Longitudinal analysis of Shigella

0.1 Introduction

Data description

- Step 2: Load the Dataset

```
df <- read_excel("3.8.2024 Compiled Shigella data.xlsx",
                  sheet = "Compiled")
```

- Step 3: Understand the Structure of the Dataset

```
# View the structure of the dataset
str(df)
```

```
tibble [2,374 x 30] (S3: tbl_df/tbl/data.frame)
 $ isotype         : num [1:2374] 1 1 1 1 1 1 1 1 1 1 ...
 $ isotype_name    : chr [1:2374] "IgG" "IgG" "IgG" "IgG" ...
 $ dilution        : num [1:2374] 1 1 1 1 1 1 1 1 1 1 ...
 $ dilution_name   : chr [1:2374] "1:1000" "1:1000" "1:1000" "1:1000" ...
 $ study_name      : chr [1:2374] "SOSAR" "SOSAR" "SOSAR" "SOSAR" ...
 $ sampleID        : chr [1:2374] "SOSAR-11001" "SOSAR-21024" "SOSAR-11001" "SOSAR-21024"
 $ sid             : chr [1:2374] "SOSAR-11001" "SOSAR-21024" "SOSAR-11001" "SOSAR-21024"
 $ unq_id          : chr [1:2374] "SOSAR-11001_d2_1:1000_IgG" "SOSAR-21024_d2_1:1000_IgG" "
 $ cohort          : num [1:2374] 4 7 4 7 4 7 4 7 4 7 ...
 $ cohort_name     : chr [1:2374] "Sf2a" "sonnei" "Sf2a" "sonnei" ...
 $ site            : num [1:2374] 2 2 2 2 2 2 2 2 2 2 ...
 $ site_name       : chr [1:2374] "Dhaka" "Dhaka" "Dhaka" "Dhaka" ...
 $ age             : num [1:2374] 4 2 4 2 4 2 4 2 4 2 ...
 $ treatment       : chr [1:2374] NA NA NA NA ...
 $ timepoint       : num [1:2374] 2 2 7 7 30 30 90 90 180 2 ...
 $ ipab_MFI        : num [1:2374] 12135 159666 165894 161022 154610 ...
```

```
$ sf3aospbsa_MFI    : num [1:2374] 5530 9448 15705 9140 7748 ...
$ sf2aospbsa_MFI    : num [1:2374] 4258 2394 36303 2243 18150 ...
$ sf6ospbsa_MFI     : num [1:2374] 4335 22139 7022 24712 2552 ...
$ sonneiospbsa_MFI  : num [1:2374] 9364 6515 8975 9149 4847 ...
$ n_ipab_MFI        : num [1:2374] 6726 88504 91956 89256 85701 ...
$ n_sf3aospbsa_MFI  : num [1:2374] 7448 12723 21149 12309 10434 ...
$ n_sf2aospbsa_MFI  : num [1:2374] 4421 2486 37692 2329 18844 ...
$ n_sf6ospbsa_MFI   : num [1:2374] 2827 14438 4579 16116 1664 ...
$ n_sonneiospbsa_MFI: num [1:2374] 12240 8516 11732 11960 6336 ...
$ ipab_BC           : num [1:2374] 111 99 177 147 147 124 150 132 78 143 ...
$ sf3aospbsa_BC     : num [1:2374] 149 106 122 95 150 172 193 142 127 97 ...
$ sf2aospbsa_BC     : num [1:2374] 95 90 89 102 135 139 134 94 97 92 ...
$ sf6ospbsa_BC      : num [1:2374] 80 55 76 61 93 83 94 62 62 59 ...
$ sonneiospbsa_BC   : num [1:2374] 50 50 32 50 50 41 50 42 34 50 ...
```

```
# View the first few rows of the dataset
head(df)
```

nt	timepoint	ipab_MFI	sf3aospbsa_MFI	sf2aospbsa_MFI	sf6ospbsa_MFI	sonneiospbsa_MFI
	2	1.21e+04	5.53e+03	4.26e+03	4.34e+03	9.36e+03
	2	1.6e+05	9.45e+03	2.39e+03	2.21e+04	6.51e+03
	7	1.66e+05	1.57e+04	3.63e+04	7.02e+03	8.98e+03
	7	1.61e+05	9.14e+03	2.24e+03	2.47e+04	9.15e+03
	30	1.55e+05	7.75e+03	1.81e+04	2.55e+03	4.85e+03
	30	1.55e+05	8.08e+03	1.84e+03	1.66e+04	7.96e+03

```
# View a summary of the dataset
summary(df)
```

```
     isotype     isotype_name         dilution dilution_name
 Min.   :1.0   Length:2374        Min.   :1   Length:2374
 1st Qu.:1.0   Class :character   1st Qu.:1   Class :character
 Median :1.5   Mode  :character   Median :1   Mode  :character
 Mean   :1.5                      Mean   :1
 3rd Qu.:2.0                      3rd Qu.:1
 Max.   :2.0                      Max.   :1
```

```
   study_name          sampleID             sid              unq_id         
 Length:2374        Length:2374        Length:2374        Length:2374        
 Class :character   Class :character   Class :character   Class :character   
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   



     cohort       cohort_name            site          site_name        
 Min.   : 4.000   Length:2374        Min.   :1.000   Length:2374        
 1st Qu.: 9.000   Class :character   1st Qu.:2.500   Class :character   
 Median : 9.000   Mode  :character   Median :7.000   Mode  :character   
 Mean   : 8.488                      Mean   :5.325                      
 3rd Qu.: 9.000                      3rd Qu.:7.000                      
 Max.   :15.000                      Max.   :7.000                      
      age            treatment           timepoint        ipab_MFI        
 Min.   : 0.08333   Length:2374        Min.   :   0    Min.   :    -1.25  
 1st Qu.: 2.00000   Class :character   1st Qu.:   0    1st Qu.:   818.58  
 Median : 3.66667   Mode  :character   Median :1825    Median :  4759.92  
 Mean   : 6.58130                      Mean   : 969    Mean   : 43525.84  
 3rd Qu.: 7.00000                      3rd Qu.:1825    3rd Qu.: 86098.81  
 Max.   :50.00000                      Max.   :1825    Max.   :189355.75  
 sf3aospbsa_MFI    sf2aospbsa_MFI     sf6ospbsa_MFI       sonneiospbsa_MFI   
 Min.   :    -9.0  Min.   :    -12.0  Min.   :    -7.75   Min.   :   -14.50  
 1st Qu.:   183.2  1st Qu.:    100.1  1st Qu.:   278.25   1st Qu.:    28.08  
 Median :   753.7  Median :    489.0  Median :  1036.29   Median :   118.12  
 Mean   :  3651.3  Mean   :   4681.1  Mean   :  4043.35   Mean   :  3007.76  
 3rd Qu.:  2956.5  3rd Qu.:   2694.0  3rd Qu.:  3714.58   3rd Qu.:   887.46  
 Max.   :83828.9   Max.   :214007.7   Max.   :166878.50   Max.   :184109.50  
   n_ipab_MFI        n_sf3aospbsa_MFI   n_sf2aospbsa_MFI    n_sf6ospbsa_MFI    
 Min.   :    -0.63  Min.   :    -12.1  Min.   :    -11.82  Min.   :    -6.84  
 1st Qu.:   361.53  1st Qu.:    320.1  1st Qu.:    135.66  1st Qu.:   231.92  
 Median :  1903.47  Median :   1273.0  Median :    621.84  Median :   845.01  
 Mean   : 24971.08  Mean   :   5702.4  Mean   :   5308.11  Mean   :  2991.34  
 3rd Qu.: 51093.85  3rd Qu.:   4743.1  3rd Qu.:   3295.17  3rd Qu.:  2826.64  
 Max.   :116048.97  Max.   :112888.9   Max.   :222195.19   Max.   :108831.02  
 n_sonneiospbsa_MFI     ipab_BC        sf3aospbsa_BC      sf2aospbsa_BC   
 Min.   :   -28.75  Min.   : 17.00   Min.   : 57.0    Min.   : 23.00   
 1st Qu.:    38.78  1st Qu.: 59.00   1st Qu.: 88.0    1st Qu.: 54.00   
 Median :   183.09  Median : 73.00   Median :100.0    Median : 63.00   
 Mean   :  3747.64  Mean   : 80.03   Mean   :103.9    Mean   : 67.03   
 3rd Qu.:  1335.17  3rd Qu.: 91.00   3rd Qu.:115.0    3rd Qu.: 76.00   
 Max.   :219024.17  Max.   :260.00   Max.   :226.0    Max.   :155.00   
  sf6ospbsa_BC     sonneiospbsa_BC 
```

```
Min.   : 27.00   Min.   : 17.00
1st Qu.: 47.00   1st Qu.: 49.00
Median : 53.00   Median : 58.00
Mean   : 56.18   Mean   : 58.37
3rd Qu.: 63.00   3rd Qu.: 68.00
Max.   :115.00   Max.   :114.00
```

- Step 4: Handle Missing Values

```
colSums(is.na(df))
```

```
            isotype        isotype_name            dilution        dilution_name
                  0                   0                   0                    0
         study_name            sampleID                 sid               unq_id
                  0                   0                   0                    0
             cohort         cohort_name                site            site_name
                  0                   0                   0                    0
                age           treatment           timepoint             ipab_MFI
                  0                1126                   0                    0
      sf3aospbsa_MFI      sf2aospbsa_MFI      sf6ospbsa_MFI     sonneiospbsa_MFI
                  0                   0                   0                    0
         n_ipab_MFI     n_sf3aospbsa_MFI    n_sf2aospbsa_MFI     n_sf6ospbsa_MFI
                  0                   0                   0                    0
  n_sonneiospbsa_MFI             ipab_BC       sf3aospbsa_BC        sf2aospbsa_BC
                  0                   0                   0                    0
       sf6ospbsa_BC       sonneiospbsa_BC
                  0                   0
```

```
# select longitudinal dataset from complied dataset

df_sosar<-df%>%filter(study_name=="SOSAR")

# split the data by isotype

df_sosar_IgA<-df_sosar%>%filter(isotype_name=="IgA")

df_sosar_IgG<-df_sosar%>%filter(isotype_name=="IgG")
```

```
# plotting how quantitative response change over time within individuals


# Check the number of observations per sid
```
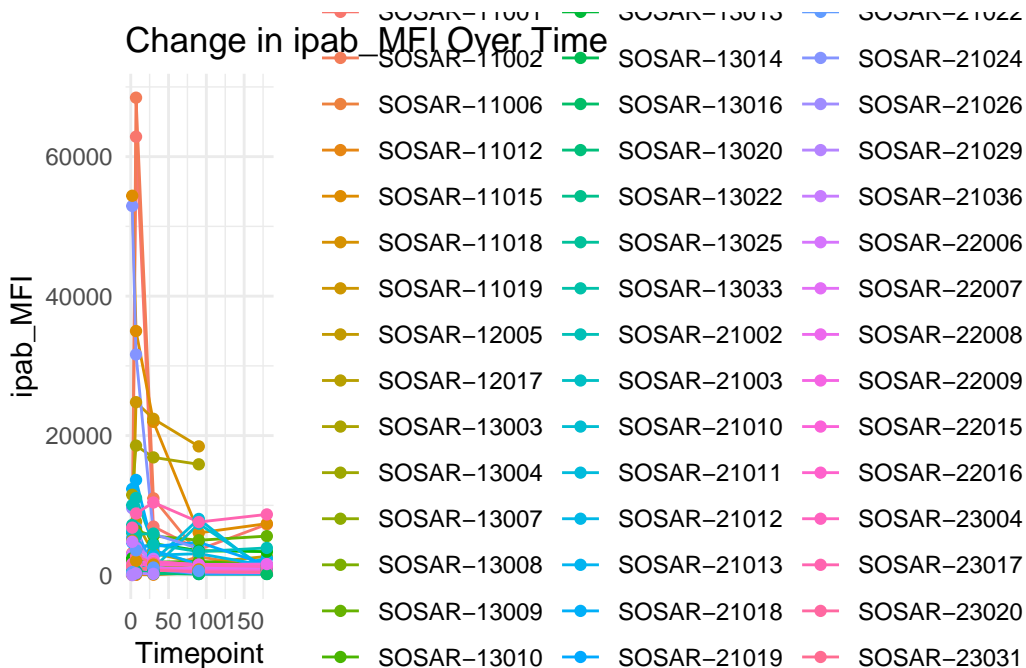
```
df_summary_IgA <- df_sosar_IgA %>%
  group_by(sid) %>%
  summarize(count = n()) %>%
  filter(count > 1)

df_summary_IgA
```

Checking unique sample identification number (sid) is important. There are 48 sid in this SOSAR study.

```
# Filter the original dataframe to include only individuals with more than one observation
df_filtered_IgA <- df_sosar_IgA %>%
  filter(sid %in% df_summary_IgA$sid)

# Plotting ipab_MFI over time for each individual
ggplot(df_filtered_IgA, aes(x = timepoint, y = ipab_MFI, group = sid, color = sid)) +
  geom_line() +
  geom_point() +
  labs(title = "Change in ipab_MFI Over Time",
       x = "Timepoint",
       y = "ipab_MFI",
       color = "Individual ID") +
  theme_minimal()
```

sid	count
SOSAR-11001	5
SOSAR-11002	4
SOSAR-11006	5
SOSAR-11012	5
SOSAR-11015	5
SOSAR-11018	4
SOSAR-11019	4
SOSAR-12005	5
SOSAR-12017	4
SOSAR-13003	4
SOSAR-13004	5
SOSAR-13007	5
SOSAR-13008	5
SOSAR-13009	5
SOSAR-13010	5
SOSAR-13011	5
SOSAR-13013	5
SOSAR-13014	5
SOSAR-13016	5
SOSAR-13020	4
SOSAR-13022	4
SOSAR-13025	3
SOSAR-13033	3
SOSAR-21002	5
SOSAR-21003	5
SOSAR-21010	5
SOSAR-21011	5
SOSAR-21012	5
SOSAR-21013	5
SOSAR-21018	5
SOSAR-21019	5