

# 1 Analysis of Clinical Trials using SAS

## 1.1 The HAMD17 study

**Context.** This is a multicenter clinical trial comparing experimental drug vs. placebo in patients with major depression disorder. The outcome is the change from baseline after 9 weeks of acute treatment, and efficacy is measured using the total score of the Hamilton depression rating scale (17 items).

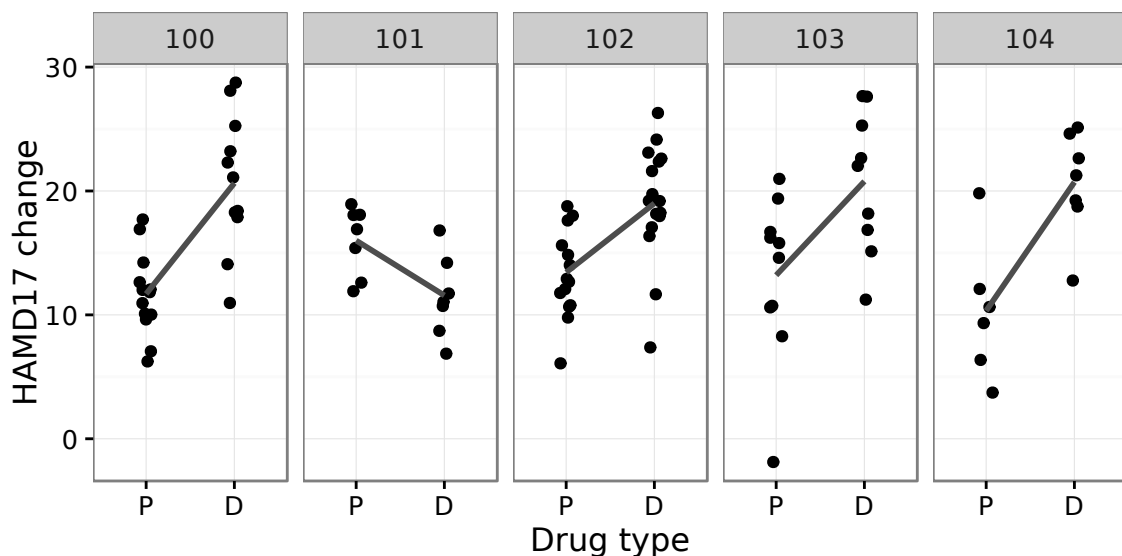
This is a classical application of unbalanced design and potential heterogeneity between clinical centres, where there is an unequal number of observations per treatment (here, drug by center).

Here are some ideas to get the data right into R:

```
raw <- textConnection("
100 P 18 100 P 14 100 D 23 100 D 18 100 P 10 100 P 17 100 D 18 100 D 22
100 P 13 100 P 12 100 D 28 100 D 21 100 P 11 100 P 6 100 D 11 100 D 25
100 P 7 100 P 10 100 D 29 100 P 12 100 P 12 100 P 10 100 D 18 100 D 14
101 P 18 101 P 15 101 D 12 101 D 17 101 P 17 101 P 13 101 D 14 101 D 7
101 P 18 101 P 19 101 D 11 101 D 9 101 P 12 101 D 11 102 P 18 102 P 15
102 P 12 102 P 18 102 D 20 102 D 18 102 P 14 102 P 12 102 D 23 102 D 19
102 P 11 102 P 10 102 D 22 102 D 22 102 P 19 102 P 13 102 D 18 102 D 24
102 P 13 102 P 6 102 D 18 102 D 26 102 P 11 102 P 16 102 D 16 102 D 17
102 D 7 102 D 19 102 D 23 102 D 12 103 P 16 103 P 11 103 D 11 103 D 25
103 P 8 103 P 15 103 D 28 103 D 22 103 P 16 103 P 17 103 D 23 103 D 18
103 P 11 103 P -2 103 D 15 103 D 28 103 P 19 103 P 21 103 D 17 104 D 13
104 P 12 104 P 6 104 D 19 104 D 23 104 P 11 104 P 20 104 D 21 104 D 25
104 P 9 104 P 4 104 D 25 104 D 19
")
d <- scan(raw, what = "character")
rm(raw)
d <- as.data.frame(matrix(d, ncol = 3, byrow = TRUE))
names(d) <- c("center", "drug", "change")
d$change <- as.numeric(as.character(d$change))
d$drug <- relevel(d$drug, ref = "P")
```

Some basic exploratory graphical analysis. In the following code, we display the raw data for each centre and highlight the difference between drug and placebo using a trend line.

```
p <- ggplot(data = d, aes(x = drug, y = change))
p <- p + geom_jitter(width = .2)
p <- p + geom_smooth(aes(group = 1), method = "lm", se = FALSE, colour = "grey30")
p + facet_grid(~ center) + labs(x = "Drug type", y = "HAMD17 change")
```



Using Hmisc package.

```
fm <- change ~ drug + center
s <- summary(fm, data = subset(d, center %in% c("100", "101", "102")),
             method = "cross", fun = smean.sd)
```

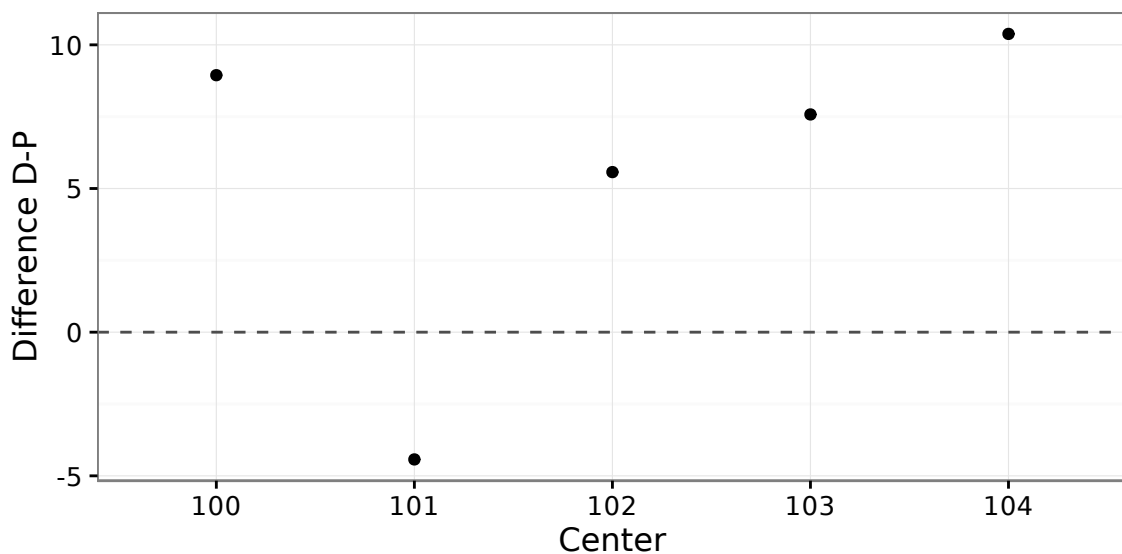
Table 1: Mean HAMD17 change by drug, center

drug	100			101			102			Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
D	11	21	5.6	7	12	3.3	16	19	4.7	34	18	5.7
P	13	12	3.4	7	16	2.7	14	13	3.6	34	13	3.6
Total	24	16	6.3	14	14	3.7	30	16	5.0	68	16	5.3

Only 3 out of 5 centres are shown.

Now, let's consider average change score by center.

```
r <- ddply(d, "center", summarize,
          delta = mean(change[drug == "D"]) - mean(change[drug == "P"]))
p <- ggplot(data = r, aes(x = center, y = delta))
p <- p + geom_point() + geom_hline(yintercept = 0, linetype = 2, colour = "grey30")
p + labs(x = "Center", y = "Difference D-P")
```



Modeling step.

By default, R computes so-called “sequential” Type I sum of squares (SS), and here is what we get when using a standard combination of `lm` and `anova`:

```
m <- lm(change ~ drug * center, data = d)
anova(m)

## Analysis of Variance Table
##
## Response: change
##          Df Sum Sq Mean Sq F value    Pr(>F)
## drug      1  888.04   888.04  40.0745 9.365e-09 ***
## center    4   87.14    21.78   0.9831 0.4209278
## drug:center 4  507.45   126.86   5.7249 0.0003761 ***
```

```
## Residuals    90 1994.38    22.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The car package allows to work with Type II and Type III SS, the later (Yates' weighted squares of means) being the default under SPSS or SAS. So, if we are interested in Type II sum of squares

```
car::Anova(m, type = "II")

## Anova Table (Type II tests)
##
## Response: change
##           Sum Sq Df F value    Pr(>F)
## drug       889.78  1 40.1528 9.109e-09 ***
## center     87.14  4  0.9831 0.4209278
## drug:center 507.45  4  5.7249 0.0003761 ***
## Residuals  1994.38 90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Type III analysis is readily obtained by replacing type = "II" with type = "III" as shown below:

```
car::Anova(m, type = "III")

## Anova Table (Type III tests)
##
## Response: change
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 1777.23  1 80.2009 4.261e-14 ***
## drug       476.64  1 21.5095 1.188e-05 ***
## center     130.55  4  1.4728 0.2170607
## drug:center 507.45  4  5.7249 0.0003761 ***
## Residuals  1994.38 90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that in the case of Type III SS, we can also use the base command drop1 and we will get similar results:

```
## options(contrasts = c("contr.sum", "contr.poly"))
drop1(m, scope = ~ ., test = "F")

## Single term deletions
##
## Model:
## change ~ drug * center
##           Df Sum of Sq    RSS    AIC F value    Pr(>F)
## <none>                 1994.4 319.29
## drug           1    476.64 2471.0 338.72 21.5095 1.188e-05 ***
## center         4    130.55 2124.9 317.63  1.4728 0.2170607
## drug:center     4    507.45 2501.8 333.96  5.7249 0.0003761 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```