

25-VanLeijsen

Simon Schwab, Audrey Yeo

11/13/2019

Reference

Van Leijsen et al. (2017). Plasma AB (Amyloid-B) Levels and Severity and Progression of Small Vessel Disease. Stroke, 16(5), 351–359. <https://doi.org/10.1161/STROKEAHA.117.019810>

We first load the appropriate packages

Notes from reading methods section

- Dependant variable: AB38
- The independant variables
- 1 or more microbleeds (n = 81)
- no microbleed (n=405)
- 1 or more lacunes (n = 132)
- no lacunes (n = 355)
- Covariate: age, sex and hypertension

```
stats.orig.IV = data.frame(  
  Fvalue = NA,  
  df1 = NA,  
  df2 = NA,  
  pvalue = 0.01, # for microbleed presence and for lacune presence groups  
  MD = NA,  
  lowerCI = NA,  
  upperCI = NA)  
  
stats.orig.CV = stats.orig.IV
```

Reading data

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

```
data = read_excel("../results/data/25-VanLeijsen/VanLeijsen-1.xlsx")  
data = data.frame(data)  
str(data)  
  
## 'data.frame': 487 obs. of 9 variables:  
## $ mb_presence_b : num 1 0 0 0 1 1 1 0 1 0 ...  
## $ lac_presence_b: num 1 1 1 1 1 1 1 1 1 0 ...  
## $ AB38 : num 23.7 29.7 26.7 33.5 16.6 38.4 27.1 31.8 24.4 30.1 ...  
## $ AB40 : num 224 225 195 243 131 ...  
## $ AB42 : num 55.8 63.6 63.9 62.3 57.4 72.1 64.8 62.7 60.2 64 ...  
## $ age : num 73.8 69 74.3 64 71.7 ...  
## $ sex : num 1 2 1 1 1 1 1 1 1 1 ...  
## $ hypertension : num 1 1 1 1 1 1 1 0 1 ...
```

```
## $ tbv_b          : num  955 1111 1069 1053 1025 ...

data$mb_presence_b = data$mb_presence_b == 1
data$lac_presence_b = data$lac_presence_b == 1
data$sex.factor = NA
data$sex.factor[data$sex == 1] = "female"
data$sex[data$sex == 2] = "0"
data$sex.factor[data$sex == 0] = "male"
data$hypertension.factor = NA
data$hypertension = data$hypertension == 1

data$Group.factor = NA # only used for table
data[data$mb_presence_b == TRUE,]$Group.factor = "microbleed"
data[data$mb_presence_b == FALSE,]$Group.factor = "no microbleed"
data[data$lac_presence_b == TRUE,]$Group.factor = "lacunes"
data[data$lac_presence_b == FALSE,]$Group.factor = "no lacunes"
unique(data$Group.factor)

## [1] "lacunes"      "no lacunes"

data$Group.factor = factor(data$Group.factor,
                           levels = c("no microbleed", "microbleed",
                                         "no lacunes", "lacunes"))
```

Descriptives

Number of samples and mean (SD) in levels of the independant variables. We reproduce the mean and sd values of Table 2 of this study

```
a = sprintf("%.1f (%.1f)", mean(data$AB38[data$mb_presence_b == 0]), sd(data$AB38[data$mb_presence_b == 0]))
b = sprintf("%.1f (%.1f)", mean(data$AB38[data$mb_presence_b == 1]), sd(data$AB38[data$mb_presence_b == 1]))

c = sprintf("%.1f (%.1f)", mean(data$AB38[data$lac_presence_b == 0]), sd(data$AB38[data$lac_presence_b == 0]))
d = sprintf("%.1f (%.1f)", mean(data$AB38[data$lac_presence_b == 1]), sd(data$AB38[data$lac_presence_b == 1]))

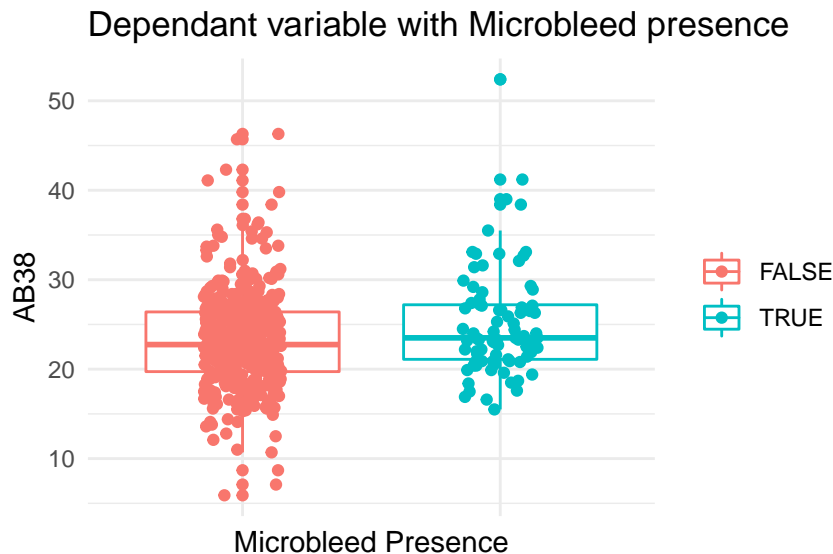
idx = c(1, 3, 2, 4) # sorting as in publication
tab.dv = array(NA, dim=c(4,3))
tab.dv[,1] = levels(data$Group.factor)
tab.dv[,2] = c("405", "81", "355", "132")
tab.dv[,3] = c(a, b, c, d)
colnames(tab.dv) = c("group", "n", "mean (SD)")
print(tab.dv)

##      group      n    mean (SD)
## [1,] "no microbleed" "405" "23.2 (5.5)"
## [2,] "microbleed"    "81"  "25.2 (6.1)"
## [3,] "no lacunes"    "355" "23.1 (5.5)"
## [4,] "lacunes"       "132" "24.7 (5.9)"
```

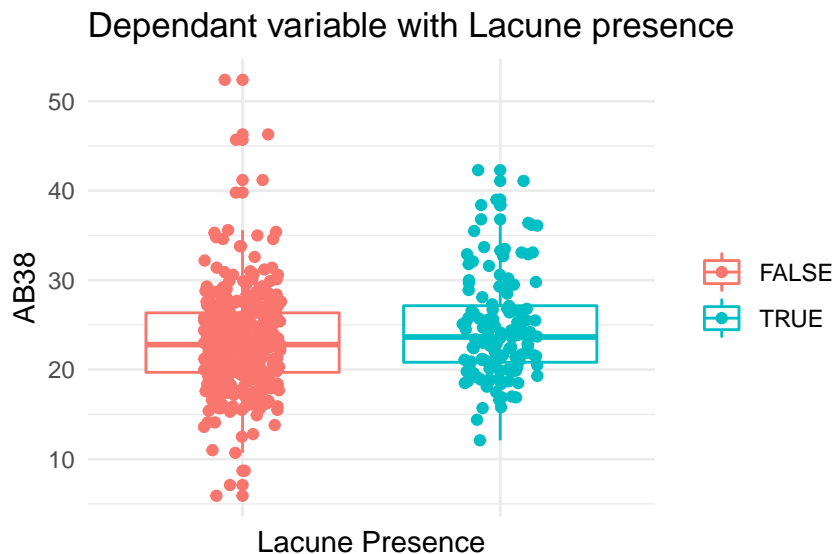
boxplot with DV

- Upon visual inspection, each level of each independant group seem to have a similar outcome effect.

```
#IV : mb_presence_b
ggplot(data, aes(y=AB38, x = mb_presence_b, color =mb_presence_b )) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  xlab("Microbleed Presence") + ylab("AB38") +
  ggtitle("Dependant variable with Microbleed presence")
```



```
#IV : lac_presence_b
ggplot(data, aes(y=AB38, x = lac_presence_b, color =lac_presence_b )) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  xlab("Lacune Presence") + ylab("AB38") +
  ggtitle("Dependant variable with Lacune presence")
```



Descriptives

COV with boxplot

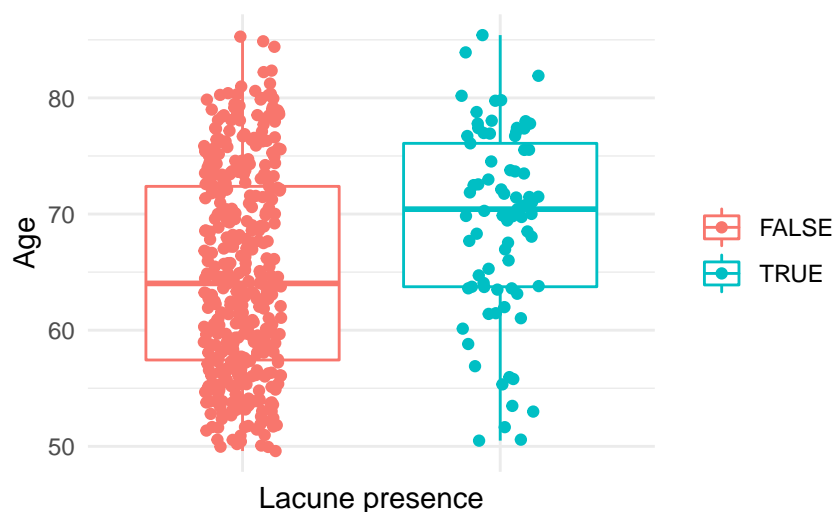
```
e = sprintf("%.1f (%.1f)", mean(data$age[data$mb_presence_b == 0]),
            sd(data$age[data$mb_presence_b == 0]))
f = sprintf("%.1f (%.1f)", mean(data$age[data$mb_presence_b == 1]),
            sd(data$age[data$mb_presence_b == 1]))
g = sprintf("%.1f (%.1f)", mean(data$age[data$lac_presence_b == 0]), sd(data$age[data$lac_presence_b == 0]))
h = sprintf("%.1f (%.1f)", mean(data$age[data$lac_presence_b == 1]), sd(data$age[data$lac_presence_b == 1]))

idx = c(1, 3, 2, 4) # sorting as in publication
tab.cv = array(NA, dim=c(4,3))
tab.cv[,1] = levels(data$Group.factor)
tab.cv[,2] = c("405", "81", "355", "132")
tab.cv[,3] = c(e,f,g,h)
colnames(tab.cv) = c("group", "n", "mean (SD)")
print(tab.cv)
```

```
##      group          n    mean (SD)
## [1,] "no microbleed" "405" "64.9 (8.7)"
## [2,] "microbleed"    "81"  "69.4 (8.1)"
## [3,] "no lacunes"    "355" "64.4 (8.6)"
## [4,] "lacunes"       "132" "69.0 (8.2)"
```

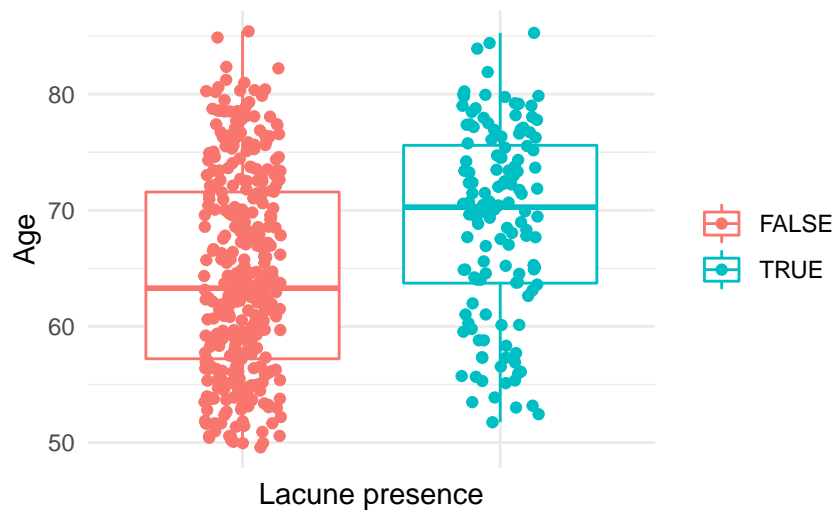
- Age upon visual inspection is similar between groups

```
ggplot(data,
        aes(y=age, x=mb_presence_b, color=mb_presence_b)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme(legend.title = element_blank()) + theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  labs(x = "Lacune presence", y = "Age", title = "" )
```



```
ggplot(data,
        aes(y=age, x=lac_presence_b, color=lac_presence_b)) +
  geom_boxplot() +
```

```
geom_point(position = position_jitter(width = 0.15, height = 0)) +
theme(legend.title = element_blank()) + theme_minimal() +
theme(axis.text.x = element_blank(), legend.title = element_blank()) +
labs(x = "Lacune presence", y = "Age", title = "" )
```



Main analysis ANCOVA

```
stats.orig.IV = data.frame(
  Fvalue = NA,
  df1 = NA,
  df2 = NA,
  pvalue = NA,
  MD = NA,
  lowerCI = NA,
  upperCI = NA)
```

```
stats.orig.CV = stats.orig.IV
```

enter all results from primary study here

```
stats.orig.IV$pvalue = 0.01
```

We verify the p values stated in Table 2 with respect to variables. * There is evidence of significance that Age as a covariate contribute to variance in outcome ($p = 8.037e-05$)

Orthogonal contrasts

```
contrasts(data$mb_presence_b) = contr.helmert(2)
```

```
contrasts(data$lac_presence_b) = contr.helmert(2)
```

#contrasts(data\$IV)<-cbind(c(-2,1,1), c(0,-1,1))

```
fit <- aov(AB38 ~ mb_presence_b*lac_presence_b + age + sex.factor + hypertension, data = data)
summary(fit)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## mb_presence_b	1	251	250.8	8.518	0.00368 **
## lac_presence_b	1	147	147.4	5.007	0.02571 *
## age	1	645	645.2	21.913	3.72e-06 ***

```
## sex.factor          1      0      0.0  0.000  0.98588
## hypertension        1    123    122.6  4.164  0.04185 *
## mb_presence_b:lac_presence_b  1     24     23.8  0.809  0.36883
## Residuals          480   14134    29.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
result = Anova(fit, type = 3) # this is not a balanced study
print(result)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: AB38
```

```
##              Sum Sq Df F value    Pr(>F)
## (Intercept)   1701.6  1  57.7885 1.547e-13 ***
## mb_presence_b    64.9  1   2.2049  0.13823
## lac_presence_b    4.5  1   0.1521  0.69673
## age            463.6  1  15.7452 8.354e-05 ***
## sex.factor        0.0  1   0.0005  0.98161
## hypertension    124.8  1   4.2371  0.04009 *
## mb_presence_b:lac_presence_b  23.8  1   0.8091  0.36883
## Residuals      14133.9 480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
stats.rep.IV = data.frame(Fvalue = sprintf("%.2f",result$`F value`[3]),
                          df1 = result$Df[3],
                          df2 = result$Df[4],
                          pvalue = formatPval(result$`Pr(>F)`[3]),
                          MD = NA,
                          lowerCI = NA,
                          upperCI = NA
)
```

```
stats.rep.CV = data.frame(Fvalue = sprintf("%.2f",result$`F value`[2]),
                          df1 = result$Df[2],
                          df2 = result$Df[4],
                          pvalue = formatPval(result$`Pr(>F)`[2]),
                          MD = NA,
                          lowerCI = NA,
                          upperCI = NA
)
```

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   NA              "0.15"
```

```
## df1      NA      " 1"
## df2      NA      " 1"
## pvalue   "0.01"   "0.70"
## MD       NA      NA
## lowerCI  NA      NA
## upperCI  NA      NA
```

Covariate

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

```
##      original Study reanalysis
## Fvalue NA      "2.20"
## df1    NA      " 1"
## df2    NA      " 1"
## pvalue NA      "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 sample sizes, the resulting F-ratio tends to be liberal and can inflate the false positive rate.
- In this study, there is homogeneity of variance

```
tapply(data$AB38, data$mb_presence_b, sd)
```

```
##      FALSE      TRUE
## 5.471873 6.069102
```

```
tapply(data$AB38, data$lac_presence_b, sd)
```

```
##      FALSE      TRUE
## 5.459909 5.870429
```

```
leveneTest(AB38 ~ lac_presence_b, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 1  1.3772 0.2412
##      485
```

```
leveneTest(AB38 ~ mb_presence_b, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##           Df F value Pr(>F)
## group    1  0.0912 0.7628
##           485
```

2. Independence between covariate and IV

- The IV of this study have evidence of significant effect to the plasma level AB38.

```
fit.cv = aov(AB38 ~ lac_presence_b + mb_presence_b, data = data)
summary(fit.cv)
```

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## lac_presence_b    1     256   256.33    8.312 0.00411 **
## mb_presence_b     1     142   141.91    4.602 0.03243 *
## Residuals       484   14926    30.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Homogeneity of regression slopes

- There is no evidence of significant interaction between covariate and independant variable. Thus we can assume homogeneity of regression slope

```
fit.hrs = aov(AB38 ~ age*mb_presence_b, data = data)
Anova(fit.hrs, type=3) # no evidence of interaction, there is homogeneity of IV levels across age
```

```
## Anova Table (Type III tests)
##
## Response: AB38
##           Sum Sq Df F value    Pr(>F)
## (Intercept)    581.2  1 19.6293 1.164e-05 ***
## age           528.9  1 17.8617 2.841e-05 ***
## mb_presence_b    16.5  1  0.5560  0.4562
## age:mb_presence_b  28.4  1  0.9603  0.3276
## Residuals    14301.0 483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

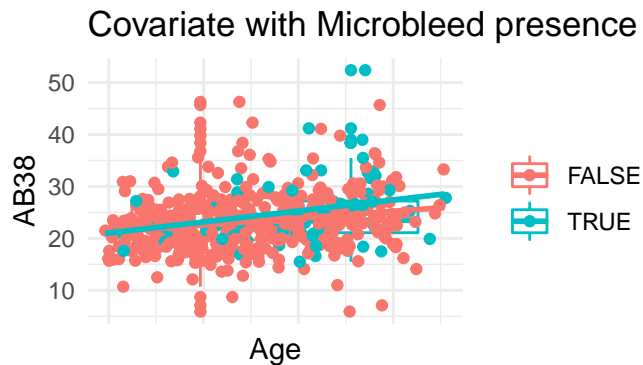
```
fit.hrs = aov(AB38 ~ age*lac_presence_b, data = data)
Anova(fit.hrs, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: AB38
##           Sum Sq Df F value    Pr(>F)
## (Intercept)   1056.2  1 35.5735 4.750e-09 ***
## age           616.0  1 20.7488 6.639e-06 ***
## lac_presence_b    4.1  1  0.1398  0.7087
## age:lac_presence_b 10.4  1  0.3490  0.5550
## Residuals    14340.2 483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Independence of Covariate with Independent Variables (Visual inspection of Homogeneity of Regression slopes)

- Visually, the two levels of each independent variable follows the same pattern - there seems to be independence of Covariate versus independent variable.

```
ggplot(data, aes(y=AB38, x= age, color= mb_presence_b)) +  
  geom_boxplot() +  
  geom_point(position = position_jitter(width = 0.15, height = 0)) +  
  geom_smooth(formula = y ~ x, method=lm, se=FALSE, fullrange=TRUE) +  
  theme_minimal() +  
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +  
  xlab("Age") + ylab("AB38") + ggtitle("Covariate with Microbleed presence")
```



```
ggplot(data, aes(y=AB38, x= age, color= lac_presence_b)) +  
  geom_boxplot() +  
  geom_point(position = position_jitter(width = 0.15, height = 0)) +  
  geom_smooth(formula = y ~ x, method=lm, se=FALSE, fullrange=TRUE) +  
  theme_minimal() +  
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +  
  xlab("Age") + ylab("AB38") + ggtitle("Covariate with Lacunes presence")
```

