



GHSUG User Group

The Means Procedure in SAS Enterprise Guide (EG)... “Summary Statistics”

Sueheir Saddik

April 15th, 2010

Outline

- The Means Procedure – Overview and Usage
- The Means Procedure – Proc Means Syntax
- SAS Enterprise Guide – Summary Statistics – Screen Shot and options
- Data Used and Information / Analysis Needed
- SAS Enterprise Guide – Summary Statistics
 - Data & Drop Down
 - Interface, Variable Selection, Statistics, Plots, Results, Tittles, Code Viewing ... Etc
 - Past Due Amount Summary
 - Data / Code Checking
 - Make it Pretty & Add Distributions
 - Profiling (Key Variables), Output, Summary
- Questions

The Means Procedure – Overview and Usage

- The Means procedure is one of the most powerful and useful procedures in SAS
- The Means procedure computes descriptive statistics for variables across the observations and within groups of observation with an option to store the statistics in a SAS dataset
 - Some of the computed statistics are:
 - Arithmetic Mean: $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
 - Variance: $Var(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$
 - Standard deviation: $SD(X) = \sqrt{Var(X)}$
- The Means Procedure can be used in; analytics, business intelligence, reporting, profiling, standard data checking for coded values created using exiting ones, checking for outliers / extreme values, missing values, data penetration, differences between population mean and certain group of observations (t-test) ... etc

The Means Procedure – Proc Means Syntax

- Simplified ...

```
proc means data = XXXX mean min max n nmiss;  
  var variable(s) name;  
run;
```

- Complicated SAS web site

PROC MEANS <option(s)> <statistic-keyword(s)>;

BY <DESCENDING> variable-1 <... <DESCENDING> variable-n><NOTSORTED>;

CLASS variable(s) </ option(s)>;

FREQ variable;

ID variable(s);

OUTPUT <OUT=SAS-data-set> <output-statistic-specification(s)> <id-group-specification(s)> <maximum-id-specification(s)> <minimum-id-specification(s)> </ option(s)> ;

TYPES request(s);

VAR variable(s) < / WEIGHT=weight-variable>;

WAYS list;

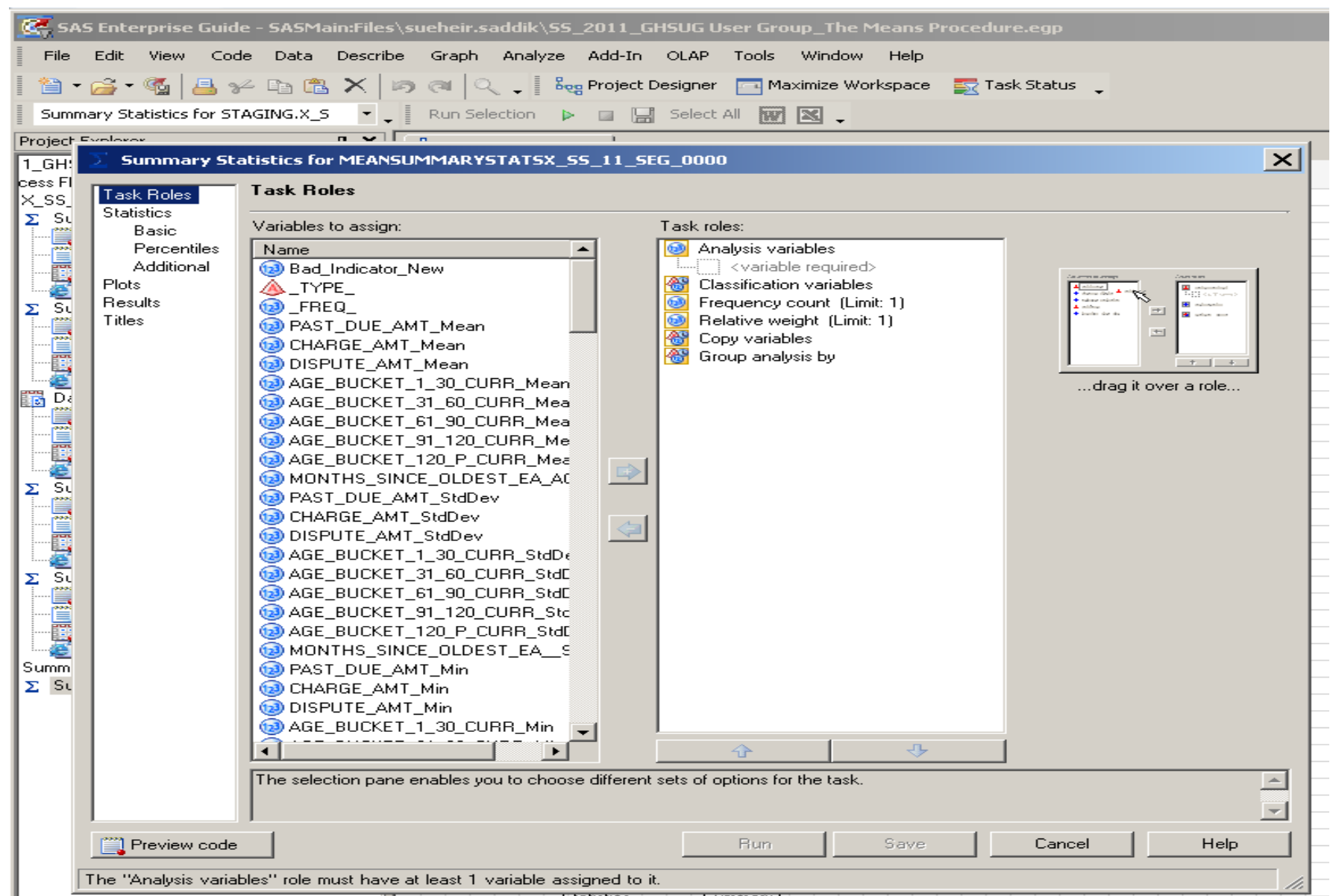
WEIGHT variable;

... Etc

RUN;

- SAS EG is user friendly, with drag and drop interface. There is no need to remember or memorize the above code ... 😊
 - More efficient
 - Less time consuming
 - Keeping it simple

SAS Enterprise Guide – Summary Statistics – Screen Shot



SAS Enterprise Guide – Summary Statistics

■ Task Roles

- Analysis variables :
 - numeric variables on which you want the statistics generated
- Classification variables:
 - no data sorting required
- Frequency count:
- Relative weight:
- Copy variables
- Group analysis by:
 - requires that the data set must be sorted ... The variables that you assign to this role are used to compute separate statistics for each distinct value or combination of values of the Group analysis by variables. The data is automatically sorted by the variables in this role before the statistics are computed.

■ Statistics

- Basic
- Percentiles
- Additional

■ Plots

- Histogram
- Box and Wisker

■ Results

- Save statistics to data set
- Show statistics

■ Titles

- Default
- specify

Statistics

■ **Basic statistics available:**

- Mean: is the arithmetic average, calculated by adding the values of a sample variable and dividing this sum by the number of observations.
- Std Dev = Standard deviation: is a statistical measure of the variability of a group of data values. This measure, which is the most widely used measure of the dispersion of a frequency distribution, is equal to the positive square root of the variance.
- Standard error: is the standard deviation of the sample mean. The standard error is defined as the ratio of the sample standard deviation to the square root of the sample size.
- Variance: is a statistical measure of dispersion of data values. This measure is an average of the total squared dispersion between each observation and the sample mean.
- Minimum: is the lowest value for an observation.
- Maximum: is the largest value for an observation.
- Range: is the difference between the largest and the smallest values in the data.
- Sum: is the sum of all observations.
- N = Number of observations: is the total number of observations that do not have a missing value.
- N Miss = Number of missing observations: is the number of observations for which no value is entered.

- You can specify the maximum number of decimal places here
 - By default, a statistic is displayed by using the best fit, which is usually 7 decimal places.

- Missing values are not included in the calculations produced by the means procedure

Statistics

- **Percentiles ... By default, no percentiles are selected.**

- 1st Pctl = 1st percentile
- 5th Pctl = 5th percentile
- 10th Pctl = 10th percentile
- Lower Quartile = 1st quartile = 25th percentile
- Median = 50th percentile
- Upper Quartile = 3rd quartile = 75th percentile
- 90th Pctl = 90th percentile
- 95th Pctl = 95th percentile
- 99th Pctl = 99th percentile

- **Additional**

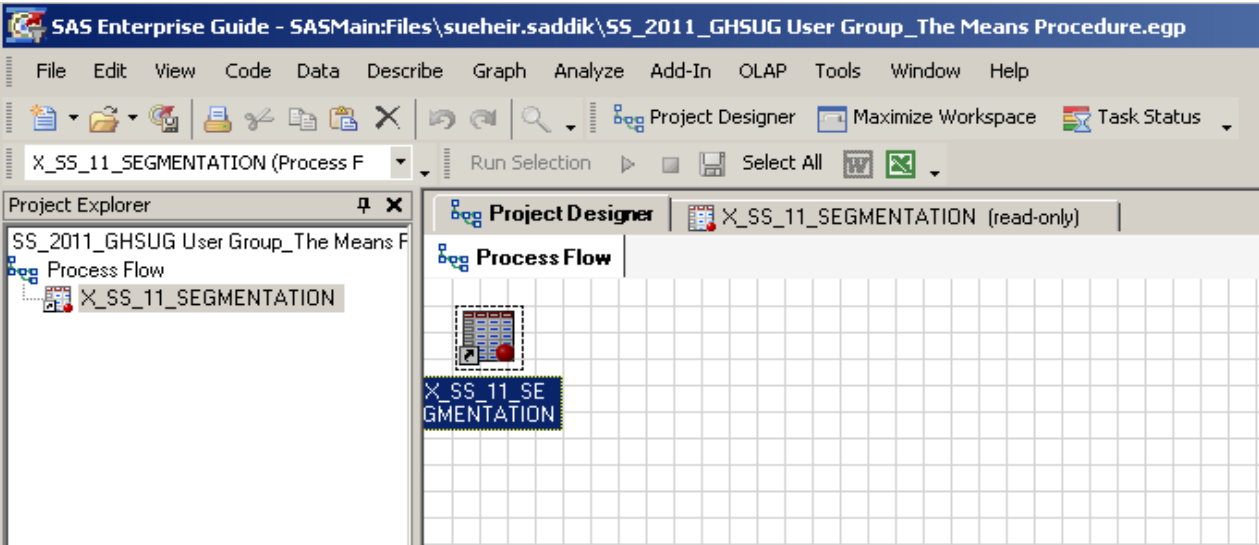
- Confidence limits of the mean
- T – statistics and prob > |t|
- Coefficient of variation
- Corrected sum of squares
- Un corrected sum of squares

Data Used and Information / Analysis Needed

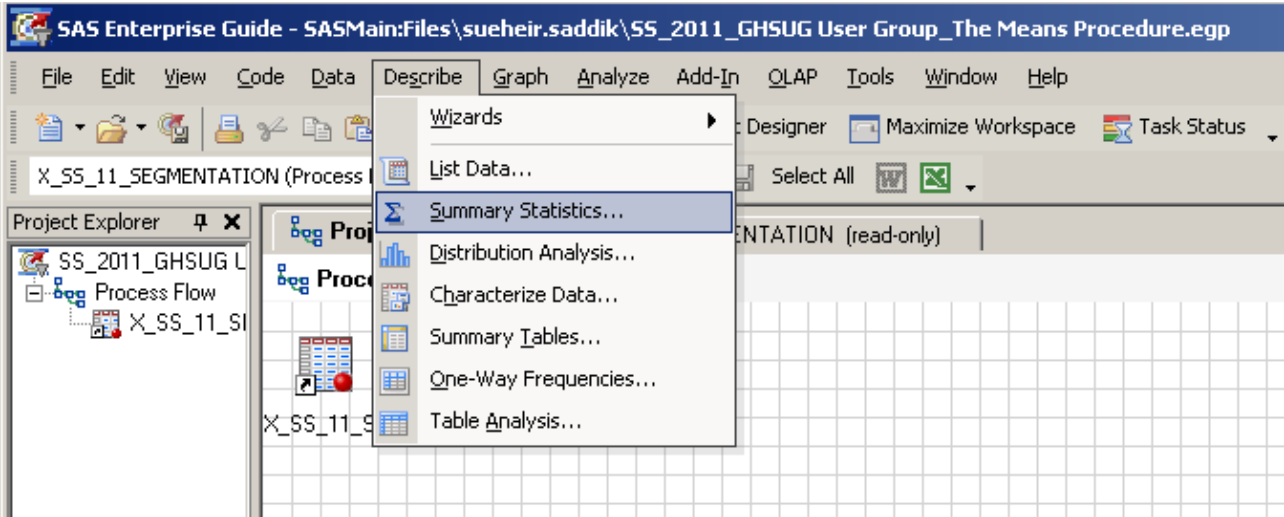
- Bell Canada data was used in the example presented in this presentation
 - Collection data with 604,009 distinct observations
 - 74 variables ... Variables used
 - Past Due Amount: amount owing by customers
 - Bad_Ind_New: an indicator with these values
 - 0 = good
 - 1 = bad
 - Aging buckets
 - Time spent in collection in days
 - Tenure in months
- Information needed - questions to answer
 - Data quality, missing values, penetration, extreme values, transformations created ... ?
 - What is the overall average past due amount?
 - How does the past due amount differ with respect to the bad / good groups?
 - What is the profile of bad customers?

SAS Enterprise Guide – Summary Statistics – Data & Drop Down

Data
Variables = columns
Observations = rows



Describe –
Summary Statistics



SAS Enterprise Guide – Summary Statistics – Interface

The screenshot displays the SAS Enterprise Guide interface. The main window is titled "SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp". The menu bar includes File, Edit, View, Code, Data, Describe, Graph, Analyze, Add-In, OLAP, Tools, Window, and Help. The toolbar contains icons for file operations, project management, and task execution. The Project Explorer on the left shows the project structure, including "GHSUG User Group_The", "ss Flow", "SS_11_SEGMENTATION", and "Summary Statistics". The Process Flow diagram shows a task named "X_SS_11_S..." connected to a "Summary Statistics" task. The "Summary Statistics for X_SS_11_SEGMENTATION" dialog box is open, showing the "Task Roles" tab. The "Variables to assign:" list includes "BAN", "EXTRACT_DATE", "BAN_STATUS", and various "AGE_BUCKET" variables. The "Task roles:" list includes "Analysis variables", "Classification variables", "Frequency count (Limit: 1)", "Relative weight (Limit: 1)", "Copy variables", and "Group analysis by". The "Analysis variables" role is selected, and the "Columns to assign" and "Columns to use" panes are visible. The "Columns to assign" pane shows "BAN" and "EXTRACT_DATE" assigned. The "Columns to use" pane shows "BAN" and "EXTRACT_DATE" selected. The "Preview code" button is visible at the bottom left of the dialog box. A status bar at the bottom indicates "The 'Analysis variables' role must have at least 1 variable assigned to it."

SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Project Designer X_SS_11_SEGMENTATION (read-only)

Process Flow

X_SS_11_S... Summary Statistics

Summary Statistics for X_SS_11_SEGMENTATION

Task Roles

Statistics
Basic
Percentiles
Additional
Plots
Results
Titles

Variables to assign:

Name
BAN
EXTRACT_DATE
BAN_STATUS
AGE_BUCKET_1_30_C
AGE_BUCKET_1_30_IN
AGE_BUCKET_31_60_C
AGE_BUCKET_31_60_IN
AGE_BUCKET_61_90_C
AGE_BUCKET_61_90_IN
AGE_BUCKET_91_120_C
AGE_BUCKET_91_120_IN
AGE_BUCKET_120_P

Task roles:

- Analysis variables
- Classification variables
- Frequency count (Limit: 1)
- Relative weight (Limit: 1)
- Copy variables
- Group analysis by

Columns to assign: BAN, EXTRACT_DATE

Columns to use: BAN, EXTRACT_DATE

...and drop it.

Preview code Run Save Cancel Help

The "Analysis variables" role must have at least 1 variable assigned to it.

SAS Enterprise Guide – Summary Statistics – Variable Selection

The screenshot displays the SAS Enterprise Guide interface. The main window shows a process flow with a task named 'Summary Statistics' for the project 'X_SS_11_SEGMENTATION'. A dialog box titled 'Summary Statistics for X_SS_11_SEGMENTATION' is open, showing the 'Task Roles' tab. The 'Variables to assign' list includes: DISPUTE_AMT, MONTHS_SINCE_OLDE, NUM_EAS, NUM_FEATURES, PAST_DUE_AMT, PRODUCT_TYPE_GRO, PROVINCE, TOTAL_DUE_AMT, SCORE_LEVEL, SCORE, MODEL_DESCRIPTION, and MODEL_TYPE. The 'Task roles' list includes: Analysis variables, Classification variables, Frequency count (Limit: 1), Relative weight (Limit: 1), Copy variables, and Group analysis by. The 'Analysis variables' role is selected, and 'PAST_DUE_AMT' is highlighted in the list. The dialog box also includes a 'Preview code' button and 'Run', 'Save', 'Cancel', and 'Help' buttons.

SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Project Designer X_SS_11_SEGMENTATION (read-only)

Process Flow

X_SS_11_SEGMENTATION (Process F)

Run Selection Select All

Project Explorer

GHSUG User Group_The

ss Flow

SS_11_SEGMENTATION

Summary Statistics

X_SS_11_S...

Summary Statistics

Summary Statistics for X_SS_11_SEGMENTATION

Task Roles

Statistics

Basic

Percentiles

Additional

Plots

Results

Titles

Task Roles

Variables to assign:

Name

DISPUTE_AMT

MONTHS_SINCE_OLDE

NUM_EAS

NUM_FEATURES

PAST_DUE_AMT

PRODUCT_TYPE_GRO

PROVINCE

TOTAL_DUE_AMT

SCORE_LEVEL

SCORE

MODEL_DESCRIPTION

MODEL_TYPE

PAST_DUE_AMT

Task roles:

Analysis variables

PAST_DUE_AMT

Classification variables

Frequency count (Limit: 1)

Relative weight (Limit: 1)

Copy variables

Group analysis by

The variables that you assign to this role are the numeric variables on which you want to have statistics

Preview code

Run

Save

Cancel

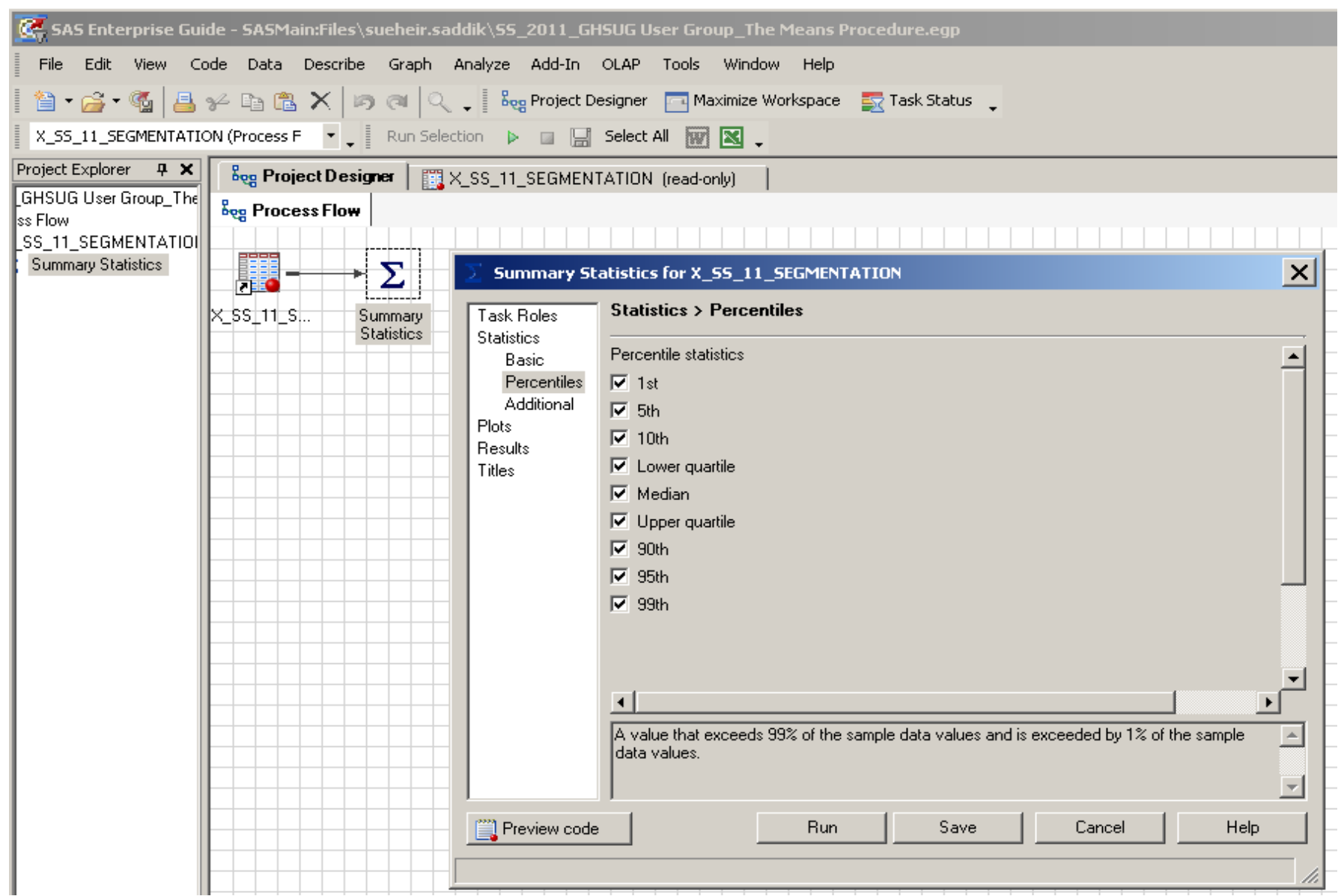
Help

SAS Enterprise Guide – Summary Statistics – Statistics – Basic Statistics

The screenshot displays the SAS Enterprise Guide interface. The title bar reads "SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egg". The menu bar includes File, Edit, View, Code, Data, Describe, Graph, Analyze, Add-In, OLAP, Tools, Window, and Help. The toolbar contains icons for file operations, project management, and execution. The Project Explorer on the left shows a project named "GHSUG User Group_The" with a process flow containing "SS_11_SEGMENTATION" and "Summary Statistics". The Process Flow diagram shows a data source "X_SS_11_S..." connected to a "Summary Statistics" task. The "Summary Statistics for X_SS_11_SEGMENTATION" dialog box is open, showing the "Statistics > Basic" tab. The "Task Roles" list on the left includes Statistics, Basic, Percentiles, Additional, Plots, Results, and Titles. The "Basic statistics" section on the right has the following options: Mean (checked), Standard deviation (checked), Standard error (unchecked), Variance (unchecked), Minimum (checked), Maximum (checked), Range (checked), Sum (checked), Sum of weights (unchecked), Number of observations (checked), and Number of missing values (checked). The "Maximum decimal places" is set to "Best fit". A description at the bottom states: "Calculates a statistical measure of the variability of a group of data values. This measure, which is the most widely used measure of the dispersion of a frequency distribution, is equal to the positive square root of the variance." The dialog box has buttons for "Preview code", "Run", "Save", "Cancel", and "Help".



SAS Enterprise Guide – Summary Statistics – Statistics – Percentiles



SAS Enterprise Guide – Summary Statistics – Statistics – Additional

The screenshot displays the SAS Enterprise Guide interface. The title bar reads "SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp". The menu bar includes File, Edit, View, Code, Data, Describe, Graph, Analyze, Add-In, OLAP, Tools, Window, and Help. The toolbar contains icons for file operations, project management, and execution. The Project Explorer on the left shows a project named "GHSUG User Group_The Means Procedure" with a process flow containing "X_SS_11_SEGMENTATION" and "Summary Statistics". The main workspace shows the "Process Flow" tab with a diagram of the process flow. A dialog box titled "Summary Statistics for X_SS_11_SEGMENTATION" is open, showing the "Statistics > Additional" tab. The "Additional statistics" section includes checkboxes for "Confidence limits of the mean", "t statistic and Prob > |t|", "Coefficient of variation", "Corrected sum of squares", and "Uncorrected sum of squares". The "Confidence level for confidence limits" is set to 95%. The dialog box also has a "Preview code" button and "Run", "Save", "Cancel", and "Help" buttons.

SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Project Designer X_SS_11_SEGMENTATION (read-only)

Process Flow

X_SS_11_S... Summary Statistics

Summary Statistics for X_SS_11_SEGMENTATION

Task Roles
Statistics
Basic
Percentiles
Additional
Plots
Results
Titles

Statistics > Additional

Additional statistics

☐ Confidence limits of the mean

☐ t statistic and Prob > |t|

☐ Coefficient of variation

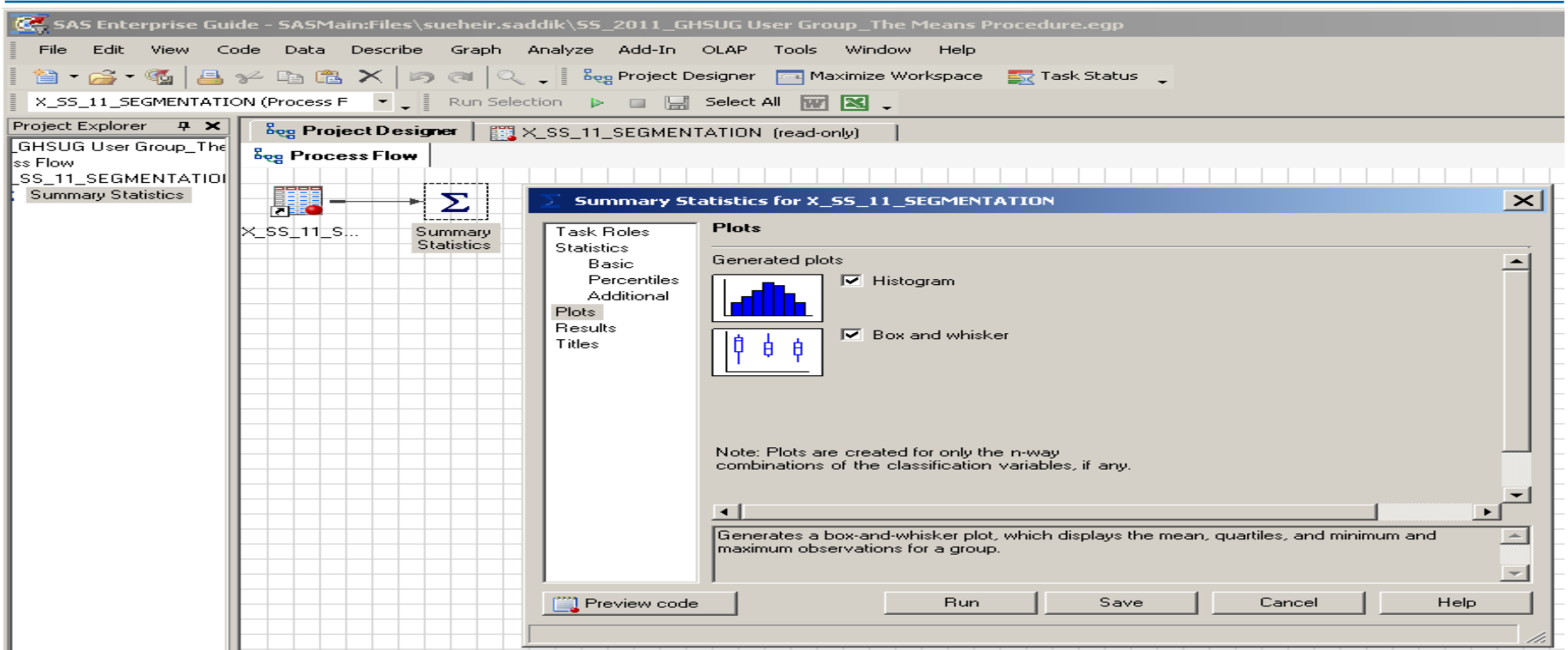
☐ Corrected sum of squares

☐ Uncorrected sum of squares

Confidence level for confidence limits
95%

Preview code Run Save Cancel Help

SAS Enterprise Guide – Summary Statistics – Plots



- No plots will be shown today ... My advice, avoid the plots option unless you have super CPU ... The process consumes lots and lots of CPU and brings the system down to it's knees ...

Email received ... **Oops** ... subject "Fw: SAS Critical Alert - Server Conso SAS VGRM1 CPU Util ..."

Image Name	User Name	CPU	Mem Usage
EXCEL.EXE	sueheir.saddik	50	265,768 K
EXCEL.EXE	sueheir.saddik	25	101,792 K
EXCEL.EXE	sueheir.saddik	25	266,344 K

SAS Enterprise Guide – Summary Statistics – Results

The screenshot displays the SAS Enterprise Guide interface. The main window shows a project titled "SS_2011_GHSUG User Group_The Means Procedure.egp". The Project Explorer on the left lists the project structure, including "Process Flow", "X_SS_11_SEGMENTATION", "PDA_G", "SEG", "Process Flow 1", and "X_SS_11_SEGMENTATION". The central workspace shows a process flow diagram with a data set icon connected to a "Summary Statistics" task. The "Summary Statistics for X_SS_11_SEGMENTATION" dialog box is open, showing the "Results" tab. The "Results" tab includes options for "Save statistics to data set" (checked), "Show statistics" (checked), and "Value to copy for Copy Variables role:" (Maximum). The "Combinations of classification variables:" is set to "N-way only". The "Specify ways:" field is empty, with examples provided below it. The dialog box also has buttons for "Preview code", "Run", "Save", "Cancel", and "Help".

SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Project Designer PDA_G SEG (read-only)

Process Flow Process Flow 1

Project Explorer

SS_2011_GHSUG User Group_The Means Pr

Process Flow

X_SS_11_SEGMENTATION

PDA_G

SEG

Process Flow 1

X_SS_11_SEGMENTATION

Summary Statistics

Summary Statistics for X_SS_11_SEGMENTATION

Task Roles

Statistics

Basic

Percentiles

Additional

Plots

Results

Titles

Results

☒ Save statistics to data set

SASMain:WORK.MEANSummaryStatsX_SS_11_SEGMENTA Browse...

☒ Show statistics

Value to copy for Copy Variables role: Maximum

Combinations of classification variables: N-way only

Specify ways: 0

Examples: 1, 3, 1...4, 0 to 4 by 2

Preview code Run Save Cancel Help



SAS Enterprise Guide – Summary Statistics – Titles

The screenshot displays the SAS Enterprise Guide interface. The main window is titled "SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egg". The menu bar includes File, Edit, View, Code, Data, Describe, Graph, Analyze, Add-In, OLAP, Tools, Window, and Help. The toolbar contains icons for file operations, project management, and execution. The Project Explorer on the left shows a project named "X_SS_11_SEGMENTATION (Process F)" with a sub-project "Summary Statistics". The Process Flow diagram in the center shows a data source "X_SS_11_S..." connected to a "Summary Statistics" task. The "Summary Statistics for X_SS_11_SEGMENTATION" dialog box is open, showing the "Titles" tab. The "Section:" list on the left includes Task Roles, Statistics, Basic, Percentiles, Additional, Plots, Results, and Titles. The "Text for section: Analysis" field is selected, and the "Use default text" checkbox is checked. The default text for the Analysis section is "Summary Statistics Results". The dialog box also has buttons for "Preview code", "Run", "Save", "Cancel", and "Help".

SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egg

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Project Designer X_SS_11_SEGMENTATION (read-only)

Process Flow

X_SS_11_S... Summary Statistics

Summary Statistics for X_SS_11_SEGMENTATION

Task Roles
Statistics
Basic
Percentiles
Additional
Plots
Results
Titles

Section:

- ☒ Analysis
- ☒ Histogram
- ☒ Box and Whisker Plot
- ☒ Footnote

Text for section: Analysis

☒ Use default text

Summary Statistics Results

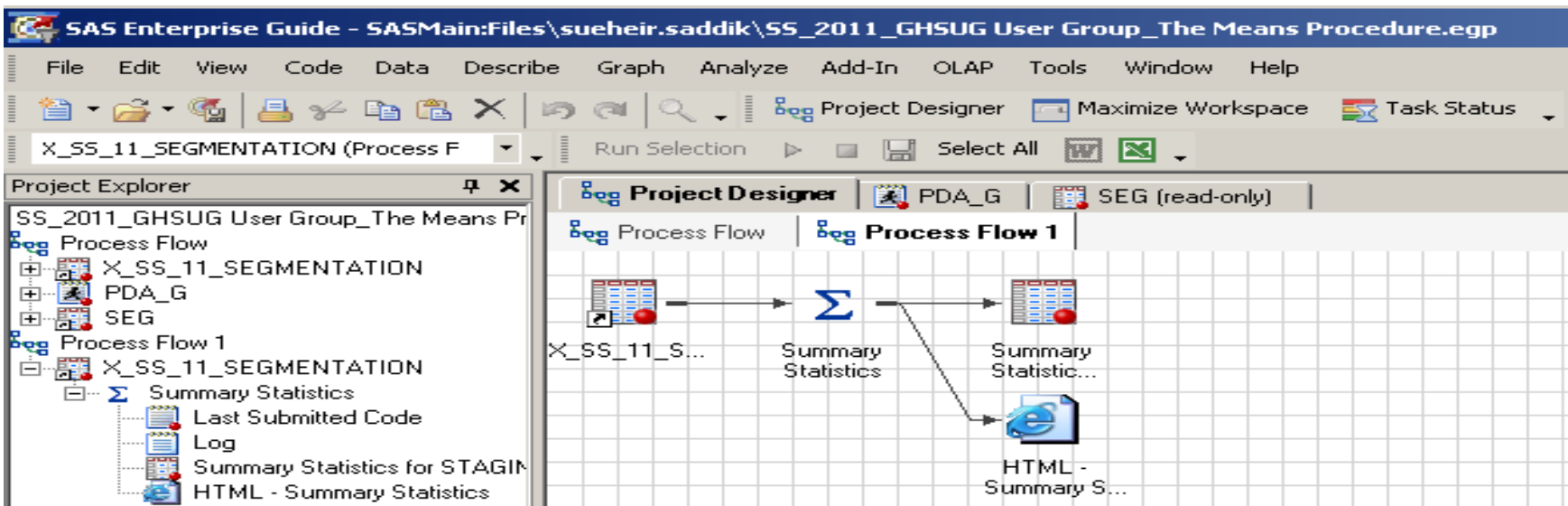
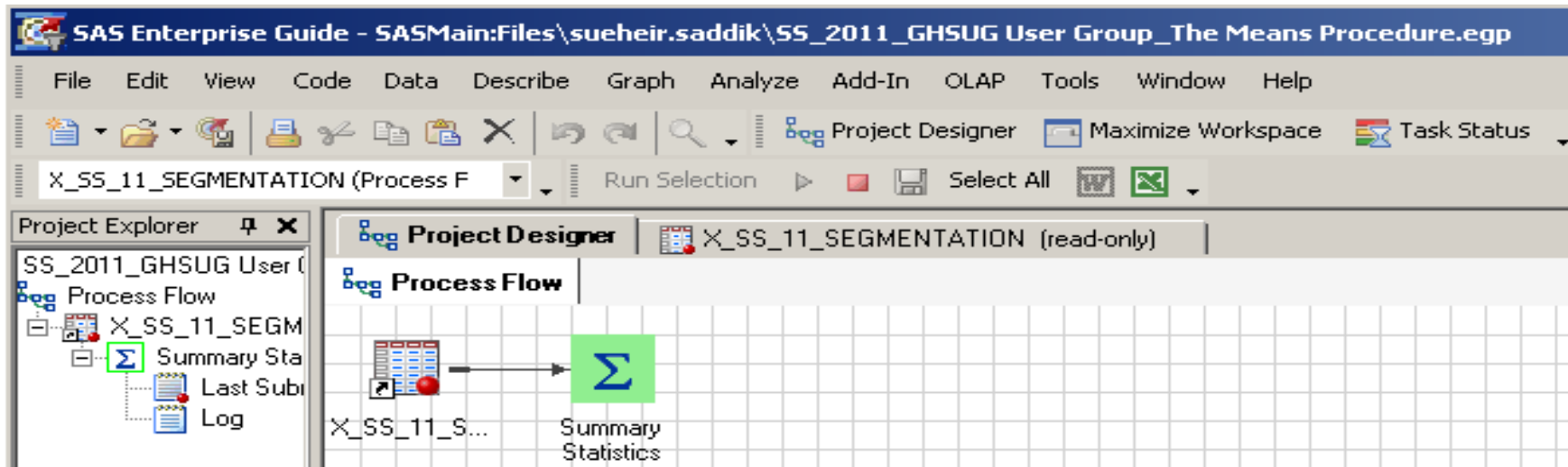
Preview code Run Save Cancel Help

SAS Enterprise Guide – Summary Statistics – Code Viewing

The screenshot displays the SAS Enterprise Guide interface. The main window shows the 'Summary Statistics for X_SS_11_SEGMENTATION' task configuration. The 'Titles' tab is active, showing the 'Section' as 'Analysis' and the 'Text for section: Analysis' as 'Summary Statistics Results'. The 'Code Preview for Task' window is open, showing the generated SAS code. The code includes a PROC SQL statement to create a view, followed by a PROC MEANS statement to calculate summary statistics. The code is as follows:

```
10  
11 PROC SQL;  
12   %_SASTASK_DROPDS(WORK.SORTTempTableSorted);  
13 QUIT;  
14  
15 /* -----  
16    Data set STAGING.X_SS_11_SEGMENTATION does not need to be sorted.  
17   ----- */  
18 PROC SQL;  
19   CREATE VIEW WORK.SORTTempTableSorted  
20     AS SELECT PAST_DUE_AMT FROM STAGING.X_SS_11_SEGMENTATION;  
21 QUIT;  
22 /* -----  
23    Run the Means Procedure  
24   ----- */  
25 TITLE;  
26 TITLE1 "Summary Statistics";  
27 TITLE2 "Results";  
28 FOOTNOTE;  
29 FOOTNOTE1 "Generated by the SAS System (&SASSERVERNAME, &SYSSCPL) on %SYSFUNC(  
30 PROC MEANS DATA=WORK.SORTTempTableSorted  
31   FW=12  
32   PRINTALLTYPES  
33   CHARTYPE  
34   QMETHOD=OS  
35   VARDEF=DF
```

SAS Enterprise Guide – Summary Statistics – Running & Done



SAS Enterprise Guide – Summary Statistics – Output

SAS output

Microsoft Excel - 6aec3a536f004dce81a01338bda3acfa

File Edit View Insert Format Tools Data SAS Window Help

Active Data: Active Selection Analyze Data Reports SAS Favorites

B16

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Summary Statistics																			
2	Results																			
3																				
4	The MEANS Procedure																			
5																				
6	Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT																			
7	Mean	Std Dev	Std Error	Minimum	Maximum	Range	Sum	N	N Miss	1st Pctl	5th Pctl	10th Pctl	Lower Quartile	Median	Upper Quartile	90th Pctl	95th Pctl	99th Pctl		
8	45.3935348	105.5042082	0.1367526	-24383.93	4133.66	28517.59	27418103.58	604009	0	-33.87	-0.53	0	0	0	68.49	137.59	198.59	358.92		
9	Generated by the SAS System (SASMain, NET_ASRV) on 13APR2011 at 12:47 PM																			
10																				
11																				
12																				

SAS output in a data set

SAS Enterprise Guide - SASMain:Files\suehir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egg

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Summary Statistics for STAGING.X_S

Run Selection | Select All

Project Explorer

- UG User Group_The Means Procedure
- 1_SEGMENTATION

	TYPE	_FREQ_	PAST_DUE_AMT_Mean	PAST_DUE_AMT_StdDev	PAST_DUE_AMT_StdErr	PAST_DUE_AMT_Min	PAST_DUE_AMT_Max
1	0	604009	45.39	105.50	0.14	-24383.93	4133.66

SAS Enterprise Guide – Summary Statistics – Using Group By Option

The screenshot displays the SAS Enterprise Guide interface. The main window shows a process flow with a 'Summary Statistics' task. A dialog box titled 'Summary Statistics for X_SS_11_SEGMENTATION' is open, showing the 'Task Roles' tab. In the 'Variables to assign' list, 'Bad_Indicator_New' is selected and circled in red. The 'Task roles' list on the right includes 'Analysis variables', 'Classification variables', 'Frequency count (Limit: 1)', 'Relative weight (Limit: 1)', 'Copy variables', and 'Group analysis by'. The 'Group analysis by' role is selected, and 'Bad_Indicator_New' is listed below it. The 'Bad_Indicator_New sort order' is set to 'Ascending'. The 'Sort by variables' checkbox is checked. The 'Run' button is visible at the bottom right of the dialog.

SAS Enterprise Guide - SASMain:Files\sueheir.saddik\SS_2011_GHSUG User Group_The Means Procedure.egp

File Edit View Code Data Describe Graph Analyze Add-In OLAP Tools Window Help

Project Designer X_SS_11_SEGMENTATION (read-only)

Process Flow

X_SS_11_SEGMENTATION (Process F)

Summary Statistics

Summary Statistics

HTML - Summary S...

Summary Statistics for X_SS_11_SEGMENTATION

Task Roles

Statistics

Basic

Percentiles

Additional

Plots

Results

Titles

Variables to assign:

Name

BAN

EXTRACT_DATE

BAN_STATUS

AGE_BUCKET_1_30_CURR

AGE_BUCKET_1_30_IND

AGE_BUCKET_31_60_CURR

AGE_BUCKET_31_60_IND

AGE_BUCKET_61_90_CURR

AGE_BUCKET_61_90_IND

AGE_BUCKET_91_120_CURR

AGE_BUCKET_91_120_IND

AGE_BUCKET_120_P_CURR

AGE_BUCKET_120_P_IND

AR_BALANCE

AR_BALANCE_0

AR_BALANCE_GROUP

AREA

AUTO_GEN_PYM_TYPE

BILL_CYCLE

BILL_MEDIA

CHARGE_AMT

DISPUTE_AMT

Task roles:

Analysis variables

PAST_DUE_AMT

Classification variables

Frequency count (Limit: 1)

Relative weight (Limit: 1)

Copy variables

Group analysis by

Bad_Indicator_New

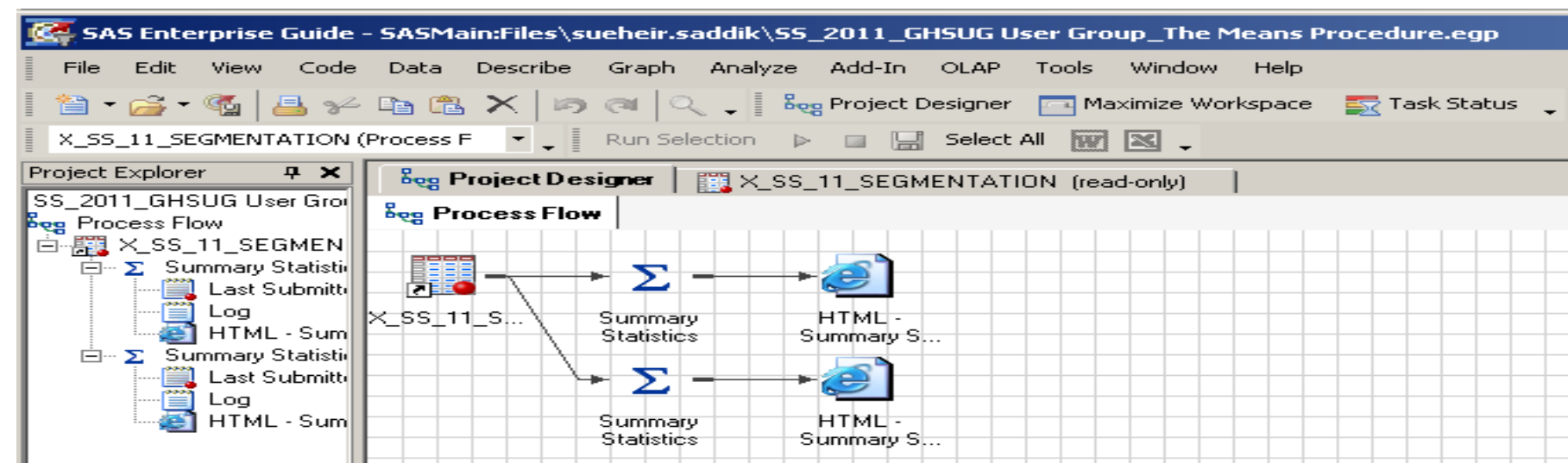
Bad_Indicator_New sort order:

Ascending

Sort by variables

Run Save Cancel Help

SAS Enterprise Guide – Summary Statistics – Output



1	Summary Statistics
2	Results
3	
4	The MEANS Procedure
5	
6	Bad_Indicator_New=0
7	
8	Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT
9	Mean Std Dev Minimum Maximum Range Sum N N Miss 1st Pctl 5th Pctl 10th Pctl Lower Quartile Median Upper Quartile 90th Pctl 95th Pctl 99th Pctl
10	39.4407213 102.0300758 -24383.93 3467.54 27851.47 20377088.09 516651 0 -36.08 -0.73 0 0 0 64.16 122.17 174.67 305.77
11	
12	Bad_Indicator_New=1
13	
14	Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT
15	Mean Std Dev Minimum Maximum Range Sum N N Miss 1st Pctl 5th Pctl 10th Pctl Lower Quartile Median Upper Quartile 90th Pctl 95th Pctl 99th Pctl
16	80.59955 118.0948121 -2921.95 4133.66 7055.61 7041015.49 87358 0 -20.79 0 0 0 50.27 114.38 226.18 308.38 514.57
17	
	Generated by the SAS System (SASMain, NET_ASRV) on 14MAR2011 at 12:49 PM



SAS Enterprise Guide – Summary Statistics – Past Due Amount Summary

	Analysis Variable : PAST_DUE_AMT																
Bad_Indicator _New	N	N Miss	Sum	Mean	Std Dev	Minimum	Maximum	Range	1st Pctl	5th Pctl	10th Pctl	Lower Quartile	Median	Upper Quartile	90th Pctl	95th Pctl	99th Pctl
0	516,651	0	\$20,377,088	\$39	\$102	-\$24,384	\$3,468	\$27,851	-\$36	-\$1	\$0	\$0	\$0	\$64	\$122	\$175	\$306
1	87,358	0	\$7,041,015	\$81	\$118	-\$2,922	\$4,134	\$7,056	-\$21	\$0	\$0	\$0	\$50	\$114	\$226	\$308	\$515
All	604,009	0	\$27,418,104	\$45	\$106	-\$24,384	\$4,134	\$28,518	-\$34	-\$1	\$0	\$0	\$0	\$68	\$138	\$199	\$359

- There are no missing past due amount values
- 604,009 observation
 - 516,651 good (86%)
 - 87,358 Bad (14%)
- Bad rate = $(87,358 / 604,009) = 14\%$
- 50% of all customers have no past due amount
- 50% of bad customers have a past due amount of \$50 or less
- Few extreme values /outliers (Min & Max) such as (-\$24,384, -\$2,922, ..., \$3,468, \$4,134)
- .
- .
- .
- There average past due amount for good is significantly lower than that of bad

SAS Enterprise Guide – Summary Statistics – Data / Code Checking

SAS Code created

```
if PAST_DUE_AMT < 0 then PAST_DUE_AMT_G = '-ive';
else if PAST_DUE_AMT = 0 then PAST_DUE_AMT_G = '$0';
else if 0 < PAST_DUE_AMT <= 75 then PAST_DUE_AMT_G = '1-$0.01 to $75';
else if 75 < PAST_DUE_AMT <= 150 then PAST_DUE_AMT_G = '2-$75.01 to $150';
else if 150 < PAST_DUE_AMT <= 250 then PAST_DUE_AMT_G = '3-$150.01 to $250';
else if 250 < PAST_DUE_AMT <= 500 then PAST_DUE_AMT_G = '4-$250.01 to $500';
else if PAST_DUE_AMT > 500 then PAST_DUE_AMT_G = '5-$500.01 +';
run;
```

Summary Statistics Interface

Task roles:

Analysis variables
PAST_DUE_AMT

Classification variables

Frequency count (Limit: 1)

Relative weight (Limit: 1)

Copy variables

Group analysis by
PAST_DUE_AMT_G

SAS output

PAST_DUE_AMT_G=\$0									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
0	0	0	276403	0	0	0	0	0	0
PAST_DUE_AMT_G=-ive									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
-32.2243993	-24383.93	-0.01	36913	0	-343.11	-100.3	-3.86	-0.09	-0.01
PAST_DUE_AMT_G=1-\$0.01 to \$75									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
46.6463012	0.01	75	160903	0	0.1	2.75	50.94	72.62	74.56
PAST_DUE_AMT_G=2-\$75.01 to \$150									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
106.2728741	75.01	150	77476	0	75.49	77.38	103.14	143.95	148.91
PAST_DUE_AMT_G=3-\$150.01 to \$250									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
190.4260796	150.01	250	35405	0	150.63	153.11	186.4	240.87	248.08
PAST_DUE_AMT_G=4-\$250.01 to \$500									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
326.9079031	250.01	500	15094	0	250.72	254.11	308.815	459.35	491.93
PAST_DUE_AMT_G=5-\$500.01 +									
Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT									
Mean	Minimum	Maximum	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl
656.8003251	500.05	4133.66	1815	0	502.24	507.38	591.94	995.9	1435.91



SAS Enterprise Guide – Data Checking Output - Make it Pretty & Add Distributions

Analysis Variable : PAST_DUE_AMT PAST_DUE_AMT													
PAST_DUE_AMT_G	Mean	Min	Max	N	N Miss	1st Pctl	5th Pctl	Median	95th Pctl	99th Pctl	Distribution	Cumulative Distribution	Inverse Cumulative Distribution
\$0	\$0	\$0	\$0	276,403	0	\$0	\$0	\$0	\$0	\$0	46%	46%	100%
-ive	-\$32	-\$24,384	-\$0	36,913	0	-\$343	-\$100	-\$4	-\$0	-\$0	6%	52%	54%
1-\$0.01 to \$75	\$47	\$0	\$75	160,903	0	\$0	\$3	\$51	\$73	\$75	27%	79%	48%
2-\$75.01 to \$150	\$106	\$75	\$150	77,476	0	\$75	\$77	\$103	\$144	\$149	13%	91%	21%
3-\$150.01 to \$250	\$190	\$150	\$250	35,405	0	\$151	\$153	\$186	\$241	\$248	6%	97%	9%
4-\$250.01 to \$500	\$327	\$250	\$500	15,094	0	\$251	\$254	\$309	\$459	\$492	2%	100%	3%
5-\$500.01 +	\$657	\$500	\$4,134	1,815	0	\$502	\$507	\$592	\$996	\$1,436	0%	100%	0%
All	\$45	-\$24,384	\$4,134	604,009	0	-\$34	-\$1	\$0	\$199	\$359	100%		

- Data created is correct ...
- 46% of customers have no past due amount
- 6% of customer have credits
- 79% of all customers have at most \$75 past due amount
- 21% of customers have at least \$75 past due amount



SAS Enterprise Guide – Summary Statistics – Profiling (Key Variables)

The screenshot displays the SAS Enterprise Guide interface. The main window is titled "Summary Statistics for X_SS_11_SEGMENTATION". The "Task Roles" pane on the left shows the "Variables to assign:" list, which includes a long list of variables such as BAN, EXTRACT_DATE, BAN_STATUS, and various AGE_BUCKET variables. The "Task roles:" pane on the right shows the "Analysis variables" list, which includes variables like PAST_DUE_AMT, CHARGE_AMT, DISPUTE_AMT, and MONTHS_SINCE_OLDEST_EA_ACTV. The "Bad_Indicator_New sort order:" dropdown is set to "Ascending", and the "Sort by variables" checkbox is checked. The "Process Flow" pane on the left shows the task being executed. The "Project Explorer" pane on the far left shows the project structure, including the "X_SS_11_SEGMENTATION" task and its sub-tasks.



SAS Enterprise Guide – Summary Statistics – Profiling Output

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Summary Statistics																	
2	Results																	
3																		
4	The MEANS Procedure																	
5																		
6	Bad_Indicator_New=0																	
7																		
8	Variable	Mean	Std Dev	Minimum	Maximum	Range	Sum	N	N Miss	1st Pctl	5th Pctl	10th Pctl	Lower Quartile	Median	Upper Quartile	90th Pctl	95th Pctl	99th Pctl
9	PAST_DUE_AMT	39	102	-24384	3468	27851	20377088	516651	0	-36	-1	0	0	0	64	122	175	306
10	CHARGE_AMT	127	134	0	57886	57886	65719403	516521	130	24	38	47	72	102	155	226	286	491
11	DISPUTE_AMT	0	1	-400	98	498	-1002	516521	130	0	0	0	0	0	0	0	0	0
12	AGE_BUCKET_1_30_CURR	25	47	0	1829	1829	12805405	516521	130	0	0	0	0	0	41	75	113	211
13	AGE_BUCKET_31_60_CURR	4	17	0	824	824	2088773	516521	130	0	0	0	0	0	0	5	28	83
14	AGE_BUCKET_61_90_CURR	1	6	0	994	994	330437	516521	130	0	0	0	0	0	0	0	0	21
15	AGE_BUCKET_91_120_CURR	0	2	0	454	454	67627	516521	130	0	0	0	0	0	0	0	0	2
16	AGE_BUCKET_120_P_CURR	0	2	0	321	321	64620	516521	130	0	0	0	0	0	0	0	0	0
	MONTHS_SINCE_OLDEST_EA_ACTV	132	116	0	359	359	68067605	516651	0	1	2	8	33	99	214	351	354	356
17																		
18																		
19	Bad_Indicator_New=1																	
20																		
21	Variable	Mean	Std Dev	Minimum	Maximum	Range	Sum	N	N Miss	1st Pctl	5th Pctl	10th Pctl	Lower Quartile	Median	Upper Quartile	90th Pctl	95th Pctl	99th Pctl
22	PAST_DUE_AMT	81	118	-2922	4134	7056	7041015	87358	0	-21	0	0	0	50	114	226	308	515
23	CHARGE_AMT	185	168	0	15140	15140	16129541	87328	30	26	41	51	78	141	238	364	481	761
24	DISPUTE_AMT	0	2	-387	0	387	-617	87328	30	0	0	0	0	0	0	0	0	0
25	AGE_BUCKET_1_30_CURR	44	69	0	3826	3826	3862311	87328	30	0	0	0	0	10	63	131	184	295
26	AGE_BUCKET_31_60_CURR	13	37	0	1477	1477	1091917	87328	30	0	0	0	0	0	0	40	80	185
27	AGE_BUCKET_61_90_CURR	2	14	0	622	622	190703	87328	30	0	0	0	0	0	0	0	6	62
28	AGE_BUCKET_91_120_CURR	0	5	0	439	439	30289	87328	30	0	0	0	0	0	0	0	0	8
29	AGE_BUCKET_120_P_CURR	0	5	0	688	688	23557	87328	30	0	0	0	0	0	0	0	0	1
	MONTHS_SINCE_OLDEST_EA_ACTV	32	61	0	356	356	2814915	87358	0	0	1	1	2	8	31	94	155	351



SAS Enterprise Guide – Summary Statistics – Profiling Summary

Bad customers profile

- Have significantly higher past due amount
- Have significantly higher charge amount
- Have significantly higher \$\$ in age bucket 1
- Tend to have low tenure
 - New customers !

Bad_Indicator_New	Variable	Mean	Std Dev	Minimum	Maximum	Range	Sum	N	N Miss	1st Pctl	5th Pctl	10th Pctl	Lower Quartile	Median	Upper Quartile	90th Pctl	95th Pctl	99th Pctl
0	PAST_DUE_AMT	\$39	102	-\$24,384	\$3,468	27,851	20,377,088	516,651	0	-36	-1	0	0	0	64	122	175	306
	CHARGE_AMT	\$127	134	\$0	\$57,886	57,886	\$65,719,403	516,521	130	\$24	\$38	\$47	\$72	\$102	\$155	\$226	\$286	\$491
	AGE_BUCKET_1	\$25	47	\$0	\$1,829	1,829	\$12,805,405	516,521	130	\$0	\$0	\$0	\$0	\$0	\$41	\$75	\$113	\$211
	AGE_BUCKET_2	\$4	17	\$0	\$824	824	\$2,088,773	516,521	130	\$0	\$0	\$0	\$0	\$0	\$0	\$5	\$28	\$83
	AGE_BUCKET_3	\$1	6	\$0	\$994	994	\$330,437	516,521	130	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$21
	AGE_BUCKET_4	\$0	2	\$0	\$454	454	\$67,627	516,521	130	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$2
	AGE_BUCKET_5	\$0	2	\$0	\$321	321	\$64,620	516,521	130	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
	Tenure	132	116	0	359	359	68,067,605	516,651	0	1	2	8	33	99	214	351	354	356
1	PAST_DUE_AMT	\$81	118	-\$2,922	\$4,134	7,056	7,041,015	87,358	0	-21	0	0	0	50	114	226	308	515
	CHARGE_AMT	\$185	168	\$0	\$15,140	15,140	\$16,129,541	87,328	30	\$26	\$41	\$51	\$78	\$141	\$238	\$364	\$481	\$761
	AGE_BUCKET_1	\$44	69	\$0	\$3,826	3,826	\$3,862,311	87,328	30	\$0	\$0	\$0	\$0	\$10	\$63	\$131	\$184	\$295
	AGE_BUCKET_2	\$13	37	\$0	\$1,477	1,477	\$1,091,917	87,328	30	\$0	\$0	\$0	\$0	\$0	\$0	\$40	\$80	\$185
	AGE_BUCKET_3	\$2	14	\$0	\$622	622	\$190,703	87,328	30	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$6	\$62
	AGE_BUCKET_4	\$0	5	\$0	\$439	439	\$30,289	87,328	30	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$8
	AGE_BUCKET_5	\$0	5	\$0	\$688	688	\$23,557	87,328	30	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$1
	Tenure	32	61	0	356	356	2,814,915	87,358	0	0	1	1	2	8	31	94	155	351
All	PAST_DUE_AMT	\$45	106	-\$24,384	\$4,134	28,518	27,418,104	604,009	0	-34	-1	0	0	0	68	138	199	359
	CHARGE_AMT	\$136	141	\$0	\$57,886	57,886	\$81,848,943	603,849	160	\$24	\$39	\$48	\$72	\$105	\$166	\$247	\$320	\$563
	AGE_BUCKET_1	\$28	51	\$0	\$3,826	3,826	\$16,667,716	603,849	160	\$0	\$0	\$0	\$0	\$0	\$45	\$82	\$125	\$230
	AGE_BUCKET_2	\$5	22	\$0	\$1,477	1,477	\$3,180,691	603,849	160	\$0	\$0	\$0	\$0	\$0	\$0	\$11	\$34	\$106
	AGE_BUCKET_3	\$1	8	\$0	\$994	994	\$521,141	603,849	160	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$25
	AGE_BUCKET_4	\$0	3	\$0	\$454	454	\$97,916	603,849	160	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$3
	AGE_BUCKET_5	\$0	3	\$0	\$688	688	\$88,178	603,849	160	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0
	Tenure	117	115	0	359	359	70,882,520	604,009	0	0	1	3	19	80	189	337	353	355



Questions ?