SASEM 2B Exercise

Fundamental Summary Analytics & Filtering

(Fall 2017)

**Sources** (adapted with permission)**-**

T. P. Cronan, Jeff Mullins, Ron Freeze, and David E. Douglas Course and Classroom Notes

Enterprise Systems, Sam M. Walton College of Business, University of Arkansas, Fayetteville

Microsoft Enterprise Consortium

IBM Academic Initiative

SAS® Multivariate Statistics Course Notes & Workshop, 2010

SAS® Advanced Business Analytics Course Notes & Workshop, 2010

Microsoft® Notes

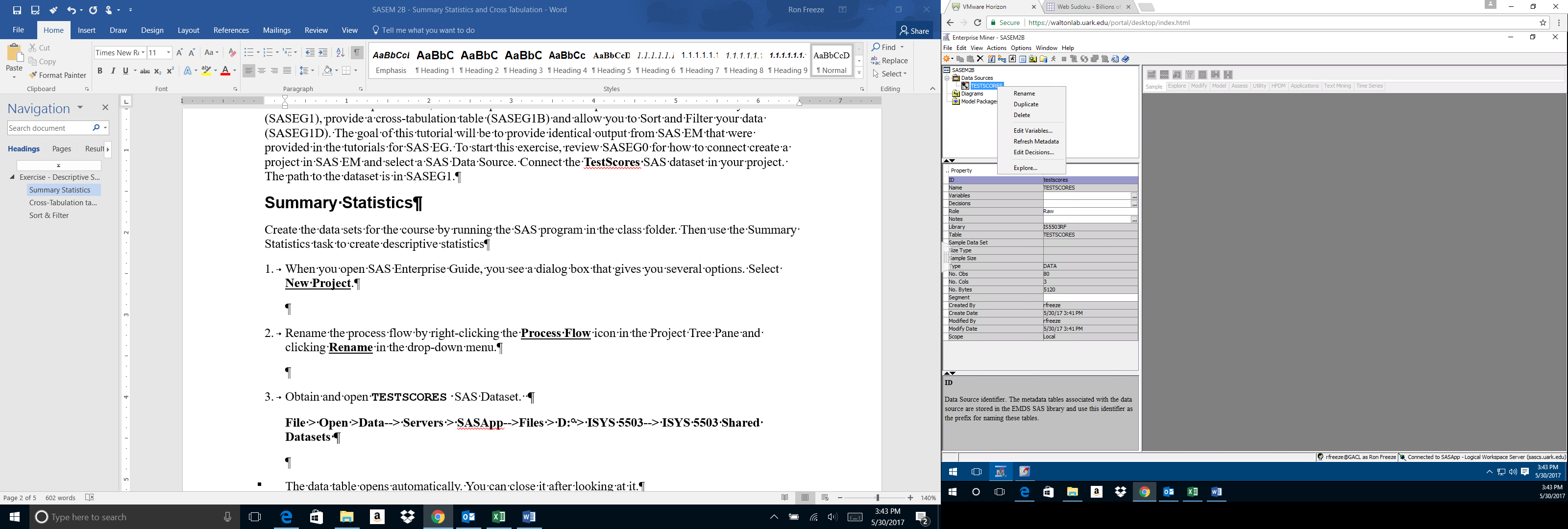
Teradata® University Network

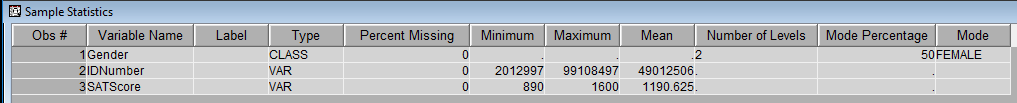
Copyright © 2013 ISYS 5503 Decision Support and Analytics, Information Systems; Timothy Paul Cronan. *For educational uses only - adapted from sources with permission. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission from the author/presenter.*

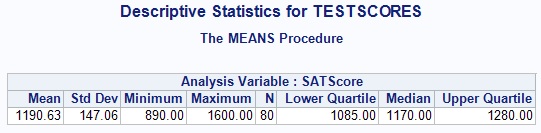
# Exercise - Descriptive Statistics

Similar to SAS Enterprise Guide, SAS Enterprise Miner can also provide a review of some summary statistics (SASEG1) and allows you to Sort and Filter your data (SASEG1D). The goal of this tutorial will be to provide identical output from SAS EM that were provided in the tutorials for SAS EG. To start this exercise, review SASEG0 for how to connect create a project in SAS EM and select a SAS Data Source. Connect the **TestScores** SAS dataset in your project. The path to the dataset is in SASEG1.

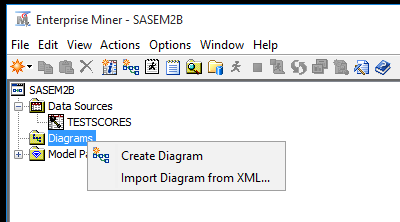
## Summary Statistics

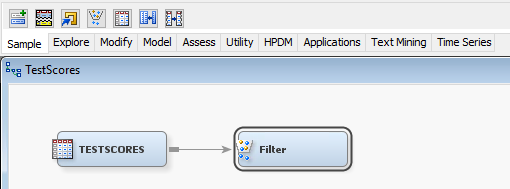
A review of the summary Statistics for a file or dataset is done with the **Explore** command in **SAS** **Enterprise** **Miner**.

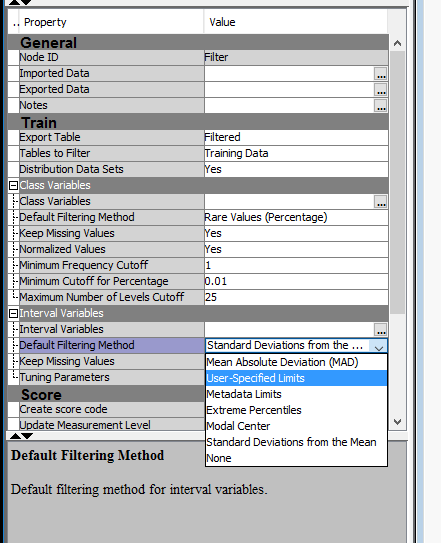
1. Access **Explore** by right clicking on the **TestScores** data set under **Data** **Sources** in the **Task** **Tree**
2. Select **Explore**
3. There are three pop-up boxes created
   1. **Sample Properties** – this box describes the dataset that you are exploring. Note that the **Rows** and the **Fetched** **Rows** are the same number at 80. Since there has been to partitioning of data, these are the same. The **Columns** rows provide an indication of the number of variables – in this case there are three variables in the dataset. There are six other descriptors that will be discussed at a later date.
   2. **IS5503RF**.**TestScores** – The name for this pop-up will change with each dataset that you **Explore**. However, this is a listing of the actual data in your dataset for you to review.
   3. **Sample** **Statistics** – This pop-up provides the initial summary statistics for the dataset. Expand the pop-up to full screen and expand the columns to be able to read the names. Note that the **Variable** **Name** Gender is a **Type** Class with the **Number** **of** **Levels** at 2. The **Mode** is FEMALE and **Mode** **Percentage** is 50%. It does not make sense to provide a **Minimum**, **Maximum** and **Mean** value for a Class **Type** and so there is none. For the **Variable** **Name** SATScore, there is a **Minimum**, **Maximum** and **Mean** value provided, but not a **Number** **of** **Levels** since this is a VAR **Type**.
4. Compare what is provided in the Explore with the output from SAS EG. SAS EM does not provide some of the descriptive aspects that SAS EG provides. The following is the output from SASEG1.

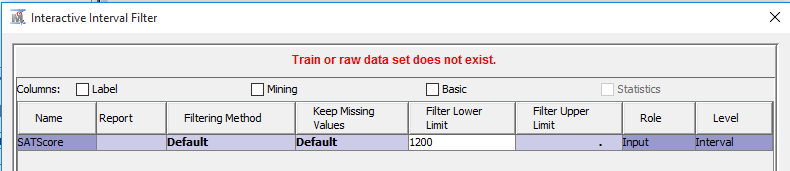
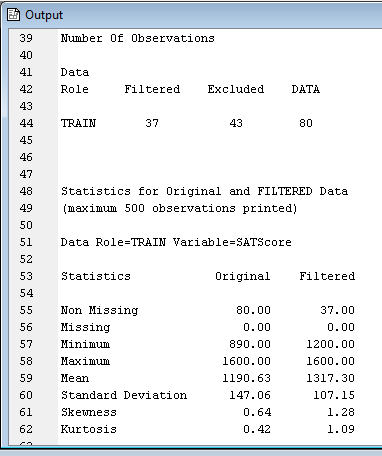


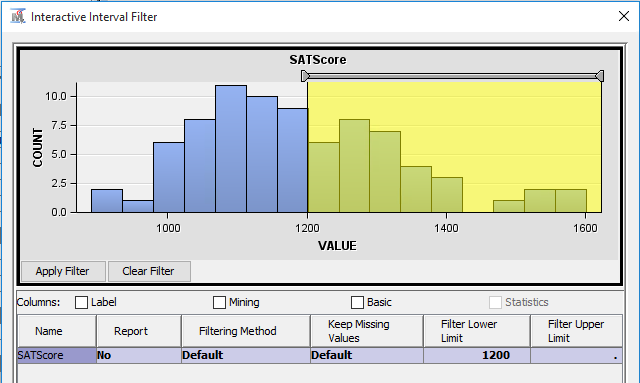
## Filter

SAS Enterprise Miner also allows the ability to **Filter** your data. In order to do this, a diagram will need to be created.

1. Right Click on **Diagrams** in the **Task** **Tree** and select **Create** **Diagram**
2. Name the **Diagram** TestScores
3. Drag and Drop the TESTSCORES Data Source to the diagram
4. Select the **Sample** tab
5. Drag and Drop the **Filter** node to the diagram
6. Connect the TESTSCORES data source node to the **Filter** node

The property panel on the left provides the specifics for filtering the data you are interested in passing along to the next node. Train is where the filtering you desire is specified. For this example we will include only those SATScores greater than 1200. Recall from our Summary Statistics that N = 80.

1. In the Property panel, select the down arrow for the following path – Train -> Interval Variables -> Default Filtering Method and select User-specified Limits
2. In the Property panel, select the for the following path – Train -> Interval Variables -> Interval Variables
3. Place 1200 in the Filter Lower Limit cell of the Interactive Interval Filter pop-up
4. Select OK
5. Right click on the Filter node and select Run – Yes
6. Once Run has completed, select Results
7. Scroll through the Output pop-up and note the following
   1. Number of Observations – 43 of the 80 observations have been excluded
   2. Statistics -> Minimum – Original SATScore was 890, Filtered SATScore is 1200
   3. Statistics -> Mean – Original SATScore was 1190.63, Filtered SATScore is 1317.30
8. Close the Results pop-up



1. Open the Interactive Interval Filter again – Step 12 above
2. Note the interactive histogram that allows a selection of SATScores via a slider
3. Close the Interactive Interval Filter
4. The TestScores diagram has a green check next to both the TESTSCORES node and the Filter node. This indicates that both nodes have been ran.
5. Save and close your project