## SASEG – Performing Cross-tabulation

(Spring 2016)

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

Ron Freeze 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

*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.*

Example:*Pelican Stores, a division of National Clothing, is a chain of women’s apparel stores operating throughout the country. The chain recently ran a promotion in which discount coupons were sent to customers of other National Clothing stores. Data collected for a sample of 100 in-store credit card transactions at Pelican Stores during one day while the promotion was running are contained in the file named PelicanStores. The Proprietary Card method of payment refers to charges made using a National Clothing charge card. Customers who made a purchase using a discount coupon are referred to as promotional customers and customers who made a purchase but did not use a discount coupon are referred to as regular customers. Because the promotional coupons were not sent to regular Pelican Stores customers, management considers the sales made to people presenting the promotional coupons as sales it would not otherwise make. Of course, Pelican also hopes that the promotional customers will continue to shop at its stores.*

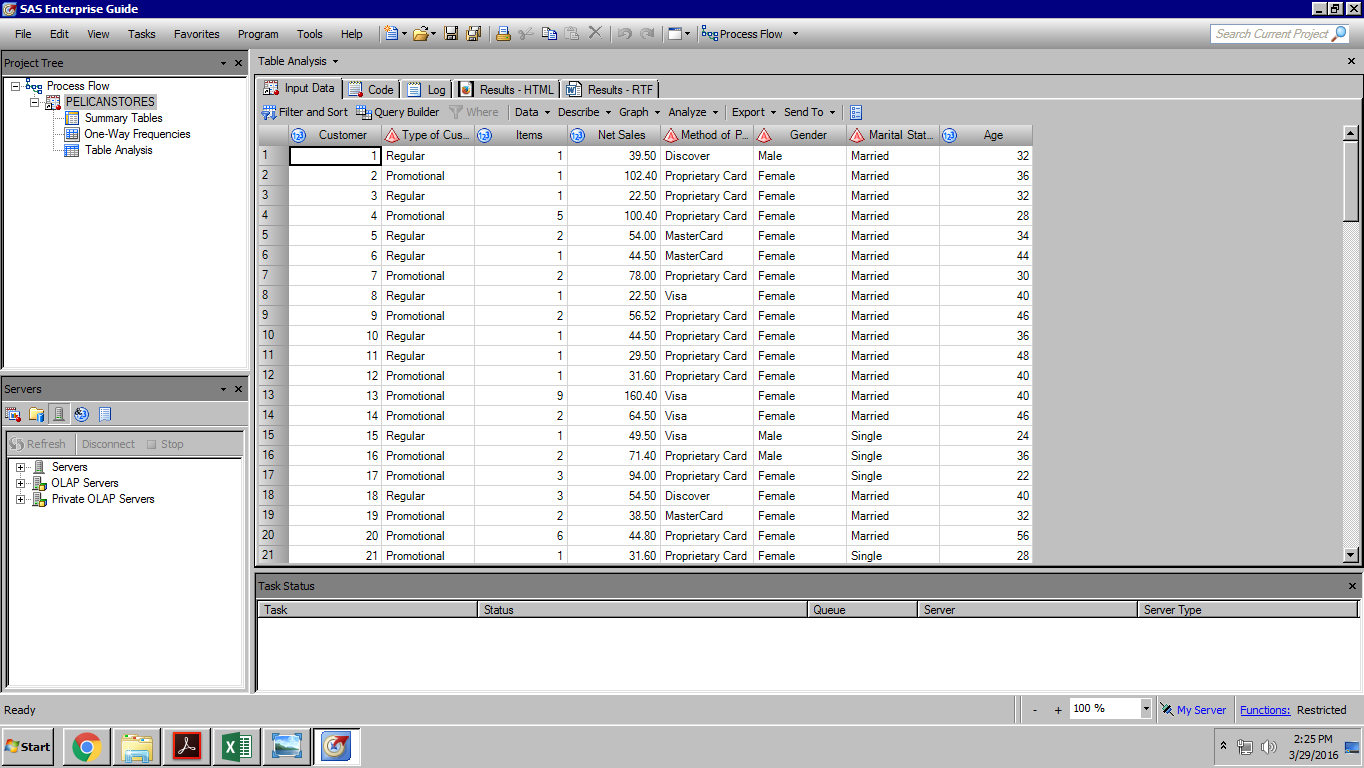
SASEG provides several tools to analyze given variables in a problem. We have used One-Way Frequencies to obtain the frequency distribution of individual variables. It always adds to the analysis if we have drilled-down information within variables. In the above case, we can obtain the number of customers and its frequency distribution by different variables (Method of Payment, Type of Customer, etc.). It would be helpful to dive deeper to get the number of customers by a particular **Method of Payment** within a particular **Type of Customer**. A simple way to achieve this is by running cross-tabulation.

## C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise – Cross-tabulation across Variables

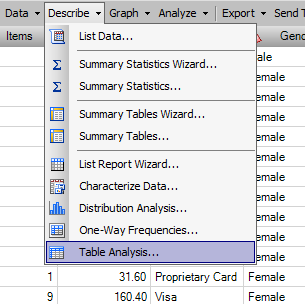
Obtain cross-tabulated information for Method of Payment vs. Type of Customers variables for the PELICANSTORES dataset.

1. Open the **PelicanStores** SAS Dataset using the following path:

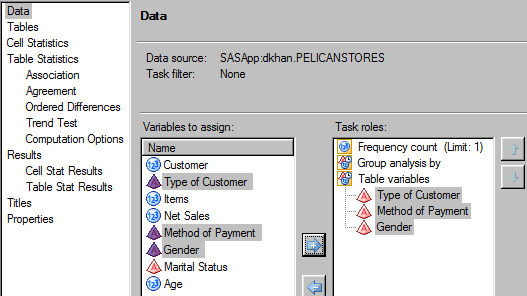
**File > Open >Data--> Servers > SASApp-->Files > D: > ISYS 5503--> ISYS 5503 Shared Datasets--> Pelican Stores**



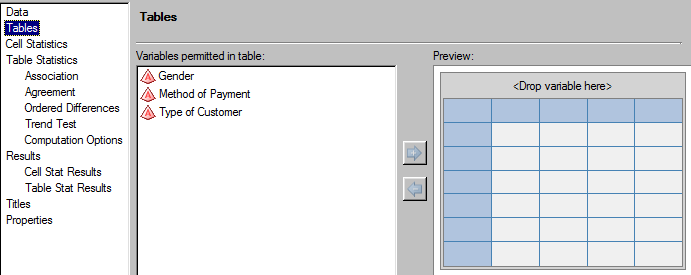
2. Select **Describe**  **Table Analysis**



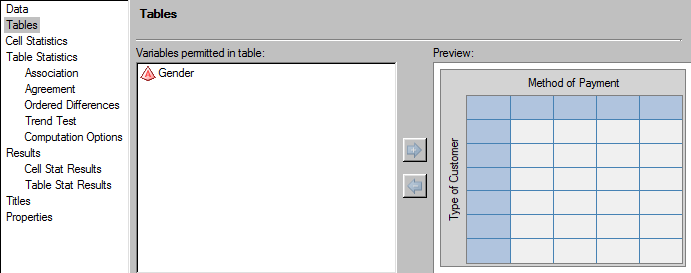
3. With **Data** selected on the left, drag and drop **Method of Payment**, **Type of Customer, and Gender** in the Table Variables field



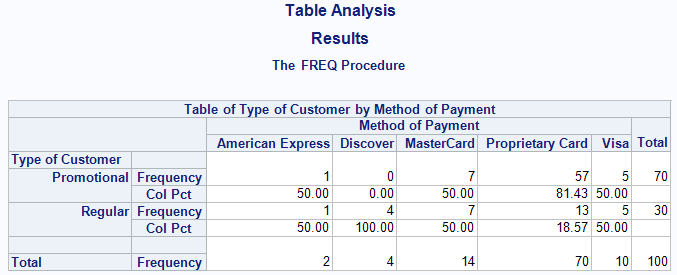
4. With **Tables** selected on the left, drag and drop **Method of Payment** and **Type of Customer** in the Preview table area. You can adjust which variable to be represented in a row/column. Please note, **Gender** variable will be used later in the exercise.



After you drag and drop the variables, the preview screen should look like below:



5. Click 



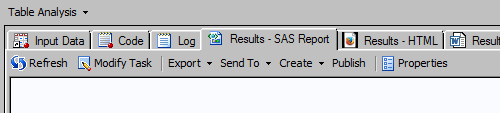
The result shows a cross-tabulation between the two variables we selected. We can now see drilled-down information within each variable. For example, among 100 customers, 70 customers paid using Proprietary Card, out of which 57 were Promotional customers.

The table also provides percentage distribution of Method of Payment variable (Col Pct or Column Percentage). We can see 81.43% of Proprietary customers are Promotional customers, while 18.57% are Regular customers. However, the table only includes percentage distribution for columns and not for rows.

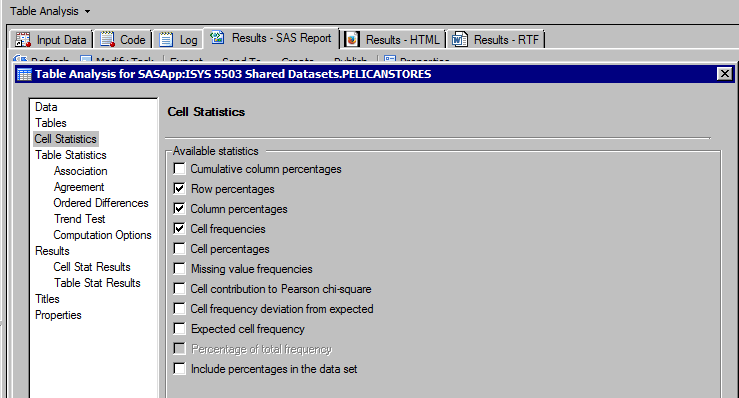
## C:\Program Files\PowerServ\CourseGraphics\demo_eye.jpgExercise – Displaying Row Percentages in Output

What if we wanted to know the percentage distribution for Type of Customer variable (or Row Pct)? For example, looking at the table we can get the percentage of Proprietary customers within Promotional category, but we do not know percentage of Promotional customers within Proprietary category. To have that information included in the table, we will modify the task.

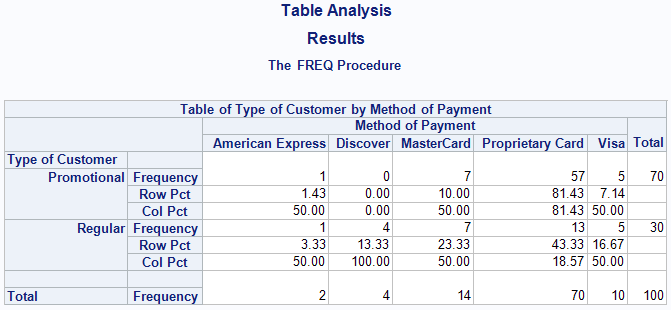
6. Click **Modify Task**



7. With **Cell Statistics** selected on the left, check the **Row percentages** box.



8. Click 

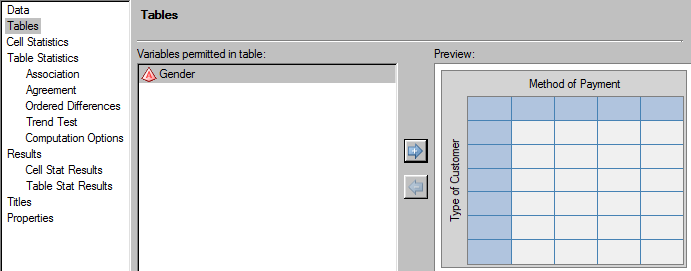


The output table now displays the row percentages (Row Pct) for Type of Customer variable. We can see the percentage of Promotional customers within Proprietary category.

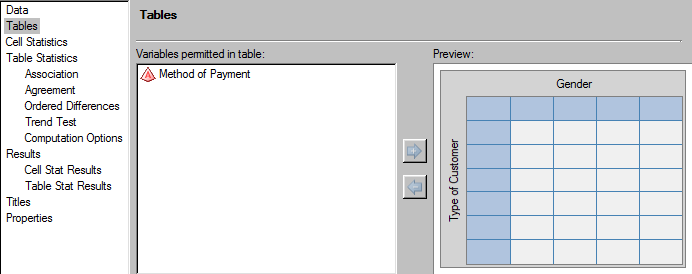
Now, suppose we want cross-tabular information for Type of Customer vs. Gender. We will modify task and replace Method of Payment with Gender variable.

9. Click ****

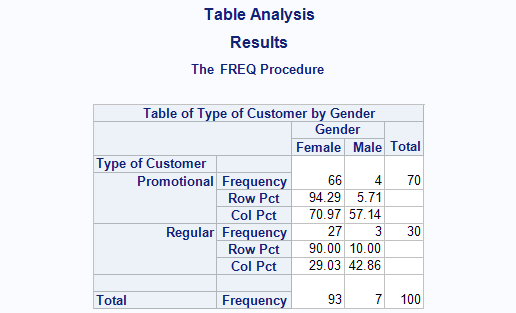
With Tables selected on the left, replace **Method of Payment** by **Gender** by dragging and dropping **Gender** over **Method of Payment** in the preview area. You could also use arrows to make the selection.



After you drag and drop the Gender over Method of Payment, the preview screen should look like below:



10. Click 



We now have the cross-tabulation results for Type of Customer vs. Gender variable, displaying both Row and Column percentages.