Validation of Hepatitis B viral count using R-computing

This is an example on how R-computing can be used for validation of an quantitative assay. In this case two assays for Hepatitis B viral count are compared.

Loading required package: nlme

In a summary. 'Zero' values have been changed to '1' in order to be able to plot in logaritmic scale. The lower limit of detection (LLD) at home-lab is 10 IU/ml and the LLD at the reference-lab os 20 IU/ml. So, if the result is <20IU/ml, the detected value could be anywhere between 1 and 20. Therefore, the lower limit of detection has been set for home-lab at '5 IU/ml' and '10 IU/ml' for the reference lab.

```
summary(HepB_Web)
```

```
##       PIN                Ref_lab              Home_lab
##  Min.   :14091022   Min.   :        1   Min.   :        1
##  1st Qu.:14104055   1st Qu.:      224   1st Qu.:      639
##  Median :14121724   Median :     1982   Median :     2168
##  Mean   :14116291   Mean   : 16447938   Mean   : 21548265
##  3rd Qu.:14132019   3rd Qu.:   151518   3rd Qu.:   842040
##  Max.   :14132394   Max.   :170000000   Max.   :288402140
```

```
head(HepB_Web)
```

```
##        PIN Ref_lab Home_lab
## 1 14091022       1      184
## 2 14091023    3473     3473
## 3 14104024    2976     2558
## 4 14104025     988     1001
## 5 14104026   96670 20892951
## 6 14104141 1526000  1048129
```

To make it more easy, the set of values from Reference-lab = 'x'. The set of values from Home-lab = 'y'

Calculate the means and difference between the two sets (x and y)

```
# derive difference
mean(x)
```

```
## [1] 16447938
```

```
mean(y)
```

```
## [1] 21548265
```

```
# mean Ref_lab - mean Home_lab
mean(x)-mean(y)
```
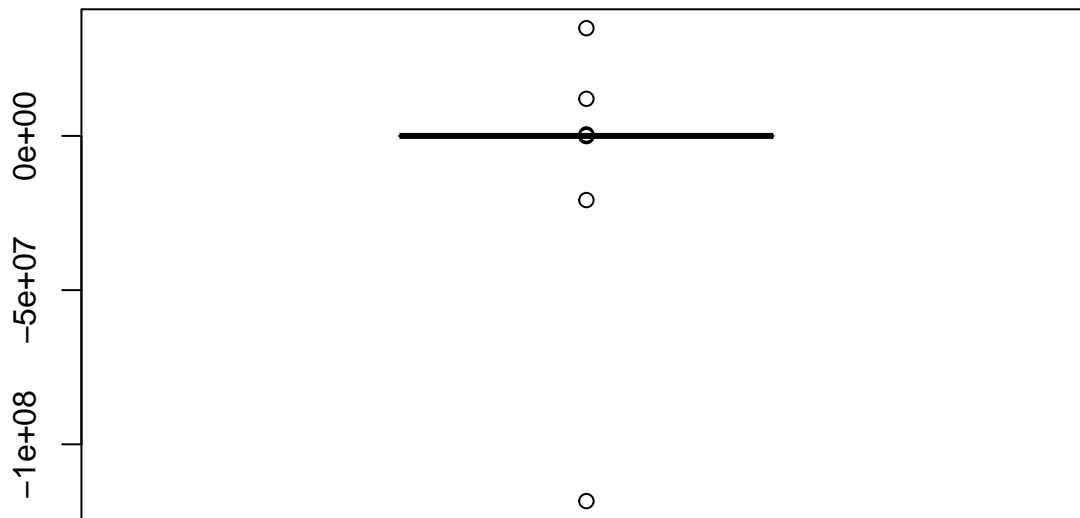
```
## [1] -5100327
```

Because n=17 is small, the distribution of the differences should be approximately normal. Check using a boxplot and QQ plot. There is some skew.

```
HepB_Web$diff <- x-y
HepB_Web$diff
```

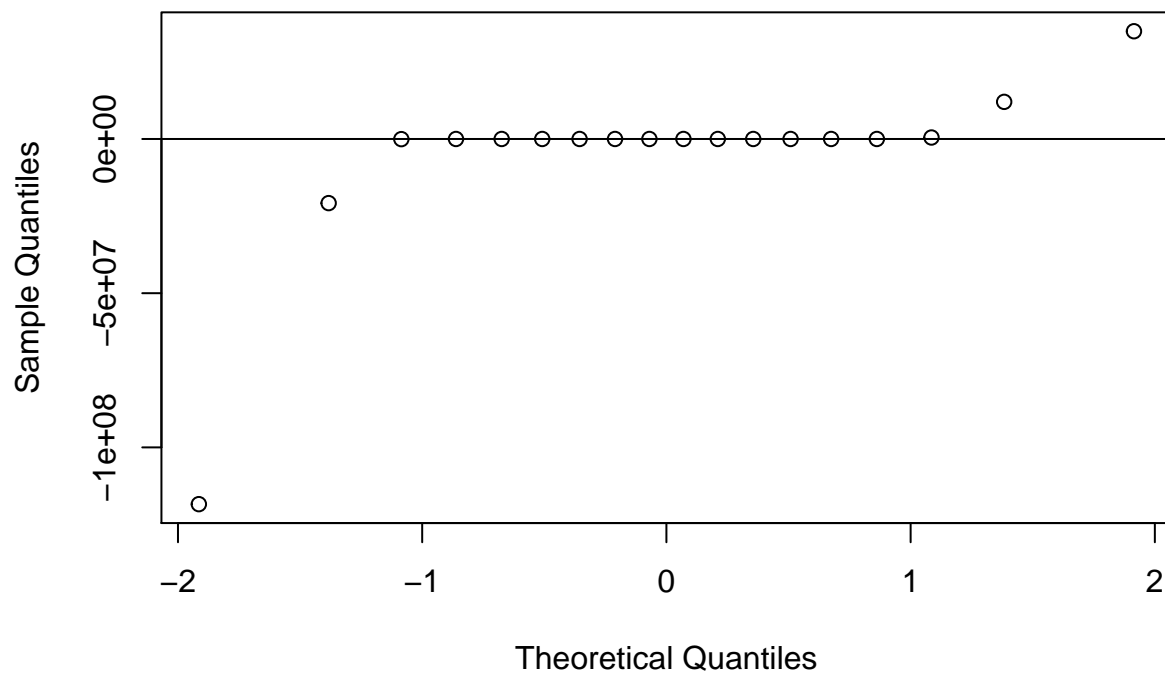```
[1]   -183   0   418   -13 -20796281   477871
[7]   77 12039815   -176 34930655 -118402140   -282
[13]   -9   -53972   -171   -1757   0   265
```

```
boxplot(HepB_Web$diff)
```



```
qqnorm(HepB_Web$diff)
qqline(HepB_Web$diff)
```

Normal Q–Q Plot

Shaphiro test of normality.

```
shapiro.test(HepB_Web$diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  HepB_Web$diff
## W = 0.47898, p-value = 5.294e-07
```

The normality test gives p < 0.003, which is small, so we reject the null hypothesis that the values are distributed normally.

This means that we cannot use the student t-test. Instead, use the Mann-Whitney-Wilcoxon Test, we can decide whether the population distributions are identical without assuming them to follow the normal distribution.

```
wilcox.test(x, y, paired = TRUE)
```

```
## Warning in wilcox.test.default(x, y, paired = TRUE): cannot compute exact
## p-value with zeroes
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x and y
## V = 59, p-value = 0.6603
## alternative hypothesis: true location shift is not equal to 0
```

$p > 0.05$ and therefore the H0 is NOT rejected. The two populations are identical.

Just to see what happens in the Student T-test. A paired t-test: one sample, two tests H0 = no difference; H1 = mean of 2 tests are different mu= a number indicating the true value of the mean (or difference in means if you are performing a two sample test).

```
t.test(x, y, mu=0, paired=T, alternative="greater")
```

```
##
##  Paired t-test
##
## data:  x and y
## t = -0.72015, df = 17, p-value = 0.7594
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -17420746        Inf
## sample estimates:
## mean of the differences
##                -5100327
```

$p = 0.759$. Because p is larger than alpha, we do NOT reject H0. In other words, it is unlikely the observed agreements happened by chance. However, because the populations do not have a normal distribution, we can not use the outcome if this test.

For correlation, three methods are used: pearson, kendall and spearman at a confidence level of 95%.

```
# correlation of the two methods
cor.test(x, y,
         alternative = c("two.sided", "less", "greater"),
         method = c("pearson", "kendall", "spearman"),
         exact = NULL, conf.level = 0.95)
```

```
##
##  Pearson's product-moment correlation
##
## data:  x and y
## t = 11.187, df = 16, p-value = 5.646e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8471778 0.9783793
```

```
## sample estimates:
##       cor
## 0.9416175
```

The correlation with the spearman test is 0.9416175. Almost perfect correlation.

Plotting the two methods using logarithmic scales.

```r
g <- ggplot(HepB_Web, aes(log(Home_lab), log(Ref_lab)))

# add layers
g +
  geom_smooth(method="lm", se=TRUE, col="steelblue", size = 1) +
  geom_point(size = 3, aes(colour = x)) +
  scale_colour_gradient("IU/ml", high = "red", low = "blue", space = "Lab") +
  labs(y = "Reference lab (log IU/ml)") +
  labs(x = "Home lab (log IU/ml)") +
  theme_bw(base_family = "Helvetica", base_size = 14) +
  scale_x_continuous(breaks=c(0,4,8,12))
```
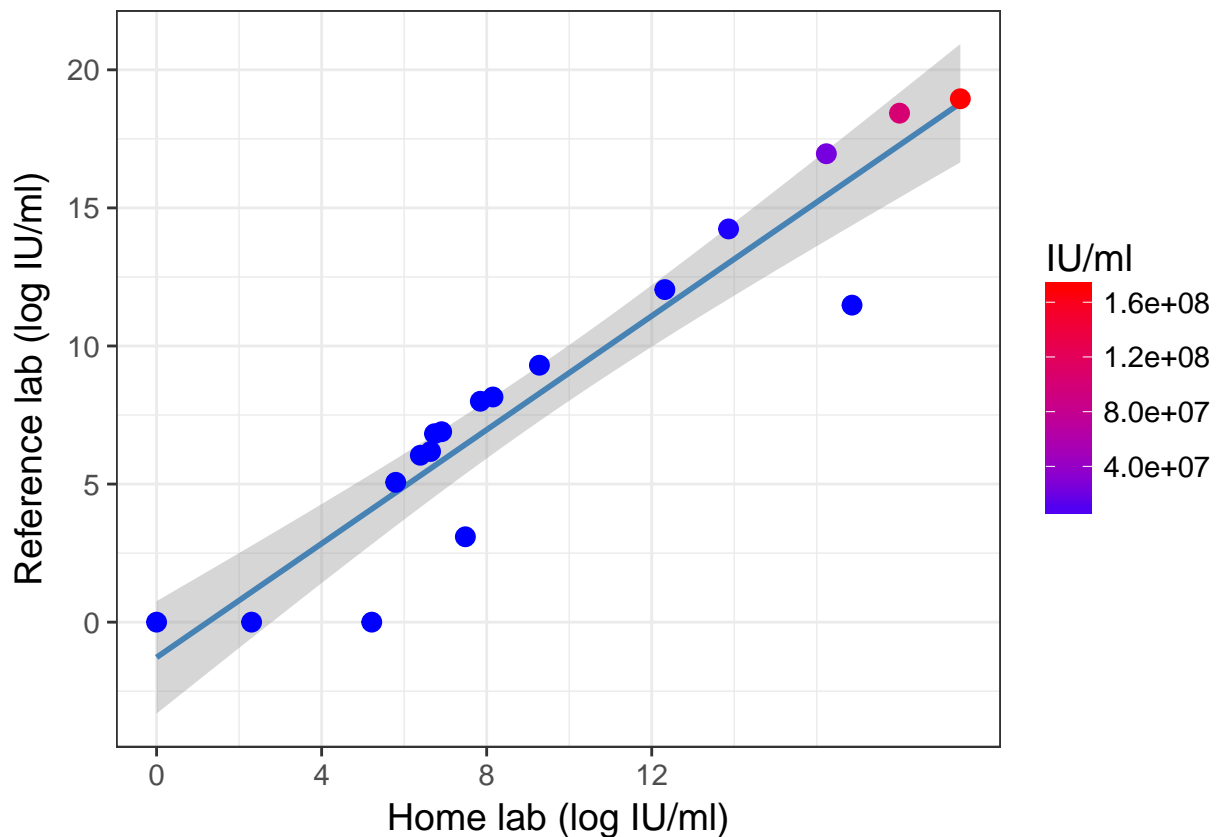

Summary data on the correlation line.

```r
regmod <- lm(y~x, data=HepB_Web)
summary(regmod)
```

```
##
## Call:
## lm(formula = y ~ x, data = HepB_Web)
##
## Residuals:
```

```
##        Min       1Q    Median       3Q       Max
## -76044958   1901358   1905277   1905580  47898082
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.905e+06  5.982e+06  -0.319    0.754
## x            1.426e+00  1.275e-01  11.187 5.65e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23770000 on 16 degrees of freedom
## Multiple R-squared:  0.8866, Adjusted R-squared:  0.8796
## F-statistic: 125.1 on 1 and 16 DF,  p-value: 5.646e-09
```

The Bland-Altman Analysis. To check if there is a bias.

```
##        Ref_lab  Home_lab        diff
## 1           1       184        -183
## 2        3473      3473           0
## 3        2976      2558         418
## 4         988      1001         -13
## 5       96670  20892951  -20796281
## 6     1526000   1048129      477871
## 7         919       842          77
## 8    23250000  11210185    12039815
## 9         421       597        -176
## 10 101000000  66069345    34930655
## 11 170000000 288402140  -118402140
## 12        483       765        -282
## 13          1        10          -9
## 14     169800    223772      -53972
## 15        158       329        -171
## 16         22      1779       -1757
## 17          1         1           0
## 18      10970     10705         265
```

```r
BlandAltman(x, y,
          x.name = "Reference lab IU/ml",
          y.name = "Home lab IU/ml",
          maintit = "Bland-Altman plot for HBV count",
          cex = 1,
          pch = 16,
          col.points = "black",
          col.lines = "blue",
          limx = NULL,
          limy = NULL,
          ymax = NULL,
          eqax = FALSE,
          xlab = NULL,
          ylab = NULL,
          print = TRUE,
          reg.line = FALSE,
          digits = 2,
          mult = FALSE)
```

```
## NOTE:
```

```
## 'AB.plot' and 'BlandAltman' are deprecated,
## and likely to disappear in a not too distant future,
## use 'BA.plot' instead.
```

Bland–Altman plot for HBV count
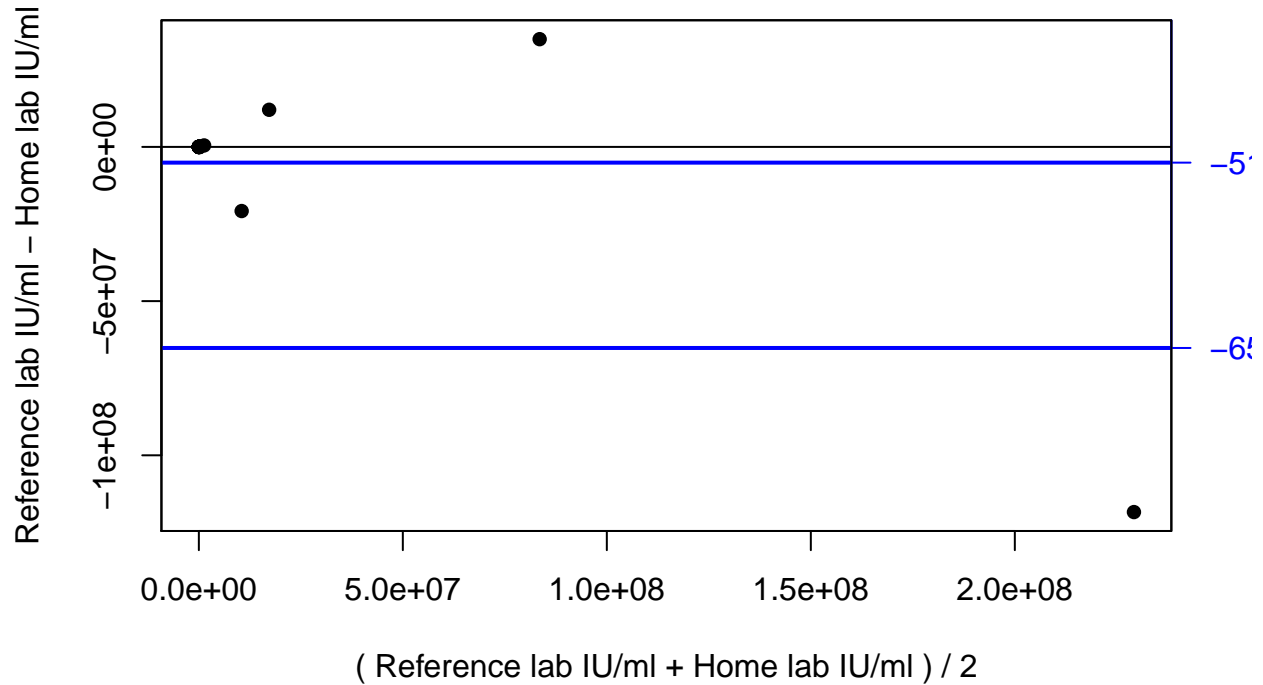

```
##
## Limits of agreement:
## Reference lab IU/ml - Home lab IU/ml                2.5% limit
##                   -5100327                         -65195645
##                97.5% limit                          SD(diff)
##                   54994992                          30047659
```

When the dots are around 0, the two test could be interchanged for a patient. So, the two test can be interchanged. There are, however, some outliners: large difference of viral count between the two labs.