

# Assignment ON2

Elmer V Villanueva

Due at 3:55 PM on Monday 09 March 2020

## Questions

```
OLD <- c(2041.7, 257.0, 524.8, 257.0, 128.8,
         128.8, 1023.3, 134.9, 257.0, 125.9,
         257.0, 123.0, 1000.0, 120.2, 128.8)
NEW <- c(12302.7, 6918.3, 4466.8, 1584.9, 933.3,
         1659.6, 9120.1, 575.4, 2630.3, 2398.8,
         1905.5, 851.1, 3467.4, 1380.4, 575.4)
TUBERCULIN <- data.frame(OLD, NEW)
str(TUBERCULIN)
```

```
## 'data.frame':   15 obs. of  2 variables:
## $ OLD: num  2042 257 525 257 129 ...
## $ NEW: num 12303 6918 4467 1585 933 ...
```

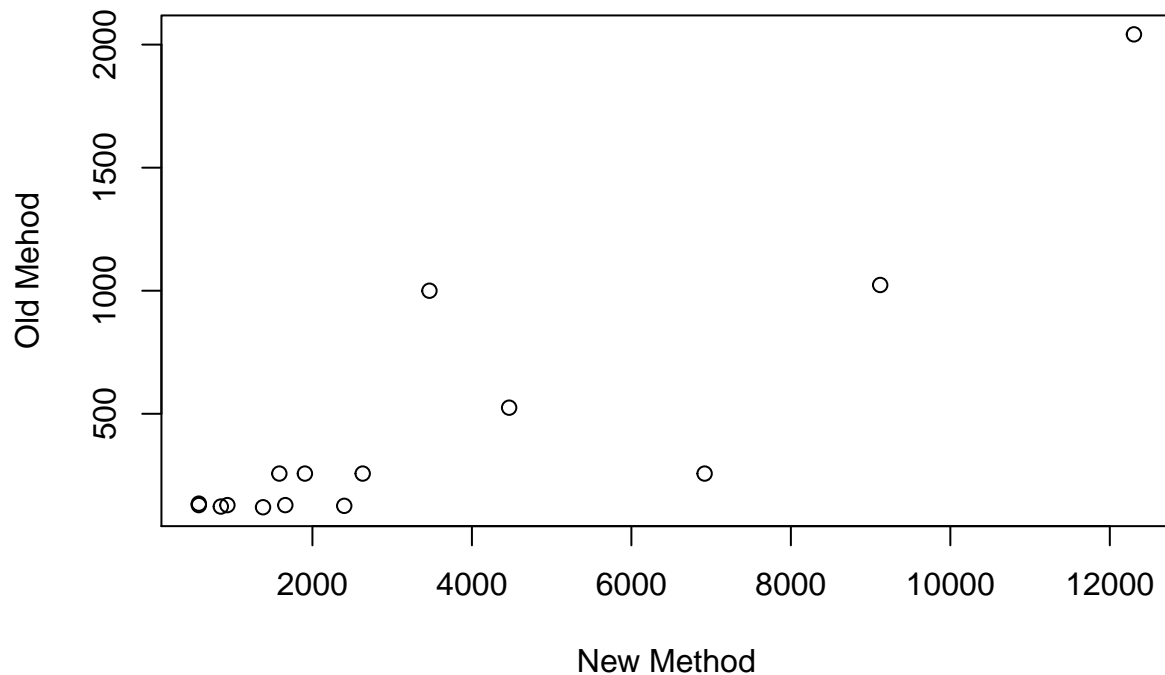
```
head(TUBERCULIN)
```

```
##      OLD      NEW
## 1 2041.7 12302.7
## 2  257.0  6918.3
## 3  524.8  4466.8
## 4  257.0  1584.9
## 5  128.8   933.3
## 6  128.8  1659.6
```

1. (10 marks) Produce a properly formatted scatterplot of the data.

```
plot(OLD ~ NEW, data = TUBERCULIN,
     main = "Estimates of tuberculin antibody levels \nunder old and new assay methods",
     ylab = "Old Mehod",
     xlab = "New Method")
```

## Estimates of tuberculin antibody levels under old and new assay methods



2. (5 marks) Regress the readings from the old method on the readings from the new method. Report estimates for the intercept and slope.

```
TUBERCULIN.LM <- lm(OLD ~ NEW, data = TUBERCULIN)
summary(TUBERCULIN.LM)
```

```
##
## Call:
## lm(formula = OLD ~ NEW, data = TUBERCULIN)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -652.41  -74.15   22.18   69.15  554.99
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -21.60568   101.69092  -0.212   0.835
## NEW          0.13457    0.02136   6.301 2.74e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 277 on 13 degrees of freedom
## Multiple R-squared:  0.7534, Adjusted R-squared:  0.7344
## F-statistic: 39.71 on 1 and 13 DF,  p-value: 2.739e-05
```

The estimated intercept is  $-21.61$ . The estimated slope is  $0.13$ .

3. (3 marks) Test the null hypothesis that the intercept is zero using  $\alpha = 0.05$ .

Under  $H_0 : b_0 = 0$ ,  $p = 0.835$ . The null hypothesis cannot be rejected.

4. (3 marks) Test the null hypothesis that the slope is zero using  $\alpha = 0.05$ .

Under  $H_1 : b_1 = 0$ ,  $p = 0.0000274$ . The null hypothesis is rejected.

5. (3 marks) Produce a 95% confidence interval for the slope.

```
confint(TUBERCULIN.LM)
```

```
##                2.5 %      97.5 %
## (Intercept) -241.2955542 198.0841887
## NEW          0.0884362   0.1807103
```

The 95% CI is (0.09, 0.18).

6. (5 marks) Interpret this 95% confidence interval.

We are 95% confident that the true mean slope lies between 0.09 and 0.18.

The readings from both methods have a wide range. It is typical in antibody concentration studies to take the logarithm of the data.

7. (5 marks) Transform the data by taking the base-10 logarithm of each datum. Present this transformation in a new, properly formatted table.

```
TUBERCULIN$OLD.LOG <- log(OLD, base = 10)
TUBERCULIN$NEW.LOG <- log(NEW, base = 10)
TUBERCULIN[, c("OLD.LOG", "NEW.LOG")]
```

```
##      OLD.LOG  NEW.LOG
## 1  3.309992 4.090000
## 2  2.409933 3.839999
## 3  2.719994 3.649997
## 4  2.409933 3.200002
## 5  2.109916 2.970021
## 6  2.109916 3.220003
## 7  3.010003 3.960000
## 8  2.130012 2.759970
## 9  2.409933 3.420005
## 10 2.100026 3.379994
## 11 2.409933 3.280009
## 12 2.089905 2.929981
## 13 3.000000 3.540004
## 14 2.079904 3.140005
## 15 2.109916 2.759970
```

The table appears below:

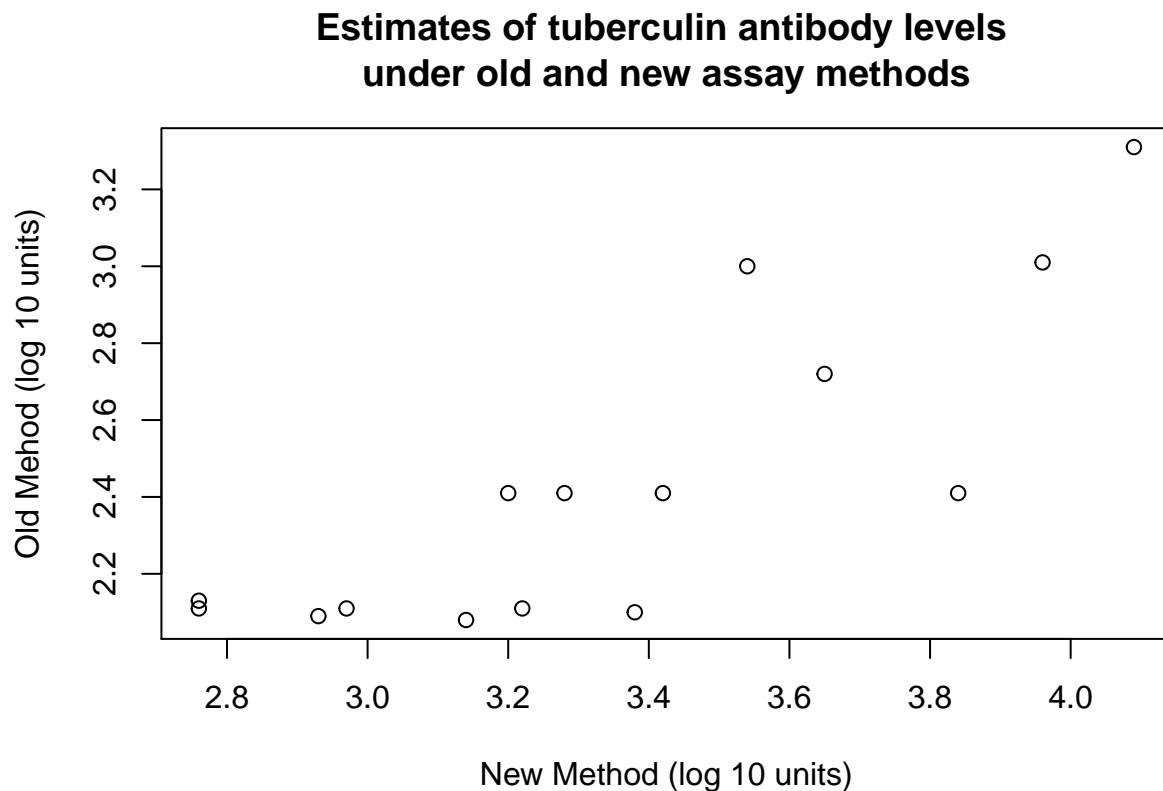
**Table. Log-transformed estimates of tuberculin antibody levels under old and new assay methods**

OLD	NEW
3.3	4.1
2.4	3.8
2.7	3.6
2.4	3.2
2.1	3.0
2.1	3.2
3.0	4.0

OLD	NEW
2.1	2.8
2.4	3.4
2.1	3.4
2.4	3.3
2.1	2.9
3.0	3.5
2.1	3.1
2.1	2.8

8. (10 marks) Produce a properly formatted scatterplot of the transformed data.

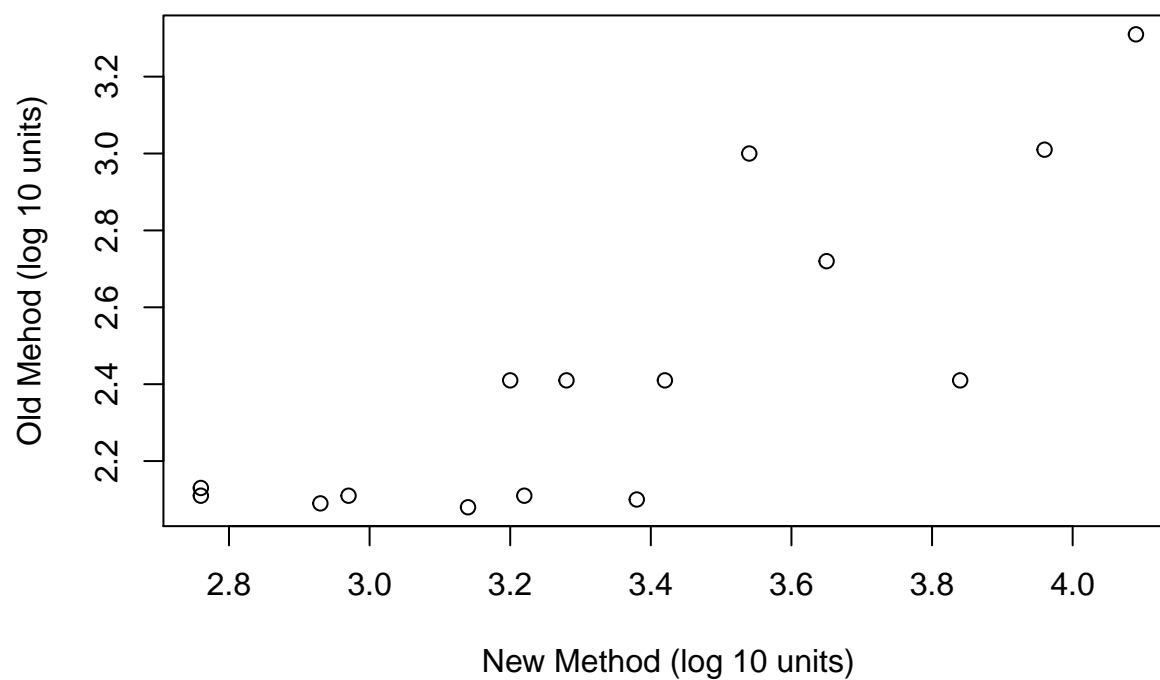
```
plot(OLD.LOG ~ NEW.LOG, data = TUBERCULIN,
     main = "Estimates of tuberculin antibody levels \nunder old and new assay methods",
     ylab = "Old Mehod (log 10 units)",
     xlab = "New Method (log 10 units)")
```



EVV Feedback. In practice, there are at least two alternative ways to include transformations. First, you don't need to create new variables, but transform on the fly. For example,

```
plot(log(OLD, base = 10) ~ log(NEW, base = 10), data = TUBERCULIN,
     main = "Estimates of tuberculin antibody levels \nunder old and new assay methods",
     ylab = "Old Mehod (log 10 units)",
     xlab = "New Method (log 10 units)")
```

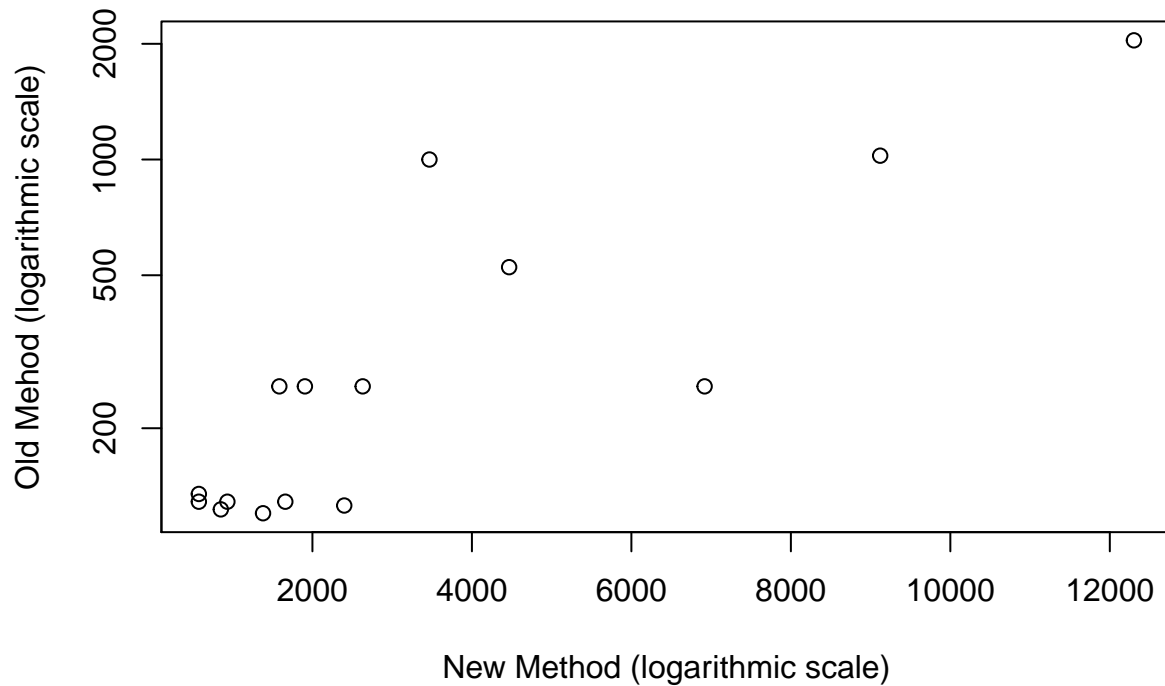
### Estimates of tuberculin antibody levels under old and new assay methods



Second, you can retain the original data, but transform the axes. For example,

```
plot(OLD ~ NEW, data = TUBERCULIN,  
     log = c("y", "x"),  
     main = "Estimates of tuberculin antibody levels \nunder old and new assay methods",  
     ylab = "Old Method (logarithmic scale)",  
     xlab = "New Method (logarithmic scale)")
```

## Estimates of tuberculin antibody levels under old and new assay methods



I leave to you to compare the unit scale and logarithmic scale graphs to show that they are, in fact, different.

9. (5 marks) Regress the transformed readings from the old method on the transformed readings from the new method. Report the regression equation.

```
TUBERCULIN.LM2 <- lm(OLD.LOG ~ NEW.LOG, data = TUBERCULIN)
summary(TUBERCULIN.LM2)
```

```
##
## Call:
## lm(formula = OLD.LOG ~ NEW.LOG, data = TUBERCULIN)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41764 -0.13194  0.03307  0.12454  0.41388
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2631     0.5094  -0.516  0.614217
## NEW.LOG       0.8049     0.1513   5.319  0.000139 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2343 on 13 degrees of freedom
## Multiple R-squared:  0.6851, Adjusted R-squared:  0.6609
## F-statistic: 28.29 on 1 and 13 DF, p-value: 0.0001393
```

The regression equation is  $\log_{10}(\text{OLD}) = -0.26 + 0.80\log_{10}(\text{NEW})$

10. (3 marks) Test the null hypothesis that the intercept of this new regression is zero using  $\alpha = 0.05$ .

The test of significance gives  $p = 0.614$ . The null hypothesis is not rejected.

11. (3 marks) Test the null hypothesis that the slope of this new regression is zero using  $\alpha = 0.05$ .

The test of significance gives  $p = 1.39 \times 10^{-4}$ . The null hypothesis is rejected.

12. (5 marks) Are your tests of hypotheses in (3) and (4) consistent with your tests in (9) and (10)? Discuss.

Despite the differences in p-values, the inferences drawn across the two analyses are the same.

13. (3 marks) Produce a 95% confidence interval for the slope of this new regression.

```
confint(TUBERCULIN.LM2)
```

```
##                2.5 %    97.5 %  
## (Intercept) -1.3636673 0.8374771  
## NEW.LOG      0.4779407 1.1317830
```

The 95% confidence interval of the slope is (0.48, 1.13).

14. (5 marks) Interpret this 95% confidence interval.

We are 95% confident that the true mean slope lies between 0.48 and 1.13.

15. (5 marks) Which of these two regression equations is more appropriate? Why?

The original data don't seem to have required the use of a transformation, so the calculation of logarithms was superfluous.

**THE END**