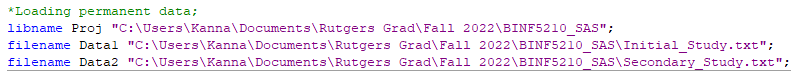
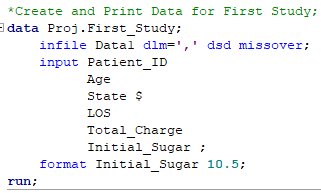
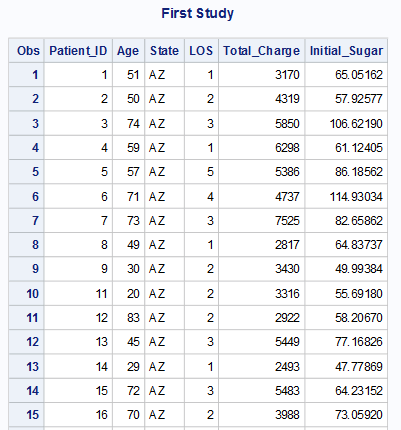
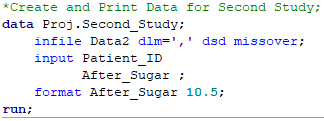
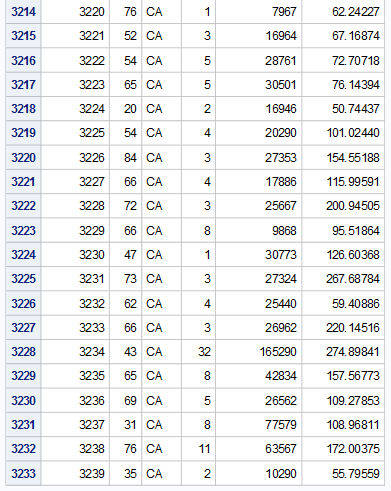
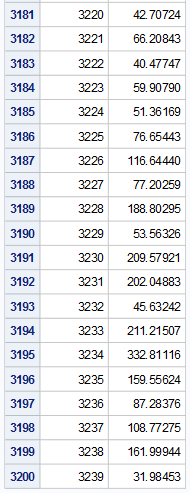
* **AN OVERVIEW OF THE PROJECT**
  + **The Study**
    - A study was conducted to determine whether Drug A helps to reduce fasting sugar level.
    - There are 3400 patients enrolled for this study and fasting sugar levels were taken before and after at a certain interval.
  + **The Objective**
    - I want to know for patients with blood sugar levels between 20-800 if there is a difference between before and after taking Drug A in blood sugar levels.  
      If so, was that difference a significant decrease in blood sugar levels?
    - I also want to know if there is a difference in after sugar levels between the age groups. Who did it affect the most? Does age group and initial sugar levels have an effect on after sugar levels?
    - Similarly, I want to know if there is a difference in after sugar levels between the states. Who did it affect the most? Does state location and initial sugar levels have an effect on after sugar levels?
    - Is there an interaction effect for age group and state?
    - Is length of stay correlated to initial sugar levels and after sugar levels?
    - If only observing normal blood sugar levels (between 70-120), is the result the same?
  + **Hypothesis: Null/Alternative**
    - H0: There is no difference between blood sugar levels before and after.  
       There is no difference between blood sugar levels between the age groups.
    - H1: There is a significant difference between blood sugar levels before and after, showing a decrease in levels.  
       There is a difference between the age groups blood sugar levels where it affects one group more than the others.
* **METHOD**

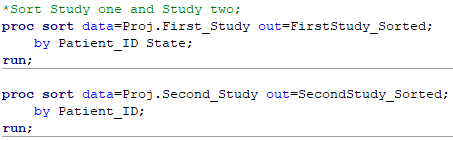


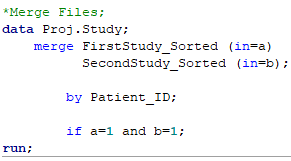
* + **How was the data cleaned up? Why?**

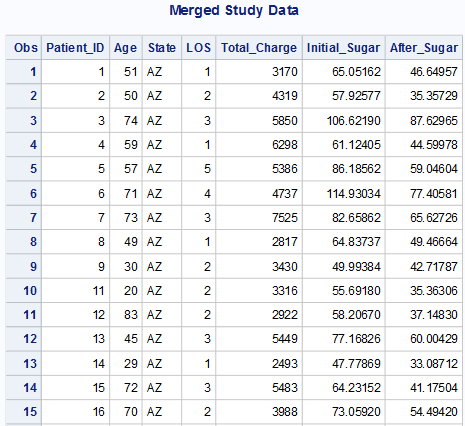
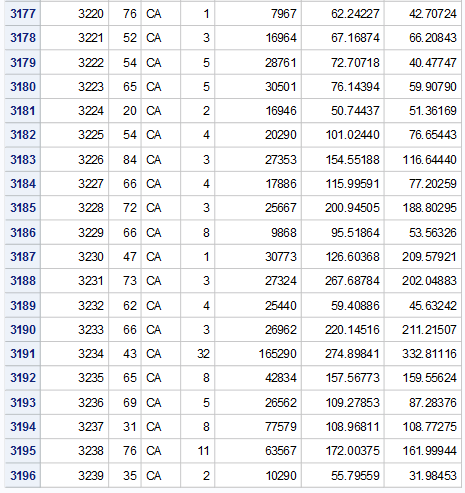
 **🡪**

 **…**   
🡪

 **…**



 **🡪**

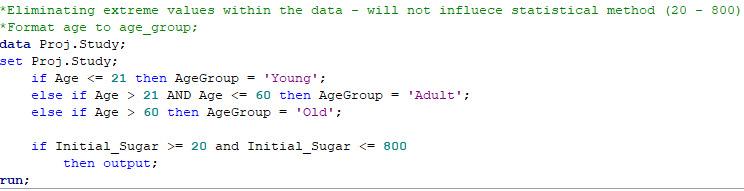
 **…**

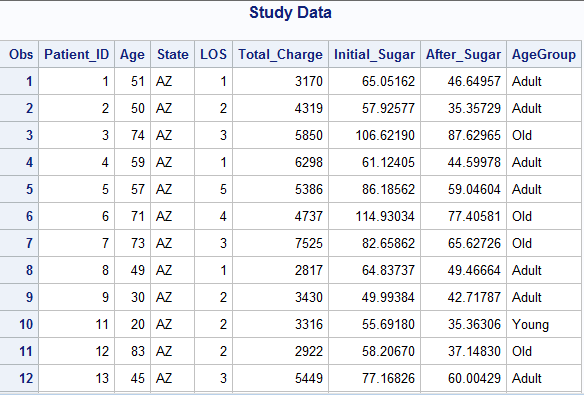
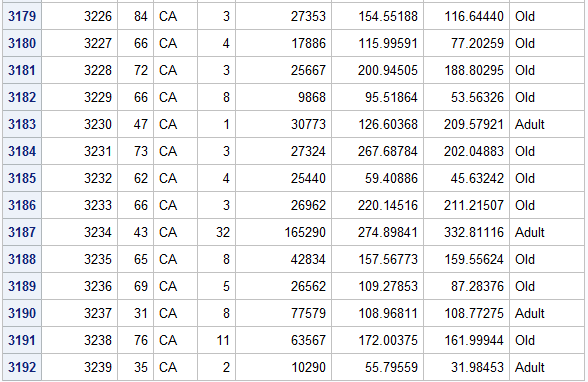
Before merging

* + - Each variable value is either numerical or a character.   
      Any other invalid values are marked as null (.) values.

ex)

* + - * STATE: 3 🡪 .
      * LOS: X 🡪 .
    - The *Missover* option will take care of values that are missing within the data
    - Both study 1 and study 2 datasets were sorted by patient\_id and state before merging
    - Data was merged by patient\_id – only if in both study 1 and study 2 datasets patient\_id matches
    - Total OBS found in the dataset is 3196.

  
  
🡪

  
…   


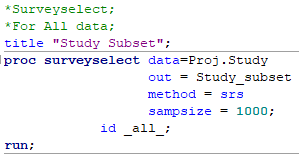
After merging

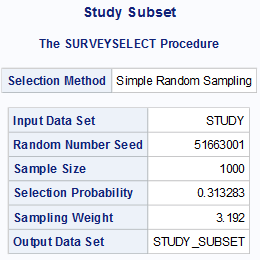
* + - I then eliminated extreme values within the data
      * Cleaning up blood sugar to levels cutoff at above 800 and below 20.
      * These numbers can create dangerous situations for patients

- above 600 can cause diabetic coma which is hyperosmolar. Anything between 600-800 is abnormal

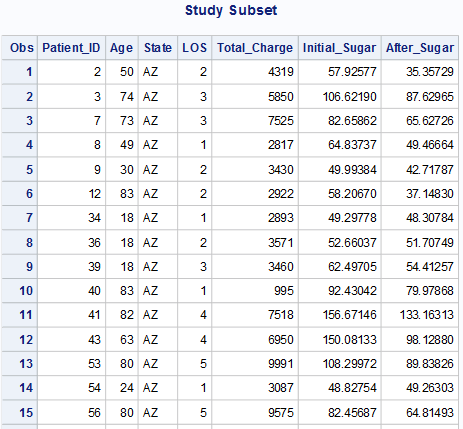
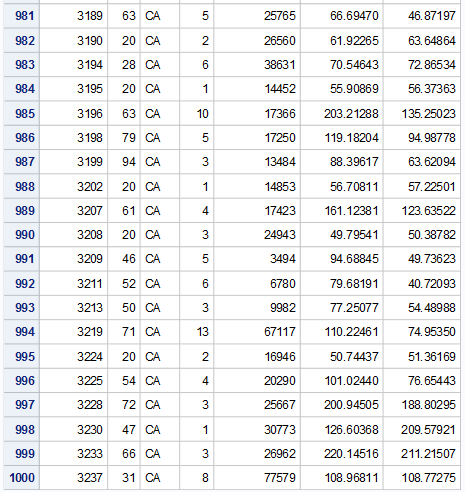
- below 40 is hypoglycemia, creating severe dehydration which can also lead to coma and/or death. Between 20-40 is abnormal

* + - Total OBS found in the dataset after elimination of extreme values is 3192.
    - Age was then formatted into 3 categories
  + **Final Subset Data for Analysis**

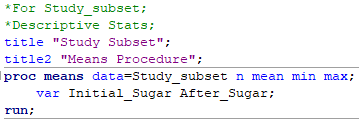
 **🡪**

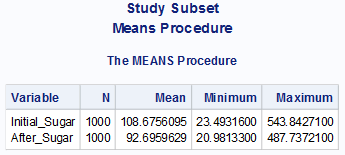


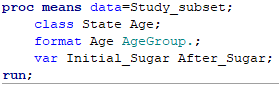
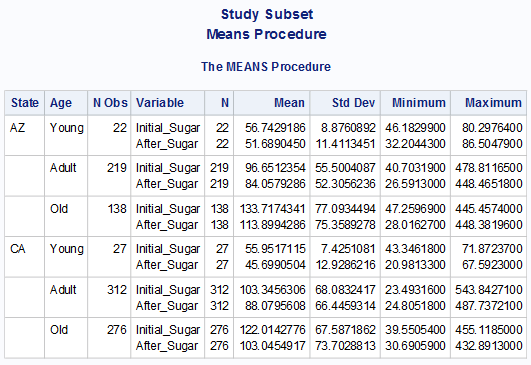
 **🡪**

 **…**

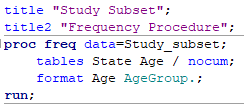
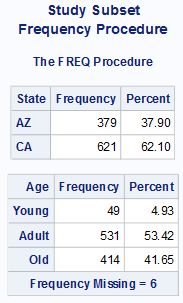
* + - A subset of the final data was taken using surveyselect, limiting the random sampling to a sample size of 1000.
    - Total OBS derived from the simple random sample is 1000.
  + **Descriptive of data**

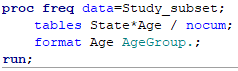
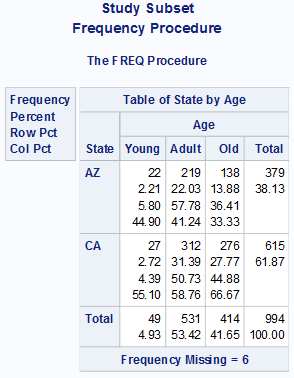
 **🡪**



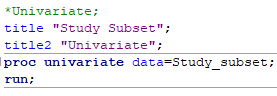
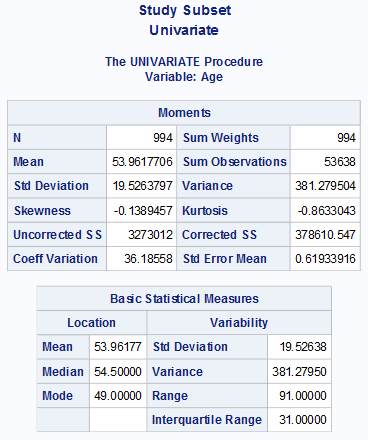
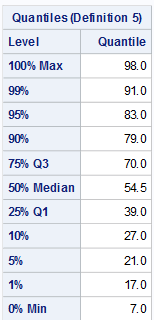
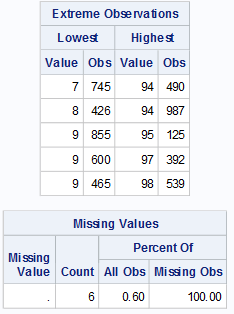
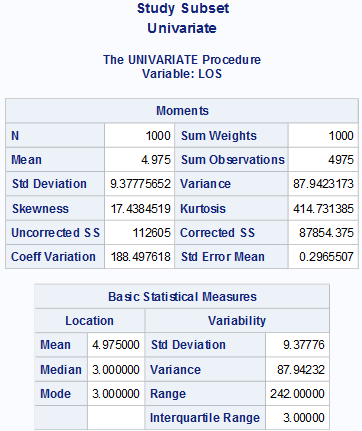
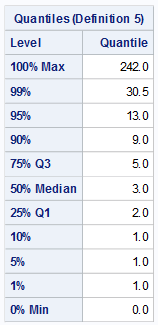
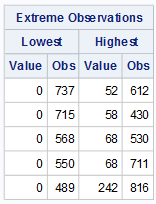
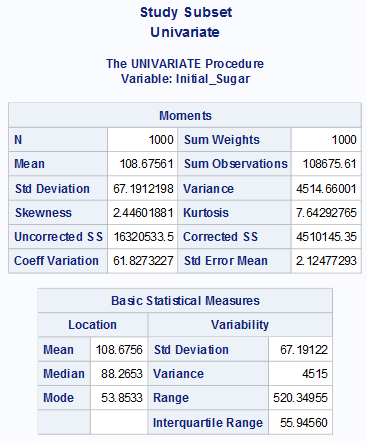
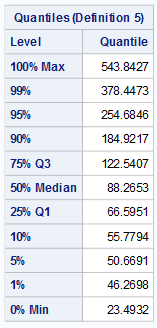
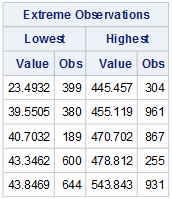
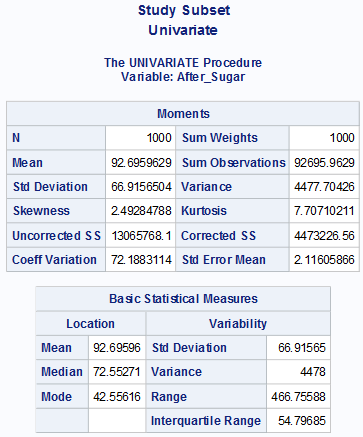
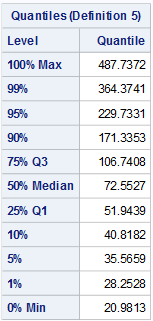
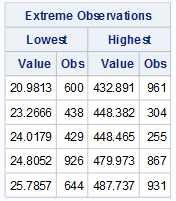
 **🡪**

* + - Procedure Means – will provide the min and max of the data, showing the spread, range, and average

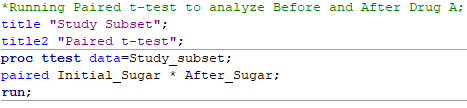
  
🡪  


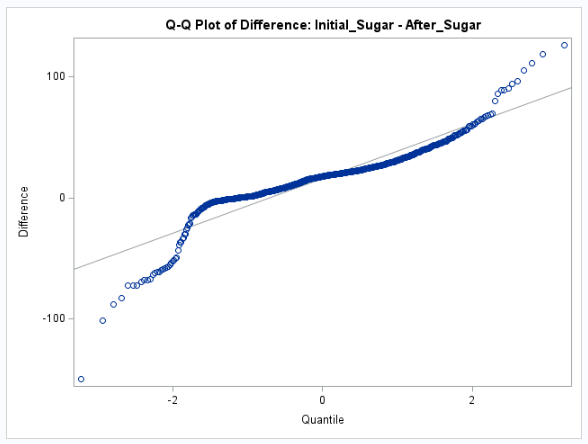
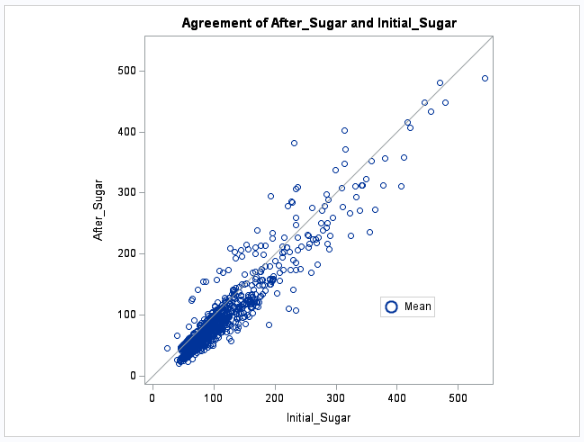
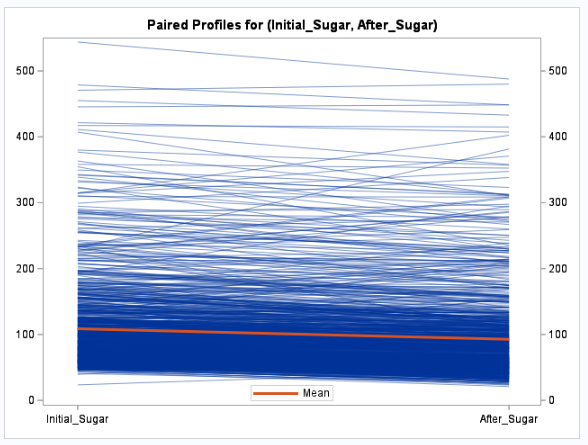
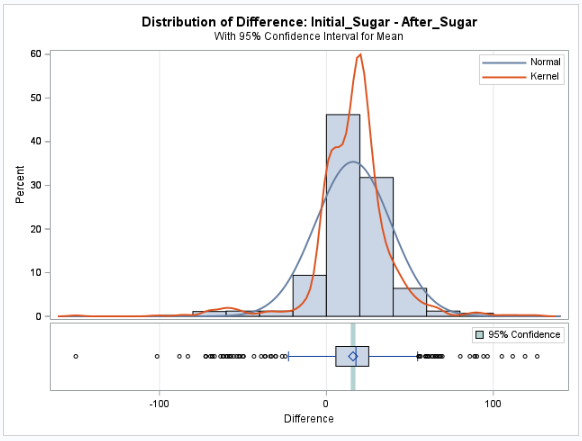
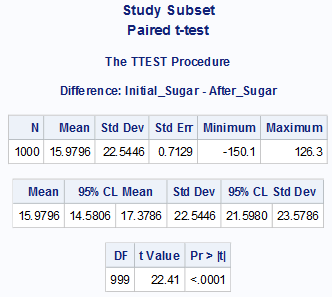
  
🡪  


* + - Procedure Freq – will show total of patients that fall in specific state and/or age group

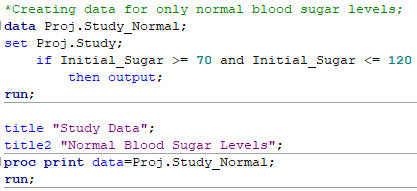
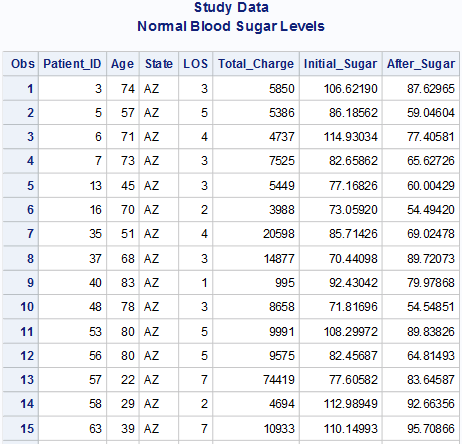
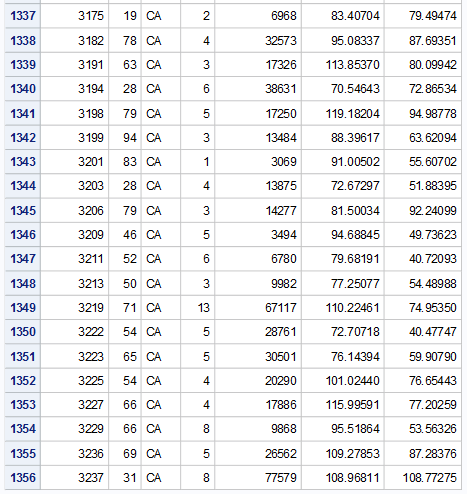
  
🡪  
  
   
  
   
  
   
  
 

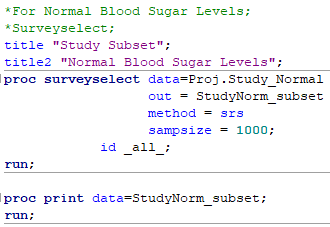
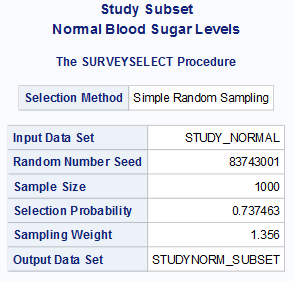
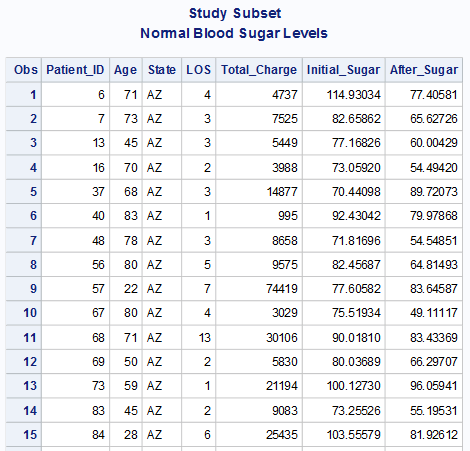
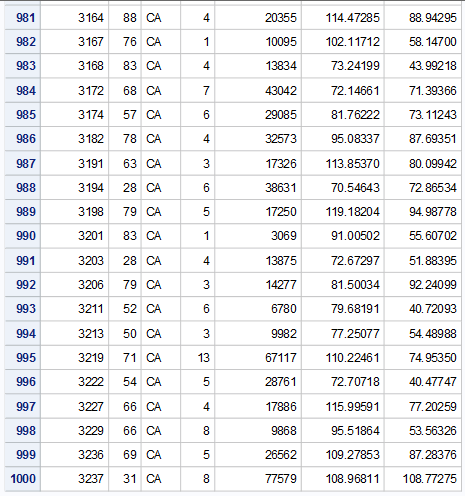
* + - Univariate – will tell us normality
  + **Paired t-test**

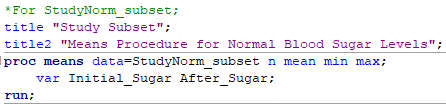
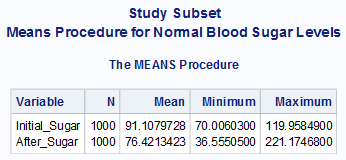
 **🡪**

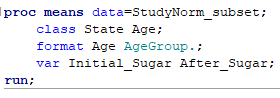
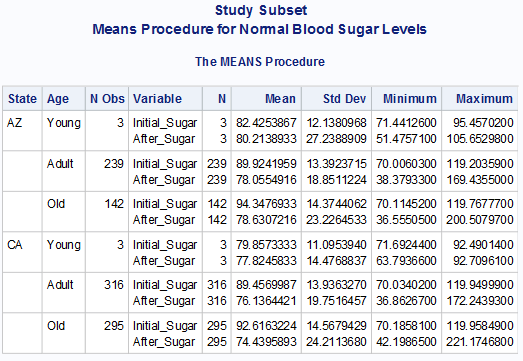


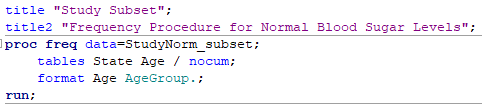
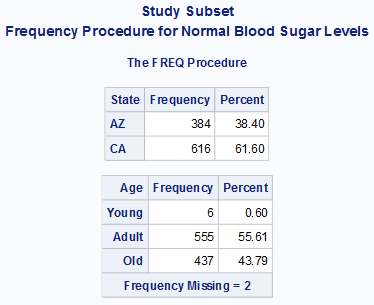
* + - Before and After Drug A
* RESULTS/FINDINGS
  + What is the result
  + What does this mean
  + What happens if the sample selection only for patients with normal blood sugar levels (70-120)? Does this change the results?

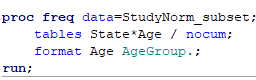
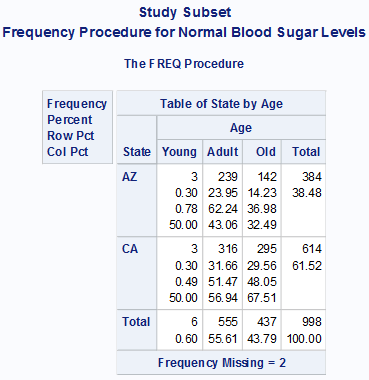
  
🡪  
  
…   


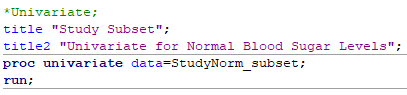
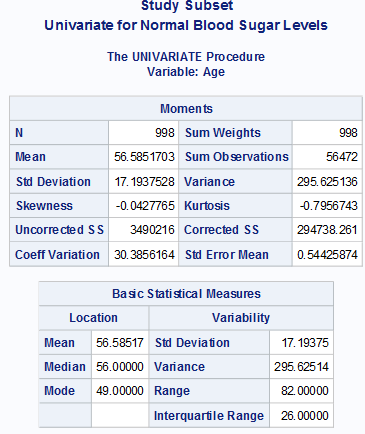
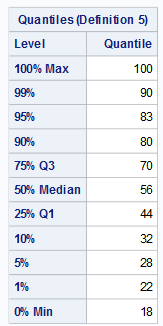
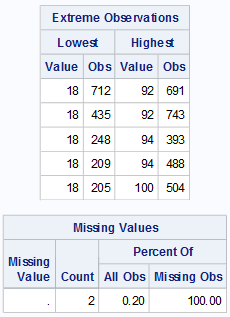
  
🡪   
  
  
…   


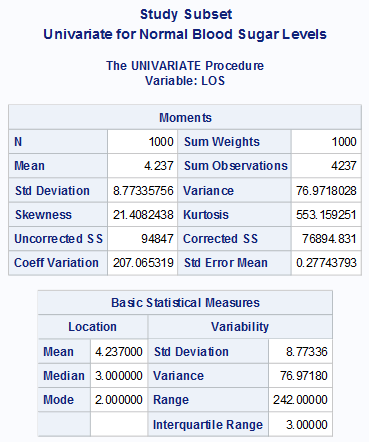
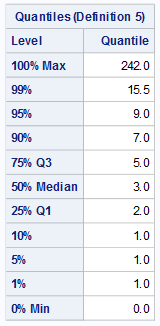
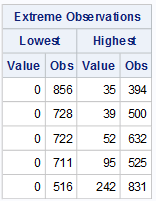
  
🡪  


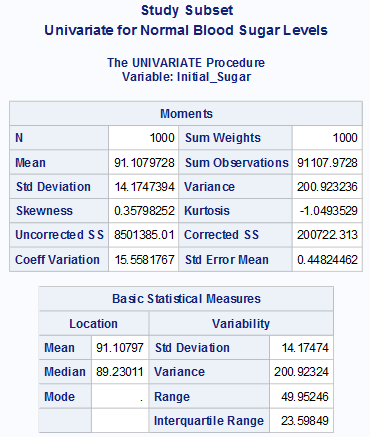
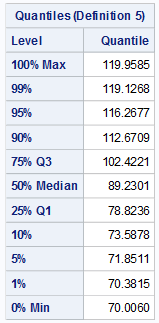
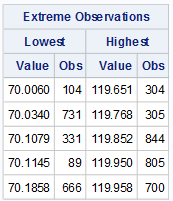
  
🡪  


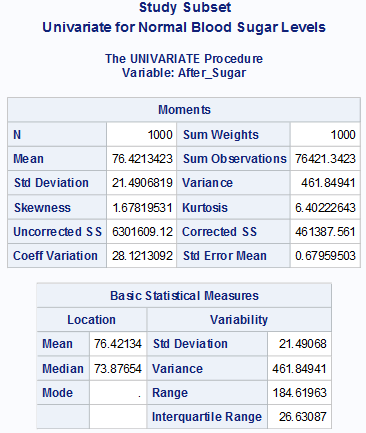
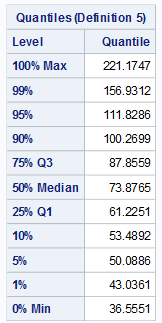
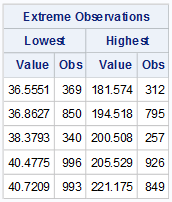
  
🡪  


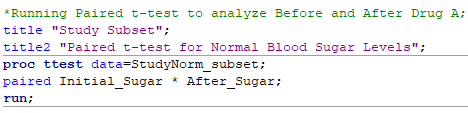
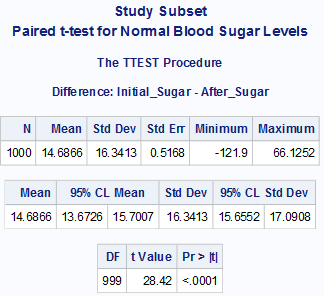
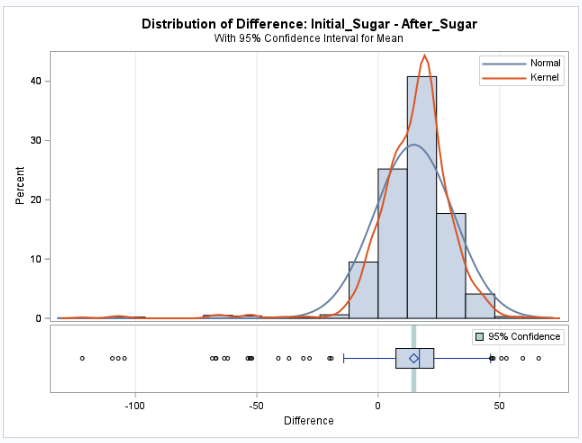
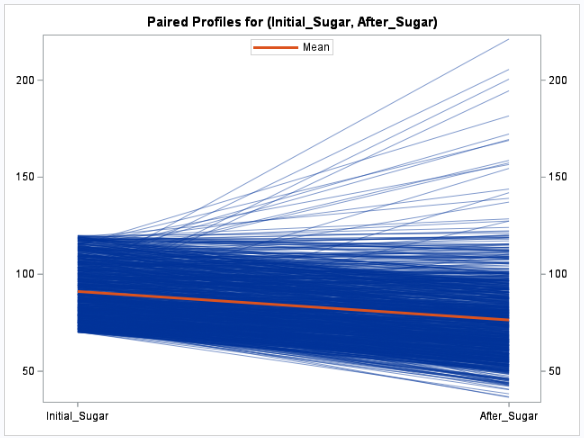
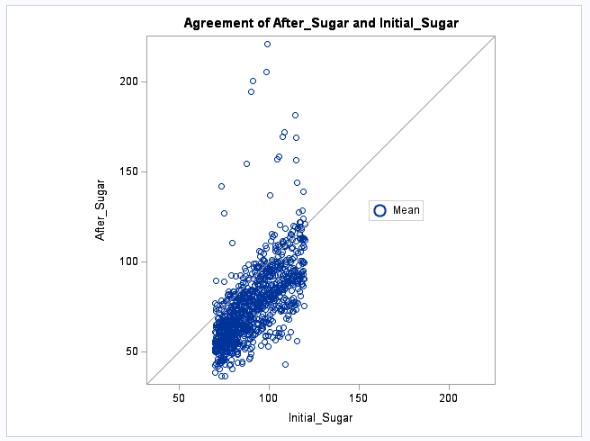
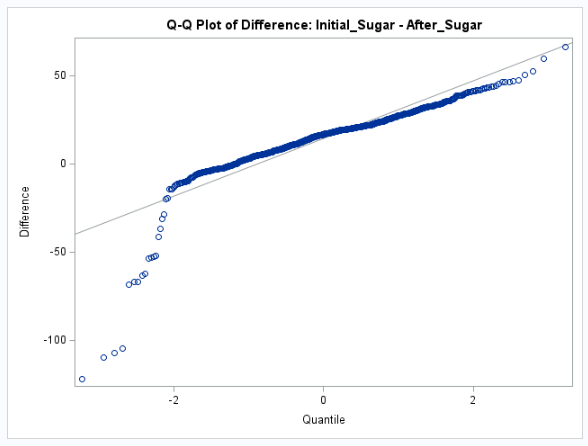
  
🡪   


  
🡪   
  
 

  
🡪  
  
  
  
  


* CONCLUSION
  + Was null/alt rejected?

For Overall Data Study: The null hypothesis was rejected…Conclusion. There was a decrease in blood sugar levels.  
For Normal Blood Sugar Levels Only Data Selection: The null hypothesis was rejected however the trend was not correctly assumed. Conclusion. There was an increase in blood sugar levels for patients with initial normal blood sugar levels.

* + What does this mean
    - The drug A was only effective for patients with range outside normal blood sugar level. If the hypothesis were accurate for normal blood sugar level Data sample, we should have seen a similar trend, showing a decrease in blood sugar levels.