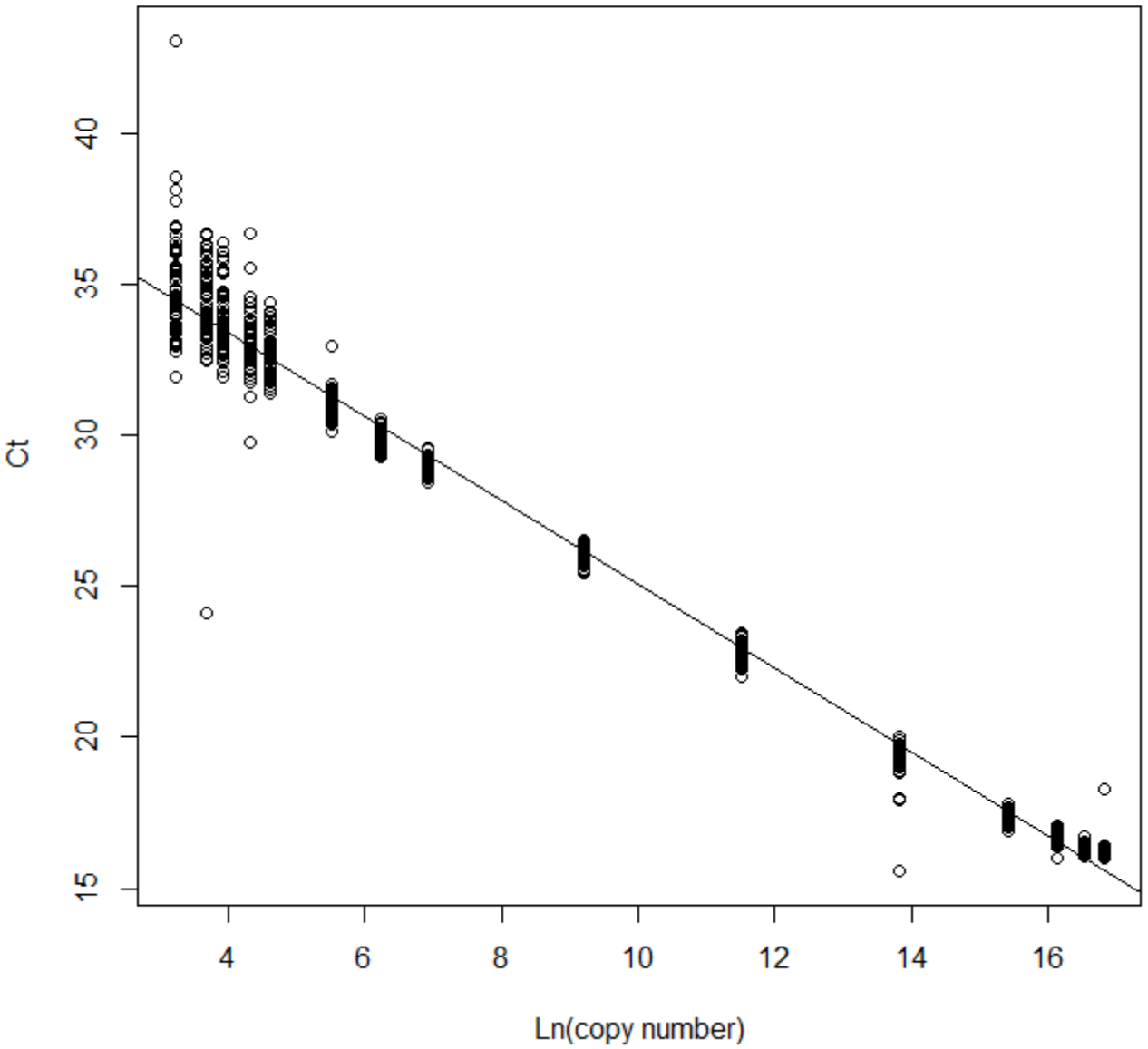
Exercise 1

- The file hiv.csv contains data on an HIV PCR assay calibration. These are dilutions of ten samples at 15 copy numbers from 25 to 20,000,000.
- In theory, the Ct value (Target in the data set) should be linear in log copy number. Fit the calibration line and look at plots to examine the assumptions of linear regression.
- What is the estimated copy number for an unknown if Ct = 25?
- The column QS is the Ct value for an in-tube standard. Consider calibrating Ct(Target) – Ct(Standard) instead. Does this work better or not? What is a good criterion?

```
> hiv.lm <- lm(Target ~ log(Nominal),data=hiv)
> summary(hiv.lm)

Call:
lm(formula = Target ~ log(Nominal), data = hiv)

Residuals:
    Min      1Q  Median      3Q     Max
-9.7150 -0.4416 -0.1037  0.3057  8.6227

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.930434   0.061413   633.9   <2e-16 ***
log(Nominal) -1.385832   0.005831  -237.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8828 on 878 degrees of freedom
Multiple R-squared:  0.9847,    Adjusted R-squared:  0.9847
F-statistic: 5.649e+04 on 1 and 878 DF,  p-value: < 2.2e-16
```

```
> anova(hiv.lm)
Analysis of Variance Table

Response: Target
              Df Sum Sq Mean Sq F value    Pr(>F)
log(Nominal)   1  44026   44026   56489 < 2.2e-16 ***
Residuals    878    684       1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

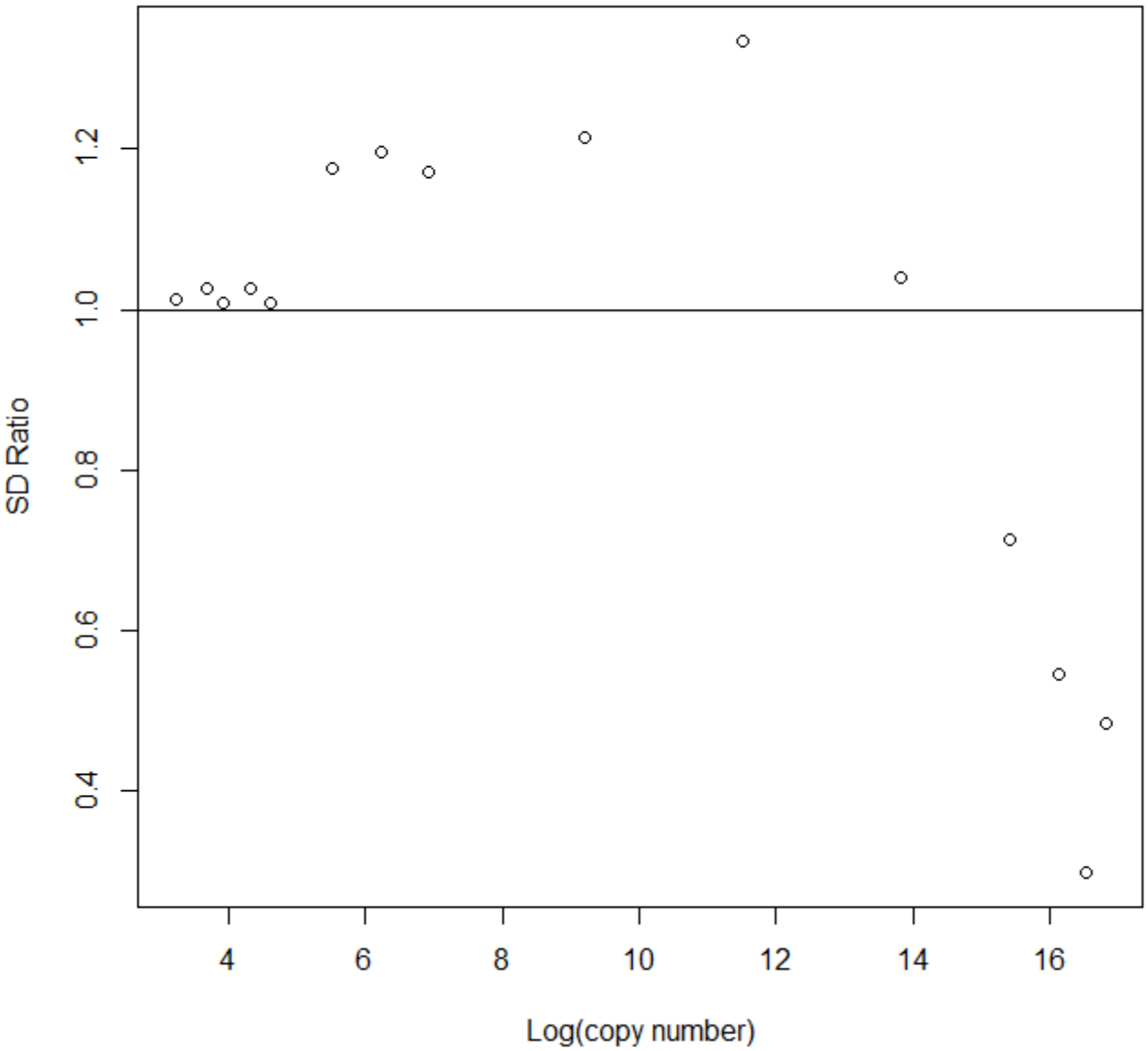

Ct
40
35
30
25
20
15
4
6
8
10
12
14
16
Ln(copy number)

Calibration Results

- Variance is not constant, being higher at higher Ct levels.
- Adding the standard helps unless copy number is very high (more than 1 million)
- Using the Target ~ regression, a Ct value of 25 corresponds to a log copy number of

  $(25 - 38.93)/(-1.386) = 10.05$

  Copy number = $\exp(10.05) = 23{,}167$



SD ratio
Target ~
vs
Target – QS ~

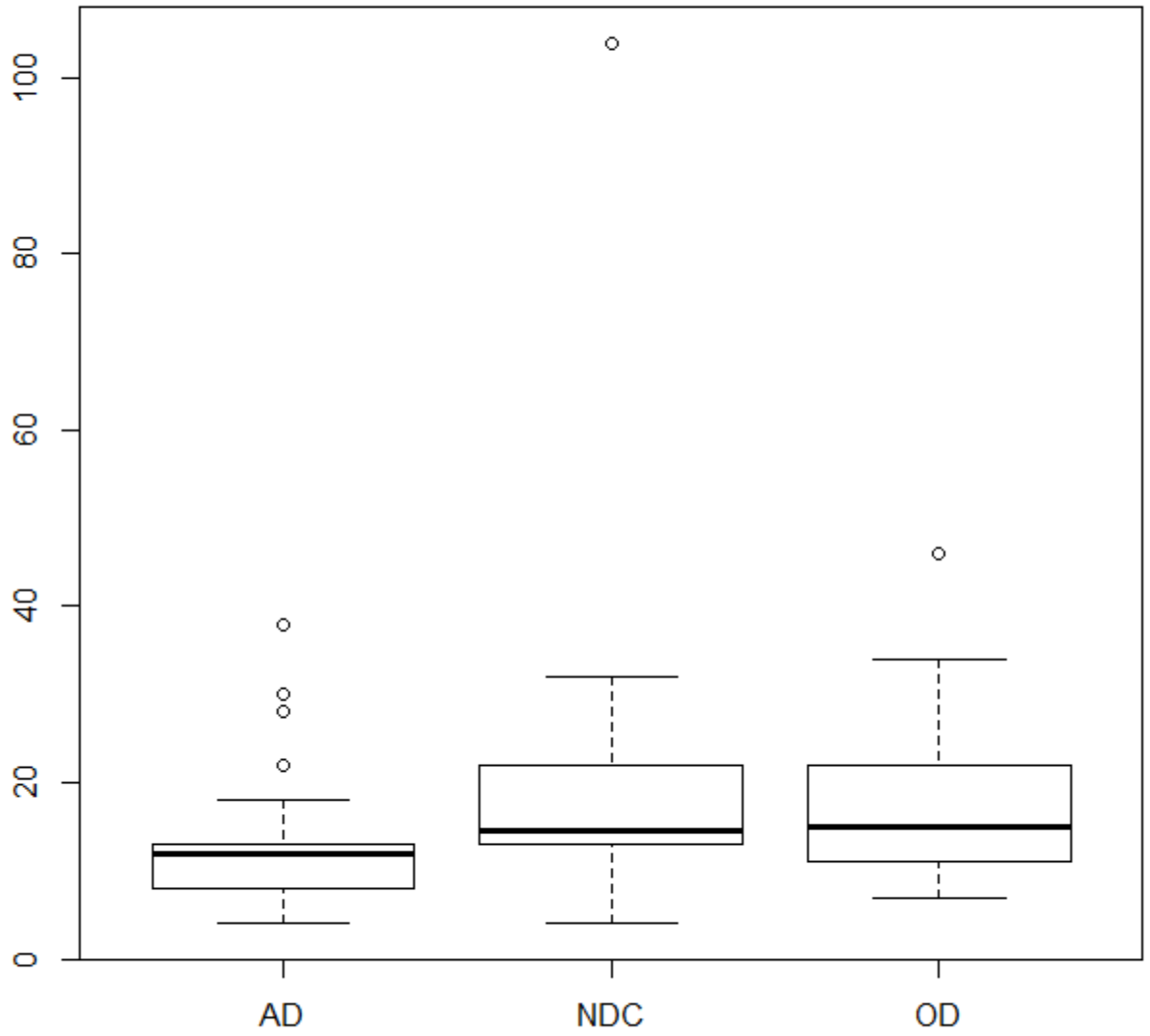
Use of standard is better when copy number > 100 and less than 5 million

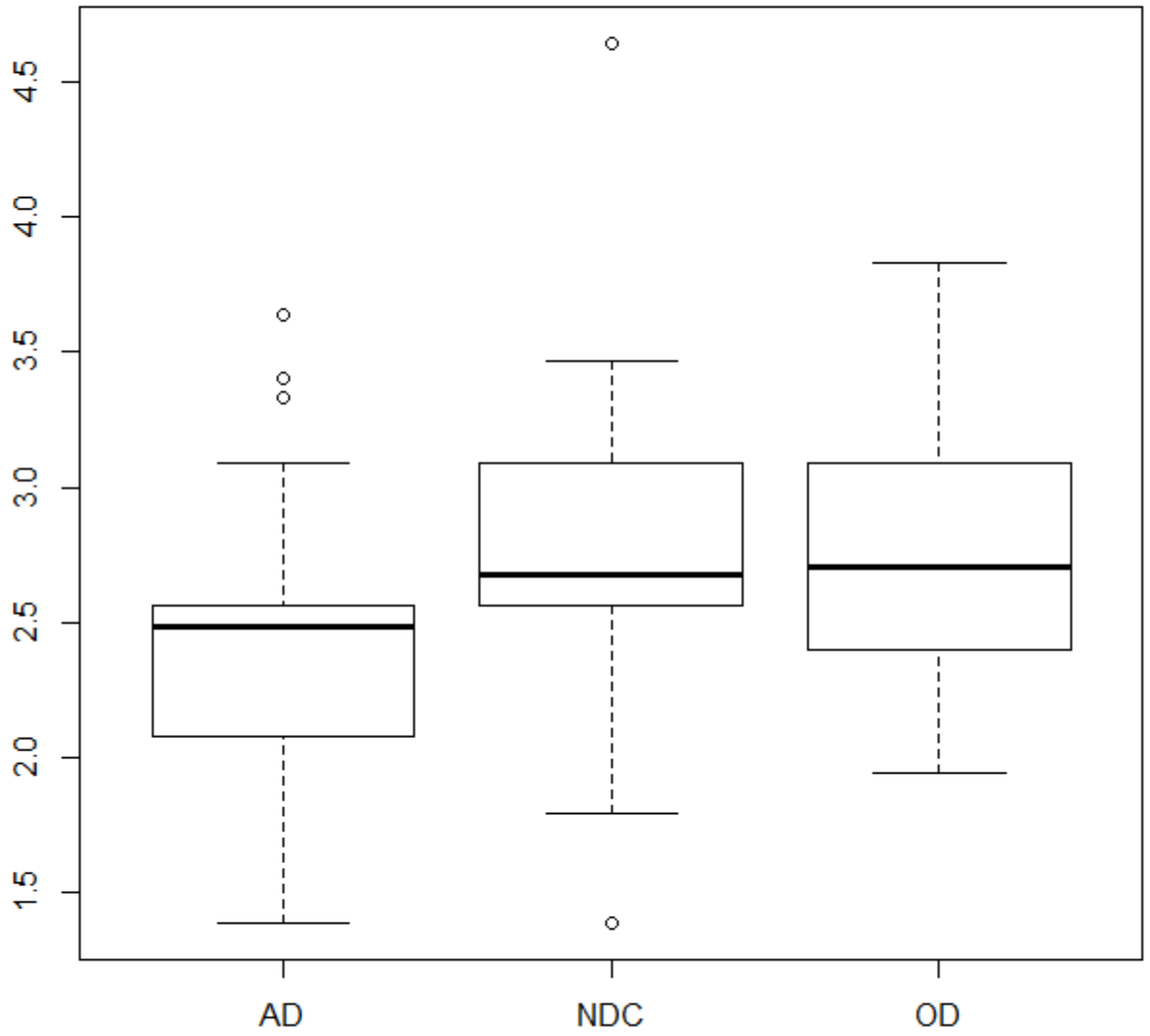
Exercise 2

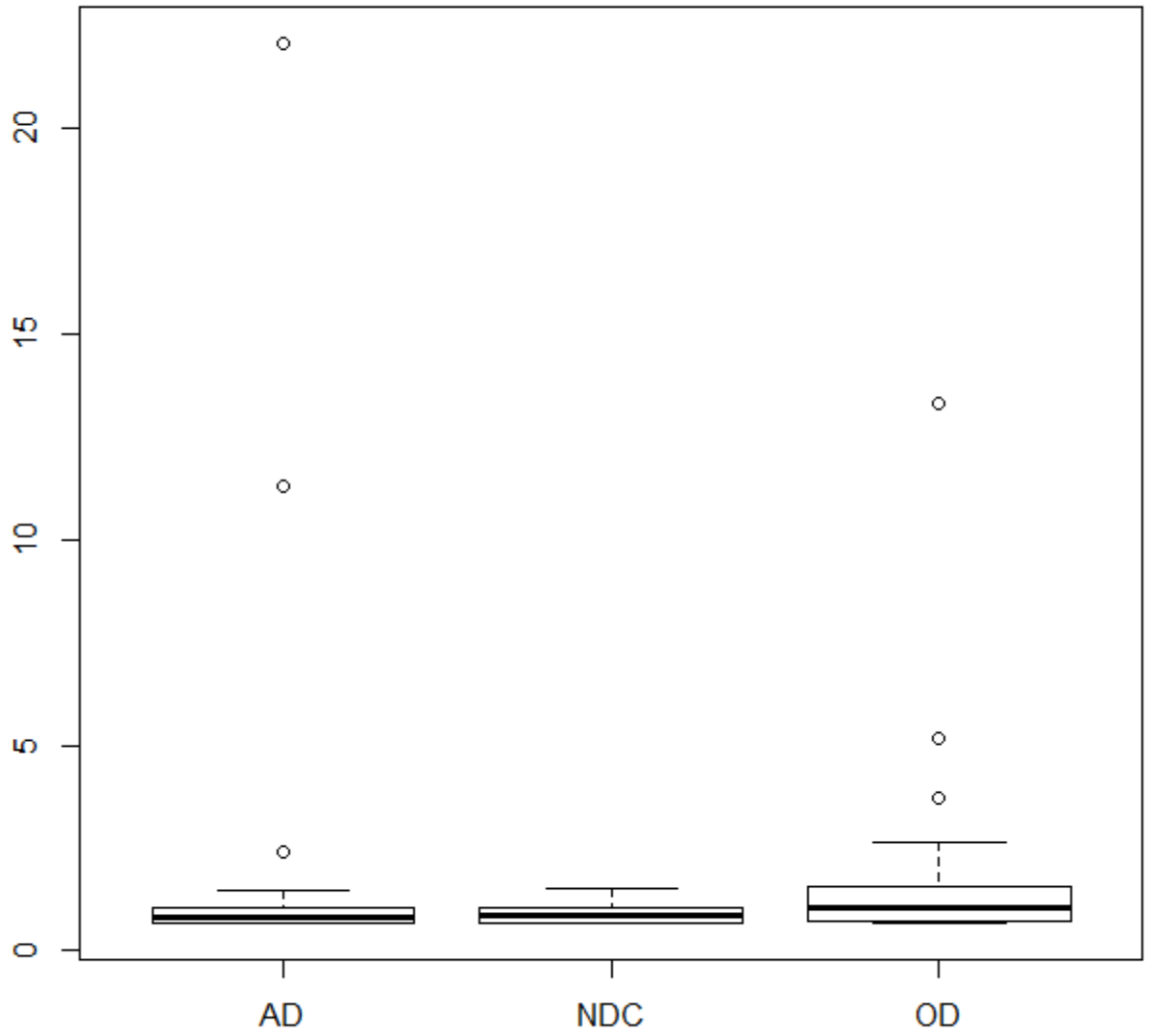
- The file AD-Luminex.csv contains Luminex protein assays for 124 proteins on 104 patients who are either AD (Alzheimer's Disease, OD (other dementia) or NDC (non-demented controls).
- See if the measured levels of ApoE are associated with diagnosis.
- See if the measured levels of IL.1beta are associated with diagnosis.

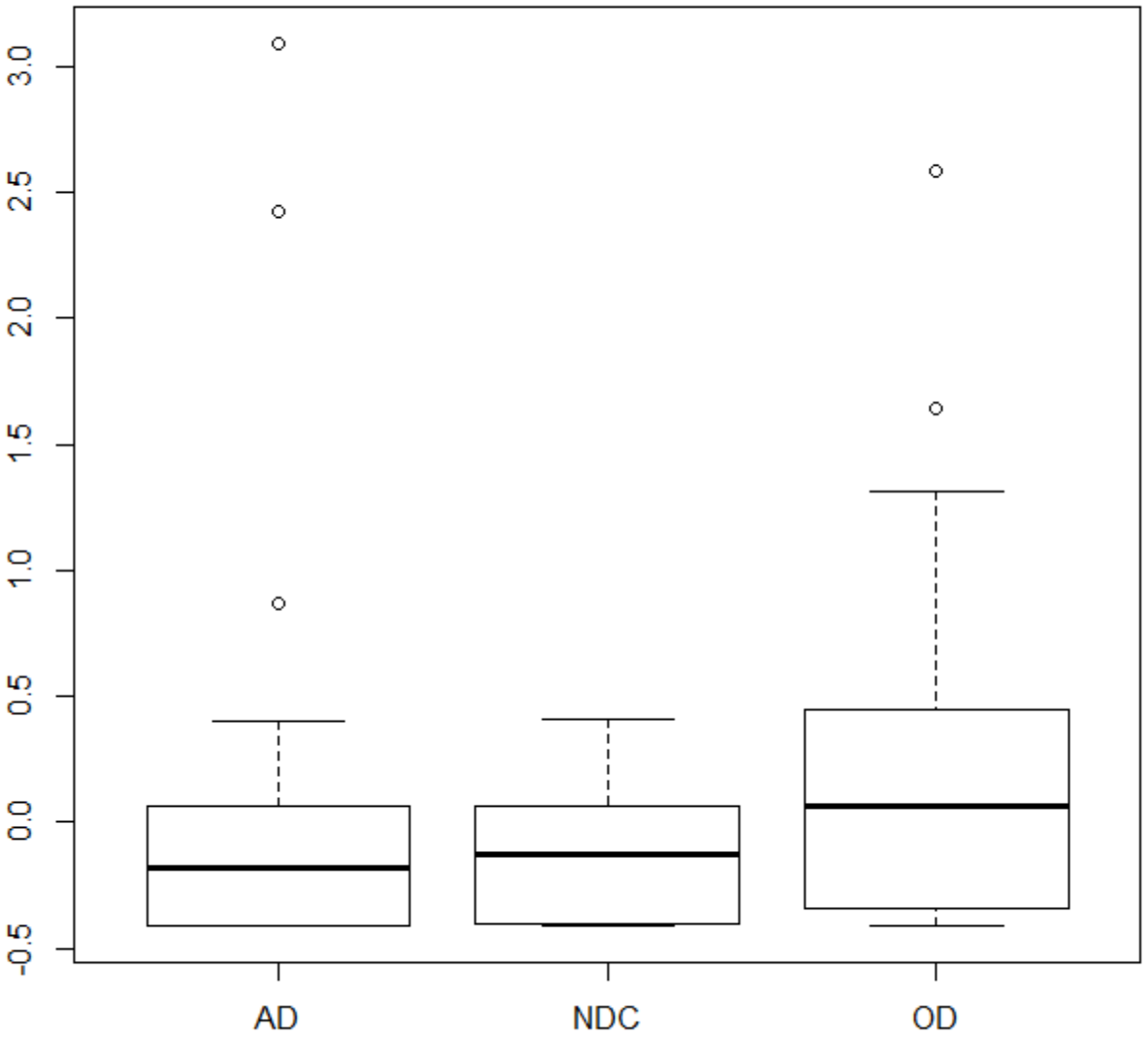
```
> ad.data <- read.csv("AD-Luminex.csv")
> anova(lm(ApoE ~ Diagnosis,data=ad.data))
Analysis of Variance Table

Response: ApoE
            Df  Sum Sq Mean Sq F value  Pr(>F)
Diagnosis    2   877.9  438.93  3.3662 0.03844 *
Residuals  101 13169.8  130.39
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(lm(log(ApoE) ~ Diagnosis,data=ad.data))
Analysis of Variance Table

Response: log(ApoE)
            Df  Sum Sq Mean Sq F value   Pr(>F)
Diagnosis    2  3.1377 1.56885  6.1503 0.003016 **
Residuals  101 25.7636 0.25508
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> anova(lm(IL.1beta ~ Diagnosis,data=ad.data))
Analysis of Variance Table

Response: IL.1beta
           Df Sum Sq Mean Sq F value Pr(>F)
Diagnosis   2  15.20  7.5992  1.0918 0.3395
Residuals 101 702.97  6.9601

> anova(lm(log(IL.1beta) ~ Diagnosis,data=ad.data))
Analysis of Variance Table

Response: log(IL.1beta)
           Df Sum Sq Mean Sq F value Pr(>F)
Diagnosis   2  1.861 0.93033  2.5717 0.0814 .
Residuals 101 36.538 0.36176
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```


0
20
40
60
80
100
AD
NDC
OD


4.5
4.0
3.5
3.0
2.5
2.0
1.5
AD
NDC
OD


0
5
10
15
20
AD
NDC
OD


3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
AD
NDC
OD

Exercise 3

- Download the 12 Affymetrix arrays from the web site
- Load the arrays into R using Read.Affy and construct the RMA expression indices