Q7

Anil Kumar Ganna

2024-04-17

#Question7 ##Write a brief report containing the analysis and visualization of the following example data generated froma hypothetical pseudovirus neutralizationassay. Analysis and visualization should be consistent with those shown in literature. State your assumptions and any formulae used in your calulcations.

##The following table represents a simplified example of raw data from a pseudovirus neutralization assay,measuring luminescence(indicative of virus entry) in response to increasing concentrations of a neutralizing antibody.Each measurement is the average luminescence from duplicate wells.

installing packages required for the operations performed
ggplot2 is used for visualization of the data
ggpmisc is used to print the regression equation and r2 and adj r2 values in the plot
dplyr is used for data manipulations
tinytex is used for get the output in pdf format
knitr is used for output format and styles
rmarkdown is used for create and genarate rmarkdown files

```r
library(ggplot2)
library(dplyr)
library(knitr)
library(rmarkdown)
library(tinytex)
library(ggpmisc)
```

Virus control(V):Wells containing virus and cells but no anti body.This serves as the maximum infection control.

Cell control (C): Wells containing cells only, without virus or antibody, to measure background luminescence.

Antibody dilutions: Different concentrations of the antibody being tested are distributed across the plate to assess their neutralizing activity.

Virus contol is 120000

Cell control is 500

```r
data <- data.frame(
  Antibody_Concentration = format(c(0, 0.01, 0.1, 1, 10, 100), scientific =
F),
  Luminescence = c(120000, 115000, 95000, 60000, 25000, 5000)
)
print(data)

  Antibody_Concentration Luminescence
1                   0.00       120000
2                   0.01       115000
3                   0.10        95000
4                   1.00        60000
5                  10.00        25000
6                 100.00         5000

data$conc <- as.numeric(data$Antibody_Concentration)
virus_control <- 120000
Cellcontrol <- 500
```

Calculation of % Inhibition: Percentage inhibition is calculated using the formula: % Inhibition = 100 x (1−(Average luminescence at a given antibody concentration/Average luminescence without antibody)

```r
data$Percent_Inhibition <- (1 - data$Luminescence / virus_control) * 100
```

##Printing the data

```r
print(data)

  Antibody_Concentration Luminescence  conc Percent_Inhibition
1                   0.00       120000 0e+00           0.000000
2                   0.01       115000 1e-02           4.166667
3                   0.10        95000 1e-01          20.833333
4                   1.00        60000 1e+00          50.000000
5                  10.00        25000 1e+01          79.166667
6                 100.00         5000 1e+02          95.833333
```

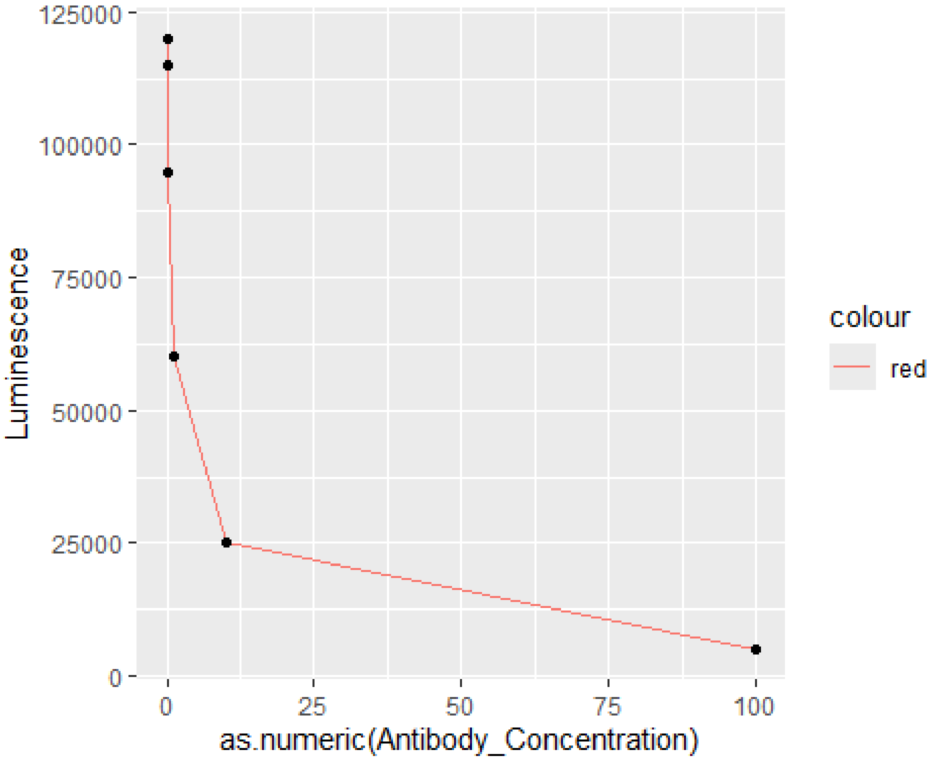
##line plot to understand the linarity between concentration and luminescence

while observing the un-transformed data, the association was non linear and it is a L shaped curve. luminescence(indicative of virus entry) decreases rapidly in response to increasing concentrations at early levels, then gradually levels off at higherlevels.

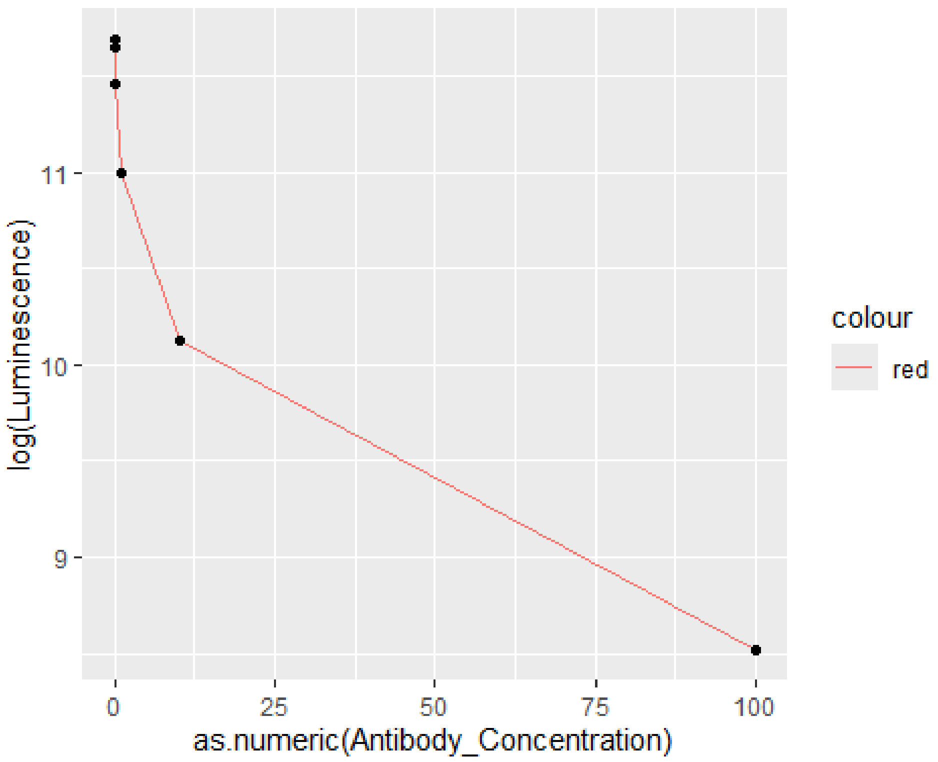
while observing semi logarithmic plot, the slope of the curve is reduced and linearity was improved.

while observing square root of concentration vs logarithmic luminescence plot, the slope of the curve is reduced and linearity was improved than others.

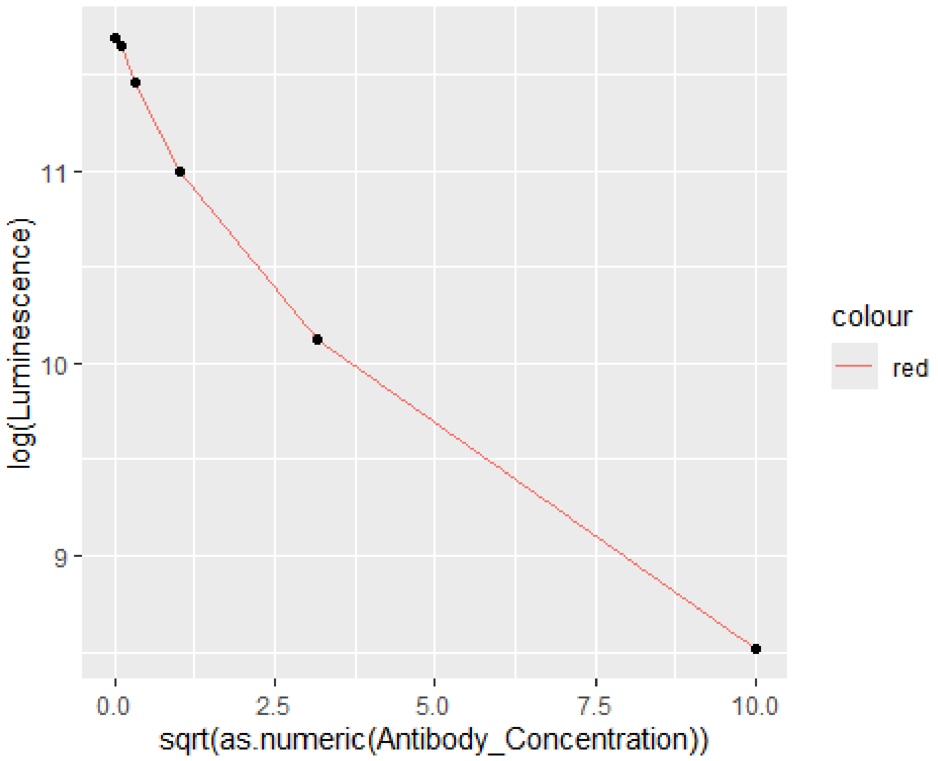
```
data %>% ggplot(aes(as.numeric(Antibody_Concentration),
Luminescence))+geom_line(aes(col="red"))+geom_point()
```



```
data %>% ggplot(aes(as.numeric(Antibody_Concentration),
log(Luminescence)))+geom_line(aes(col="red"))+geom_point()
```



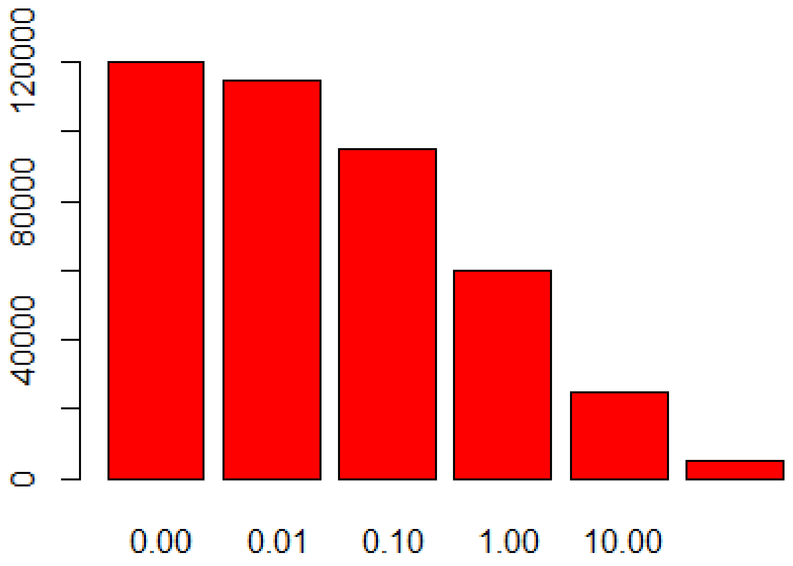
```
data %>% ggplot(aes(sqrt(as.numeric(Antibody_Concentration)),
log(Luminescence)))+geom_line(aes(col="red"))+geom_point()
```



Bar plot to understand the average luminescence from duplicate wells for increasing levels of concentration where the x-axis represents the concentration of neutralizing antibodies, and the y-axis represents the percentage inhibition of virus entry.

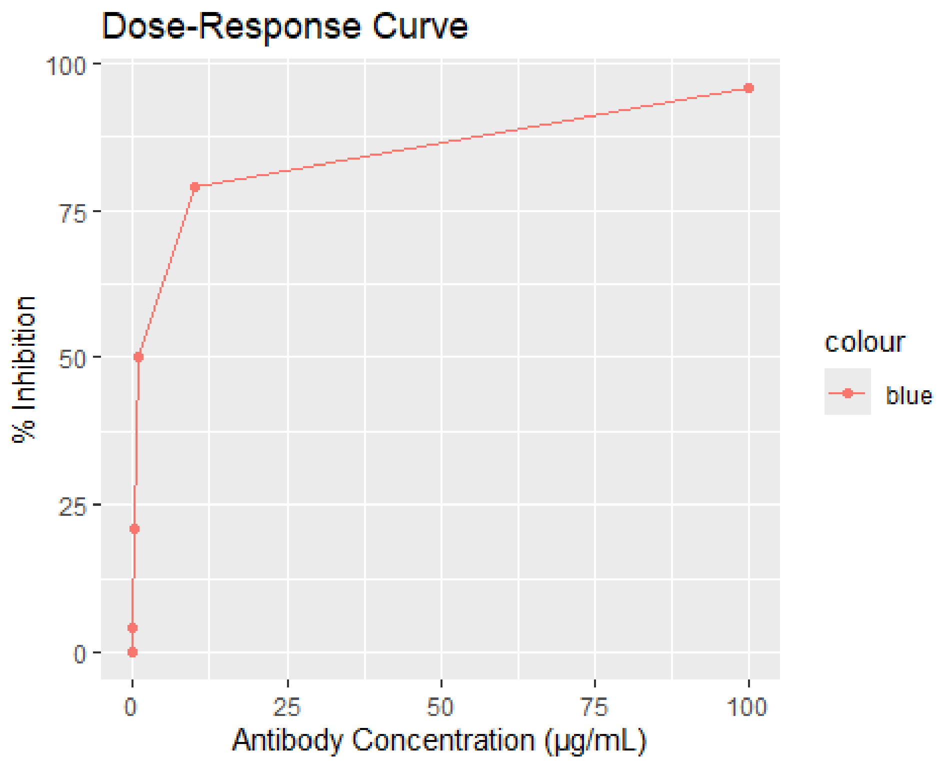
From the following bar plot, it is showing that the average luminescence from duplicate wells was decreasing while the increasing concentrations of a neutralizing antibody.

```
barplot(data$Luminescence, names.arg=data$Antibody_Concentration, col =
"red")
```



The data can be visualized using a dose-response curve, where the x-axis represents the concentration of neutralizing antibodies, and the y-axis represents the percentage inhibition of virus entry. By examining the shape of the curve and the position of the data points relative to the curve, we can assess the efficacy and potency of the neutralizing antibodies.

```
ggplot(data, aes(x = conc, y = Percent_Inhibition, color="blue")) +
  geom_point() +
  geom_line() +
  labs(
    title = "Dose-Response Curve",
    x = "Antibody Concentration (µg/mL)",
    y = "% Inhibition",
  )
```



From the above plot, the curve says that the inhibition increases rapidly at lower antibody concentrations up to a concentration level of 10. At higher concentration levels, the inhibition rate rises slowly and becomes constant. It slowed down after a concentration level of 10 and a higher concentration of 100.

##Simple linear regression assumptions

1. Linearity: The relationship between the independent variable (X) and the dependent variable (Y) is linear. This means that changes in Y are directly proportional to changes in X, and the data points should roughly form a straight line pattern when plotted.

2. Independence of Observations: The observations in the data set are independent of each other. In other words, the value of one data point should not be influenced by the value of another data point.

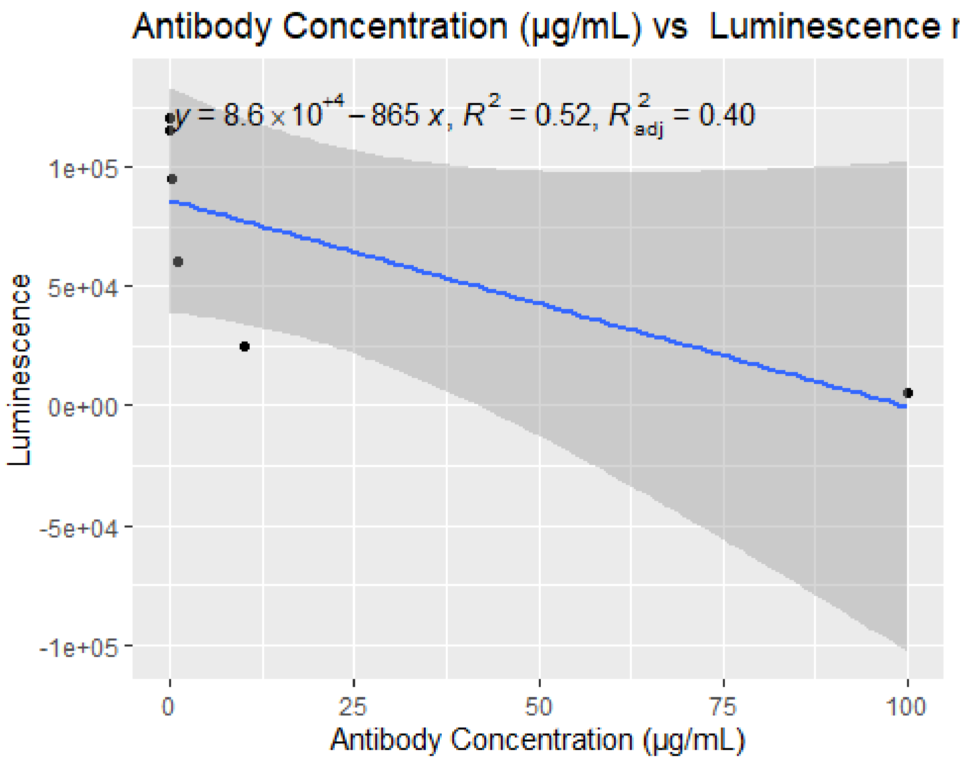
3. Normality of Residuals: The residuals should follow a normal distribution. This means that when you plot the residuals (the differences between observed and predicted Y values) against the predicted Y values, the distribution of residuals should approximate a bell-shaped curve.

4. Independence of Errors: The errors (residuals) should be independent of each other. There should be no pattern or correlation in the residuals when plotted against the independent variable or any other variable.

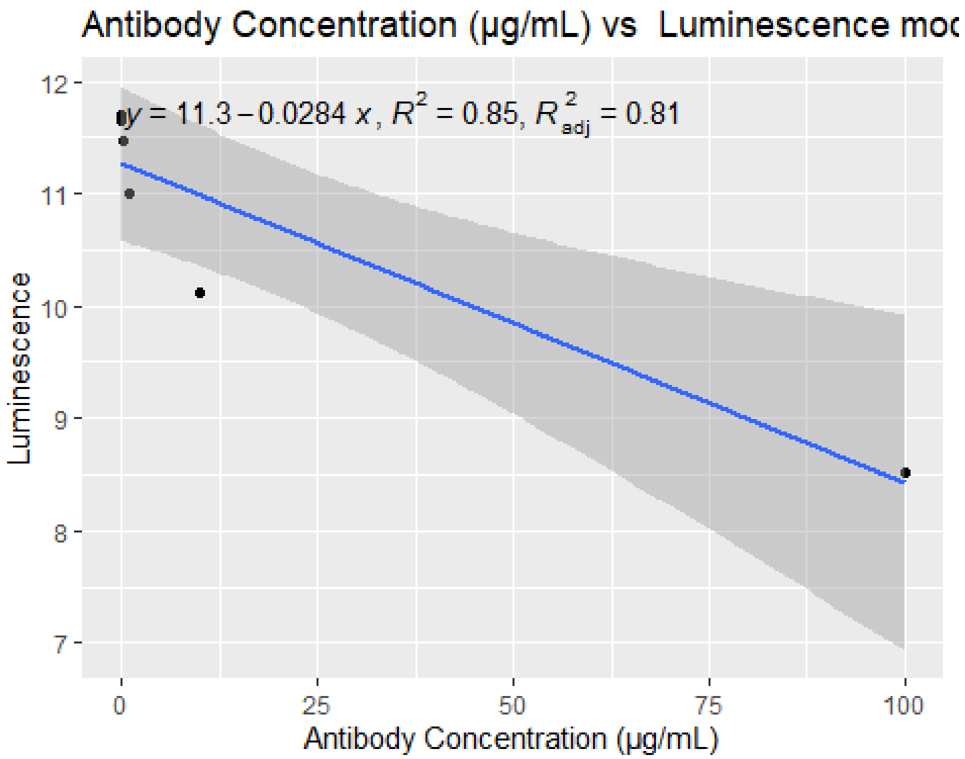
```r
reg1 <- lm(Luminescence~conc, data)
print(summary(reg1))
```

```
Call:
lm(formula = Luminescence ~ conc, data = data)

Residuals:
     1      2      3      4      5      6 
 33983  28991   9069 -25152 -52368   5476 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  86017.2    16937.5   5.079  0.00709 **
conc          -864.9      412.8  -2.095  0.10421   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37020 on 4 degrees of freedom
Multiple R-squared:  0.5233,	Adjusted R-squared:  0.4041 
F-statistic:  4.39 on 1 and 4 DF,  p-value: 0.1042
```

```r
ggplot(data, aes(x = conc, y = as.numeric(Luminescence))) +
  geom_point() + stat_poly_line() +
  stat_poly_eq(use_label(c("eq", "R2", "adj.R2"))) +
  labs(
    title = "Antibody Concentration (μg/mL) vs  Luminescence model1",
    x = "Antibody Concentration (μg/mL)",
    y = "Luminescence",
  )
```

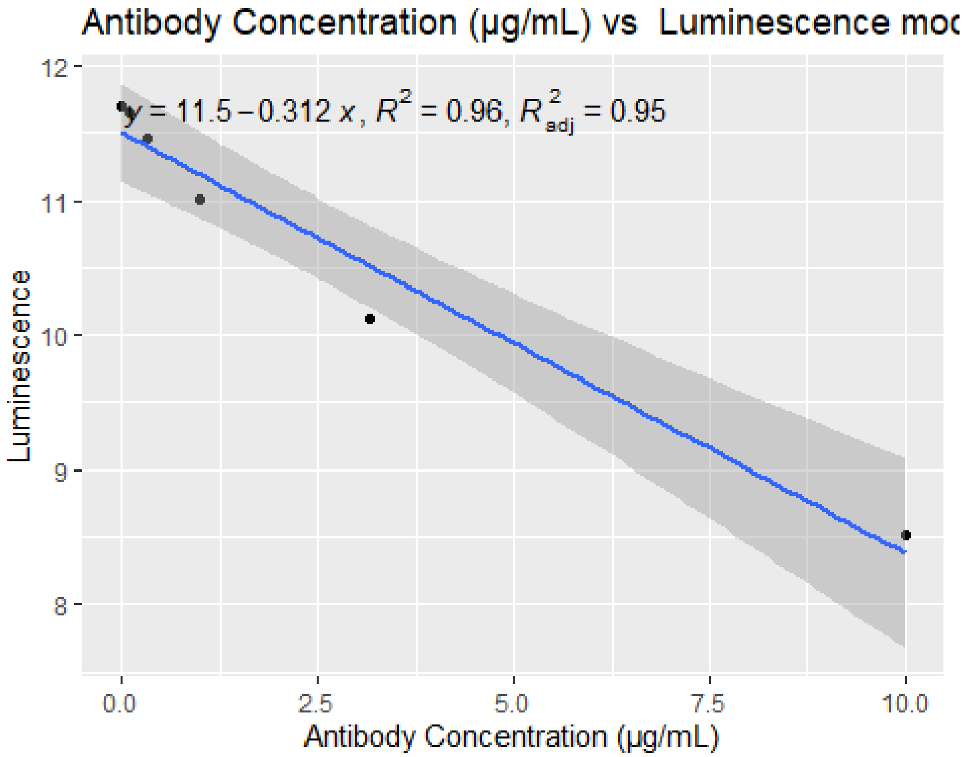


```
reg2 <- lm(log(Luminescence)~conc, data)
print(summary(reg2))


Call:
lm(formula = log(Luminescence) ~ conc, data = data)

Residuals:
       1        2        3        4        5        6 
 0.42692  0.38464  0.19614 -0.23784 -0.85779  0.08792 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11.268328   0.247300   45.57 1.39e-06 ***
conc        -0.028391   0.006027   -4.71  0.00924 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5406 on 4 degrees of freedom
Multiple R-squared:  0.8473,	Adjusted R-squared:  0.8091 
F-statistic: 22.19 on 1 and 4 DF,  p-value: 0.009237

ggplot(data, aes(x = conc, y = log(Luminescence))) +
  geom_point() + stat_poly_line() +
  stat_poly_eq(use_label(c("eq", "R2", "adj.R2"))) +
  labs(
    title = "Antibody Concentration (µg/mL) vs  Luminescence model2",
    x = "Antibody Concentration (µg/mL)",
    y = "Luminescence",
  )
```



```
reg3 <- lm(log(Luminescence)~sqrt(conc), data)
print(summary(reg3))


Call:
lm(formula = log(Luminescence) ~ sqrt(conc), data = data)

Residuals:
       1        2        3        4        5        6 
 0.19429  0.18294  0.05938 -0.18674 -0.38731  0.13745 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11.50096    0.13052   88.11 9.94e-08 ***
sqrt(conc)  -0.31212    0.03033  -10.29 0.000503 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2639 on 4 degrees of freedom
Multiple R-squared:  0.9636,    Adjusted R-squared:  0.9545 
F-statistic: 105.9 on 1 and 4 DF,  p-value: 0.000503
```

```r
ggplot(data, aes(x = sqrt(conc), y = log(Luminescence))) +
  geom_point() + stat_poly_line() +
  stat_poly_eq(use_label(c("eq", "R2", "adj.R2"))) +
  labs(
    title = "Antibody Concentration (µg/mL) vs  Luminescence model3",
    x = "Antibody Concentration (µg/mL)",
    y = "Luminescence",
  )
```



From the above three models, model 3 performs well(adjusted r square value:0.95); the concentrations are transformed into its square roots, but while transforming into logarithmic, it was generating Na's because of 0 concentrations. The dependent variable, luminescence, was log-transformed.

Hence, from the models, we conclude that the luminescence (indicative of virus entry) decreases while antibody concentrations increase.And antibody concentrations causes to the luminescence (indicative of virus entry) decreases. there was statistical significant negative linear association between the luminescence (indicative of virus entry) and concentrations.