The Importance of Accounting for Clustered Data

Grouped and Individual Data

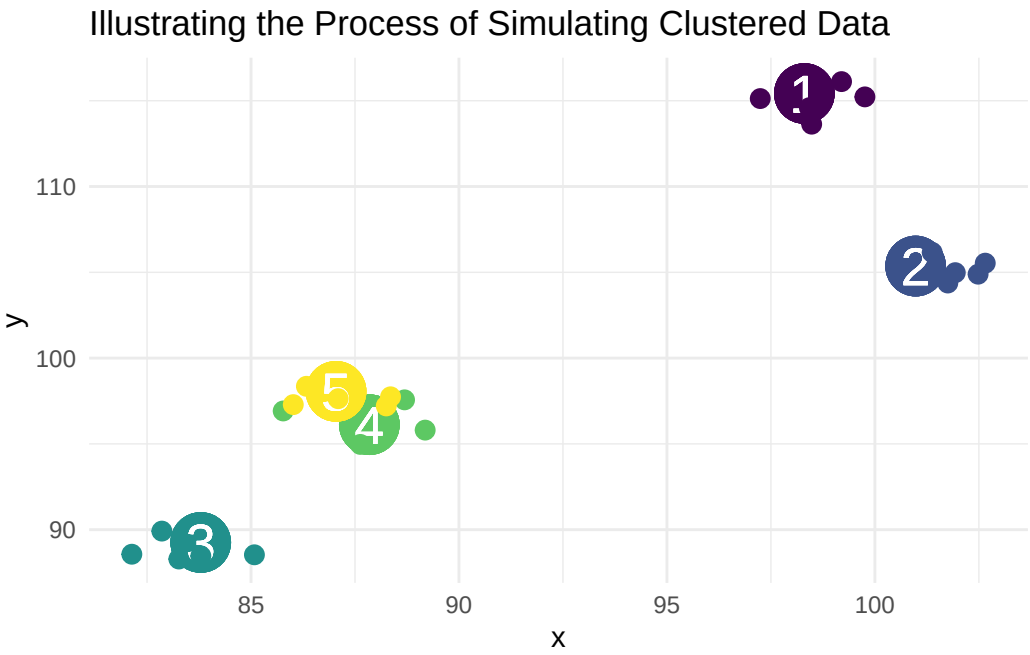
@BlandAltman1994 suggested the following procedure for simulating some data:

> "The data were generated from random numbers, and there is no relation between X and Y at all. Firstly, values of X and Y were generated for each 'subject,' then a further random number was added to make the individual observation." [@BlandAltman1994]

So... we follow their procedure.

> **Simulating The Data**
>
> The graph below illustrates the process of simulating the data.

```
ggplot(mydata,
       aes(color = factor(group_num),
           label = group_num)) +
  geom_point(aes(x = x_group,
                 y = y_group),
             size = 10,
             show.legend = FALSE) +
  geom_text(aes(x = x_group,
                y = y_group),
            color = "white",
            size = 7,
            show.legend = FALSE) +
  geom_point(aes(x = x_individual,
                 y = y_individual),
             size = 3,
             show.legend = FALSE) +
  labs(title = "Illustrating the Process of Simulating Clustered Data",
       x = "x",
       y = "y") +
  theme_minimal() +
  scale_color_viridis_d(name = "group")
```



Analyses

OLS

```
myOLS <- lm(y_individual ~ x_individual, data = mydata)

# summary(myOLS)

sjPlot::tab_model(myOLS,
                  dv.labels = c("OLS"),
                  show.se = TRUE,
                  show.ci = FALSE,
                  show.stat = TRUE)
```

```
# pander(tidy(myOLS))
```

MLM

```
myMLM <- lmer(y_individual ~ x_individual + (1 | group_num),
              data = mydata)

# summary(myMLM)

sjPlot::tab_model(myMLM,
                  dv.labels = c("MLM"),
                  show.se = TRUE,
                  show.ci = FALSE,
                  show.stat = TRUE)
```

```
# pander(tidy(myMLM))
```

Compare OLS and MLM

```
tab_model(myOLS, myMLM,
          dv.labels = c("OLS", "MLM"),
          show.se = TRUE,
          show.ci = FALSE,
          show.stat = TRUE)
```