```r
library(tidyverse)
library(summarytools)
library(lme4)
library(lmerTest)
library(gridExtra)
knitr::opts_chunk$set(tidy.opts=list(width.cutoff=40), tidy=TRUE)
knitr::opts_chunk$set(echo = TRUE, results = 'hide')


data <- read.csv("D:/P8157_FINAL/BMI_IOS_SCD_Asthma.csv")
# check variables
data_grouped <- data |>
    filter(Group == "C-SCD") |>
    arrange(Subject.ID, Observation_number)

subjects_always_yes <- data_grouped |>
    group_by(Subject.ID) |>
    summarise(All_Yes = all(Asthma == "Yes",
        na.rm = TRUE)) |>
    filter(All_Yes == TRUE)

subjects_always_no <- data_grouped |>
    group_by(Subject.ID) |>
    summarise(All_Yes = all(Asthma == "No",
        na.rm = TRUE)) |>
    filter(All_Yes == TRUE)
nrow(subjects_always_yes)
nrow(subjects_always_no)

# rename variables to snake_space
data <- data |>
    mutate(SCD = ifelse(Group == "C-SCD",
        1, ifelse(Group == "C-Asthma", 0,
            NA)))

names(data) <- gsub("\\.\\.", "_", names(data))
names(data) <- tolower(gsub("\\.", "", names(data)))

data <- data |>
    rename(id = subjectid) |>
    select(-group) |>
    mutate(across(c(asthma, laba, ics, hydroxyurea),
        ~ifelse(. == "Yes", 1, ifelse(. ==
            "No", 0, NA))), gender = ifelse(tolower(gender) ==
        "female", 0, ifelse(tolower(gender) ==
        "male", 1, NA))  # Convert gender to lowercase before comparison
) |>
    filter(asthma == 1)


# mutate r5-20
data <- data |>
    mutate(r520hz_pp = r5hz_pp - r20hz_pp)
```

```r
# baseline summary statistics
baseline_df <- data |>
    group_by(id) |>
    filter(observation_number == min(observation_number)) |>
    ungroup()

summary(baseline_df)


# show imbalanced feature
observation_counts <- data |>
    group_by(id) |>
    summarise(number_of_observations = n())

summary_of_observations <- observation_counts |>
    group_by(number_of_observations) |>
    summarise(ids_with_this_many_observations = n())

print(summary_of_observations)

# check consistance of variables

consistent_gender <- data |>
    group_by(id) |>
    summarise(n_distinct_gender = n_distinct(gender)) |>
    filter(n_distinct_gender > 1)
print(consistent_gender)
treatment_changes <- data |>
    group_by(id) |>
    summarise(n_distinct_ics = n_distinct(ics),
        n_distinct_laba = n_distinct(laba),
        n_distinct_hydroxyurea = n_distinct(hydroxyurea)) |>
    filter(n_distinct_ics > 1 | n_distinct_laba >
        1 | n_distinct_hydroxyurea > 1)
print(treatment_changes)

height_trends <- data |>
    arrange(id, age_months) |>
    group_by(id) |>
    mutate(height_change = c(NA, diff(height_cm))) |>
    summarise(downward_trend = all(height_change <
        0, na.rm = TRUE), num_observations = n()) |>
    filter(downward_trend == TRUE & num_observations >
        1)


# plots
p1 = ggplot(data, aes(x = age_months, y = height_cm,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and Height",
        x = "Age (months)", y = "Height (cm)",
        color = "Gender") + theme_minimal()
```

```
p2 = ggplot(data, aes(x = age_months, y = bmi,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and BMI",
        x = "Age (months)", y = "BMI", color = "Gender") +
    theme_minimal()

P1 = grid.arrange(p1, p2, ncol = 2)

p3 = ggplot(data, aes(x = age_months, y = r5hz_pp,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and R5",
        x = "Age (months)", y = "R5", color = "Gender") +
    theme_minimal()

p4 = ggplot(data, aes(x = age_months, y = r20hz_pp,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and R20",
        x = "Age (months)", y = "R5_20",
        color = "Gender") + theme_minimal()

p5 = ggplot(data, aes(x = age_months, y = r520hz_pp,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and R5_20",
        x = "Age (months)", y = "R5_20",
        color = "Gender") + theme_minimal()

p6 = ggplot(data, aes(x = age_months, y = x5hz_pp,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and X5",
        x = "Age (months)", y = "X5", color = "Gender") +
    theme_minimal()

p7 = ggplot(data, aes(x = age_months, y = fres_pp,
    group = id, color = factor(gender))) +
    geom_point() + geom_line() + scale_color_manual(values = c("#FA8072",
    "#56B4E9"), labels = c("Male", "Female")) +
    labs(title = "Relationship between Age and Fres",
        x = "Age (months)", y = "Fres", color = "Gender") +
    theme_minimal()

P2 = grid.arrange(p3, p4, p5, p6, p7, ncol = 3,
    nrow = 2)
```

```
# exclude outliers
data_clean <- data |>
    mutate(across(c(r5hz_pp, r20hz_pp, x5hz_pp,
        fres_pp, r520hz_pp), ~ifelse(abs(scale(.)) >=
        3, NA, .)))

summary(data_clean)

# models
model_cr5 <- lmer(r5hz_pp ~ age_months +
    gender * bmi + scd + ics + laba + hydroxyurea +
    (1 | id), data = data_clean)

summary(model_cr5)
qqnorm(residuals(model_cr5), main = "QQ Plot for R5")
qqline(residuals(model_cr5))
```
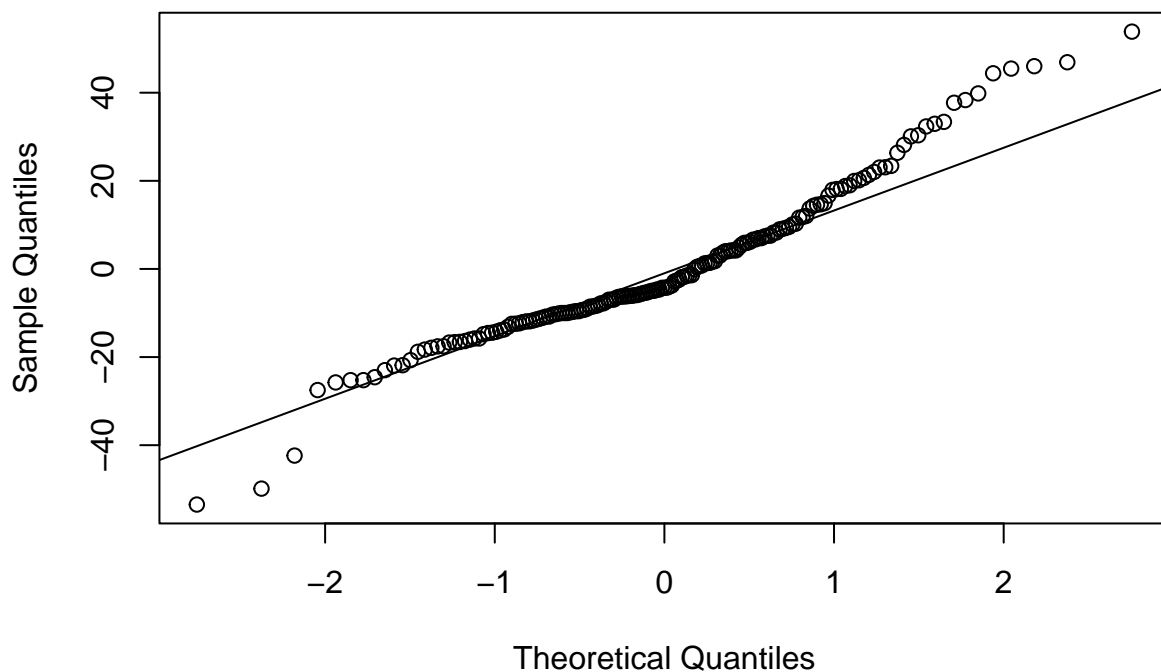
QQ Plot for R5


```
confint(model_cr5, level = 0.95)


## Computing profile confidence intervals ...

model_cr20 <- lmer(r20hz_pp ~ age_months +
    gender * bmi + scd + ics + laba + hydroxyurea +
    (1 | id), data = data_clean)
```
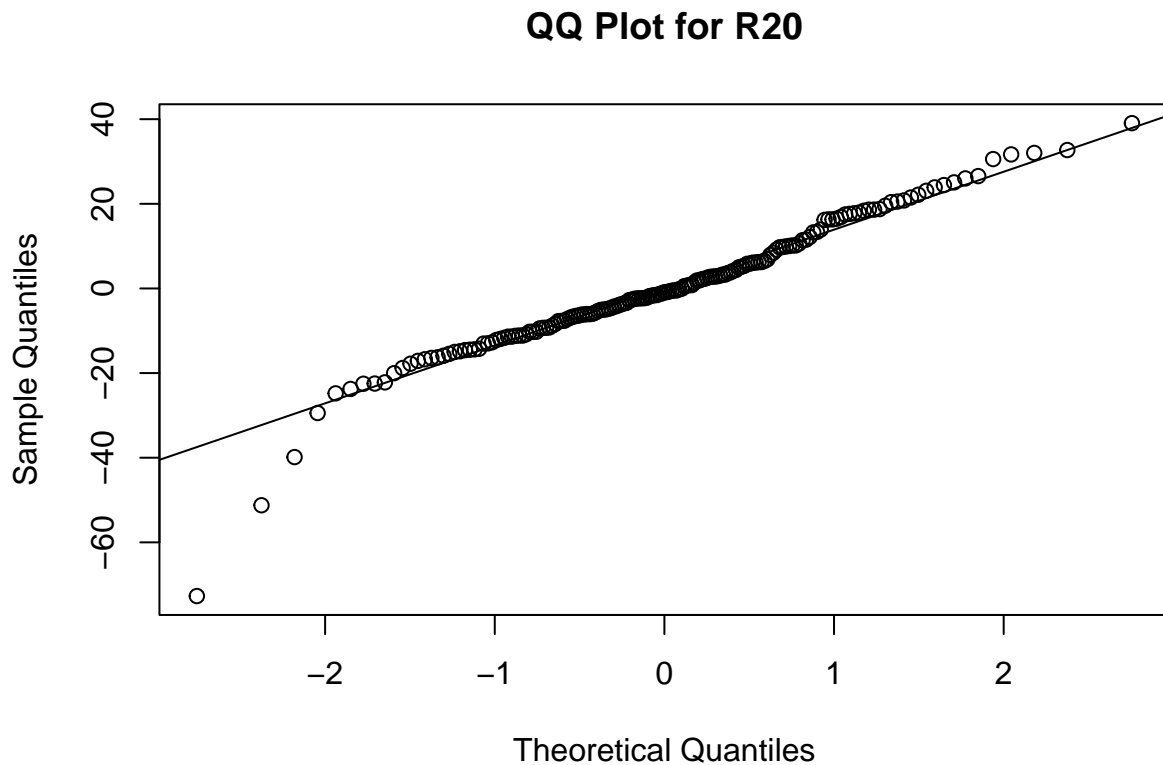
```
summary(model_cr20)
qqnorm(residuals(model_cr20), main = "QQ Plot for R20")
qqline(residuals(model_cr20))
```
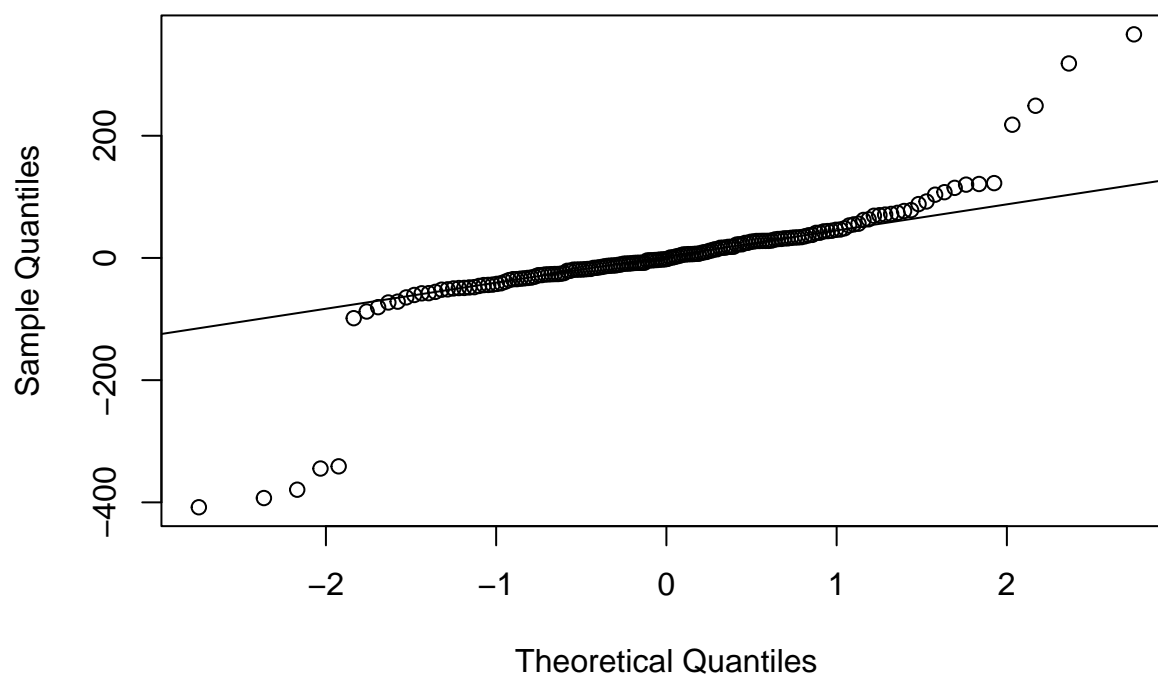
QQ Plot for R20


```
confint(model_cr20, level = 0.95)
```

```
## Computing profile confidence intervals ...
```

```
model_cx5 <- lmer(x5hz_pp ~ age_months +
    gender * bmi + scd + ics + laba + hydroxyurea +
    (1 | id), data = data_clean)
```

```
summary(model_cx5)
qqnorm(residuals(model_cx5), main = "QQ Plot for X5")
qqline(residuals(model_cx5))
```
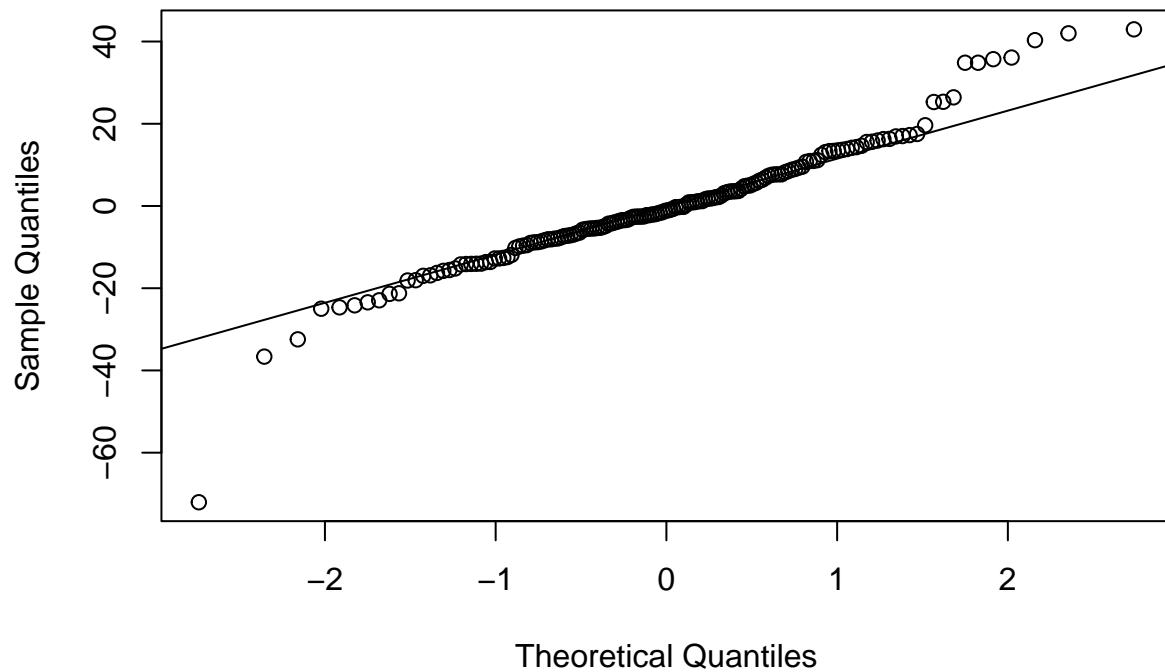
QQ Plot for X5


```r
confint(model_cx5, level = 0.95)
```

```
## Computing profile confidence intervals ...
```

```r
model_cfres <- lmer(fres_pp ~ age_months +
    gender * bmi + scd + ics + laba + hydroxyurea +
    (1 | id), data = data_clean)

summary(model_cfres)
qqnorm(residuals(model_cfres), main = "QQ Plot for Fres")
qqline(residuals(model_cfres))
```
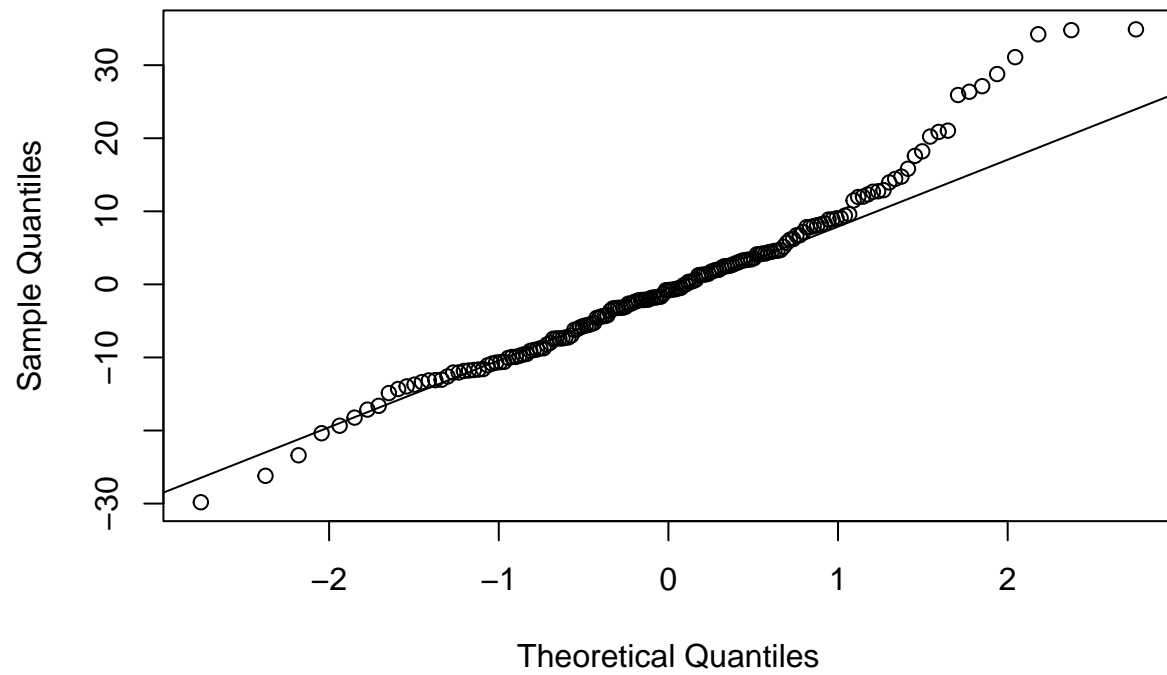
QQ Plot for Fres


```
confint(model_cfres, level = 0.95)
```

```
## Computing profile confidence intervals ...
```

```
model_cr520 <- lmer(r520hz_pp ~ age_months +
    gender * bmi + scd + ics + laba + hydroxyurea +
    (1 | id), data = data_clean)
```

```
summary(model_cr520)
qqnorm(residuals(model_cr520), main = "QQ Plot for R5-20")
qqline(residuals(model_cr520))
```

QQ Plot for R5–20


```r
confint(model_cr520, level = 0.95)
```

```
## Computing profile confidence intervals ...
```