This exploratory data analysis examined the changes in the existence of women in the summer Olympics. The debut of the Olympics was in 1896, and four years later in 1900, women were allowed to participate in the games as well. The first year, there were two events women could participate in, tennis and golf, and a total of 22 women participated that year. Since then, more and more women have joined the Olympic games and numerous sports and events followed. The International Olympic Committee(IOC) oversees all the other departments that are involved in putting on the Olympic Games. It was created in 1894, and currently is 30.8% female. The initial presence of women in the International Olympic Committee (IOC) began in 1981 when Flor Isava-Fonseca and Pirjo Haeggman were included. Has the presence of women in the summer Olympics changed since 1981 when women began sitting on the IOC?

The data involved seven variables related to women in the Olympics. This includes the year, sports, women's events, total events, percent of women's events, women participants, and percent of women participants. These events spanned from 1900 to 2016, the last summer Olympic games. Observations increased in each variable over the time span, making the PMF difficult to interpret since where were hardly any repeats if any. The CDF showed that before women were in the IOC, there was not a rapid increase of participants, however, after 1981 there was a rapid increase over a short period of time, making the Postioc CDF more horizontal. The CCDF of women participants indicated that the analytical distribution was exponential. Each of the variables when plotted against each other had some sort of positive relationship with points going from the bottom left to top right corner of the graph. Correlation and covariance of two situations, women participants vs. year and women events vs. women participants, were very extreme to support a conclusion of a strong positive relationship for each situation. The hypothesis test was run on women participant data from 1900-1980 and 1981-2016, labeled Preioc and Postioc. The alternative hypothesis was that the means for these two groups were not equal, hypothesizing that there was some change in women participants from before women's inclusion in the IOC and after. The p-value from this test was less than 0.0 so we can reject the null hypothesis that the means of each group are equal. Regression was the most difficult and posed the greatest challenge. Overall, the linear regression analysis resulted in an R^2 of 0.94 and p-value of less than 0.000 and standard error or 0.0013.

I felt like I needed more data after 1981, making the pre and post time more event and less unbalanced. This aspect gave me trouble when trying the hypothesis test most of all. I felt the variables that I had were sufficient and explained a lot. The areas I did not fully understand completely were how to truly find the best analytical distribution, which I felt my assumptions for the exponential distribution may be incorrect, as well as the regression analysis. I looked at numerous different ways to perform the regression analysis that I think I confused myself on how to fully analyze the regression using only one method.