# 25-VanLeijsen

*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

```
## Loading required package: survival
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:biostatUZH':
##
##     logit
```

## 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


stats.orig.allCV =  data.frame(
  Fvalue =  NA,
  df1 = NA,
  df2 = NA,
  pvalue = NA)
```

## 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 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$sex == 1] = "male"
data$sex.factor[data$sex == 2] = "female"
data$sex.factor[data$sex.factor == "male" ] = 1
data$sex.factor[data$sex.factor == "female" ] = 0
data$sex.factor = as.factor(data$sex.factor)
data$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 ==
b = sprintf("%.1f (%.1f)", mean(data$AB38[data$mb_presence_b == 1]), sd(data$AB38[data$mb_presence_b ==

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

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")
```

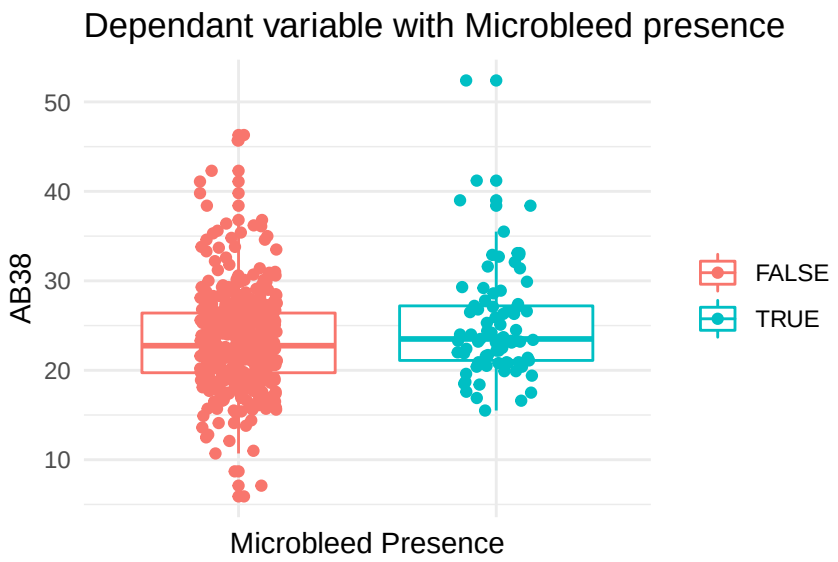
```r
tab.dv[,3] = c(a, b, c, d)
colnames(tab.dv) = c("group", "n", "mean (SD)")
print(tab.dv)
```

```
##      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)"
```

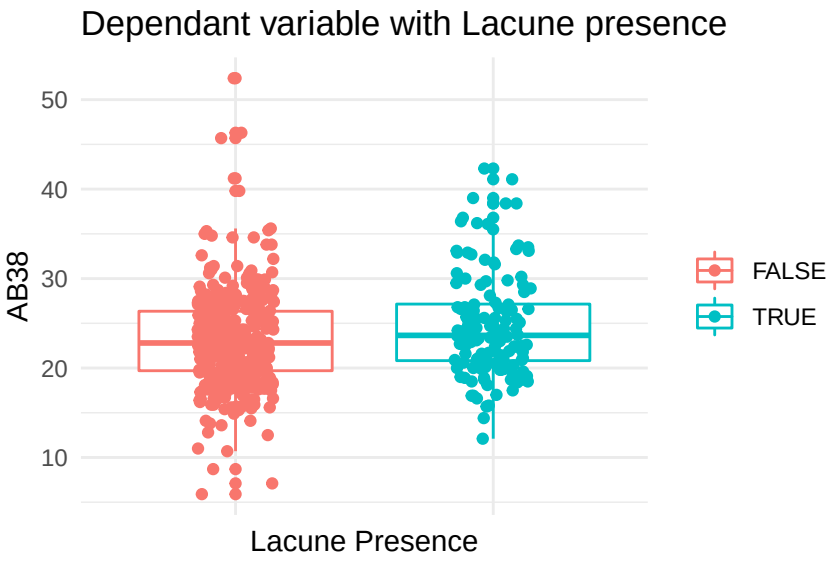
## boxplot with DV

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

```r
#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")
```




```r
#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

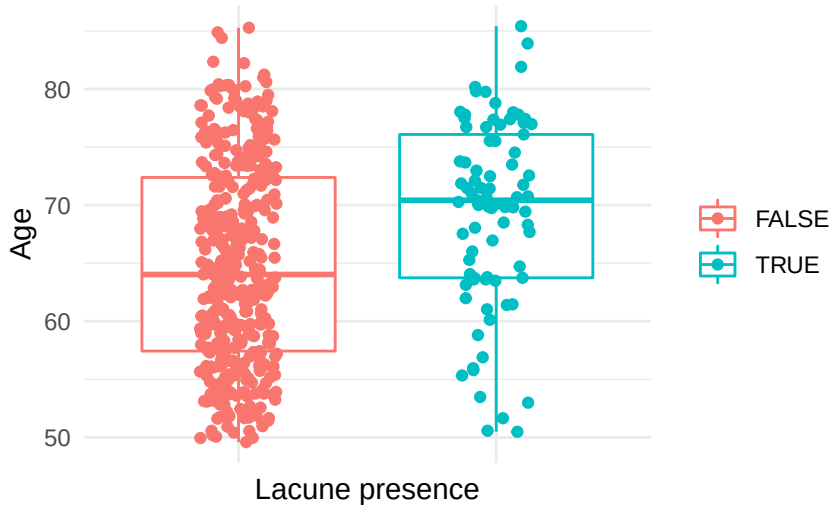
```
e = sprintf("%.1f (%.1f)", mean(data$age[data$mb_presence_b == 0]),
            sd(data$age[data$mb_presence_b == 0]))
f = sprintf("%.1f (%.1f)", mean(data$age[data$mb_presence_b == 1]),
            sd(data$age[data$mb_presence_b == 1]))
g = sprintf("%.1f (%.1f)", mean(data$age[data$lac_presence_b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

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)
```

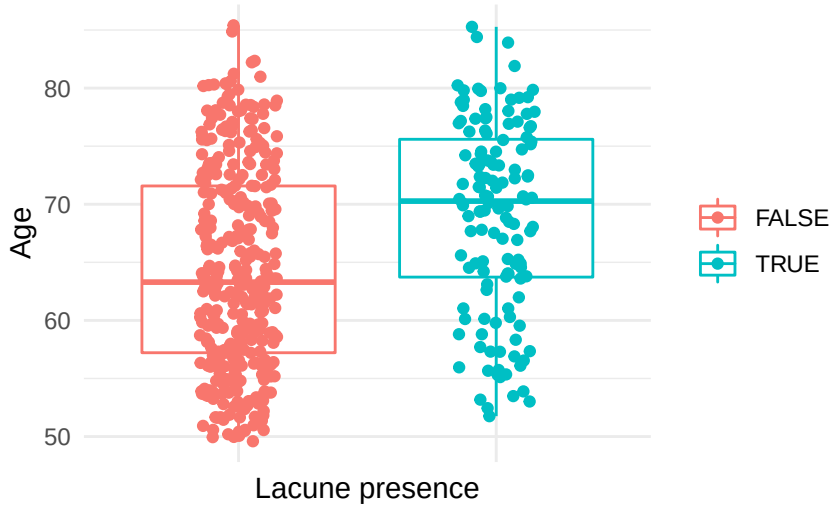
```
##      group           n     mean (SD)
## [1,] "no microbleed" "405" "64.9 (8.7)"
## [2,] "microbleed"    "81"  "69.4 (8.1)"
## [3,] "no lacunes"    "355" "64.4 (8.6)"
## [4,] "lacunes"       "132" "69.0 (8.2)"
```

- 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.IVmicrobleed = data.frame(
  Fvalue =  NA,
  df1 = NA,
  df2 = NA,
  pvalue = "<0.01")

stats.orig.IVlacunes = data.frame(
  Fvalue =  NA,
  df1 = NA,
```

```
  df2 = NA,
  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 = 3.825e-13)

```
# 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 ~ age + sex.factor + hypertension + mb_presence_b*lac_presence_b , data = data)
summary(fit) # we use interaction term because we have lacunes yes microbleeds no
```

```
##                               Df Sum Sq Mean Sq F value   Pr(>F)
## age                            1    887   887.0  30.123 6.58e-08 ***
## sex.factor                     1      0     0.5   0.016   0.8982
## hypertension                   1    147   146.9   4.990   0.0260 *
## mb_presence_b                  1    100    99.8   3.388   0.0663 .
## lac_presence_b                 1     32    31.9   1.084   0.2983
## mb_presence_b:lac_presence_b   1     24    23.8   0.809   0.3688
## 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, since ordering of variables matter, we us
print(result)
```

```
## Anova Table (Type III tests)
##
## Response: AB38
##                               Sum Sq  Df F value    Pr(>F)
## (Intercept)                   1675.1   1 56.8892 2.327e-13 ***
## 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                   64.9   1  2.2049   0.13823
## lac_presence_b                   4.5   1  0.1521   0.69673
## 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.IVmicrobleeds = data.frame(Fvalue = sprintf("%.2f",result$`F value`[5]),
                          df1 = result$Df[5],
                          df2 = result$Df[8],
                          pvalue = formatPval(result$`Pr(>F)`[5]))

stats.rep.IVlacunes = data.frame(Fvalue = sprintf("%.2f",result$`F value`[6]),
                          df1 = result$Df[6],
                          df2 = result$Df[8],
                          pvalue = formatPval(result$`Pr(>F)`[6]))

stats.rep.CVage = data.frame(Fvalue = sprintf("%.2f",result$`F value`[2]),
                          df1 = result$Df[2],
                          df2 = result$Df[8],
```

```
                          pvalue = formatPval(result$`Pr(>F)`[2]))

stats.rep.CVsex = data.frame(Fvalue = sprintf("%.2f",result$`F value`[3]),
                          df1 = result$Df[3],
                          df2 = result$Df[8],
                          pvalue = formatPval(result$`Pr(>F)`[3]))

stats.rep.CVhypertension = data.frame(Fvalue = sprintf("%.2f",result$`F value`[4]),
                          df1 = result$Df[4],
                          df2 = result$Df[8],
                          pvalue = formatPval(result$`Pr(>F)`[4]))
```

# Comparing ANCOVA in original study with reanalysis

## Independant variable

```
tab.IV = rbind(stats.orig.IVmicrobleed, stats.orig.IVlacunes, stats.rep.IVmicrobleeds, stats.rep.IVlacu
rownames(tab.IV) = c("original study microbleed", "original Study lacunes", "reanalysis for IV microble
print(t(tab.IV))
```

```
##        original study microbleed original Study lacunes
## Fvalue NA                        NA
## df1    NA                        NA
## df2    NA                        NA
## pvalue "<0.01"                   "<0.01"
##        reanalysis for IV microbleeds presence
## Fvalue "2.20"
## df1    " 1"
## df2    "480"
## pvalue "0.14"
##        reanalysis for IV lacunes presence
## Fvalue "0.15"
## df1    " 1"
## df2    "480"
## pvalue "0.70"
```

## Covariate

- The study does not report results from covariate of age

```
tab.CV = rbind(stats.orig.allCV, stats.rep.CVage, stats.rep.CVsex, stats.rep.CVhypertension)
rownames(tab.CV) = c("original CV age,sex,hypertension", "reanalysis age", "reanalysis sex", "reanalysi
print(t(tab.CV))
```

```
##        original CV age,sex,hypertension reanalysis age reanalysis sex
## Fvalue NA                               "15.75"        "0.00"
## df1    NA                               " 1"           " 1"
## df2    NA                               "480"          "480"
## pvalue NA                               "< 0.0001"     "0.98"
##        reanalysis hypertension
## Fvalue "4.24"
```

```
## df1    " 1"
## df2    "480"
## pvalue "0.04"
```

# 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.
- In this study, statistical descriptives show that the highest and lowest variances seem close. Also, there is homogeneity of variance as p values exceed 0.05 for the Levene's test

```
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 Independant variable of this study have evidence of significant effect to the covariate Age and Hypertension (for Lacune presence only). The assumption thus does not hold for Age and for Hypertension for the lacunes group which may influence the interpretation of the results from the ANCOVA analysis. The variable sex is independant to the independant variable and is an appropriate covariate to include.
- As a way of comparison, we perform a two way anova to assess the main effect of two independant variables without covariates and find that there is evidence of significant of influence of Independant variables to Dependant variable.

```
fit.cvage = aov(age ~ lac_presence_b + mb_presence_b, data = data)
summary(fit.cvage)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## lac_presence_b   1   2039  2039.3  28.649 1.34e-07 ***
## mb_presence_b    1    638   637.9   8.962   0.0029 **
## Residuals      484  34453    71.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit.cvsex = glm(sex.factor ~ lac_presence_b + mb_presence_b, family = binomial, data = data)
summary(fit.cvsex)
```

```
##
## Call:
## glm(formula = sex.factor ~ lac_presence_b + mb_presence_b, family = binomial,
##     data = data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.438  -1.261   0.995   1.096   1.096
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.39432    0.12815   3.077  0.00209 **
## lac_presence_b1  0.12588    0.10835   1.162  0.24532
## mb_presence_b1   0.07466    0.12967   0.576  0.56476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 665.32  on 486  degrees of freedom
## Residual deviance: 663.09  on 484  degrees of freedom
## AIC: 669.09
##
## Number of Fisher Scoring iterations: 4
```

```
fit.cvhypertension = glm(hypertension ~ lac_presence_b + mb_presence_b, family = binomial, data = data)
summary(fit.cvhypertension )
```

```
##
## Call:
## glm(formula = hypertension ~ lac_presence_b + mb_presence_b,
##     family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0765  -1.5009   0.5632   0.8853   0.8853
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.3837     0.1705   8.118 4.75e-16 ***
## lac_presence_b1   0.5133     0.1426   3.599  0.00032 ***
## mb_presence_b1    0.1359     0.1600   0.849  0.39573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 567.12  on 486  degrees of freedom
## Residual deviance: 548.58  on 484  degrees of freedom
## AIC: 554.58
##
## Number of Fisher Scoring iterations: 4
```

```
fit.main <- aov(AB38 ~ mb_presence_b + lac_presence_b, data = data) #checking DV and IV alone
summary(fit.main)
```

```
##                 Df Sum Sq Mean Sq F value  Pr(>F)
## mb_presence_b    1    251  250.81   8.133 0.00453 **
## lac_presence_b   1    147  147.43   4.781 0.02926 *
## 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.

```
#alternatively Anova(aov(AB38 ~age*mb_presence_b + age + mb_presence_b, data = data), type = 3)
fit.hrs_mic = aov(AB38 ~ age*mb_presence_b, data = data)
Anova(fit.hrs_mic, 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
```

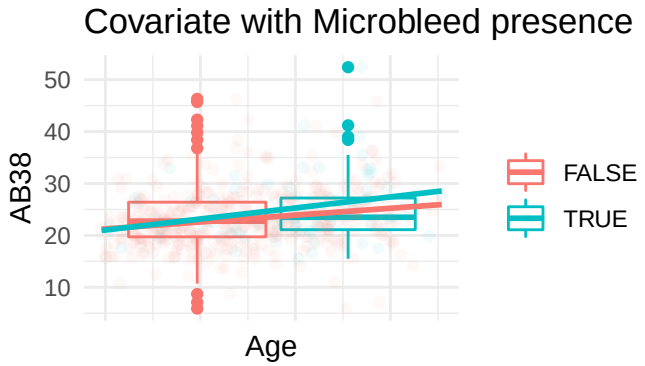
```
fit.hrs_lac = aov(AB38 ~ age*lac_presence_b, data = data)
Anova(fit.hrs_lac, type=3)
```

```
## 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
```

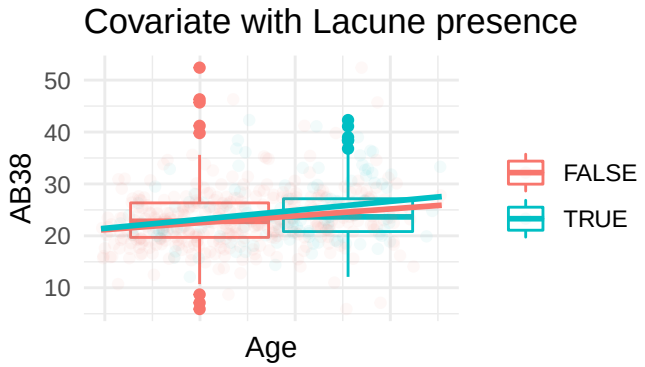
## Independance of Covariate with Independant Variables (Visual inspection of Homogeneity of Regression slopes)

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

```
ggplot(data, aes(y=AB38, x= age, color= mb_presence_b)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0), alpha = 0.05) +
  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), alpha = 0.05) +
  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 Lacune presence")
```




Reference : Field, Miles & Miles (2012), Discovering Statistics, Published by Sage Pub, UK.