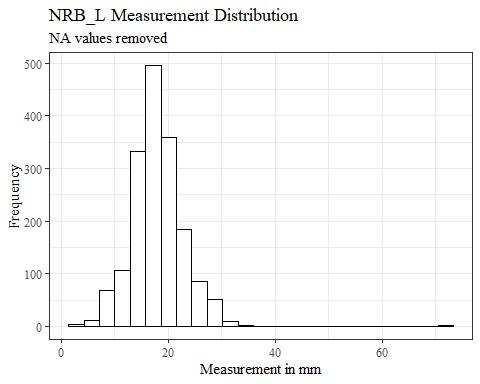
| **ID** | **NRB\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201123-009 | 72 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(NRB\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(NRB\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("NRB\_L Extreme Outliers (mean=17.98mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : NRB\_L Extreme Outliers (mean=17.98mm)

| **ID** | **NRB\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201123-009 | 72 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(NRB\_L = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=NRB\_L))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="NRB\_L Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#ProS\_L  
ProS\_L\_out <- SA3\_data2 %>%   
 identify\_outliers(ProS\_L)  
  
ProS\_L\_out$extreme\_overall <- ProS\_L\_out$is.extreme  
  
ProS\_L\_out <- ProS\_L\_out[c(1,7,19)]  
  
  
  
ProS\_L\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(ProS\_L)  
  
ProS\_L\_out1$extreme\_race\_eth <- ProS\_L\_out1$is.extreme  
  
ProS\_L\_out1 <- ProS\_L\_out1[c(2,8,19)]  
  
  
  
  
ProS\_L\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(ProS\_L)  
  
ProS\_L\_out2$extreme\_gender <- ProS\_L\_out2$is.extreme  
  
ProS\_L\_out2 <- ProS\_L\_out2[c(2,8,19)]  
  
  
  
ProS\_L\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(ProS\_L)  
  
ProS\_L\_out3$extreme\_age <- ProS\_L\_out3$is.extreme  
  
ProS\_L\_out3 <- ProS\_L\_out3[c(2,8,19)]  
  
  
ProS\_L\_extreme <- full\_join(ProS\_L\_out, ProS\_L\_out1, by=c('ID', 'ProS\_L'))  
ProS\_L\_extreme <- full\_join(ProS\_L\_extreme, ProS\_L\_out2, by=c('ID', 'ProS\_L'))  
ProS\_L\_extreme <- full\_join(ProS\_L\_extreme, ProS\_L\_out3, by=c('ID', 'ProS\_L'))  
  
ProS\_L\_extreme <-   
 filter(ProS\_L\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
ProS\_L\_extreme$extreme\_overall <- as.character(ProS\_L\_extreme$extreme\_overall)  
ProS\_L\_extreme$extreme\_race\_eth <- as.character(ProS\_L\_extreme$extreme\_race\_eth)  
ProS\_L\_extreme$extreme\_gender <- as.character(ProS\_L\_extreme$extreme\_gender)  
ProS\_L\_extreme$extreme\_age <- as.character(ProS\_L\_extreme$extreme\_age)  
  
ProS\_L\_extreme[is.na(ProS\_L\_extreme)]="Not Outlier"   
  
#Size 12 Table TNR  
flextable(ProS\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("ProS\_L Extreme Outliers (mean=19.16mm)")

**Table** : ProS\_L Extreme Outliers (mean=19.16mm)

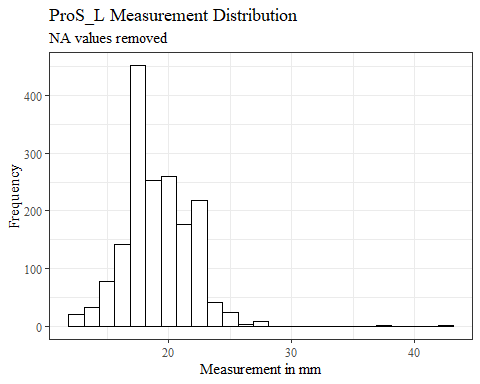
| **ID** | **ProS\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201013-003 | 42 | TRUE | TRUE | TRUE | TRUE |
| 400-20210317-009 | 42 | TRUE | TRUE | TRUE | TRUE |
| 400-20210122-007 | 38 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(ProS\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(ProS\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("ProS\_L Extreme Outliers (mean=19.16mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : ProS\_L Extreme Outliers (mean=19.16mm)

| **ID** | **ProS\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201013-003 | 42 | TRUE | TRUE | TRUE | TRUE |
| 400-20210317-009 | 42 | TRUE | TRUE | TRUE | TRUE |
| 400-20210122-007 | 38 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(ProS\_L = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=ProS\_L))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="ProS\_L Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



SelP\_L

#SelP\_L  
SelP\_L\_out <- SA3\_data2 %>%   
 identify\_outliers(SelP\_L)  
  
SelP\_L\_out$extreme\_overall <- SelP\_L\_out$is.extreme  
  
SelP\_L\_out <- SelP\_L\_out[c(1,8,19)]  
  
  
  
SelP\_L\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(SelP\_L)  
  
SelP\_L\_out1$extreme\_race\_eth <- SelP\_L\_out1$is.extreme  
  
SelP\_L\_out1 <- SelP\_L\_out1[c(2,9,19)]  
  
  
  
  
SelP\_L\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(SelP\_L)  
  
SelP\_L\_out2$extreme\_gender <- SelP\_L\_out2$is.extreme  
  
SelP\_L\_out2 <- SelP\_L\_out2[c(2,9,19)]  
  
  
  
SelP\_L\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(SelP\_L)  
  
SelP\_L\_out3$extreme\_age <- SelP\_L\_out3$is.extreme  
  
SelP\_L\_out3 <- SelP\_L\_out3[c(2,9,19)]  
  
  
SelP\_L\_extreme <- full\_join(SelP\_L\_out, SelP\_L\_out1, by=c('ID', 'SelP\_L'))  
SelP\_L\_extreme <- full\_join(SelP\_L\_extreme, SelP\_L\_out2, by=c('ID', 'SelP\_L'))  
SelP\_L\_extreme <- full\_join(SelP\_L\_extreme, SelP\_L\_out3, by=c('ID', 'SelP\_L'))  
  
SelP\_L\_extreme <-   
 filter(SelP\_L\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
SelP\_L\_extreme$extreme\_overall <- as.character(SelP\_L\_extreme$extreme\_overall)  
SelP\_L\_extreme$extreme\_race\_eth <- as.character(SelP\_L\_extreme$extreme\_race\_eth)  
SelP\_L\_extreme$extreme\_gender <- as.character(SelP\_L\_extreme$extreme\_gender)  
SelP\_L\_extreme$extreme\_age <- as.character(SelP\_L\_extreme$extreme\_age)  
  
SelP\_L\_extreme[is.na(SelP\_L\_extreme)] = "Not Outlier"   
  
#Size 12 Table TNR  
flextable(SelP\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("SelP\_L Extreme Outliers (mean=44.53mm)")

**Table** : SelP\_L Extreme Outliers (mean=44.53mm)

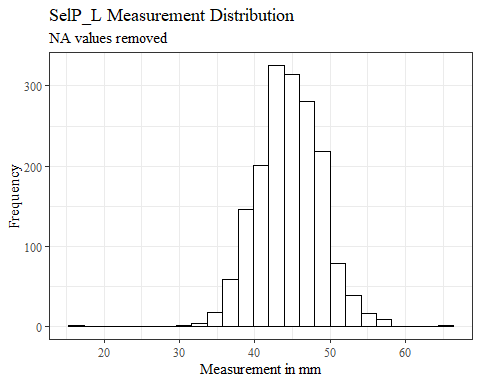
| **ID** | **SelP\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201202-019 | 16 | TRUE | TRUE | TRUE | TRUE |
| 400-20210201-002 | 65 | FALSE | TRUE | TRUE | FALSE |
| 400-20201216-014 | 51 | Not Outlier | Not Outlier | TRUE | Not Outlier |
| 400-20210126-012 | 52 | Not Outlier | Not Outlier | TRUE | Not Outlier |

#%>% set\_header\_labels(values = list(SelP\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(SelP\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("SelP\_L Extreme Outliers (mean=44.53mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : SelP\_L Extreme Outliers (mean=44.53mm)

| **ID** | **SelP\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201202-019 | 16 | TRUE | TRUE | TRUE | TRUE |
| 400-20210201-002 | 65 | FALSE | TRUE | TRUE | FALSE |
| 400-20201216-014 | 51 | Not Outlier | Not Outlier | TRUE | Not Outlier |
| 400-20210126-012 | 52 | Not Outlier | Not Outlier | TRUE | Not Outlier |

#%>% set\_header\_labels(values = list(SelP\_L = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=SelP\_L))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="SelP\_L Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#SelM\_L  
SelM\_L\_out <- SA3\_data2 %>%   
 identify\_outliers(SelM\_L)  
  
SelM\_L\_out$extreme\_overall <- SelM\_L\_out$is.extreme  
  
SelM\_L\_out <- SelM\_L\_out[c(1,9,19)]  
  
  
  
SelM\_L\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(SelM\_L)  
  
SelM\_L\_out1$extreme\_race\_eth <- SelM\_L\_out1$is.extreme  
  
SelM\_L\_out1 <- SelM\_L\_out1[c(2,10,19)]  
  
  
  
  
SelM\_L\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(SelM\_L)  
  
SelM\_L\_out2$extreme\_gender <- SelM\_L\_out2$is.extreme  
  
SelM\_L\_out2 <- SelM\_L\_out2[c(2,10,19)]  
  
  
  
SelM\_L\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(SelM\_L)  
  
SelM\_L\_out3$extreme\_age <- SelM\_L\_out3$is.extreme  
  
SelM\_L\_out3 <- SelM\_L\_out3[c(2,10,19)]  
  
  
SelM\_L\_extreme <- full\_join(SelM\_L\_out, SelM\_L\_out1, by=c('ID', 'SelM\_L'))  
SelM\_L\_extreme <- full\_join(SelM\_L\_extreme, SelM\_L\_out2, by=c('ID', 'SelM\_L'))  
SelM\_L\_extreme <- full\_join(SelM\_L\_extreme, SelM\_L\_out3, by=c('ID', 'SelM\_L'))  
  
SelM\_L\_extreme <-   
 filter(SelM\_L\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
SelM\_L\_extreme$extreme\_overall <- as.character(SelM\_L\_extreme$extreme\_overall)  
SelM\_L\_extreme$extreme\_race\_eth <- as.character(SelM\_L\_extreme$extreme\_race\_eth)  
SelM\_L\_extreme$extreme\_gender <- as.character(SelM\_L\_extreme$extreme\_gender)  
SelM\_L\_extreme$extreme\_age <- as.character(SelM\_L\_extreme$extreme\_age)  
  
SelM\_L\_extreme[is.na(SelM\_L\_extreme)] = "Not Outlier"   
  
#Size 12 Table TNR  
flextable(SelM\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("SelM\_L Extreme Outliers (mean=116.2mm)")

**Table** : SelM\_L Extreme Outliers (mean=116.2mm)

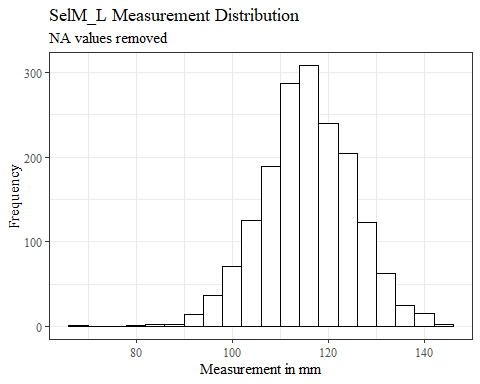
| **ID** | **SelM\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201202-019 | 69 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(SelM\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(SelM\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("SelM\_L Extreme Outliers (mean=116.2mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : SelM\_L Extreme Outliers (mean=116.2mm)

| **ID** | **SelM\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201202-019 | 69 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(SelM\_L = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=SelM\_L))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="SelM\_L Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#SnasM\_C  
SnasM\_C\_out <- SA3\_data2 %>%   
 identify\_outliers(SnasM\_C)  
  
SnasM\_C\_out$extreme\_overall <- SnasM\_C\_out$is.extreme  
  
SnasM\_C\_out <- SnasM\_C\_out[c(1,10,19)]  
  
  
  
SnasM\_C\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(SnasM\_C)  
  
SnasM\_C\_out1$extreme\_race\_eth <- SnasM\_C\_out1$is.extreme  
  
SnasM\_C\_out1 <- SnasM\_C\_out1[c(2,11,19)]  
  
  
  
  
SnasM\_C\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(SnasM\_C)  
  
SnasM\_C\_out2$extreme\_gender <- SnasM\_C\_out2$is.extreme  
  
SnasM\_C\_out2 <- SnasM\_C\_out2[c(2,11,19)]  
  
  
  
SnasM\_C\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(SnasM\_C)  
  
SnasM\_C\_out3$extreme\_age <- SnasM\_C\_out3$is.extreme  
  
SnasM\_C\_out3 <- SnasM\_C\_out3[c(2,11,19)]  
  
  
SnasM\_C\_extreme <- full\_join(SnasM\_C\_out, SnasM\_C\_out1, by=c('ID', 'SnasM\_C'))  
SnasM\_C\_extreme <- full\_join(SnasM\_C\_extreme, SnasM\_C\_out2, by=c('ID', 'SnasM\_C'))  
SnasM\_C\_extreme <- full\_join(SnasM\_C\_extreme, SnasM\_C\_out3, by=c('ID', 'SnasM\_C'))  
  
SnasM\_C\_extreme <-   
 filter(SnasM\_C\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
SnasM\_C\_extreme$extreme\_overall <- as.character(SnasM\_C\_extreme$extreme\_overall)  
SnasM\_C\_extreme$extreme\_race\_eth <- as.character(SnasM\_C\_extreme$extreme\_race\_eth)  
SnasM\_C\_extreme$extreme\_gender <- as.character(SnasM\_C\_extreme$extreme\_gender)  
SnasM\_C\_extreme$extreme\_age <- as.character(SnasM\_C\_extreme$extreme\_age)  
  
SnasM\_C\_extreme[is.na(SnasM\_C\_extreme)] = "Not Outlier"   
  
#Size 12 Table TNR  
flextable(SnasM\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("SnasM\_C Extreme Outliers (mean=75.1mm)")

**Table** : SnasM\_C Extreme Outliers (mean=75.1mm)

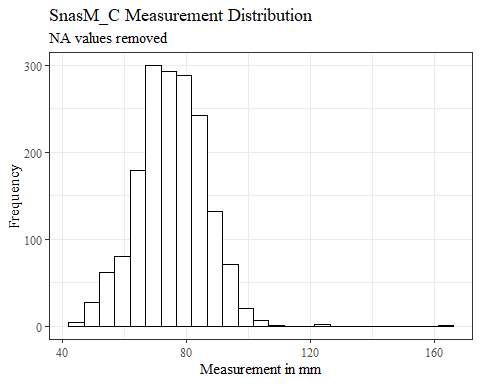
| **ID** | **SnasM\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201013-003 | 124 | FALSE | FALSE | TRUE | FALSE |
| 400-20210317-009 | 125 | FALSE | TRUE | TRUE | FALSE |
| 400-20210203-001 | 163 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(SnasM\_C = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(SnasM\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("SnasM\_C Extreme Outliers (mean=75.1mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : SnasM\_C Extreme Outliers (mean=75.1mm)

| **ID** | **SnasM\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20201013-003 | 124 | FALSE | FALSE | TRUE | FALSE |
| 400-20210317-009 | 125 | FALSE | TRUE | TRUE | FALSE |
| 400-20210203-001 | 163 | TRUE | TRUE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(SnasM\_C = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=SnasM\_C))+  
 geom\_bar(stat="bin", bins=25, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="SnasM\_C Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#TrSman\_C  
TrSman\_C\_out <- SA3\_data2 %>%   
 identify\_outliers(TrSman\_C)  
  
TrSman\_C\_out$extreme\_overall <- TrSman\_C\_out$is.extreme  
  
TrSman\_C\_out <- TrSman\_C\_out[c(1,11,19)]  
  
  
  
TrSman\_C\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(TrSman\_C)  
  
TrSman\_C\_out1$extreme\_race\_eth <- TrSman\_C\_out1$is.extreme  
  
TrSman\_C\_out1 <- TrSman\_C\_out1[c(2,12,19)]  
  
  
  
  
TrSman\_C\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(TrSman\_C)  
  
TrSman\_C\_out2$extreme\_gender <- TrSman\_C\_out2$is.extreme  
  
TrSman\_C\_out2 <- TrSman\_C\_out2[c(2,12,19)]  
  
  
  
TrSman\_C\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(TrSman\_C)  
  
TrSman\_C\_out3$extreme\_age <- TrSman\_C\_out3$is.extreme  
  
TrSman\_C\_out3 <- TrSman\_C\_out3[c(2,12,19)]  
  
  
TrSman\_C\_extreme <- full\_join(TrSman\_C\_out, TrSman\_C\_out1, by=c('ID', 'TrSman\_C'))  
TrSman\_C\_extreme <- full\_join(TrSman\_C\_extreme, TrSman\_C\_out2, by=c('ID', 'TrSman\_C'))  
TrSman\_C\_extreme <- full\_join(TrSman\_C\_extreme, TrSman\_C\_out3, by=c('ID', 'TrSman\_C'))  
  
TrSman\_C\_extreme <-   
 filter(TrSman\_C\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
TrSman\_C\_extreme$extreme\_overall <- as.character(TrSman\_C\_extreme$extreme\_overall)  
TrSman\_C\_extreme$extreme\_race\_eth <- as.character(TrSman\_C\_extreme$extreme\_race\_eth)  
TrSman\_C\_extreme$extreme\_gender <- as.character(TrSman\_C\_extreme$extreme\_gender)  
TrSman\_C\_extreme$extreme\_age <- as.character(TrSman\_C\_extreme$extreme\_age)  
  
TrSman\_C\_extreme[is.na(TrSman\_C\_extreme)] = "Not Outlier"   
  
#Size 12 Table TNR  
flextable(TrSman\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("TrSman\_C Extreme Outliers (mean=153.4mm)")

**Table** : TrSman\_C Extreme Outliers (mean=153.4mm)

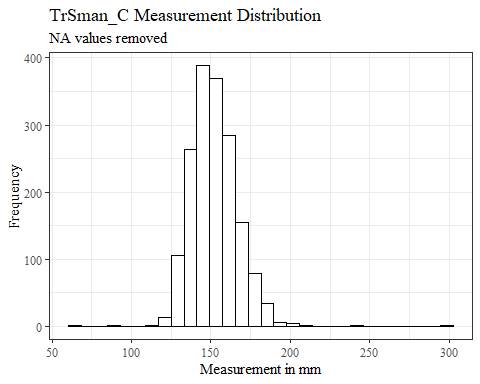
| **ID** | **TrSman\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20210216-006 | 298 | TRUE | TRUE | TRUE | TRUE |
| 400-20210120-002 | 64 | TRUE | TRUE | TRUE | TRUE |
| 400-20210129-009 | 246 | TRUE | TRUE | TRUE | TRUE |
| 400-20210203-009 | 87 | FALSE | FALSE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(TrSman\_C = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(TrSman\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("TrSman\_C Extreme Outliers (mean=153.4mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : TrSman\_C Extreme Outliers (mean=153.4mm)

| **ID** | **TrSman\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |
| 400-20210216-006 | 298 | TRUE | TRUE | TRUE | TRUE |
| 400-20210120-002 | 64 | TRUE | TRUE | TRUE | TRUE |
| 400-20210129-009 | 246 | TRUE | TRUE | TRUE | TRUE |
| 400-20210203-009 | 87 | FALSE | FALSE | TRUE | TRUE |

#%>% set\_header\_labels(values = list(TrSman\_C = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=TrSman\_C))+  
 geom\_bar(stat="bin", bins=30, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="TrSman\_C Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#TrTr\_C  
TrTr\_C\_out <- SA3\_data2 %>%   
 identify\_outliers(TrTr\_C)  
  
TrTr\_C\_out$extreme\_overall <- TrTr\_C\_out$is.extreme  
  
TrTr\_C\_out <- TrTr\_C\_out[c(1,12,19)]  
  
  
  
TrTr\_C\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(TrTr\_C)  
  
TrTr\_C\_out1$extreme\_race\_eth <- TrTr\_C\_out1$is.extreme  
  
TrTr\_C\_out1 <- TrTr\_C\_out1[c(2,13,19)]  
  
  
  
  
TrTr\_C\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(TrTr\_C)  
  
TrTr\_C\_out2$extreme\_gender <- TrTr\_C\_out2$is.extreme  
  
TrTr\_C\_out2 <- TrTr\_C\_out2[c(2,13,19)]  
  
  
  
TrTr\_C\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(TrTr\_C)  
  
TrTr\_C\_out3$extreme\_age <- TrTr\_C\_out3$is.extreme  
  
TrTr\_C\_out3 <- TrTr\_C\_out3[c(2,13,19)]  
  
  
TrTr\_C\_extreme <- full\_join(TrTr\_C\_out, TrTr\_C\_out1, by=c('ID', 'TrTr\_C'))  
TrTr\_C\_extreme <- full\_join(TrTr\_C\_extreme, TrTr\_C\_out2, by=c('ID', 'TrTr\_C'))  
TrTr\_C\_extreme <- full\_join(TrTr\_C\_extreme, TrTr\_C\_out3, by=c('ID', 'TrTr\_C'))  
  
TrTr\_C\_extreme <-   
 filter(TrTr\_C\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
TrTr\_C\_extreme$extreme\_overall <- as.character(TrTr\_C\_extreme$extreme\_overall)  
TrTr\_C\_extreme$extreme\_race\_eth <- as.character(TrTr\_C\_extreme$extreme\_race\_eth)  
TrTr\_C\_extreme$extreme\_gender <- as.character(TrTr\_C\_extreme$extreme\_gender)  
TrTr\_C\_extreme$extreme\_age <- as.character(TrTr\_C\_extreme$extreme\_age)  
  
TrTr\_C\_extreme[is.na(TrTr\_C\_extreme)] = "Not Outlier"   
  
#Size 12 Table TNR  
flextable(TrTr\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("TrTr\_C Extreme Outliers (mean=282.7)")

**Table** : TrTr\_C Extreme Outliers (mean=282.7)

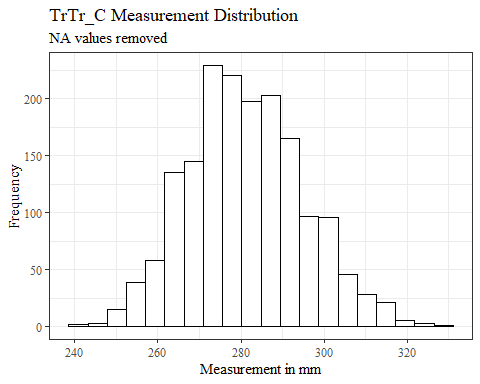
| **ID** | **TrTr\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |

#%>% set\_header\_labels(values = list(TrTr\_C = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(TrTr\_C\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("TrTr\_C Extreme Outliers (mean=282.7)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : TrTr\_C Extreme Outliers (mean=282.7)

| **ID** | **TrTr\_C** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |

#%>% set\_header\_labels(values = list(TrTr\_C = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=TrTr\_C))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="TrTr\_C Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



#TrTr\_L  
TrTr\_L\_out <- SA3\_data2 %>%   
 identify\_outliers(TrTr\_L)  
  
TrTr\_L\_out$extreme\_overall <- TrTr\_L\_out$is.extreme  
  
TrTr\_L\_out <- TrTr\_L\_out[c(1,13,19)]  
  
  
  
TrTr\_L\_out1 <- SA3\_data2 %>%   
 group\_by(race\_eth) %>%   
 identify\_outliers(TrTr\_L)  
  
TrTr\_L\_out1$extreme\_race\_eth <- TrTr\_L\_out1$is.extreme  
  
TrTr\_L\_out1 <- TrTr\_L\_out1[c(2,14,19)]  
  
  
  
  
TrTr\_L\_out2 <- SA3\_data2 %>%   
 group\_by(gender) %>%   
 identify\_outliers(TrTr\_L)  
  
TrTr\_L\_out2$extreme\_gender <- TrTr\_L\_out2$is.extreme  
  
TrTr\_L\_out2 <- TrTr\_L\_out2[c(2,14,19)]  
  
  
  
TrTr\_L\_out3 <- SA3\_data2 %>%   
 group\_by(age\_group) %>%   
 identify\_outliers(TrTr\_L)  
  
TrTr\_L\_out3$extreme\_age <- TrTr\_L\_out3$is.extreme  
  
TrTr\_L\_out3 <- TrTr\_L\_out3[c(2,14,19)]  
  
  
TrTr\_L\_extreme <- full\_join(TrTr\_L\_out, TrTr\_L\_out1, by=c('ID', 'TrTr\_L'))  
TrTr\_L\_extreme <- full\_join(TrTr\_L\_extreme, TrTr\_L\_out2, by=c('ID', 'TrTr\_L'))  
TrTr\_L\_extreme <- full\_join(TrTr\_L\_extreme, TrTr\_L\_out3, by=c('ID', 'TrTr\_L'))  
  
TrTr\_L\_extreme <-   
 filter(TrTr\_L\_extreme,   
 extreme\_overall == "TRUE" |   
 extreme\_race\_eth == "TRUE" |  
 extreme\_gender == "TRUE" |  
 extreme\_age == "TRUE" )  
  
TrTr\_L\_extreme$extreme\_overall <- as.character(TrTr\_L\_extreme$extreme\_overall)  
TrTr\_L\_extreme$extreme\_race\_eth <- as.character(TrTr\_L\_extreme$extreme\_race\_eth)  
TrTr\_L\_extreme$extreme\_gender <- as.character(TrTr\_L\_extreme$extreme\_gender)  
TrTr\_L\_extreme$extreme\_age <- as.character(TrTr\_L\_extreme$extreme\_age)  
  
TrTr\_L\_extreme[is.na(TrTr\_L\_extreme)] = "Not Outlier"   
  
#Size 12 Table TNR  
flextable(TrTr\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("TrTr\_L Extreme Outliers (mean=146.5mm)")

**Table** : TrTr\_L Extreme Outliers (mean=146.5mm)

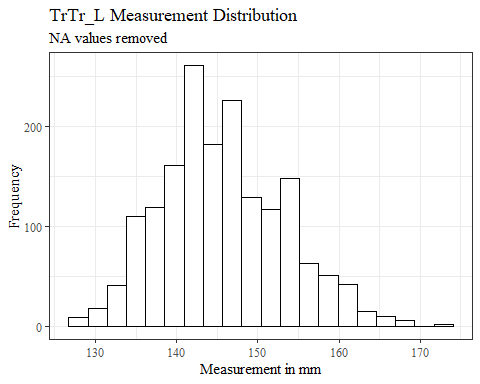
| **ID** | **TrTr\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |

#%>% set\_header\_labels(values = list(TrTr\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(TrTr\_L\_extreme) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("TrTr\_L Extreme Outliers (mean=146.5mm)") %>%   
 fit\_to\_width(7.5) %>%   
 autofit

**Table** : TrTr\_L Extreme Outliers (mean=146.5mm)

| **ID** | **TrTr\_L** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** |
| --- | --- | --- | --- | --- | --- |

#%>% set\_header\_labels(values = list(TrTr\_L = "Alare/AlareCont"))  
  
ggplot(data=SA3\_data2, aes(x=TrTr\_L))+  
 geom\_bar(stat="bin", bins=20, color= "black", fill = "white")+  
 theme\_bw() + theme(text=element\_text(family= "Times New Roman"))+  
 labs(title="TrTr\_L Measurement Distribution",  
 subtitle = "NA values removed",  
 y="Frequency",  
 x="Measurement in mm")



##outliers  
mult\_out <- SA3\_data2  
  
  
#cutoff <- qchisq(1-.001, ncol(mult\_out[ , -c(1,14,15,16)]))  
  
  
mult\_out$mahal <- mahalanobis(mult\_out[ , -c(1,14,15,16)],  
 colMeans(mult\_out[ , -c(1,14,15,16)]),  
 cov(mult\_out[ , -c(1,14,15,16)]))  
  
mult\_out$p <- pchisq(mult\_out$mahal, df=11, lower.tail=FALSE)  
  
  
  
mult\_out$multv\_outlier <- with(mult\_out, ifelse(p < 0.001, 'TRUE', 'FALSE'))  
  
mult\_out <- mult\_out %>%   
 filter(multv\_outlier == TRUE)  
  
  
##df  
#summary(mahal < cutoff)  
  
#mult\_out2 <- SA3\_data2 %>%  
 #mahalanobis\_distance() %>%  
 #filter(is.outlier == TRUE) %>%  
 #as.data.frame()  
  
#mult\_out2  
  
##exclude outliers  
#noout = subset(nomiss, mahal < cutoff)

AA\_C\_extreme$offending <- "AA\_C univar"  
BiW\_C\_extreme$offending <- "BiW\_C univar"  
BiW\_L\_extreme$offending <- "BiW\_L univar"  
GoSub\_C\_extreme$offending <- "GoSub\_C univar"  
NRB\_L\_extreme$offending <- "NRB\_L univar"  
ProS\_L\_extreme$offending <- "ProS\_L univar"  
SelP\_L\_extreme$offending <- "SelP\_L univar"  
SelM\_L\_extreme$offending <- "SelM\_L univar"  
SnasM\_C\_extreme$offending <- "SnasM\_C univar"  
TrSman\_C\_extreme$offending <- "TrSman\_C univar"  
TrTr\_C\_extreme$offending <- "TrTr\_C univar"  
TrTr\_L\_extreme$offending <- "TrTr\_L univar"

all\_extreme <- full\_join(AA\_C\_extreme, BiW\_C\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, BiW\_L\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, GoSub\_C\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, NRB\_L\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, ProS\_L\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, SelP\_L\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, SelM\_L\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, SnasM\_C\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, TrSman\_C\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, TrTr\_C\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))  
  
all\_extreme <- full\_join(all\_extreme, TrTr\_L\_extreme, by=c('ID',   
 'extreme\_overall',   
 'extreme\_race\_eth',   
 'extreme\_gender',  
 'extreme\_age',  
 'offending'))

combin\_out <- full\_join(mult\_out, all\_extreme, by=c('ID'))  
  
str(combin\_out)

## tibble [28 × 36] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:28] "400-20201013-003" "400-20201013-003" "400-20210210-009" "400-20210317-009" ...  
## $ AA\_C.x : num [1:28] 63 63 50 66 66 66 71 52 68 59 ...  
## $ BiW\_C.x : num [1:28] 143 143 131 128 128 136 156 164 161 133 ...  
## $ BiW\_L.x : num [1:28] 117 117 121 106 106 117 131 120 132 110 ...  
## $ GoSub\_C.x : num [1:28] 114 114 137 93 93 125 124 106 110 93 ...  
## $ NRB\_L.x : num [1:28] 18 18 18 16 16 72 26 19 27 20 ...  
## $ ProS\_L.x : num [1:28] 42 42 16 42 42 20 24 17 17 17 ...  
## $ SelP\_L.x : num [1:28] 43 43 45 42 42 40 46 43 43 16 ...  
## $ SelM\_L.x : num [1:28] 113 113 121 105 105 129 121 107 114 69 ...  
## $ SnasM\_C.x : num [1:28] 124 124 80 125 125 87 102 67 101 79 ...  
## $ TrSman\_C.x : num [1:28] 167 167 196 138 138 182 185 144 156 141 ...  
## $ TrTr\_C.x : num [1:28] 297 297 313 270 270 305 296 277 275 266 ...  
## $ TrTr\_L.x : num [1:28] 152 152 167 139 139 167 167 139 143 140 ...  
## $ gender : Factor w/ 3 levels "Other","Female",..: 2 2 3 2 2 3 3 2 3 2 ...  
## $ race\_eth : Factor w/ 5 levels "Other","Asian",..: 3 3 2 2 2 5 5 5 3 5 ...  
## $ age\_group : Factor w/ 3 levels "18-36","37-54",..: 2 2 1 2 2 1 3 3 2 2 ...  
## $ mahal : num [1:28] 235.4 235.4 31.3 269.6 269.6 ...  
## $ p : num [1:28] 3.17e-44 3.17e-44 9.90e-04 2.12e-51 2.12e-51 ...  
## $ multv\_outlier : chr [1:28] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ AA\_C.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ extreme\_overall : chr [1:28] "TRUE" "FALSE" NA "TRUE" ...  
## $ extreme\_race\_eth: chr [1:28] "TRUE" "FALSE" NA "TRUE" ...  
## $ extreme\_gender : chr [1:28] "TRUE" "TRUE" NA "TRUE" ...  
## $ extreme\_age : chr [1:28] "TRUE" "FALSE" NA "TRUE" ...  
## $ offending : chr [1:28] "ProS\_L univar" "SnasM\_C univar" NA "ProS\_L univar" ...  
## $ BiW\_C.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ BiW\_L.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ GoSub\_C.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ NRB\_L.y : num [1:28] NA NA NA NA NA 72 NA NA NA NA ...  
## $ ProS\_L.y : num [1:28] 42 NA NA 42 NA NA NA NA NA NA ...  
## $ SelP\_L.y : num [1:28] NA NA NA NA NA NA NA NA NA 16 ...  
## $ SelM\_L.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ SnasM\_C.y : num [1:28] NA 124 NA NA 125 NA NA NA NA NA ...  
## $ TrSman\_C.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ TrTr\_C.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...  
## $ TrTr\_L.y : num [1:28] NA NA NA NA NA NA NA NA NA NA ...

combin\_out1 <- combin\_out[c(1,19,21,22,23,24,25)]  
  
combin\_out1$offending[is.na(combin\_out1$offending)] = "Not Univar Outlier"   
  
combin\_out1[is.na(combin\_out1)] = "Not Outlier"   
  
str(combin\_out1)

## tibble [28 × 7] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:28] "400-20201013-003" "400-20201013-003" "400-20210210-009" "400-20210317-009" ...  
## $ multv\_outlier : chr [1:28] "TRUE" "TRUE" "TRUE" "TRUE" ...  
## $ extreme\_overall : chr [1:28] "TRUE" "FALSE" "Not Outlier" "TRUE" ...  
## $ extreme\_race\_eth: chr [1:28] "TRUE" "FALSE" "Not Outlier" "TRUE" ...  
## $ extreme\_gender : chr [1:28] "TRUE" "TRUE" "Not Outlier" "TRUE" ...  
## $ extreme\_age : chr [1:28] "TRUE" "FALSE" "Not Outlier" "TRUE" ...  
## $ offending : chr [1:28] "ProS\_L univar" "SnasM\_C univar" "Not Univar Outlier" "ProS\_L univar" ...

combin\_out1$multv\_outlier <- factor(combin\_out1$multv\_outlier, levels = c("TRUE", "FALSE", "Not Outlier"))  
combin\_out1$extreme\_overall <- factor(combin\_out1$extreme\_overall, levels = c("TRUE", "FALSE", "Not Outlier"))  
combin\_out1$extreme\_race\_eth <- factor(combin\_out1$extreme\_race\_eth, levels = c("TRUE", "FALSE", "Not Outlier"))  
combin\_out1$extreme\_gender <- factor(combin\_out1$extreme\_gender, levels = c("TRUE", "FALSE", "Not Outlier"))  
combin\_out1$extreme\_age <- factor(combin\_out1$extreme\_age, levels = c("TRUE", "FALSE", "Not Outlier"))  
  
  
arrange(combin\_out1, multv\_outlier + extreme\_overall)

## Warning in Ops.factor(multv\_outlier, extreme\_overall): '+' not meaningful for  
## factors

## # A tibble: 28 × 7  
## ID multv\_outlier extreme\_over…¹ extre…² extre…³ extre…⁴ offen…⁵  
## <chr> <fct> <fct> <fct> <fct> <fct> <chr>   
## 1 400-20201013-003 TRUE TRUE TRUE TRUE TRUE ProS\_L…  
## 2 400-20201013-003 TRUE FALSE FALSE TRUE FALSE SnasM\_…  
## 3 400-20210210-009 TRUE Not Outlier Not Ou… Not Ou… Not Ou… Not Un…  
## 4 400-20210317-009 TRUE TRUE TRUE TRUE TRUE ProS\_L…  
## 5 400-20210317-009 TRUE FALSE TRUE TRUE FALSE SnasM\_…  
## 6 400-20201123-009 TRUE TRUE TRUE TRUE TRUE NRB\_L …  
## 7 400-20201209-011 TRUE Not Outlier Not Ou… Not Ou… Not Ou… Not Un…  
## 8 400-20201214-021 TRUE Not Outlier Not Ou… Not Ou… Not Ou… Not Un…  
## 9 400-20201215-009 TRUE Not Outlier Not Ou… Not Ou… Not Ou… Not Un…  
## 10 400-20201202-019 TRUE TRUE TRUE TRUE TRUE SelP\_L…  
## # … with 18 more rows, and abbreviated variable names ¹​extreme\_overall,  
## # ²​extreme\_race\_eth, ³​extreme\_gender, ⁴​extreme\_age, ⁵​offending  
## # ℹ Use `print(n = ...)` to see more rows

#combin\_out1[with(combin\_out1, order("multv\_outlier", "extreme\_overall", "extreme\_race\_eth", "extreme\_gender", "extreme\_age")), ]

#Size 12 Table TNR  
flextable(combin\_out1) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Univariate and Multivariate Outlier Status") %>%   
 autofit

**Table** : Univariate and Multivariate Outlier Status

| **ID** | **multv\_outlier** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** | **offending** |
| --- | --- | --- | --- | --- | --- | --- |
| 400-20201013-003 | TRUE | TRUE | TRUE | TRUE | TRUE | ProS\_L univar |
| 400-20201013-003 | TRUE | FALSE | FALSE | TRUE | FALSE | SnasM\_C univar |
| 400-20210210-009 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20210317-009 | TRUE | TRUE | TRUE | TRUE | TRUE | ProS\_L univar |
| 400-20210317-009 | TRUE | FALSE | TRUE | TRUE | FALSE | SnasM\_C univar |
| 400-20201123-009 | TRUE | TRUE | TRUE | TRUE | TRUE | NRB\_L univar |
| 400-20201209-011 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201214-021 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201215-009 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201202-019 | TRUE | TRUE | TRUE | TRUE | TRUE | SelP\_L univar |
| 400-20201202-019 | TRUE | TRUE | TRUE | TRUE | TRUE | SelM\_L univar |
| 400-20201203-004 | TRUE | TRUE | TRUE | TRUE | TRUE | GoSub\_C univar |
| 400-20201204-018 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201208-024 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20210216-006 | TRUE | TRUE | TRUE | TRUE | TRUE | GoSub\_C univar |
| 400-20210216-006 | TRUE | TRUE | TRUE | TRUE | TRUE | TrSman\_C univar |
| 400-20210120-002 | TRUE | TRUE | TRUE | TRUE | TRUE | TrSman\_C univar |
| 400-20210122-007 | TRUE | TRUE | TRUE | TRUE | TRUE | ProS\_L univar |
| 400-20210128-015 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20210129-009 | TRUE | TRUE | TRUE | TRUE | TRUE | GoSub\_C univar |
| 400-20210129-009 | TRUE | TRUE | TRUE | TRUE | TRUE | TrSman\_C univar |
| 400-20210201-002 | TRUE | FALSE | TRUE | TRUE | FALSE | SelP\_L univar |
| 400-20210203-001 | TRUE | TRUE | TRUE | TRUE | TRUE | SnasM\_C univar |
| 400-20210203-009 | TRUE | FALSE | FALSE | TRUE | TRUE | TrSman\_C univar |
| 400-20201012-012 | Not Outlier | Not Outlier | TRUE | Not Outlier | Not Outlier | AA\_C univar |
| 400-20210108-006 | Not Outlier | FALSE | FALSE | TRUE | FALSE | BiW\_C univar |
| 400-20201216-014 | Not Outlier | Not Outlier | Not Outlier | TRUE | Not Outlier | SelP\_L univar |
| 400-20210126-012 | Not Outlier | Not Outlier | Not Outlier | TRUE | Not Outlier | SelP\_L univar |

#%>% set\_header\_labels(values = list(TrTr\_L = "Alare/AlareCont"))  
  
#Autofit Width Table TNR  
flextable(combin\_out1) %>%  
 my\_ft\_theme()%>%   
 bold(part = "header") %>%   
 set\_caption("Univariate and Multivariate Outlier Status") %>%   
 autofit() %>%   
 fit\_to\_width(7.5)

**Table** : Univariate and Multivariate Outlier Status

| **ID** | **multv\_outlier** | **extreme\_overall** | **extreme\_race\_eth** | **extreme\_gender** | **extreme\_age** | **offending** |
| --- | --- | --- | --- | --- | --- | --- |
| 400-20201013-003 | TRUE | TRUE | TRUE | TRUE | TRUE | ProS\_L univar |
| 400-20201013-003 | TRUE | FALSE | FALSE | TRUE | FALSE | SnasM\_C univar |
| 400-20210210-009 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20210317-009 | TRUE | TRUE | TRUE | TRUE | TRUE | ProS\_L univar |
| 400-20210317-009 | TRUE | FALSE | TRUE | TRUE | FALSE | SnasM\_C univar |
| 400-20201123-009 | TRUE | TRUE | TRUE | TRUE | TRUE | NRB\_L univar |
| 400-20201209-011 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201214-021 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201215-009 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201202-019 | TRUE | TRUE | TRUE | TRUE | TRUE | SelP\_L univar |
| 400-20201202-019 | TRUE | TRUE | TRUE | TRUE | TRUE | SelM\_L univar |
| 400-20201203-004 | TRUE | TRUE | TRUE | TRUE | TRUE | GoSub\_C univar |
| 400-20201204-018 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20201208-024 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20210216-006 | TRUE | TRUE | TRUE | TRUE | TRUE | GoSub\_C univar |
| 400-20210216-006 | TRUE | TRUE | TRUE | TRUE | TRUE | TrSman\_C univar |
| 400-20210120-002 | TRUE | TRUE | TRUE | TRUE | TRUE | TrSman\_C univar |
| 400-20210122-007 | TRUE | TRUE | TRUE | TRUE | TRUE | ProS\_L univar |
| 400-20210128-015 | TRUE | Not Outlier | Not Outlier | Not Outlier | Not Outlier | Not Univar Outlier |
| 400-20210129-009 | TRUE | TRUE | TRUE | TRUE | TRUE | GoSub\_C univar |
| 400-20210129-009 | TRUE | TRUE | TRUE | TRUE | TRUE | TrSman\_C univar |
| 400-20210201-002 | TRUE | FALSE | TRUE | TRUE | FALSE | SelP\_L univar |
| 400-20210203-001 | TRUE | TRUE | TRUE | TRUE | TRUE | SnasM\_C univar |
| 400-20210203-009 | TRUE | FALSE | FALSE | TRUE | TRUE | TrSman\_C univar |
| 400-20201012-012 | Not Outlier | Not Outlier | TRUE | Not Outlier | Not Outlier | AA\_C univar |
| 400-20210108-006 | Not Outlier | FALSE | FALSE | TRUE | FALSE | BiW\_C univar |
| 400-20201216-014 | Not Outlier | Not Outlier | Not Outlier | TRUE | Not Outlier | SelP\_L univar |
| 400-20210126-012 | Not Outlier | Not Outlier | Not Outlier | TRUE | Not Outlier | SelP\_L univar |

#%>% set\_header\_labels(values = list(TrTr\_L = "Alare/AlareCont"))

##exclude outliers  
SA3\_noout <- anti\_join(SA3\_data2, mult\_out, by="ID")  
  
1710-19

## [1] 1691

str(SA3\_noout)

## tibble [1,691 × 16] (S3: tbl\_df/tbl/data.frame)  
## $ ID : chr [1:1691] "400-20201012-002" "400-20201012-003" "400-20201012-004" "400-20201012-005" ...  
## $ AA\_C : num [1:1691] 65 55 70 58 67 60 59 59 65 65 ...  
## $ BiW\_C : num [1:1691] 130 127 143 140 137 130 141 138 143 150 ...  
## $ BiW\_L : num [1:1691] 115 108 121 109 104 106 109 111 113 116 ...  
## $ GoSub\_C : num [1:1691] 93 93 115 93 103 100 79 106 85 102 ...  
## $ NRB\_L : num [1:1691] 17 18 19 21 19 14 17 18 16 17 ...  
## $ ProS\_L : num [1:1691] 17 18 14 13 20 20 18 12 24 22 ...  
## $ SelP\_L : num [1:1691] 42 41 51 44 47 48 46 41 46 44 ...  
## $ SelM\_L : num [1:1691] 122 99 130 115 119 126 117 112 117 117 ...  
## $ SnasM\_C : num [1:1691] 82 55 84 74 73 80 78 76 64 75 ...  
## $ TrSman\_C : num [1:1691] 177 145 178 147 157 164 149 159 151 160 ...  
## $ TrTr\_C : num [1:1691] 296 276 292 273 279 300 283 275 307 286 ...  
## $ TrTr\_L : num [1:1691] 155 141 156 149 146 146 147 151 157 144 ...  
## $ gender : Factor w/ 3 levels "Other","Female",..: 3 2 3 3 3 3 3 3 3 3 ...  
## $ race\_eth : Factor w/ 5 levels "Other","Asian",..: 3 5 5 5 5 5 3 3 5 5 ...  
## $ age\_group: Factor w/ 3 levels "18-36","37-54",..: 1 2 2 1 2 3 1 1 1 1 ...

write\_xlsx(SA3\_noout, "C:\\Users\\19177\\OneDrive - Colostate\\Desktop\\Dissertation\\headscan\_dissertation\\SA3\_noout.xlsx")