Reanalysis of 15-Marcato

Simon Schwab*

19/02/2021

Contents

Reference

Marcato, S., Kleinbub, J. R., Querin, G., Pick, E., Martinelli, I., Bertolin, C., Cipolletta, S., Pegoraro, E., Sorarù, G., & Palmieri, A. (2018). Unimpaired Neuropsychological Performance and Enhanced Memory Recall in Patients with Sbma: A Large Sample Comparative Study. Scientific Reports, 8(1), 13627. https://doi.org/10.1038/s41598-018-32062-5

Notes from reading methods section

- Dependant variable: PM (Prose memory test also known as Babcock story recall test)
- Independent variable: Group
  - Patients (n=64)

*University of Zurich, simon.schwab@uzh.ch

  - Controls (n=78)
- Covariate: age, education
- Design: 1-way ANCOVA with group (2 levels) as IV and age, education as covariates

Reading data

Data is loaded, reshaped if necessary, and factors are specified.

```
PATH = file.path(path.expand("~"), "Data", "ancova") # ancova project folder
data = read.csv2(file.path(PATH, "dataPrimaryStudies", "15-Marcato", "marcato_ds_export.csv"))

# rename Italian variable names
names(data)[1] = "group"
names(data)[2] = "age"
names(data)[3] = "education"
names(data)[4] = "pm"

data$group = as.factor(data$group)
data$pm = as.numeric(data$pm)

# summary(data)
```

Descriptives

Dependent variable

Number of samples and mean (SD) in levels of the independent variables. We reproduce Table 2.

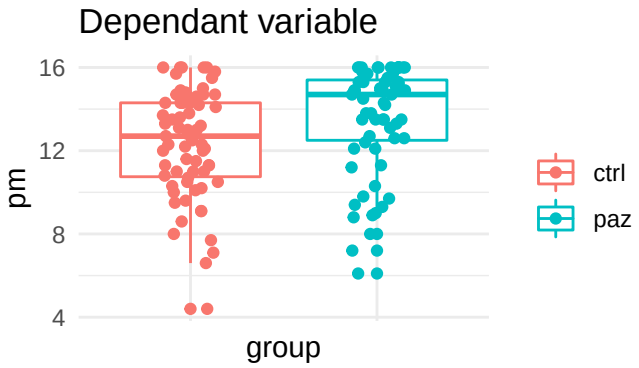
```
tab.dv = tapply(data$pm, data$group,
                  function (x) sprintf("%0.2f (%0.2f)", mean(x, na.rm = T), sd(x, na.rm = T)))
print(tab.dv)
```

```
##           ctrl            paz
## "12.35 (2.57)" "13.54 (2.58)"
```

```
ggplot(data, aes(y=pm, x=group, color=group)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  ggtitle("Dependant variable")
```

```
## Warning: Removed 12 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```



Covariate(s)

Descriptives in Table 2

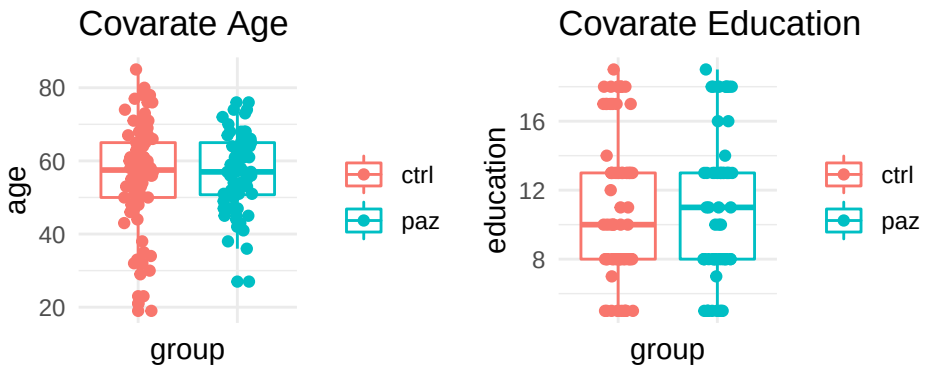
```
tab.cv = array(NA, dim=c(2,2))
tab.cv[1,] = tapply(data$age, data$group,
                    function (x) sprintf("%0.1f (%0.1f)", mean(x), sd(x)))
tab.cv[2,] = tapply(data$education, data$group,
                    function (x) sprintf("%0.1f (%0.1f)", mean(x), sd(x)))
rownames(tab.cv) = c("age", "education")
colnames(tab.cv) = levels(data$group)
print(tab.cv)
```

```
##           ctrl          paz
## age       "56.2 (14.1)" "57.3 (10.5)"
## education "11.1 (4.2)"  "11.1 (4.2)"
```

```
p1 = ggplot(data, aes(y=age, x=group, color=group)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  ggtitle("Covarate Age")

p2 = ggplot(data, aes(y=education, x=group, color=group)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  ggtitle("Covarate Education")

plot_grid(p1, p2, ncol = 2, nrow = 1)
```



Main analysis ANCOVA

```
# Orthogonal contrasts
contrasts(data$group) = contr.helmert(2)

fit.ancova = aov(pm ~ age + education + group, data = data)
# result = summary(fit) # Type I
result = Anova(fit.ancova, type=3) # Type III
print(result)
```

```
## Anova Table (Type III tests)
##
## Response: pm
##             Sum Sq  Df F value    Pr(>F)
## (Intercept) 605.03   1 94.9630 < 2.2e-16 ***
## age          15.40   1  2.4171  0.122524
## education    13.81   1  2.1673  0.143470
## group        50.42   1  7.9136  0.005694 **
## Residuals   802.78 126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing ANCOVA in original study with reanalysis

Independant variable

```
tab.IV = rbind(stats.orig.IV, stats.rep.IV)
rownames(tab.IV) = c("original Study", "reanalysis")
print(t(tab.IV))
```

```
##         original Study reanalysis
## Fvalue  "7.91"         "7.91"
## df1     "1"            "1"
## df2     "126"          "126"
## pvalue  "0.006"        "0.006"
## MD      NA             NA
## lowerCI NA             NA
## upperCI NA             NA
```

Covariate

```
tab.CV = rbind(stats.orig.CV.age, stats.rep.CV.age)
rownames(tab.CV) = c("original Study", "reanalysis")
print(t(tab.CV))
```

```
##         original Study reanalysis
## Fvalue  "2.42"         "2.42"
## df1     "1"            "1"
## df2     "126"          "126"
## pvalue  "0.123"        "0.12"
## MD      NA             NA
## lowerCI NA             NA
## upperCI NA             NA
```

```
tab.CV = rbind(stats.orig.CV.edu, stats.rep.CV.edu)
rownames(tab.CV) = c("original Study", "reanalysis")
print(t(tab.CV))
```

```
##         original Study reanalysis
## Fvalue  "2.17"         "2.17"
## df1     "1"            "1"
## df2     "126"          "126"
## pvalue  "0.143"        "0.14"
## MD      NA             NA
## lowerCI NA             NA
## upperCI NA             NA
```

Assumptions

1. Homogeneity of variance

- ANOVA/ANCOVA is fairly robust in terms of the error rate when sample sizes are equal.
- When groups with larger sample sizes have larger variances than the groups with smaller sample sizes, the resulting F-ratio tends to be conservative. That is, it's more likely to produce a non-significant result when a genuine difference does exist in the population.
- Conversely, when the groups with larger sample sizes have smaller variances than the groups with smaller samples sizes, the resulting F-ratio tends to be liberal and can inflate the false positive rate.

```
tapply(data$pm, data$group, function (x) sd(x, na.rm = TRUE))
```

```
##     ctrl      paz
## 2.566233 2.577590
```

```
leveneTest(pm ~ group, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   1  0.1503 0.6989
##       128
```

2. Independence between covariate and IV

When the covariate and the experimental effect (independent variable) are not independent the treatment effect is obscured, spurious treatment effects can arise and the interpretation of the ANCOVA is seriously compromised.

We test whether our groups differ on the CV. If the groups do not significantly differ then is appropriate to use the covariate.

```
fit.cv = aov(age ~ group, data = data)
Anova(fit.cv, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: age
##             Sum Sq  Df   F value Pr(>F)
## (Intercept) 452232   1 2856.2069 <2e-16 ***
## group           43   1    0.2744 0.6012
## Residuals    22167 140
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit.cv = aov(education ~ group, data = data)
Anova(fit.cv, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: education
##              Sum Sq  Df  F value Pr(>F)
## (Intercept) 17340.0   1 972.0522 <2e-16 ***
## group           0.0   1   0.0009 0.9758
## Residuals    2497.4 140
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Homogeneity of regression slopes

- We test the interaction between the IV and the CV

```
fit.hrs = aov(pm ~ age*group, data = data)
Anova(fit.hrs, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: pm
##              Sum Sq  Df  F value  Pr(>F)
## (Intercept) 1248.61   1 193.0037 < 2e-16 ***
## age           30.99   1   4.7907 0.03046 *
## group          6.98   1   1.0789 0.30093
## age:group      1.45   1   0.2238 0.63698
## Residuals    815.14 126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit.hrs = aov(pm ~ education*group, data = data)
Anova(fit.hrs, type=3)
```

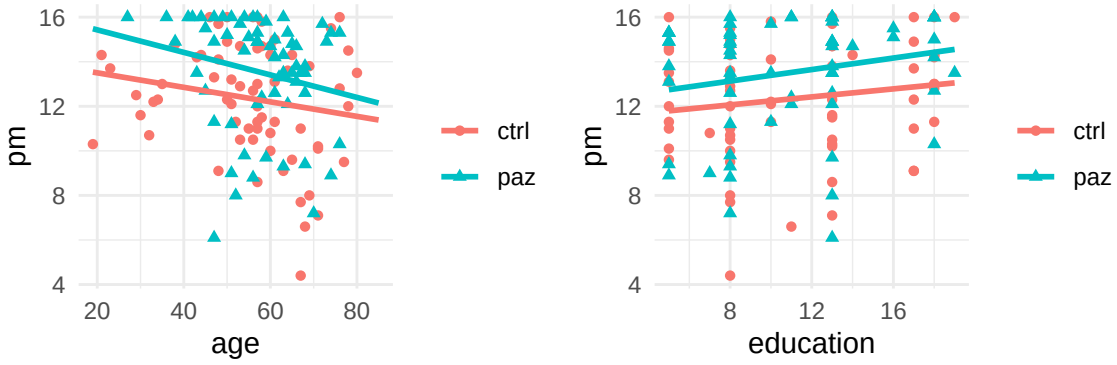
```
## Anova Table (Type III tests)
##
## Response: pm
##                  Sum Sq  Df  F value  Pr(>F)
## (Intercept)     2327.33   1 358.8273 < 2e-16 ***
## education         29.07   1   4.4823 0.03621 *
## group              2.39   1   0.3677 0.54533
## education:group    0.95   1   0.1468 0.70223
## Residuals        817.23 126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
p1 = ggplot(data, aes(y=pm, x=age, color=group, shape=group)) +
  geom_point() +
  geom_smooth(formula = y ~ x,method=lm, se=FALSE, fullrange=TRUE) +
  theme_minimal() +
  theme(legend.title = element_blank())

p2 = ggplot(data, aes(y=pm, x=education, color=group, shape=group)) +
  geom_point() +
  geom_smooth(formula = y ~ x,method=lm, se=FALSE, fullrange=TRUE) +
  theme_minimal() +
  theme(legend.title = element_blank())

plot_grid(p1, p2, ncol = 2, nrow = 1)
```

```
## Warning: Removed 12 rows containing non-finite values (stat_smooth).
## Warning: Removed 12 rows containing missing values (geom_point).
## Warning: Removed 12 rows containing non-finite values (stat_smooth).
## Warning: Removed 12 rows containing missing values (geom_point).
```



Notes

- 12 missings in pm not mentioned in the paper.
- The first reported ANCOVA was reproduced
- F-values, dfs and p-value were completely reported, also for covariates
- Assumptions were all met
- There were 4 other outcomes, and in total 5 ANCOVAs were performed. Multiplicity was not addresses.

Data was analyzed according to recommendations by Field, Miles, & Field (2012).