# model.rmd

Yanran Li, Yijin Wang, Shubo Zhang

2024-03-20

```
library(mice)
library(tidyr)
library(dplyr)
library(visdat)
library(naniar)
library(ggplot2)
library(lme4)
library(broom.mixed)
library(kableExtra)
library(gtsummary)
```

```
data <- read.csv("data.csv") %>%
  janitor::clean_names() %>%
  mutate(day = as.factor(day),
         treatment_group = as.factor(treatment_group))
```

```
# pre-processing: expand design matrix and fill in NAs
full_data <- data %>%
  complete(day,
           nesting(subject_id, treatment_group)) %>%
  relocate(day, .after = subject_id) %>%
  relocate(treatment_group, .after = gender) %>%
  group_by(subject_id) %>%
  fill(age, gender, .direction = "downup") %>%
  arrange(subject_id, day) %>%
  group_by(subject_id) %>%
  # calculate difference
  mutate(diff_mem_comp = mem_comp - first(mem_comp)) %>%
  filter(day != 0) %>%
  select(-mem_comp) %>%
  ungroup()
```

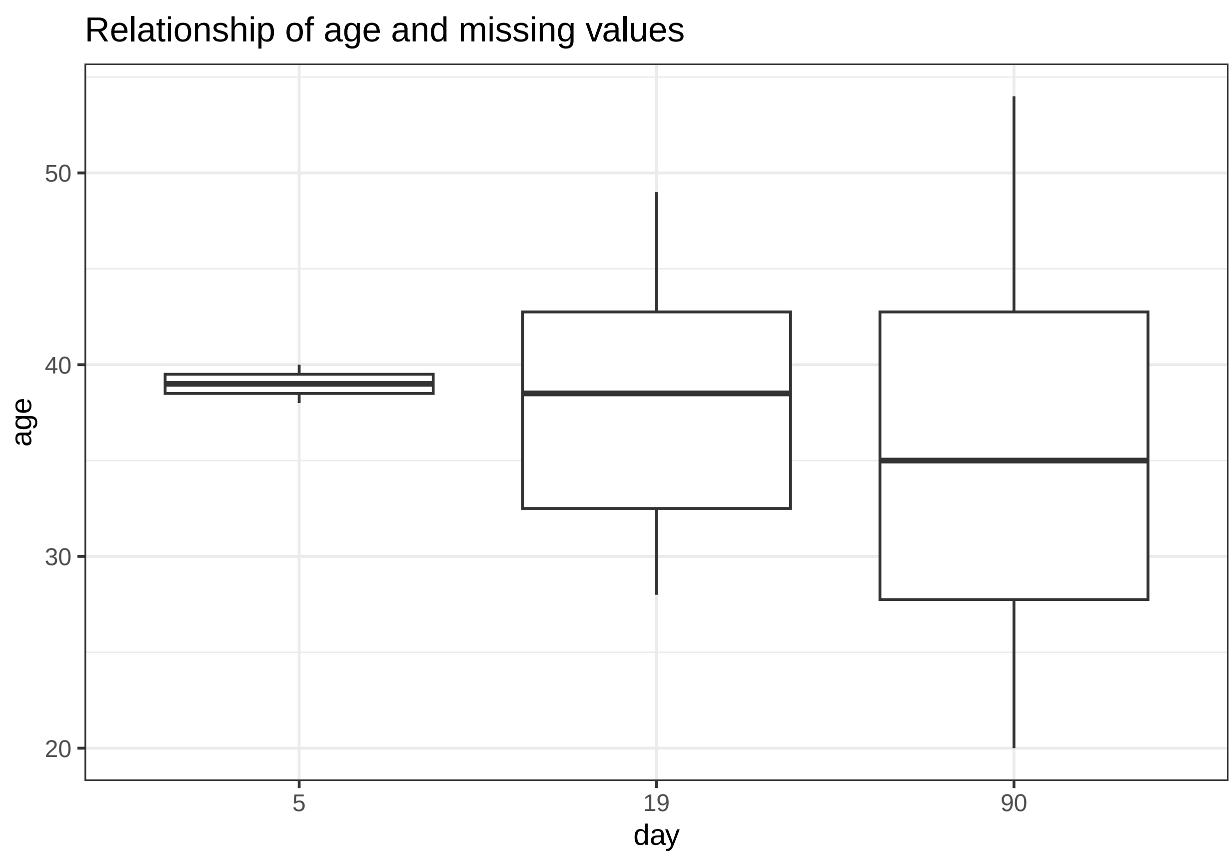
```
full_data %>%
  group_by(day) %>%
  summarise(`Missing Data Percentage` = sum(is.na(diff_mem_comp))/n()*100) %>%
  rename(Day= day) %>%
  kable()
```

| Day | Missing Data Percentage |
|---|---:|
| 5 | 4.255319 |
| 19 | 25.531915 |
| 90 | 34.042553 |

```
vis_miss <- vis_miss(full_data)
```

```
vis_miss_facet <- vis_miss(full_data,facet = day)

gg_miss_var <- gg_miss_var(full_data,
            facet = day)
```

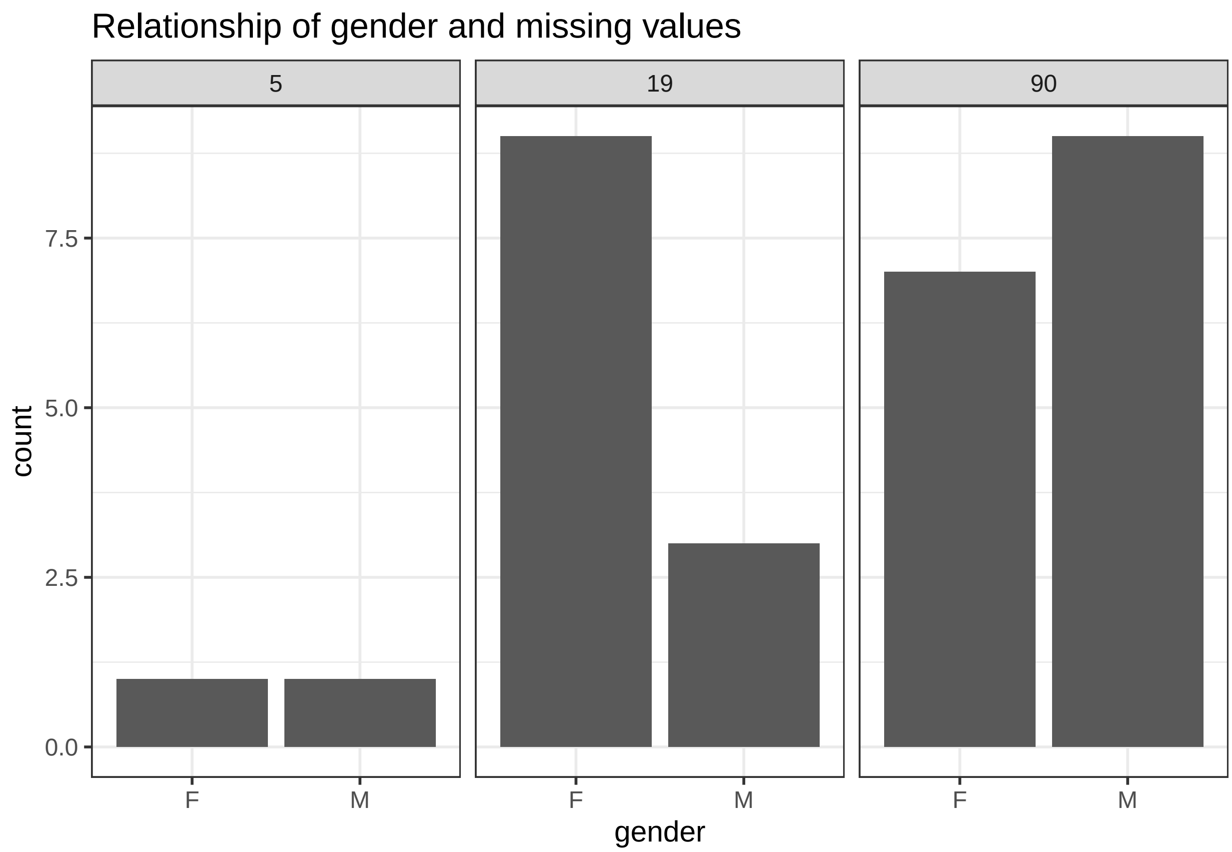
```
full_data %>% filter(is.na(diff_mem_comp)) %>%
  ggplot(aes(x=as.factor(day)))+
  geom_boxplot(aes(y=age))+
  theme_bw()+
  xlab("day")+
  ggtitle("Relationship of age and missing values")
```




```
full_data %>% filter(is.na(diff_mem_comp)) %>%
  ggplot()+
  geom_bar(aes(x=gender, stat = "count"))+
  facet_wrap(day~.)+
```

```
  theme_bw()+
  xlab("gender")+
  ggtitle("Relationship of gender and missing values")
```

```
## Warning in geom_bar(aes(x = gender, stat = "count")): Ignoring unknown
## aesthetics: stat
```




```
# Rigorous test for MCAR: Little's Test.
# HO: data is MCAR
# HA: data is not MCAR ==> MAR or MNAR
# reject HO in this case
mcar_test(full_data)
```

```
## # A tibble: 1 x 4
##   statistic    df p.value missing.patterns
##       <dbl> <dbl>   <dbl>            <int>
## 1      15.4     5 0.00867                2
```

```
mar_data <- full_data %>%
      filter(!is.na(diff_mem_comp)) %>%
  mutate(day = factor(day,levels = c(5,19,90)),
         treatment_group = factor(treatment_group,levels = c("A","B","C")))
```

```r
# MAR assumption model
trt_model <- with(mar_data, lmer(diff_mem_comp ~ age + gender + treatment_group+day + (1|subject_id)))

## Aim 1: test separately for two treatment effect

summary_output <- summary(trt_model)
coef_table <- as.data.frame(summary_output$coefficients)
df <- df.residual(trt_model)
p_values <- 2 * pt(abs(coef_table[, "t value"]), df = df, lower.tail = FALSE)
coef_table$p_value <- p_values
coef_table %>% kable(caption = "Under MAR Assumption Model Summary Table")
```

Table 2: Under MAR Assumption Model Summary Table

| | Estimate | Std. Error | t value | p_value |
|---|---|---|---|---|
| (Intercept) | 0.2294536 | 0.3701469 | 0.6198987 | 0.5367069 |
| age | -0.0114407 | 0.0087688 | -1.3047029 | 0.1949301 |
| genderM | 0.1520745 | 0.1578286 | 0.9635424 | 0.3375542 |
| treatment_groupB | 0.0516028 | 0.1918193 | 0.2690179 | 0.7884593 |
| treatment_groupC | 0.4271182 | 0.1828673 | 2.3356733 | 0.0214653 |
| day19 | 0.0456282 | 0.0900300 | 0.5068113 | 0.6133811 |
| day90 | 0.0215896 | 0.0949081 | 0.2274790 | 0.8205066 |

```r
## Aim 2: jointly test interaction terms
full_model <- with(mar_data, lmer(diff_mem_comp ~ age + gender + treatment_group+day  +treatment_group:d
car::Anova(full_model, type= "III") %>% kable(caption="Type III Anova Test")
```

Table 3: Type III Anova Test

| | Chisq | Df | Pr(>Chisq) |
|---|---|---|---|
| (Intercept) | 0.7466533 | 1 | 0.3875379 |
| age | 1.9966452 | 1 | 0.1576478 |
| gender | 1.0351356 | 1 | 0.3089555 |
| treatment_group | 5.5578423 | 2 | 0.0621055 |
| day | 0.7412677 | 2 | 0.6902967 |
| treatment_group:day | 6.0043155 | 4 | 0.1988262 |

```r
trt_model_table <-
  trt_model %>%
  tbl_regression(
    estimate_fun = function(x) style_number(x, digits = 3),
    exponentiate = F,
    label = list(
      age ~ "Age",
      gender ~ "Gender",
      treatment_group ~ "Treatment",
      day ~ "Day"
    ))

gt.trt_model_table <- trt_model_table %>% as_gt()
gt::gtsave(gt.trt_model_table, file = "./pic/tbl_MAR_trt_model_table.png")
```

```r
# sensitivity analysis for each data missing assumptions
# mcar - fit data with completers data
completers_data <- full_data %>%
  group_by(subject_id) %>%
  mutate(missing = any(is.na(diff_mem_comp))) %>%
  ungroup() %>%
  filter(missing == FALSE)

completers_trt_model <- with(completers_data, lmer(diff_mem_comp ~ age + gender + treatment_group+day +

## Aim 1: test separately for two treatment effect

summary_output <- summary(completers_trt_model)
coef_table <- as.data.frame(summary_output$coefficients)
df <- df.residual(completers_trt_model)
p_values <- 2 * pt(abs(coef_table[, "t value"]), df = df, lower.tail = FALSE)
coef_table$p_value <- p_values
coef_table %>% kable(caption = "Completers' Data Model Summary Table")
```

Table 4: Completers' Data Model Summary Table

| | Estimate | Std. Error | t value | p_value |
|---|---|---|---|---|
| (Intercept) | -0.6003684 | 0.5348537 | -1.1224909 | 0.2659136 |
| age | 0.0067441 | 0.0121918 | 0.5531654 | 0.5821079 |
| genderM | 0.0640661 | 0.1979345 | 0.3236730 | 0.7472577 |
| treatment_groupB | 0.2868187 | 0.2428841 | 1.1808872 | 0.2420866 |
| treatment_groupC | 0.4370294 | 0.2212894 | 1.9749219 | 0.0526630 |
| day19 | 0.0329321 | 0.1169854 | 0.2815061 | 0.7792448 |
| day90 | -0.0056296 | 0.1169854 | -0.0481223 | 0.9617710 |

```r
## Aim 2: jointly test interaction terms
completers_full_model <- with(completers_data, lmer(diff_mem_comp ~ age + gender + treatment_group+day
car::Anova(full_model, type= "III") %>% kable(caption="Type III Anova Test")
```

Table 5: Type III Anova Test

| | Chisq | Df | Pr(>Chisq) |
|---|---|---|---|
| (Intercept) | 0.7466533 | 1 | 0.3875379 |
| age | 1.9966452 | 1 | 0.1576478 |
| gender | 1.0351356 | 1 | 0.3089555 |
| treatment_group | 5.5578423 | 2 | 0.0621055 |
| day | 0.7412677 | 2 | 0.6902967 |
| treatment_group:day | 6.0043155 | 4 | 0.1988262 |

```r
completers_trt_model_table <-
  completers_trt_model %>%
  tbl_regression(
    estimate_fun = function(x) style_number(x, digits = 3),
    exponentiate = F,
    label = list(
```

```
      age ~ "Age",
      gender ~ "Gender",
      treatment_group ~ "Treatment",
      day ~ "Day"
    ))

gt.completers_trt_model_table <- completers_trt_model_table %>% as_gt()
gt::gtsave(gt.completers_trt_model_table, file = "./pic/tbl_completers_trt_model_table.png")
```

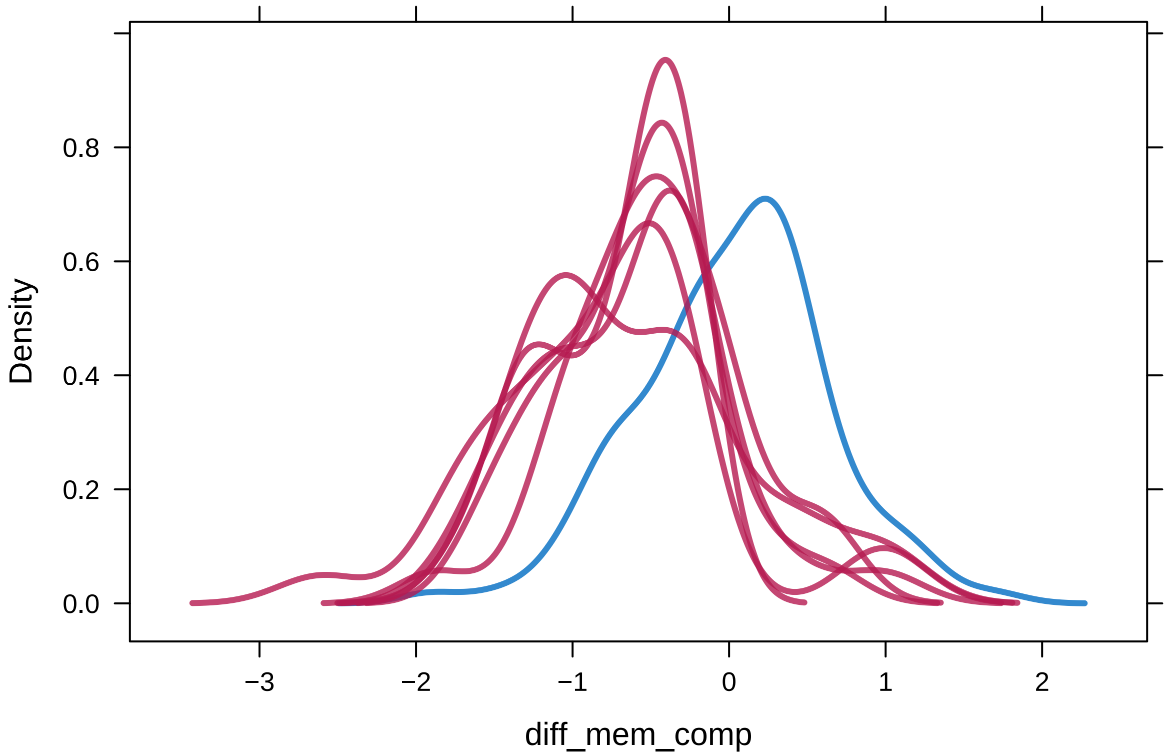
```
# mnar - fill na's with different delta --> fit model --> CI's
# multiple imputations
imputed_data <- mice::mice(data = full_data, m = 6, seed =2024, print = FALSE)
```

```
## Warning: Number of logged events: 31
```

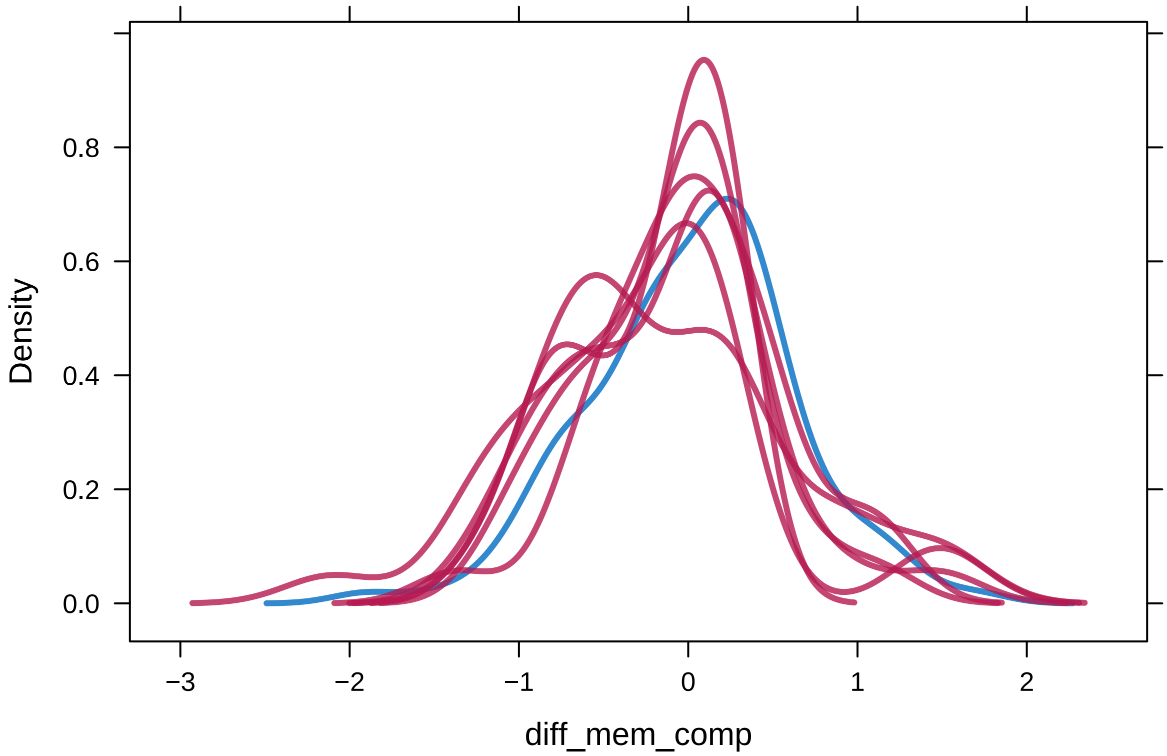
```
# Create a delta vector that represent the following adjustment values for mmHg: 0 for MAR, and -0.7, -
delta <- c(-0.7, -0.2, 0.2, 0.7)

imp.all <- vector("list", length(delta))
post <- imputed_data$post
for (i in 1:length(delta)){
  d <- delta[i]
  cmd <- paste("imp[[j]][,i] <- imp[[j]][,i] +", d)
  post["diff_mem_comp"] <- cmd
  imp <- mice::mice(full_data, post = post,m = 6,seed = 2024, print = FALSE)
  imp.all[[i]] <- imp
}
```

```
## Warning: Number of logged events: 31

## Warning: Number of logged events: 31

## Warning: Number of logged events: 31

## Warning: Number of logged events: 31
```

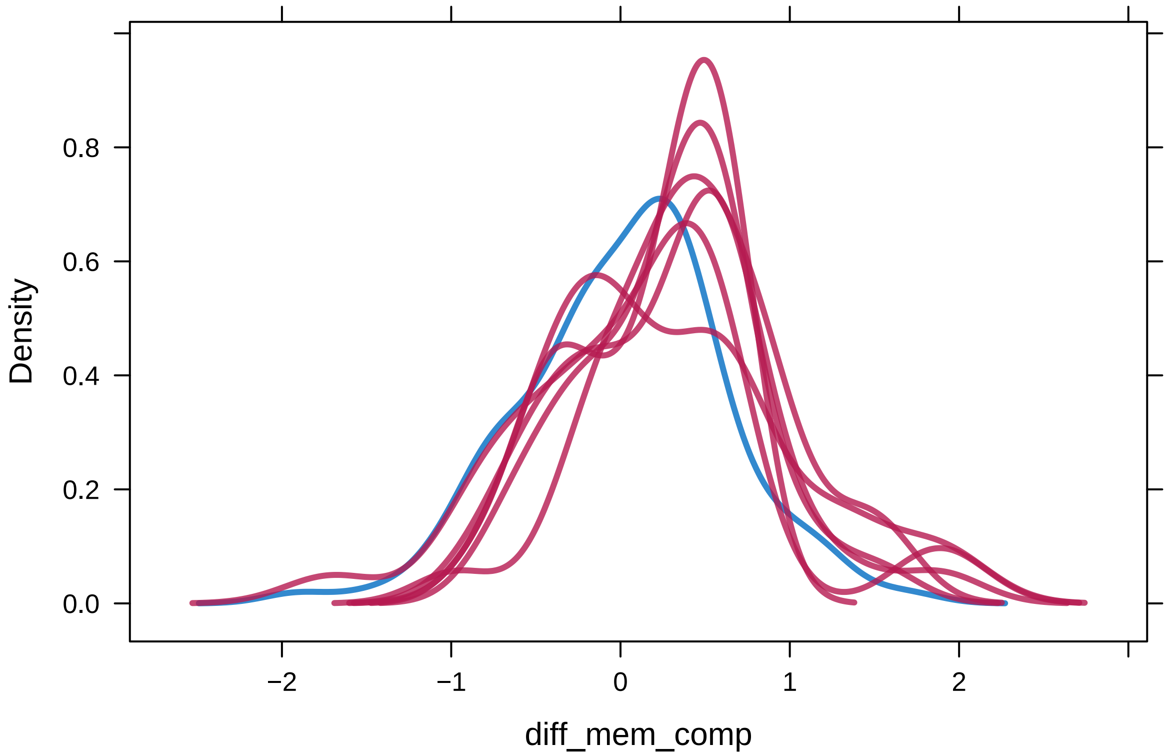
```
# imputation with no adjustment (delta = -0.7)
densityplot(imp.all[[1]], lwd = 3)
```



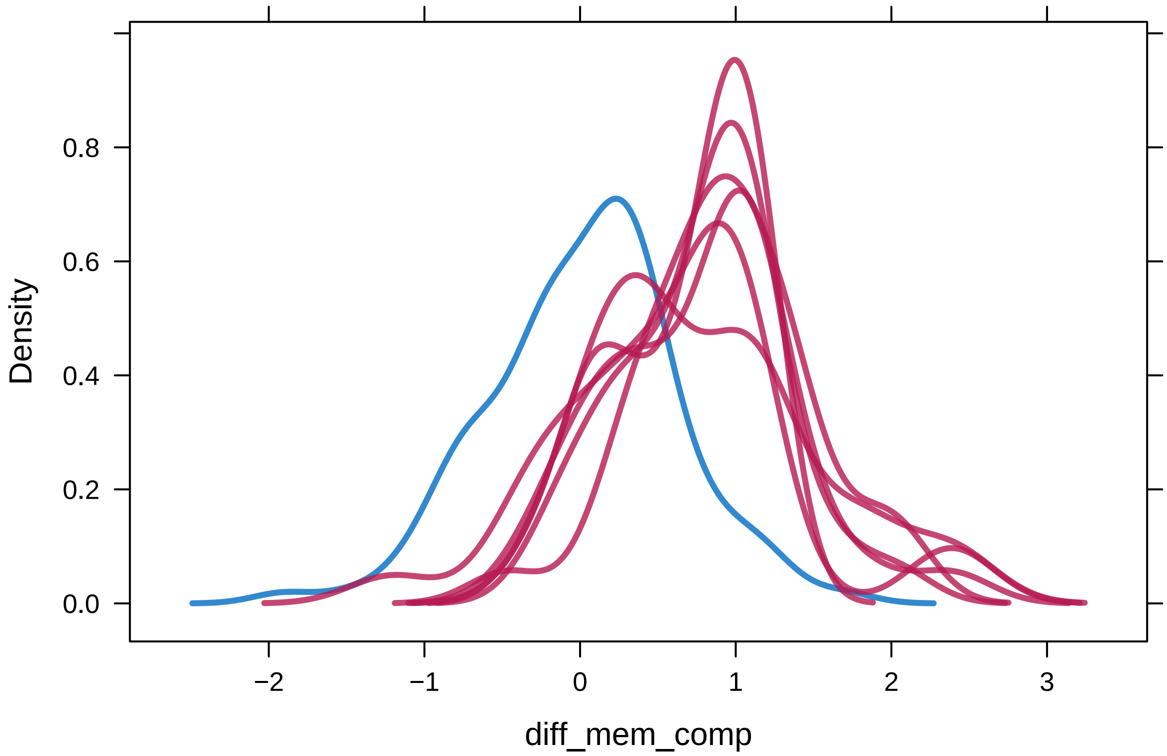

```r
# imputation with adjustment (delta = -0.2)
densityplot(imp.all[[2]], lwd = 3)
```



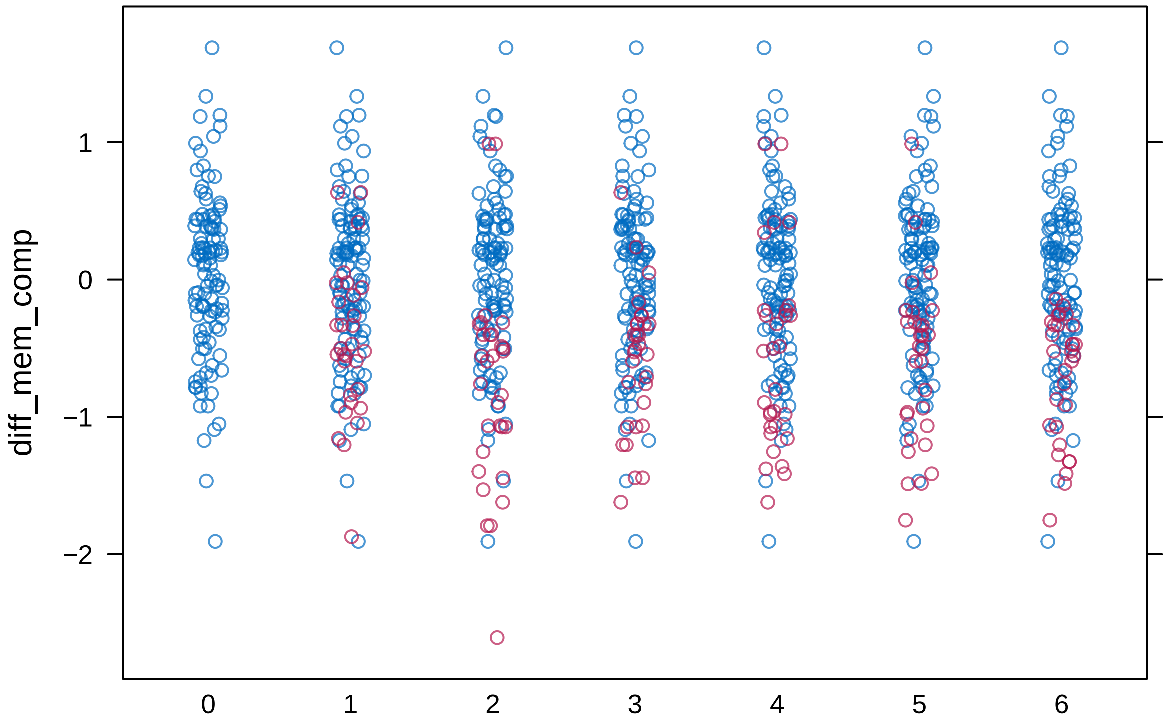

```
# imputation with adjustment (delta = 0.2)
densityplot(imp.all[[3]], lwd = 3)
```



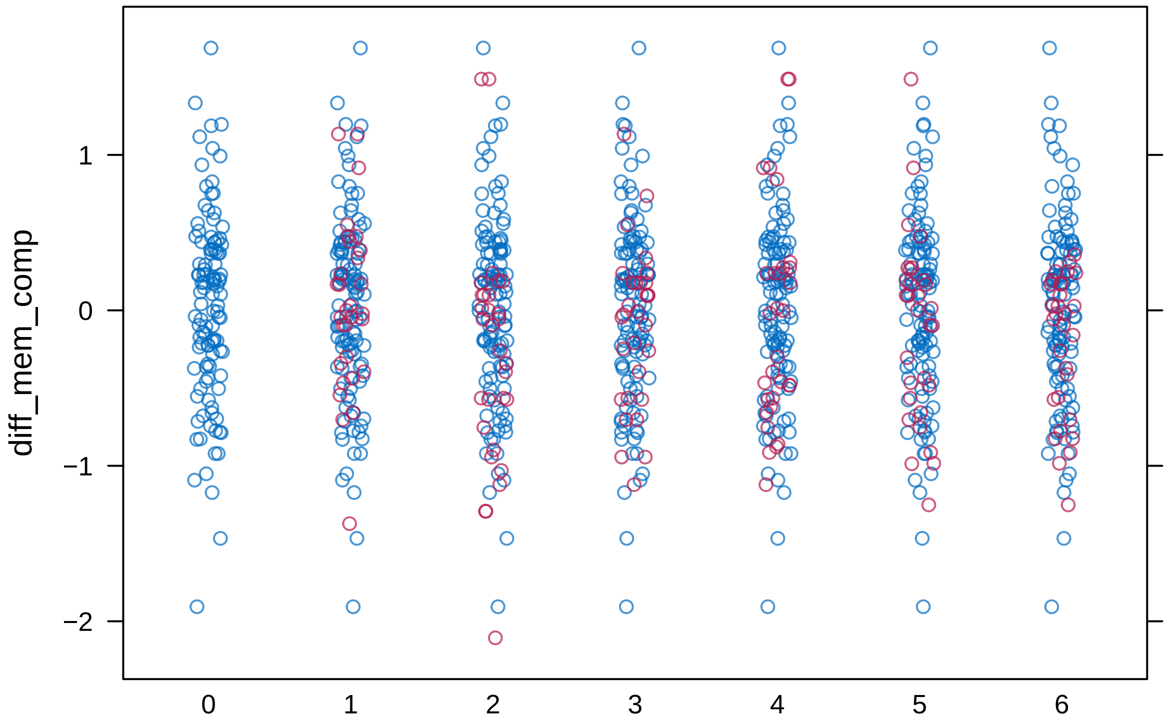

```
# imputation with adjustment (delta = 0.7)
densityplot(imp.all[[4]], lwd = 3)
```

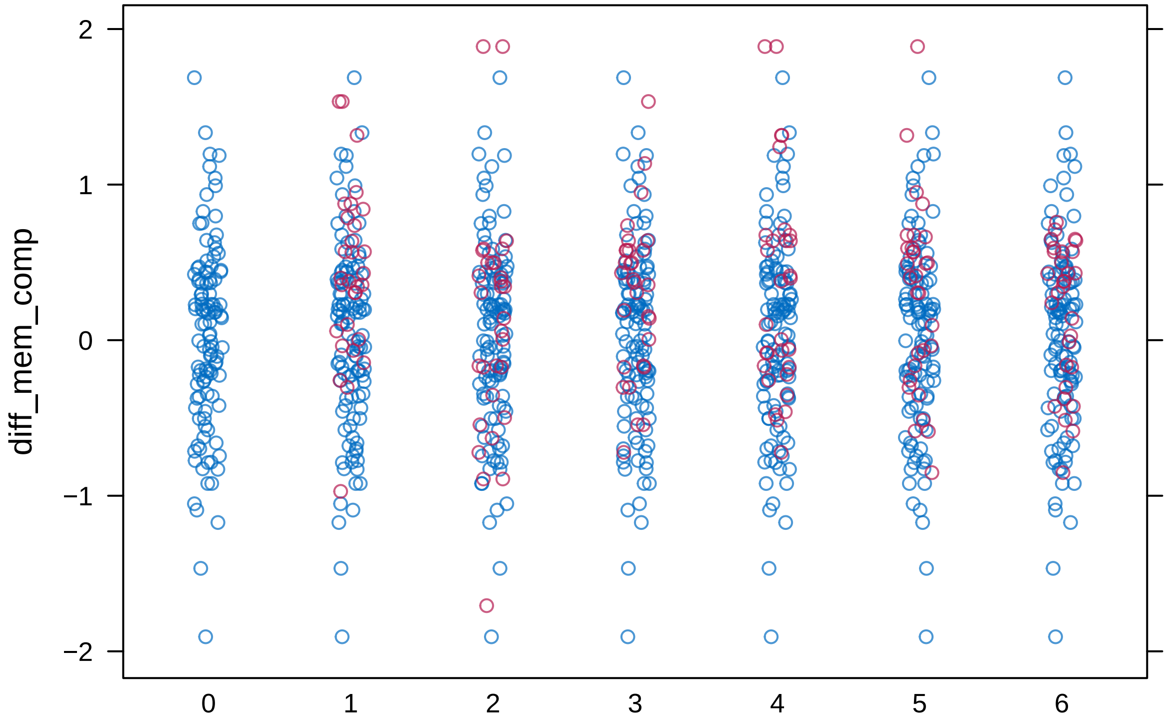



```
for (i in 1:length(delta)){
  curr_data <- imp.all[[i]]
  print(mice::stripplot(curr_data, diff_mem_comp))
}
```

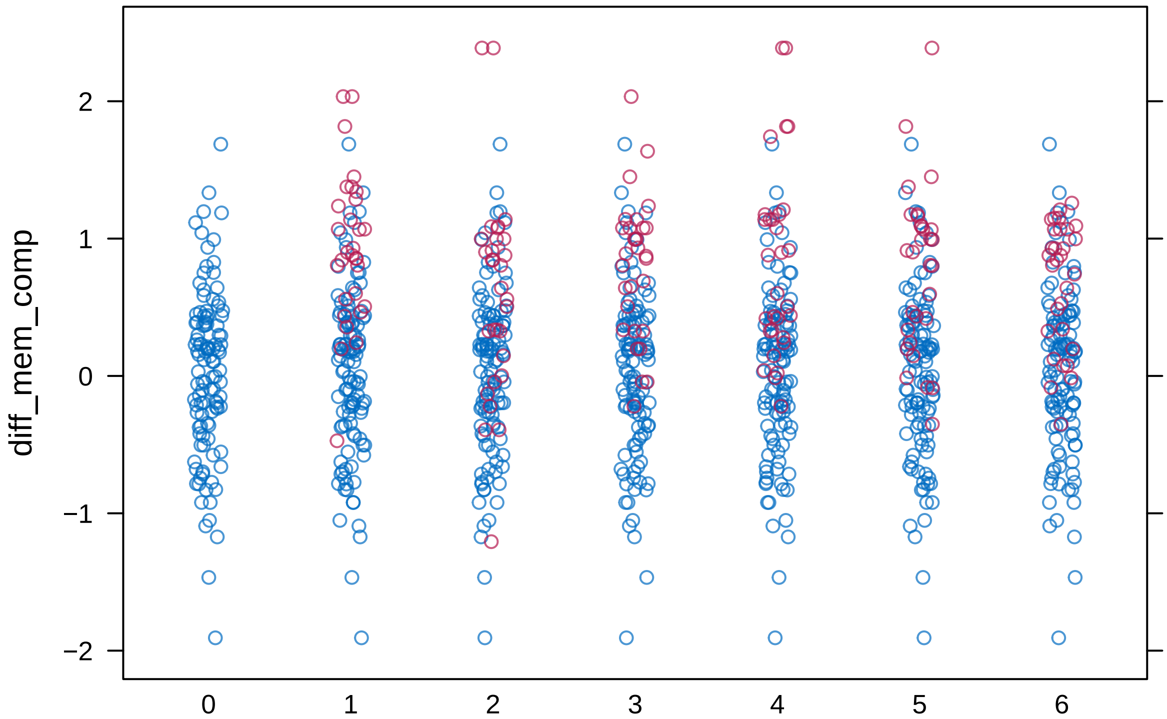


diff_mem_comp
1
0
−1
−2
0
1
2
3
4
5
6


diff_mem_comp
1
0
−1
−2
0
1
2
3
4
5
6



diff_mem_comp
2
1
0
−1
−2
0
1
2
3
4
5
6




```
int_results<-tibble(
  statistic = numeric(),
  df1 = numeric(),
  df2 = numeric(),
  p.value=numeric(),
  riv = numeric()
)


# calculate results for imputation+adjustment datasets
for (i in 1:length(delta)){
  curr_data <- imp.all[[i]]

  # fit model
  curr_full_model <- with(curr_data,lmer(diff_mem_comp ~ age + gender + treatment_group+day+treatment_g

  curr_trt_model <- with(curr_data,lmer(diff_mem_comp ~ age + gender + treatment_group+day + (1|subject_

  curr.pool.trt.fit <- pool(curr_trt_model) %>% summary

  print(curr.pool.trt.fit %>% kable(caption = paste0("delta=", delta[i]," Summary Table")))

  # anova based on likelihood ratio
  curr_int_result<-anova(curr_trt_model, curr_full_model,method = "D3")[["out"]][["2 ~~ 1"]][["result"]]
    t() %>%
    as.data.frame()
```

```
  colnames(curr_int_result) <-c("statistic","df1","df2","p.value","riv")

  int_results<- int_results %>% rbind(curr_int_result)
}
```

Table 6: delta=-0.7 Summary Table

| term | estimate | std.error | statistic | df | p.value |
|---|---:|---:|---:|---:|---:|
| (Intercept) | 0.0085035 | 0.3642237 | 0.0233469 | 31.02879 | 0.9815231 |
| age | -0.0063985 | 0.0086018 | -0.7438579 | 28.45655 | 0.4630608 |
| genderM | 0.1378534 | 0.1329736 | 1.0366972 | 87.83756 | 0.3027226 |
| treatment_groupB | 0.0226668 | 0.1634136 | 0.1387081 | 79.87091 | 0.8900299 |
| treatment_groupC | 0.4094935 | 0.1498087 | 2.7334428 | 112.77782 | 0.0072790 |
| day19 | -0.0323330 | 0.1251122 | -0.2584322 | 65.37118 | 0.7968853 |
| day90 | -0.2413337 | 0.1286710 | -1.8755880 | 51.63986 | 0.0663705 |

Table 7: delta=-0.2 Summary Table

| term | estimate | std.error | statistic | df | p.value |
|---|---:|---:|---:|---:|---:|
| (Intercept) | 0.1237174 | 0.3555619 | 0.3479488 | 28.89259 | 0.7304010 |
| age | -0.0085214 | 0.0084294 | -1.0109152 | 26.75466 | 0.3211127 |
| genderM | 0.1025427 | 0.1293774 | 0.7925858 | 85.02797 | 0.4302251 |
| treatment_groupB | 0.0413311 | 0.1590897 | 0.2597978 | 76.87298 | 0.7957143 |
| treatment_groupC | 0.4170558 | 0.1454923 | 2.8665138 | 111.30631 | 0.0049642 |
| day19 | 0.0740499 | 0.1114839 | 0.6642212 | 51.68766 | 0.5095011 |
| day90 | -0.0923976 | 0.1154634 | -0.8002324 | 39.61403 | 0.4283492 |

Table 8: delta=0.2 Summary Table

| term | estimate | std.error | statistic | df | p.value |
|---|---:|---:|---:|---:|---:|
| (Intercept) | 0.2158885 | 0.3633203 | 0.5942097 | 30.80381 | 0.5567112 |
| age | -0.0102197 | 0.0086205 | -1.1855139 | 28.64291 | 0.2455567 |
| genderM | 0.0742941 | 0.1333613 | 0.5570886 | 88.12956 | 0.5788790 |
| treatment_groupB | 0.0562626 | 0.1638799 | 0.3433163 | 80.18489 | 0.7322591 |
| treatment_groupC | 0.4231056 | 0.1502736 | 2.8155691 | 112.92717 | 0.0057472 |
| day19 | 0.1591563 | 0.1094868 | 1.4536572 | 49.59765 | 0.1523401 |
| day90 | 0.0267514 | 0.1135364 | 0.2356195 | 37.87463 | 0.8149974 |

Table 9: delta=0.7 Summary Table

| term | estimate | std.error | statistic | df | p.value |
|---|---:|---:|---:|---:|---:|
| (Intercept) | 0.3311023 | 0.3901022 | 0.8487579 | 37.64723 | 0.4013808 |
| age | -0.0123426 | 0.0092409 | -1.3356456 | 35.04209 | 0.1902760 |
| genderM | 0.0389833 | 0.1461012 | 0.2668241 | 96.61744 | 0.7901732 |
| treatment_groupB | 0.0749270 | 0.1792172 | 0.4180791 | 89.51610 | 0.6768905 |
| treatment_groupC | 0.4306678 | 0.1655055 | 2.6021357 | 117.00162 | 0.0104620 |

| term | estimate | std.error | statistic | df | p.value |
|---|---|---|---|---|---|
| day19 | 0.2655393 | 0.1187795 | 2.2355642 | 59.17023 | 0.0291664 |
| day90 | 0.1756876 | 0.1225223 | 1.4339226 | 46.04623 | 0.1583507 |

```r
int_results %>% kable(caption = "Pooled Anova Test For Interaction Term")
```

Table 10: Pooled Anova Test For Interaction Term

| statistic | df1 | df2 | p.value | riv |
|---|---|---|---|---|
| 0.9890056 | 4 | 234.4496 | 0.4142293 | 0.3219873 |
| 0.7956954 | 4 | 156.7185 | 0.5296437 | 0.4307285 |
| 0.5579189 | 4 | 149.2497 | 0.6935537 | 0.4470963 |
| 0.2868625 | 4 | 202.5440 | 0.8862853 | 0.3567690 |