Biostatistics - Dr. Patrick
BMEN 350; Section 201
Saurabh Dhole 626 002 135
October 29th, 2021

Project 4

A. Read and Reflect

Upon reviewing the article called "Simple linear regression" by Dr. Altman, I was intrigued by one of the concepts of linear regression that was discussed eloquently by Dr. Altman. In his knowledgeable article, Dr. Altman emphasizes that regression is a specific kind of association that may be linear or nonlinear. All this time before I had read this article, I had thought that regression could only be linear. This is because I would often pair the words regression and linear together. I now know that simple linear regression is the most basic type of regression relationship, and that non-linear relationships do exist. As a Biomedical Engineer, my new found knowledge of regression is very important to me. I say this because in internships and research experiences I will often be presented with plots that describe a certain relationship between variables. Instead of trying to fit those plots to a linear model, I can now experiment with different non-linear models to ensure that I can acquire the best fitting line. In a general sense, this new found information is essential to my understanding of plotting and data analysis. I say this because I perform quite a bit of data analysis in my part-time research job, and being able to accustom the plots to a line that provides the best association is very valuable to the neuronal drug addiction projects that I am working on. I plan on holding on to this information when I need to establish a new equation of fit for a plot of data in my research projects, or in the BMEN 350 course module projects.

After reading the knowledgeable article called "Association, correlation and causation" by Dr. Altman, I was absolutely satisfied upon reinforcing my interpretation of the concept about correlation does not imply causation. In previous statistics courses, this concept was made clear to me and I understood it very well. This article allowed me to reinforce this concept as it has made me think about confounding variables. Allow me to explain what I mean by this. In the article, Dr. Altman provides the example of people who drink more than 4 cups of coffee daily have a decreased chance of developing skin cancer. One would say that there is a correlation between drinking coffee and not developing skin cancer, however there exists a confounding variable! People who drink a lot of coffee work indoors for long periods of time and thus have reduced exposure to sunlight. Therefore, in this case, the amount of time spent outdoors is indeed the confounding variable. This information is indeed very important to me as a Biomedical Engineer. I say this because I frequently have to present my findings from the lymphatics research lab to the principal investigator. Often times my presentation is faulty as I sometimes pair a correlation with a causation. The information in this article will remind me not to do that anymore and to search for any confounding variables. In a general sense this information is very valuable because we are frequently bombarded with ads stating that there is a causation between two variables and to mitigate such a causation, we must buy a certain product. Such ads are often on television and are often manipulative. To be able to avoid such manipulative ads, I will be sure to brain storm any confounding variables. I plan to keep this information at the back of my mind whenever I am presented with sets of data in a research or clinical setting. This will prevent me from supporting false causations.

Following my review of the article called "Understanding Bland Altman analysis" by Dr. Giavarina, I found myself satisfied with the information I had learned about the technicalities of the Bland Altman analysis. Dr. Giavarina makes it very clear in the article that Bland and Altman wanted to be able to "describe the agreement between two quantitative measurements". Dr. Giavarina further states that the method to quantify agreement between two quantitative measurements is to construct limits of agreement. I was initially puzzled on how these limits of agreement could be constructed. Upon reading further I found out that these statistical limits are found by using the mean and standard deviation of the differences between the two measurements. As a Biomedical Engineer, this information is crucial for my work in undergraduate research and industrial experiences. I say this because I will most likely have to apply the Bland Altman analysis in research settings and in quality engineering internships. In a more general sense, this information is essential for my understanding of limits of agreement and how they relate to the Bland Altman analysis. I say this because I can now apply this technical skill in many software platforms such as Excel, Matlab, and R for the purposes of personal projects or research projects. I plan on hoarding this key piece of technical information for future reference when I will need it in BMEN 350 module projects and during quality engineering internships.

After reading the informative article called "Receiver-Operating Characteristic Analysis for Evaluation Diagnostic Tests and Predictive Models" by Dr. Zou, I was absolutely taken aback by the evaluative power of ROC analysis. Dr. Zou makes it clear that ROC analysis is an extremely useful tool for evaluating the performance of diagnostic tests and for evaluating the accuracy of statistical models. Dr. Zou gives the example that ROC analysis can be used to evaluate a predictive model to estimate expected outcomes such as mortality or adverse cardiac events. As a Biomedical Engineer, this information is very important to me. I say this because there will be times in my research studies during which I will have to evaluate how well our lab's models are able to predict adverse cancer-lymphatic events in certain cell lines. In those situations, I will certainly have to use evaluative methods such as the ROC analysis to test diagnostic power. In a more general sense, this information is essential for me as it gives me relief that there is a method to hold these diagnostic tests accountable. If there was no method such as the ROC to test the diagnostic power of some of the published diagnostic tests, we would be in a fairly bad place in terms of scientific research. I plan to use ROC analysis in my BMEN 350 course projects as well as in R&D internships.

Upon reading the instructive article called "What is an ROC curve?" by Dr. Hoo, I was truly enlightened on some of the technical aspects of the ROC curve and its uses. I would like to reflect on my observations here. In the article, Dr. Hoo makes it very clear that the graphical ROC curve is produced by plotting sensitivity, also known as true positive rate, on the y-axis against 1-specificity, also known as false positive rate, on the x-axis. Dr. Hoo explains that prior to this step, the sensitivities and specificities for different values of a continuous test measure have to be tabulated. As a Biomedical Engineer, this information is indeed paramount to my utilization of ROC analysis. I say this because before reading this article, I really did not know how to tabulate or plot an ROC curve. In R&D internships and quality engineering internships, I will have to apply the ROC analysis towards new products and new procedures. The information in this article will guide me on how I can do this. In a general sense, this information is indeed important for me as I now have a new tool to add to my statistical analysis toolbox. I can unleash this tool whenever I need to perform an ROC analysis whether that is in a clinic, research, or industrial setting. I plan to use the information on how to plot an ROC curve in my BMEN 350 course module projects. I will also hoard this information for future reference in internships.

B. Problem 1

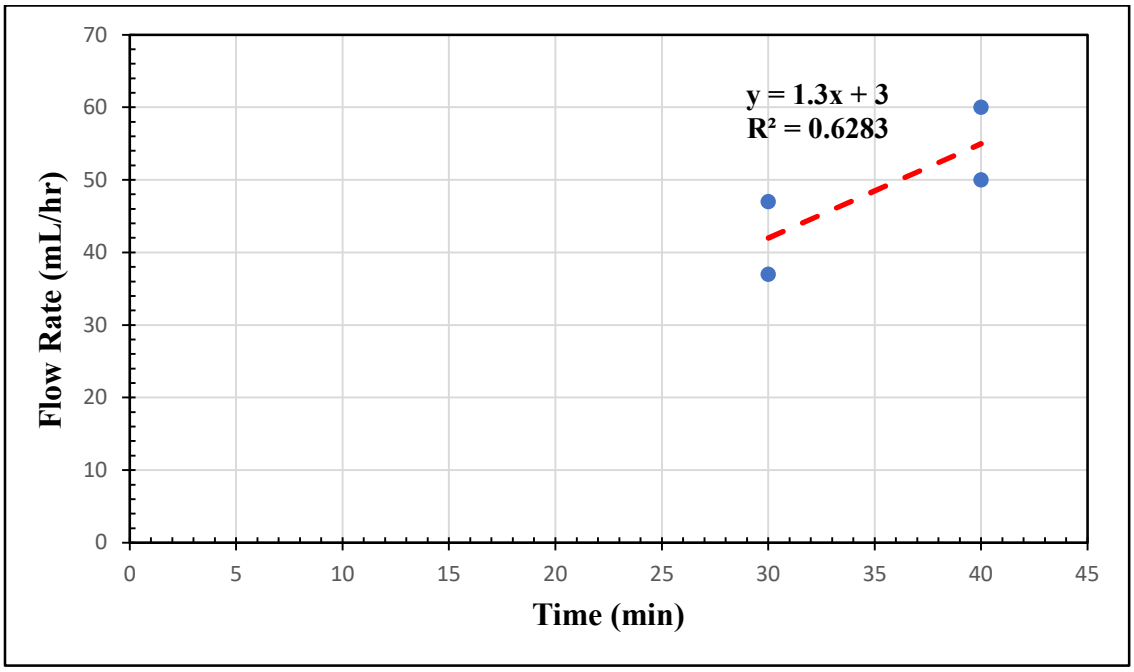


Figure 1: **Flow Rate (mL/hr) vs Time (min) for runs 1 – 4. It can be observed that the line of regression or line of best fit has the equation y = 1.3x + 3 and that the R^2 value or the coefficient of determination is 0.63.**

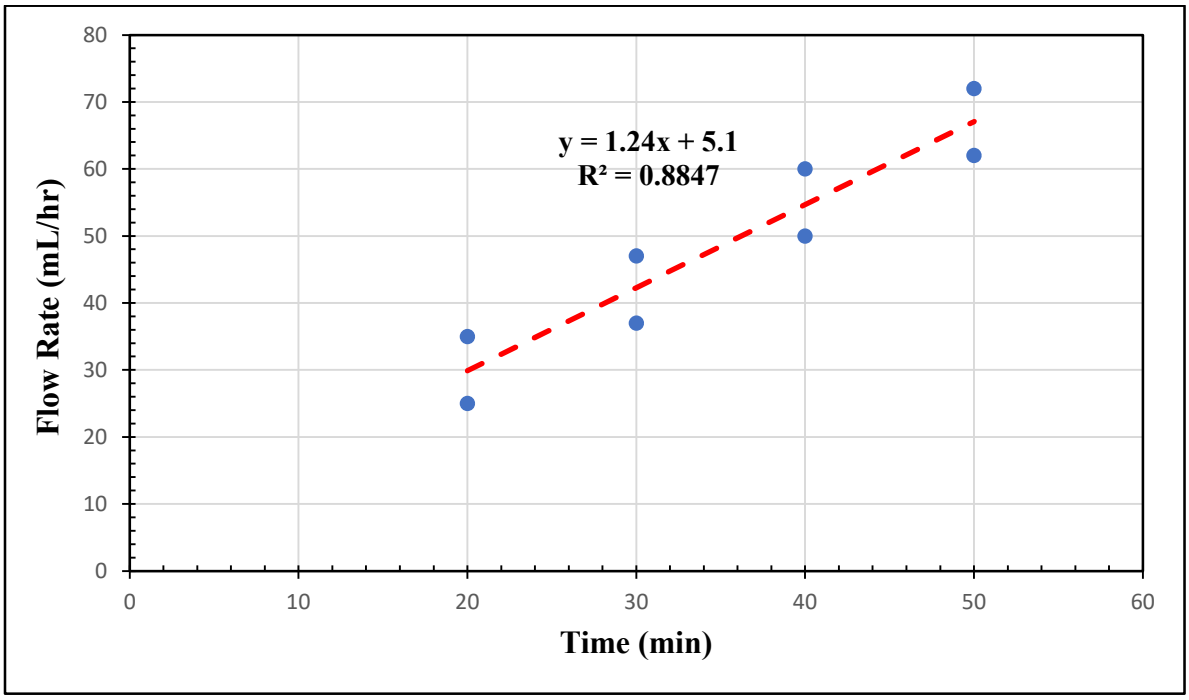


Figure 2: **Flow Rate (mL/hr) vs Time (min) for runs 1 – 8. It can be observed that the line of regression or line of best fit has the equation y = 1.2x + 5.1 and that the R^2 value or the coefficient of determination is 0.88.**

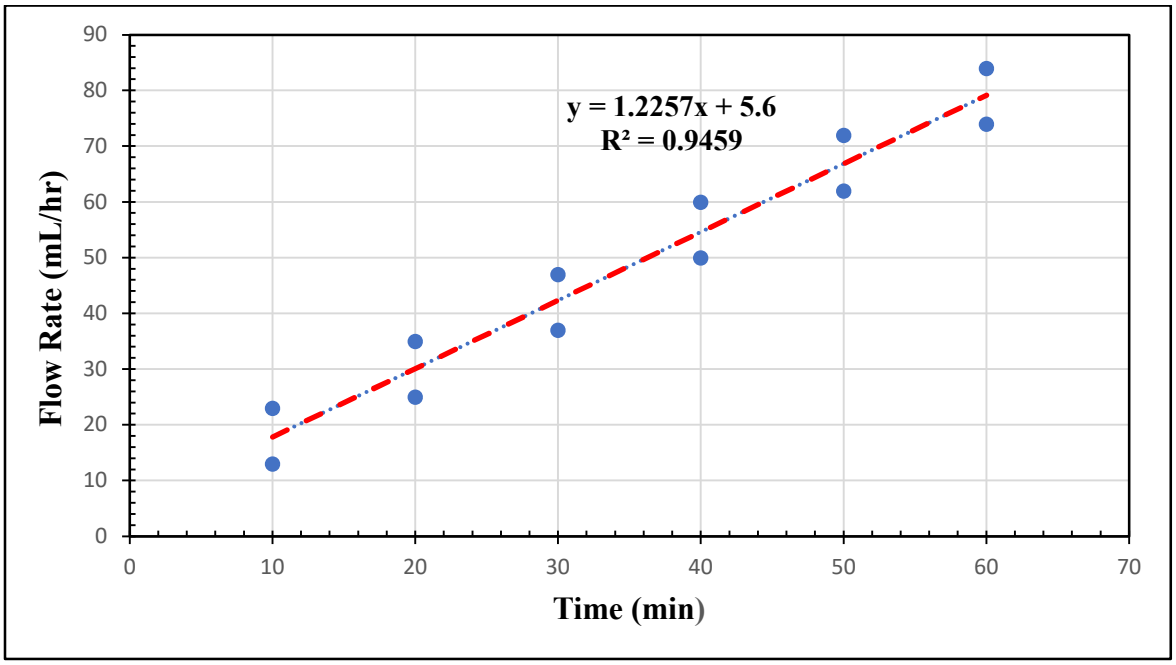


Figure 3: *Flow Rate (mL/hr) vs Time (min) for runs 1 – 12.* It can be observed that the line of regression or line of best fit has the equation y = 1.2x + 5.6 and that the R^2 value or the coefficient of determination is 0.94.

The flow rate vs time graphs are plotted in Figures 1 – 3 for runs 1 – 4, 1 – 8, and 1 – 12 respectively. The coefficient of determination or R^2 for the three figures increased from 0.63 (Figure 1), to 0.88 (Figure 2), to 0.94 (Figure 3). As the range of data increased the coefficient of determination (R^2) also increased. This means that as the range of the data increased, the goodness of fit of the line of best fit also increased. Goodness of fit however, does not imply agreement.

C. Problem 2

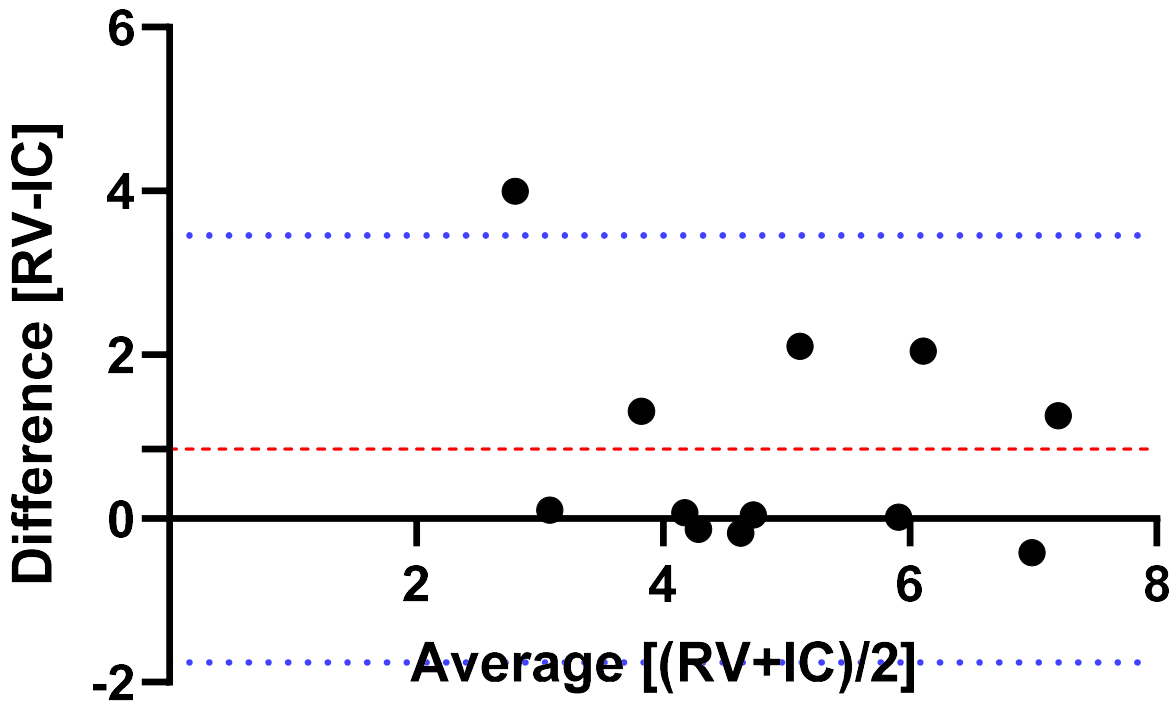


Figure 4: **Bland Altman Plot for Cardiac Ejection Fraction. The difference in cardiac ejection fraction (%) between RV and IC is plotted against the average of ejection fraction (%) of RV and IC. The upper limit of agreement is 3.456 (as indicated by the dotted blue line) and the lower limit of agreement is -1.754 (as indicated by the dotted blue line). The bias is 0.8508 (as indicated by the dashed red line). The standard deviation of the bias is 1.329. These values are provided in the Appendix as well. 95% limits of agreement were used here.**

The idea of a Bland Altman analysis is to investigate the agreement between two quantitative measurements. It can be easily deciphered from the Bland Altman plot above that most of the dots are indeed in close proximity to the bias line (0.8508) and in between the upper and lower limits of agreement. The mean difference (bias) itself is indeed close to 0 (0.8508 as represented by the red dashed line). The closer the mean difference (bias) is closer to zero, the better the agreement between the two measurements. So, it can be seen from Figure 4 above that there is indeed a decent amount of agreement between the two measurements as the difference mean (bias) is fairly close to 0. Since 95% limits of agreement were established in Figure 4, we can say that 95% of the differences between the two measurements is expected to be within the values described by the 95% limits of agreement. Very wide limits of agreement suggest less precision and less agreement between two measurements.

D. Problem 3

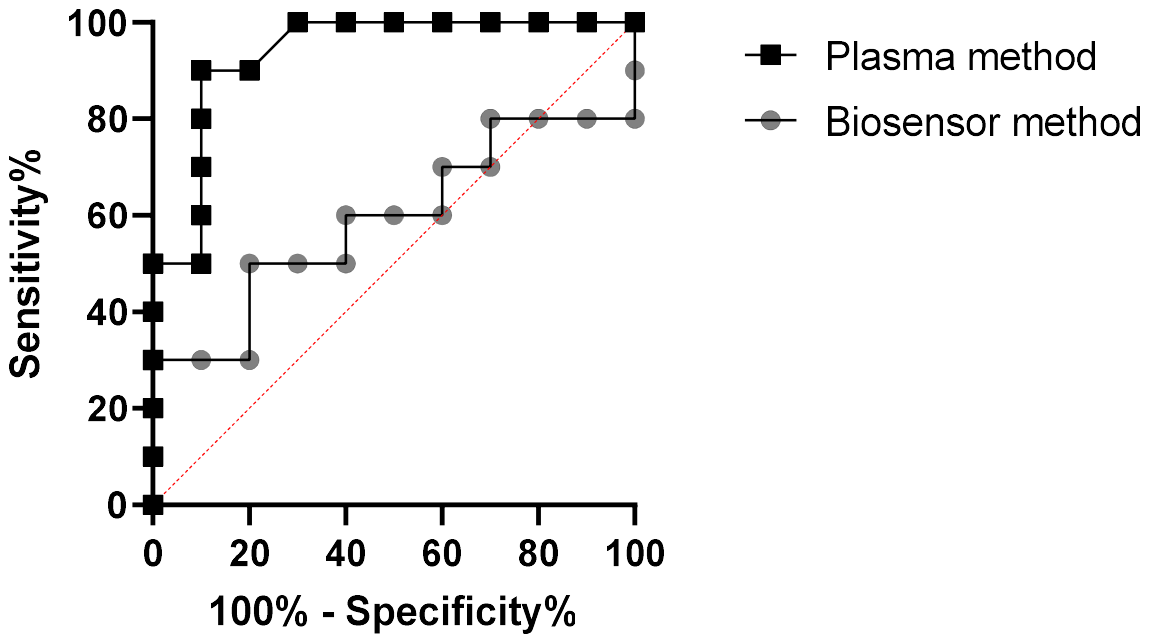


Figure 5: **ROC Curve of Plasma method and Biosensor method. It can be deciphered from the Appendix that the AUC for the ROC curve for Biosensor method is 0.590 (or 59%). It can also be deciphered from the Appendix that the AUC for the ROC curve for Plasma method is 0.935 (or 93.5%). The "random chance" curve is shown as a red dotted line. The AUC of the "random chance" curve is 0.50 (or 50%).**

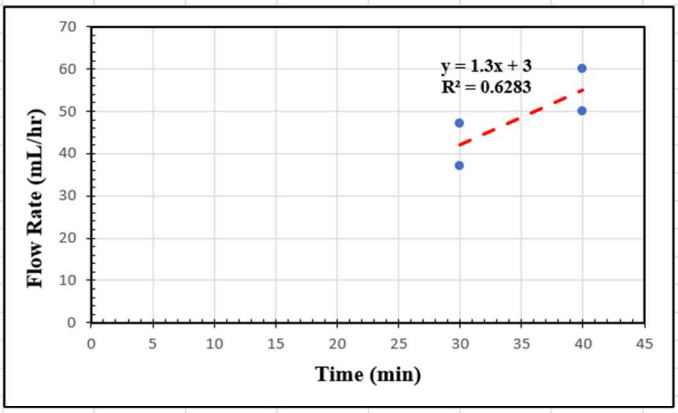
It has been established from Figure 5 and from the Appendix that the AUC for the Biosensor method ROC curve is 0.590 (or 59%). It has also been established from Figure 5 and the Appendix that the AUC for the Plasma method ROC curve is 0.935 (or 93.5%). Therefore, the AUC for the Plasma method ROC curve is greater than that of the Biosensor method ROC curve. For this reason, the overall diagnostic accuracy of the Plasma method is greater than that of the Biosensor method. For this reason, I would prefer to use the Plasma method in the clinic.

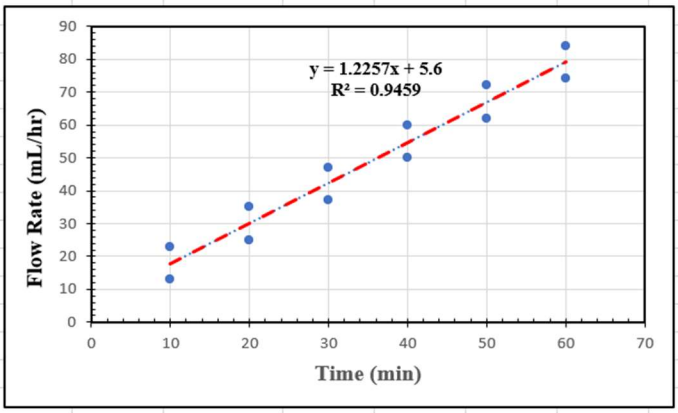
Appendix begins on next page …

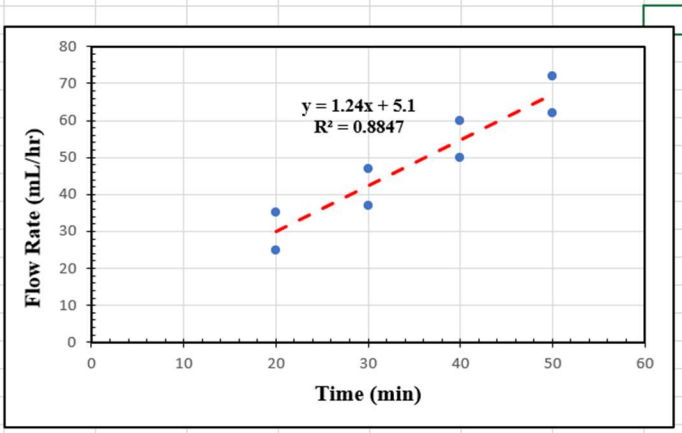
E. Appendix

Problem 1: Excel was used to plot the graphs in Figures 1 – 3.

Run	Time (min)	Flow Rate (mL/hr)
1	30	37
2	30	47
3	40	50
4	40	60
5	20	25
6	20	35
7	50	62
8	50	72
9	10	13
10	10	23
11	60	74
12	60	84



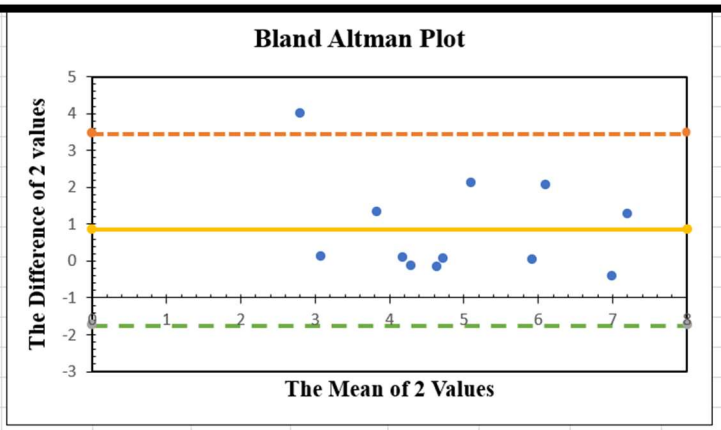




Appendix continues on next page …

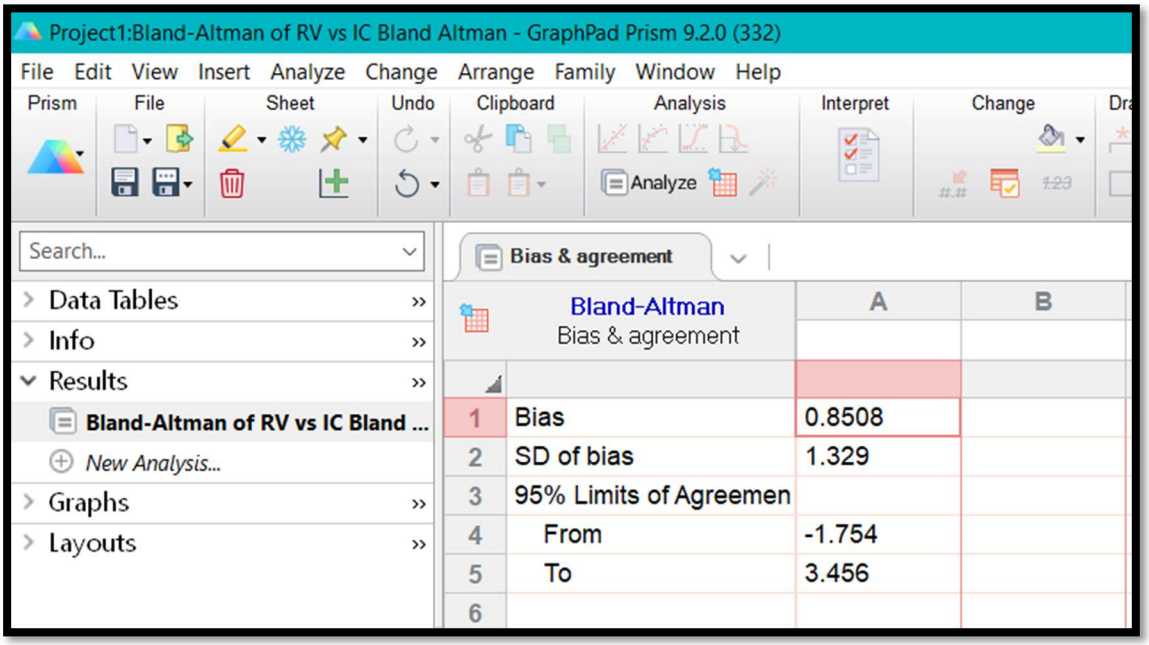
Problem 2: I used Microsoft Excel initially to find the Bland Altman plot, and to do initial calculations. I later switched to Graph Pad Prism.

member	RV method	IC method		Difference	mean
1	7.83	6.57		1.26	7.2
2	6.16	4.06		2.1	5.11
3	4.75	4.71		0.04	4.73
4	4.21	4.14		0.07	4.175
5	3.13	3.03		0.1	3.08
6	5.92	5.9		0.02	5.91
7	7.13	5.09		2.04	6.11
8	4.54	4.72		-0.18	4.63
9	4.48	3.17		1.31	3.825
10	4.22	4.35		-0.13	4.285
11	6.78	7.2		-0.42	6.99
12	4.8	0.8		4	2.8



	Bias:	0.850833	
	Stand. Dev:	1.329077	
Bias - 1.96(std)	Lower LOA	-1.75416	
Bias+ 1.96(std)	Upper LOA	3.455824	

x axis	Upper LOA	Lower LOA	Bias
0	3.455	-1.75	0.8508
8	3.455	-1.75	0.8508



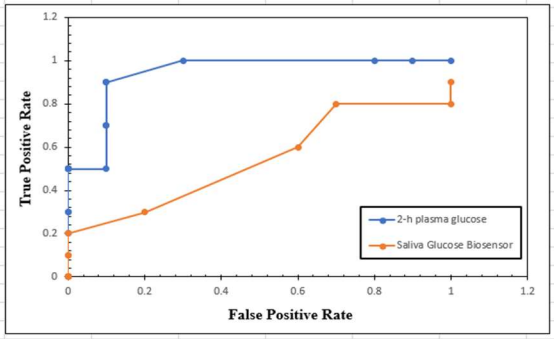
Initially Excel was used to calculate Bias, SD of bias, lower LOA, and Upper LOA. It can be seen that the values from Excel and the values from Graph Pad agree with each

These are the results of the Bland Altman analysis for RV vs IC Cardiac Ejection fraction. Bias, SD of bias, lower LOA and Upper LOA are all provided. From Graph Pad prism

Appendix continues on next page …

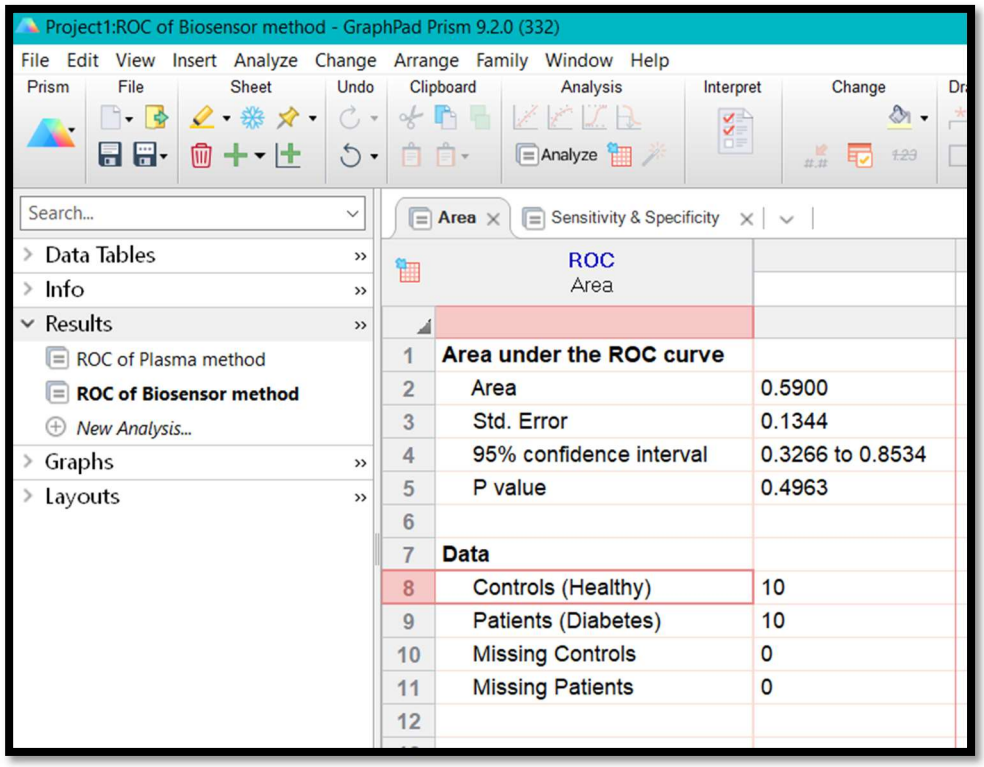
Problem 3: I used Microsoft Excel initially to find the ROC curves and initial calculations. I later switched to Graph Pad Prism.

2-h plasma glucose (mmol/L)		Counts		Cumulative				
mmol/L		Healthy	Diabetes	Healthy	Diabetes	FPR	TPR	AUC
0		0	0					
<4		0	0	0	0	1	1	
4 - 4.99		1	0	1	0	0.9	1	
5 - 5.99		1	0	2	0	0.8	1	
6 - 6.99		5	0	7	0	0.3	1	
7 - 7.99		2	1	9	1	0.1	0.9	
8 - 8.99		0	0	9	1	0.1	0.9	
9 - 9.99		0	2	9	3	0.1	0.7	
10 - 10.99		0	0	9	3	0.1	0.7	
11 - 11.99		0	2	9	5	0.1	0.5	
12 - 12.99		1	0	10	5	0	0.5	
13 - 13.99		0	0	10	5	0	0.5	
14 - 14.99		0	0	10	5	0	0.5	
15 - 15.99		0	0	10	5	0	0.5	
16 - 16.99		0	0	10	5	0	0.5	
17 - 17.99		0	0	10	5	0	0.5	
18 - 18.99		0	2	10	7	0	0.3	
19 -19.99		0	0	10	7	0	0.3	
20 - 20.99		0	1	10	8	0	0.2	
>21		0	2	10	10	0	0	
		10	10					



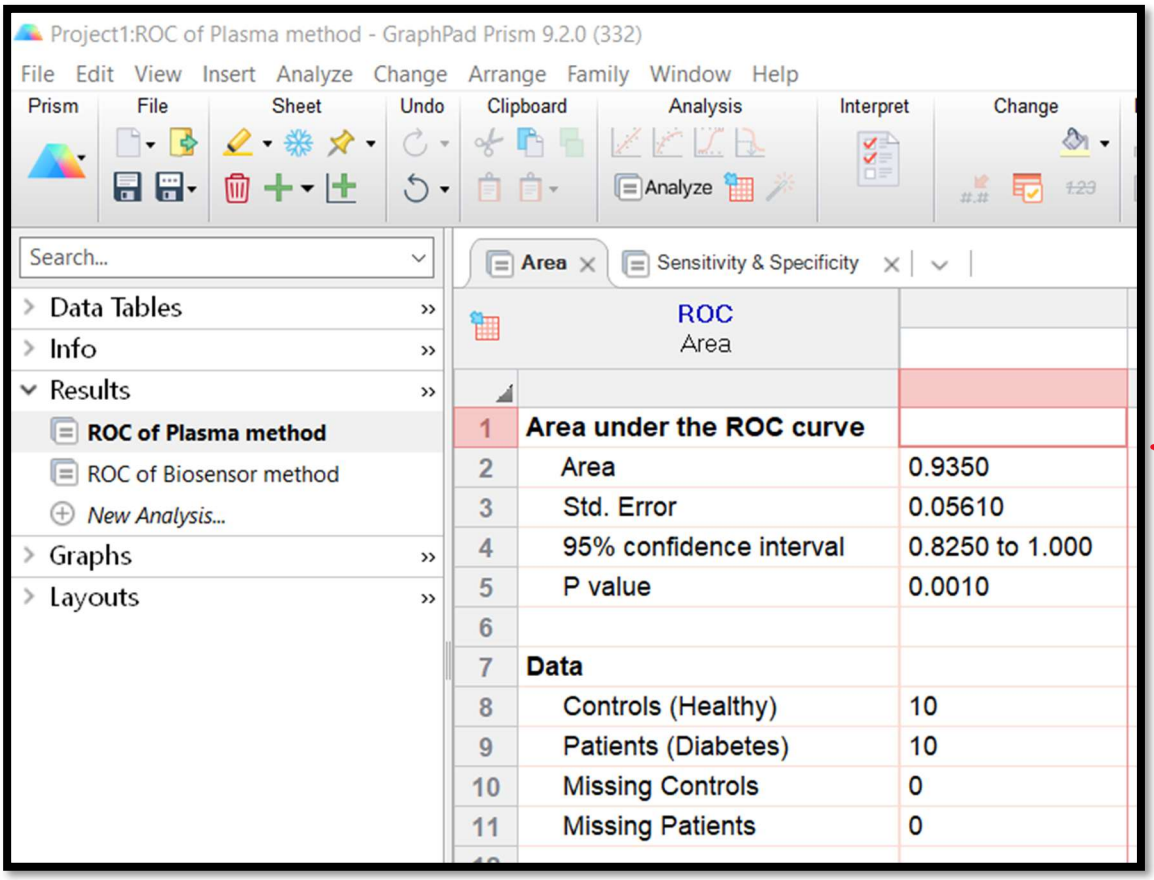
Saliva Glucose Biosensor		Counts		Cumulative				
mmol/L		Healthy	Diabetes	Healthy	Diabetes	FPR	TPR	AUC
0		0	0					
<4		0	1	0	1	1	0.9	
4 - 4.99		0	1	0	2	1	0.8	
5 - 5.99		3	0	3	2	0.7	0.8	
6 - 6.99		1	2	4	4	0.6	0.6	
7 - 7.99		4	3	8	7	0.2	0.3	
8 - 8.99		2	1	10	8	0	0.2	
9 - 9.99		0	1	10	9	0	0.1	
10 - 10.99		0	0	10	9	0	0.1	
11 - 11.99		0	1	10	10	0	0	
12 - 12.99		0	0	10	10	0	0	
13 - 13.99		0	0	10	10	0	0	
14 - 14.99		0	0	10	10	0	0	

Excel being used initially to plot the ROC curves.



GraphPad Prism used to find AUC of Biosensor method. AUC of Biosensor method = 0.590

Appendix continues on next page …



GraphPad Prism used to find AUC of Plasma method. AUC of plasma method = 0.935