Report of Sample Analysis Results

Steven Allen

2020-02-08

# Introduction

Developed an R/Shiny app: Sample submission system (with integrated RMarkdown report) for Capstone Project, tools include git, SQLite, RStudio;

Used the tool to submit samples for analysis, track analysis result, and submission date with the role of submitter, download submission data, create data linear regression model, load analysis result, data stored in a database with the role of the analyst.

* **Submitter**
  + Able to track different analytes analysis result
  + Able to download query result
  + Submitter only see own data
* **Staff**
  + Able to submit samples analysis expected raw results to database
  + Able to track analysis result from database
  + Able to download analysis result
* **Analyst**
  + Able to load standard data, produce expected raw data versus result plot
  + Able to track raw result
  + Perform linear regression of controls
  + Able to calculate sample result
  + Able to download sample result
  + Able to download final report in PDF/HTML/Word

# Methods

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

Before attempting to fit a linear model to observed data, a modeler should first determine whether or not there is a relationship between the variables of interest. In this report, standard data will be used. In the standard, the relationship between analyte’s experimental result and actual result is linearly related.

In the report, we will build linear regression models to relate the experimental results of analyte HDL, LDL, and Triglycerides to their actual results for standard data. Then, the models will be used to predict actual results for corresponding analytes.

## Linear Regression Model of HDL

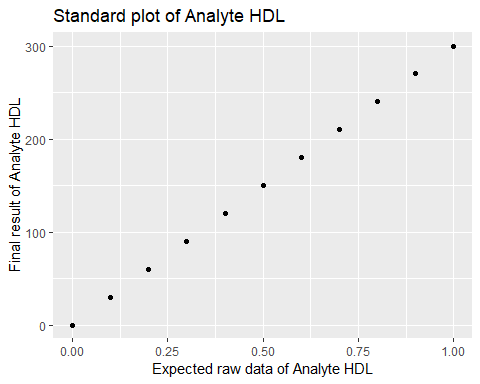
Summary of the linear regression model for analyte *HDL*

##   
## Call:  
## lm(formula = result ~ expected\_raw, data = stddata\_HDL)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.664e-14 -1.346e-14 -4.184e-15 7.440e-15 3.446e-14   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.714e-14 1.077e-14 1.591e+00 0.146   
## expected\_raw 3.000e+02 1.821e-14 1.648e+16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.91e-14 on 9 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 2.714e+32 on 1 and 9 DF, p-value: < 2.2e-16

The equation of the linear regression model for analyte HDL is:

where x = HDL;

The scatter plot for analyte HDL is shown in Figure 1.



Standard Plot of Analyte HDL

From Figure 1 we see that states with higher experimental results of analyte HDL tend to have higher actual results. The relationship between an explanatory variable and a response variable is linearly related.

## Linear Regression Model of LDL

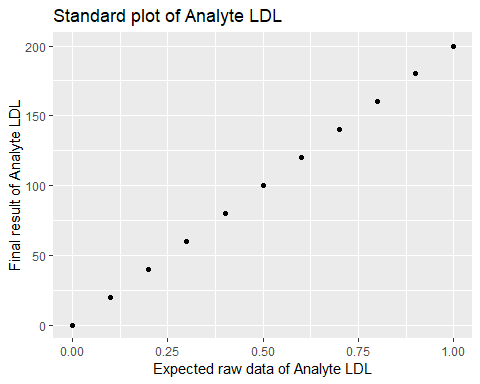
Summary of the linear regression model for analyte *LDL*

##   
## Call:  
## lm(formula = result ~ expected\_raw, data = stddata\_LDL)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.705e-14 -6.846e-15 -2.426e-15 1.011e-14 1.708e-14   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.428e-14 6.579e-15 5.210e+00 0.000556 \*\*\*  
## expected\_raw 2.000e+02 1.112e-14 1.798e+16 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.166e-14 on 9 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 3.234e+32 on 1 and 9 DF, p-value: < 2.2e-16

The equation of the linear regression model for analyte LDL is:

where x = LDL;

The scatter plot for analyte LDL is shown in Figure 2.



Standard Plot of Analyte LDL

From Figure 2 we see that states with higher experimental results of analyte LDL tend to have higher actual results. The relationship between an explanatory variable and a response variable is linearly related.

## Linear Regression Model of Triglycerides

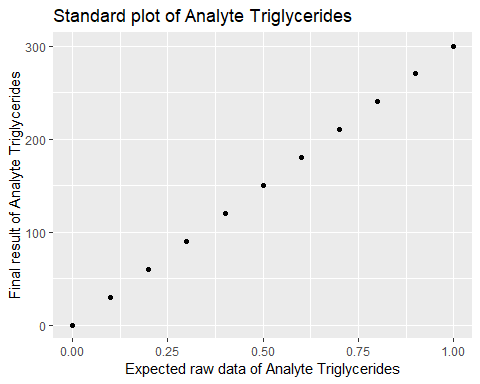
Summary of the linear regression model for analyte *Triglycerides*

##   
## Call:  
## lm(formula = result ~ expected\_raw, data = stddata\_Triglycerides)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.664e-14 -1.346e-14 -4.184e-15 7.440e-15 3.446e-14   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.714e-14 1.077e-14 1.591e+00 0.146   
## expected\_raw 3.000e+02 1.821e-14 1.648e+16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.91e-14 on 9 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 2.714e+32 on 1 and 9 DF, p-value: < 2.2e-16

The equation of the linear regression model for analyte Triglycerides is:

where x = Triglycerides;

The scatter plot for analyte Triglycerides is shown in Figure 3.

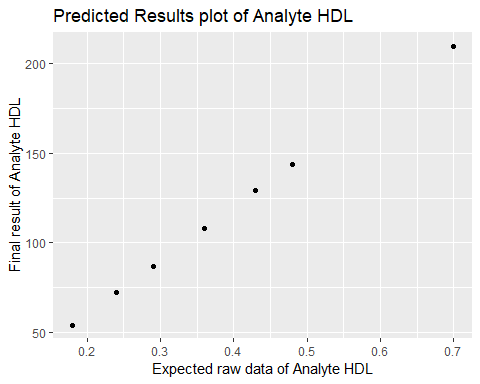


Standard Plot of Analyte Triglycerides

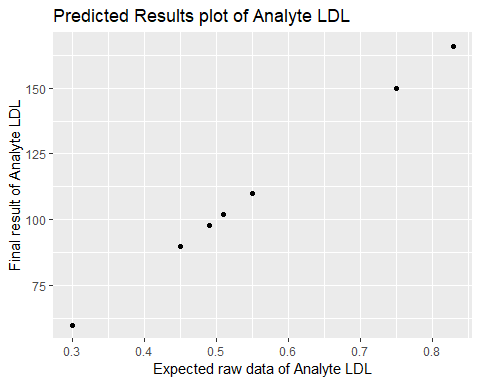
From Figure 3 we see that states with higher experimental results of analyte Triglycerides tend to have higher actual results. The relationship between an explanatory variable and a response variable is linearly related.

# Results

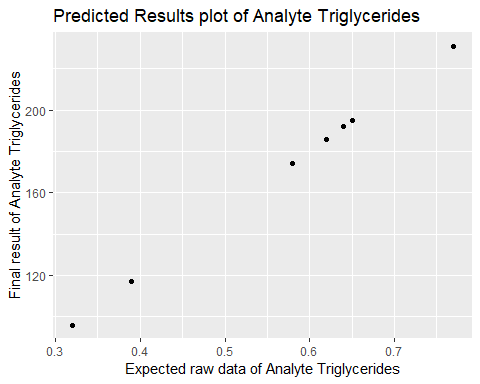
The predicted results for analytes HDL, LDL and Triglycerides are shown in Figures 4 to 6.



Predicted Results Plot of Analyte HDL



Predicted Results Plot of Analyte LDL



Predicted Results Plot of Analyte Triglycerides

The predicted results for analytes HDL, LDL and Triglycerides are shown in Table 1.

# Conclusion

For each of our three analytes HDL, LDL and Triglycerides, we found that there was a strong linearly relationship between the explanatory variable and actual value.

# References

“Shiny from RStudio” by RStudio Inc. Weblink:<https://shiny.rstudio.com/>.

“R Markdown from RStudio” by RStudio Inc. Weblink:<https://rmarkdown.rstudio.com/index.html>.