

SAS Enterprise Guide

Summary Analytics & Hypothesis Testing

(June, 2018)

<u>Sources</u> (adapted with permission)-Ron Freeze Course and Classroom Notes Enterprise Systems, Sam M. Walton College of Business, University of Arkansas, Fayetteville SAS[®] Multivariate Statistics Course Notes & Workshop, 2010 SAS[®] Advanced Business Analytics Course Notes & Workshop, 2010 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.

Here's the Story...

This tutorial walks you through features and functions of SAS Enterprise Guide 7.1. The basis of this tutorial is a data set from Dillards that was extracted from the University of Arkansas Enterprise Systems group in the Information Systems Department. There are two exercises that include: 1) summary statistics for the Midwest Area (AR, KS, MO, OK) Sales Revenue, and 2) a hypothesis test determining the stores that are statistically different in sales from the State average. The focus of this analysis will be Sales Revenue for the year 2015.

<u>NOTE</u>: Use of this tutorial requires access to SAS Enterprise Guide 7. Obtaining the data for this tutorial requires an extracted Dillards 2016 file from the Enterprise Systems at the University of Arkansas (<u>https://walton.uark.edu/enterprise/</u>). The data is extracted using the file "Dillards 2016 Data Extraction-TUNtoExcel-KPIs.pdf" (available on the Enterprise Systems website and the Teradata University Network website) and should not be downloaded to your personal drives. The file should remain on the Remote Desktop S: drive provided by the University of Arkansas. This is due to our agreement with the data providers. Questions can be directed to Ron Freeze at <u>rfreeze@walton.uark.edu</u>.

ESTIMATED COMPLETION TIME: 20-40 minutes

Connect to your data

The following section is duplicated from the Pre-Conference Preparation.docx document for connecting to your data. The end result is a project file (.egp) that has the data already identified for this tutorial. If this has been accomplished, please skip to <u>Create Summary Statistics</u>.

Verify SAS-EG Access and set up project file

For this step, be sure you know where the **DillardsTUN-Export-KPIs.csv** file is loaded on your local computer or the remote desktop.

- a. Open SAS Enterprise Guide and close any pop-ups
- b. Select File ⇒ Import Data
- c. Navigate to the location where you placed **DillardsTUN-Export-KPIs.csv**
- d. Click on your file and then click Open

The Import Data wizard is displayed with your filename.

e. Accept the defaults for step 1 and click Next>



f. Accept the defaults in step 2



- g. Click Next> to advance to step 3
 Define Field Attributes. Change the amount format to Currency DOLLARw.d by selecting the BEST10. in the Output Format column. Note to include two decimal places.
- h. Select OK and Next>

ielec	t columns and	define attribut	66:					
Inc	Source Name	Name	Label	Туре	Source	Len.	Output Format	Output Informat
	TRAN_DA	TRAN_DA	TRAN_DATE	Date	MMDDYY	8	MMDDYY	MMDDYY.
	CITY	CITY	CITY	String	SCHAR17.	17	SCHAR17.	\$CHAR17.
	ZIP_CODE	ZIP_CODE	ZIP_CODE	Number	BEST5.	8	BEST5.	BEST5.
	STATE	STATE	STATE	String	SCHAR2.	2	SCHAR2.	SCHAR2.
	STORE	STORE	STORE	Number	BEST4.	8	BEST4.	BEST4.
\sim	amount	amount	amount	Number	COMMA10.	8	BEST10	BEST10.
	Catego Numer Date	utput Data Se ories: ic /	t Format Formats: DOLLARw.d DOLLARXw.d	1	_ ×	^		
	Catego Numer Date Date/	utput Data Se pries: ic /	t Format Formats: DOLLARW.d DOLLARXw.d EURFRATSw.d EURFRBEFw.d EURFRCHFw.d			^		
	Catego Numer Date Time Date/ Currer Attrib Ove Dec	utput Data Se ories: ic Time (cy utes rall width: imal places:	Format Formats: DOLLARW d DOLLARW d EURFRATSW d EURFRATSW d EURFRCHFw. d	Min: 2 Max: Min: 0 Max:	□ × 32 9	~		
Se	Catego Numer Date Time Date/ Correct - Attrib Ove Dec	utput Data Se ric Time Sy Utes rall width: imal places: ription sign, commas	t Format Formats: DOLLARW.d DOLLARW.d EURFRATSw.d EURFRATSw.d EURFRATSw.d EURFRCHF.w 10 0 1 R 0		□ × 32 9	 		Modify

- i. Accept the defaults for step 4
- j. Click Finish

lect	columns and def	ine attributes:						
nc	Source Name	Name	Label	Туре	Source Informat	Len.	Output Format	Output Informat
~	TRAN_DATE	TRAN_DATE	TRAN_DATE	Date	MMDDYY10.		MMDDYY10.	MMDDYY10.
2	CITY	CITY	CITY	String	SCHAR17.	17	SCHAR17.	SCHAR17.
7	ZIP_CODE	ZIP_CODE	ZIP_CODE	Number	BEST5.	8	BEST5.	BEST5.
2	STATE	STATE	STATE	String	SCHAR2.	2	\$CHAR2.	SCHAR2.
2	STORE	STORE	STORE	Number	BEST4.	8	BEST4.	BEST4.
1	amount	amount	amount	Number	COMMA10.	8	BEST10.	BEST10.



🕼 SAS Enterprise Guide	
File Edit View Tasks Favorites Program Tools He	🎦 • 🚰 🔛 🚰 📇 94 🗈 ட X 19 (제 🖃 + 🗞 Process Flow 📲
Project Tree	+ x Import Data (DillardsTUN-Export-KPIs.csv) +
Bog Process Flow	🛄 Code 📋 Log 💹 Output Data
Dillards TUN-Export-KPIs.csv Import Data (Dillards TUN-Export-KPIs.csv)	🚯 🔍 Modify Task 🐺 Filter and Sort 🏨 Query Builder 🍸 Where Data 🗸 Describe 🝷 Graph 🖌 Analyze 🖌 Export 👻 Send To 🖛 📳
	🛅 TRAN_DATE 💩 CITY 🔞 ZIP_CODE 💩 STATE 🔞 STORE 📆 amount
	1 01/01/2014 JACKSONVILL 32225 FL 511 \$5.145.40
	2 01/01/2014 LONGVIEW 75605 TX 748 \$79,596.14
	3 01/01/2014 HUNTSVILLE 35801 AL 460 \$74,033.21
	4 01/01/2014 MIDLAND 79705 TX 746 \$77,607.09
	5 01/01/2014 CEDAR PARK 78613 TX 735 \$79.348.10
	6 01/01/2014 HUMBLE 77338 TX 780 \$54,484.41
	7 01/01/2014 AURORA 80012 CO 982 \$29.735.11
	8 01/01/2014 ABILENE 79606 TX 744 \$66.863.06
	9 01/01/2014 LAKE HAVAS 86404 AZ 912 \$20,109,53

The Import Data wizard processes and produces a SAS view of the data.

9

- k. Switch to the Process Flow and note that the CSV file is now in a temporary location as a SAS Dataset and ready for analysis
- 1. Select File ⇒ Save Project As....
- m. Navigate to your S: drive folder
- n. Name the Project (DillardsTUN-KPI.egp in the example) and



The project can now be opened and analysis performed on the dataset.

NOTE: you may need to refresh your data connection if you have reopened your project file. To do this, select one of the nodes in your project and select Run Project.

Save

Create Summary Statistics

These reports use the DillardsTUN-Export-KPIs data set.

Filter Data to 2015

The initial step will be to filter your data to only 2015 using the Filter and Sort task. The benefit of using the Filter and Sort is that your original file remains untouched with all of the initial data

- Select the Data node 1.
- 2. Select <u>Tasks</u> \Rightarrow <u>Data</u> \Rightarrow <u>Filter and Sort...</u> from the drop-down menus.



🔯 DillardsTUN-KPI - SAS Enterprise Guide



×

✓ Browse.

~ X

Clear All

.... AND 🗸 🗙

3. Move all six variables from Available to Selected

Available (0): Selected (6 of 6): Name Type Label Image: TRAN_DATE Date TRAN_DATE Image: Transport Transport Transport Transport Transport Trans	📰 Variables	🍾 Filter	🛃 Sort	Results			
Name Type Label	Available (0):			_	Selected (6 of 6):		
	Name				Name TRAN_DATE CITY 2IP_CODE STATE STORE	Type Date Char Num Char Num	Label TRAN_DATI CITY ZIP_CODE STATE STORE

~ 12/31/2015

~ 1/1/2015

- 4. Select the *Filter Data* Tab
 - a. Select **TRAN_DATE** for the Task Filter:
 - b. Select Less than or equal to and enter 12/31/2015
 - c. Select AND
 - d. Select **TRAN_DATE** for the 2nd Task Filter:
 - e. Select Greater than or equal to and enter 1/1/2015
- 5. Select OK
- 6. Replace the default text description for the Query Builder to Filter to 2015

Midwest Area Summary Statistics

Now we can create a *Summary Statistic* report for the four States in the Midwest Area. Make sure you are on the *Process Flow* view to start the *Summary Statistic* Task.

OLAP

Edit Data and Filter

TRAN_DATE

TRAN_DATE

Display labels instead of variable names

Task filter:

Data source: SASApp:WORK.DILLARDSTUN_EXPORT_KPIS_0000

✓ Less than or equal to

Greater than or equal to

- 7. Select the Data node
- Select <u>Tasks</u> ⇒ <u>Describe</u> ⇒ <u>Summary Statistics...</u> from the drop-down menus.
- 9. With **<u>Data</u>** selected on the left...
 - a. Select *Edit*... at the top right (This will be to filter the data to the Midwest Area)
 - b. Select **STATE** for the Task Filter:
 - c. Select equal to and enter AR (Arkansas)
 - d. Select OR
 - e. Select **STATE** for the 2nd Task Filter:
 - f. Select equal to and enter KS (Kansas)
 - g. Repeat for MO (Missouri) and OK (Oklahoma)
 - h. Select OK

ata source: SASApp: 'ask filter:	WORK.FILTER_FOR_DILLARDSTUN_	EXPORT_KP	~ Browse
STATE	✓ Equal to	~ AR	OR ~
STATE	 Equal to 	✓ KS	OR 🗸 🕻
STATE	 Equal to 	∽ MO	OR 🗸
STATE	Equal to Equal to	~ ОК	· · · · · ·
STATE	✓ Equal to	νIoκ	





- i. Drag the variable **amount** from the *Variables to assign:* pane to the *Analysis variables* role in the *Task roles:* pane.
- j. Drag the variable **STATE** from the Variables to assign: pane to the Classification variables in the Task roles: pane
- Select <u>Basic</u> under Statistics on the left. Leave the default basic statistics. Change Maximum decimal places to 2.
- View the selections on both <u>Percentiles</u> and <u>Additional</u>, but no changes will be made.
- 12. Select <u>Titles</u> on the left. Deselect Use default text. Select the default text in the box and type Midwest Area Descriptive Statistics for Dillards Stores in 2015. Leave the default footnote text.



14. The SAS Output is shown in the screen shot and provides the

starting values for each state's average revenue per day. The following are the values from the SAS Output.

- Arkansas \$ 32,242.55
- Kansas \$ 8,854.01
- Missouri \$ 6,705.22
- Oklahoma \$10,085.74
- 15. Save the project by selecting <u>File</u> \Rightarrow <u>Save</u>

cs	Data			
Percentiles Additional ts	Data source: SASApp:WORK.FILTER_FOR_DI Task filter: STATE = 'AR' OR STATE = 'KS' O	Edt		
3	Variables to assign:	Task roles:		Class level STATE
003	Name TRAN_DATE © ZP_CODE © STATE © STORE	Anayles vatables Anayles vatables Anayles vatables Anayles vatables Group analysis by	0	Sat by: Descending frequencies Sot order: Descending Descending Exclude Alow multi-tabel formats
	The variables that you assign to this role are charact all selected analysis variables for each unique comb	ter or discrete numeric variables that are used to divide th ination of classification variables.	ne input data into categories or subgroups.	The statistics will be calculated on

Σ Summary Statistics for SASApp:WORK.DILLARDSTUN_EXPORT_KPIS_0000

Data Statistics	S	Statisti	cs > Basic							
Perc Add Plots Results	ic centiles itional	Basic sta	tistics ard deviation ard error			Maximu 2	ım decimal		~	
Midwest Area fo	r SASApp:WORK.FI	LTER_FOR_DILLA	RDSTUN_EXPORT_KP							×
Data Statistics	Titles									
Balo-verdes Adotonal Pota Resulta Titles Properties	Sector:	er Plot	Test for section: Analysis — Use default test Miderest Area Descriptive St	atatics for Dillards Store	e in (2015					
Preview code						Run 💌	Save	Cancel	Help	

Midwest Area Descriptive Statistics for Dillards Stores in 2015

The MEANS Procedure

	Analysis Variable : amount										
STATE	N Obs	Mean	Std Dev	Minimum	Maximum	N					
OK	3620	10085.74	10348.08	374.85	202162.71	3620					
MO	3620	6705.22	6664.10	472.15	133749.00	3620					
AR	3515	32242.55	81743.62	6.75	2473019.15	3515					
KS	2172	8854.01	8744.91	900.27	155416.00	2172					

Generated by the SAS System ('SASApp',X64_ES08R2) on May 29, 2018 at 1:01:22 PM

Summary Statistics on your own

Midwest Area Stores per State

Add the variable store as a Classification variable in order to see the Summary Statistics for each store. The screen shot is the SAS Output of this modification. Then means of each store will be the comparison's made in the hypothesis testing of the next section.

Midwest Area Descriptive Statistics for Dillards Stores in 2015 The MEANS Procedure

	07005		Analysis Va	anable . dillo	Juilt		
STATE	STORE	N Obs	Mean	Std Dev	Minimum	Maximum	N
OK	790	362	3580.65	2660.70	1080.11	34155.70	362
	791	362	9982.80	7151.66	1523.51	94671.45	362
	792	362	20028.10	12524.68	4234.24	154918.77	362
	793	362	4010.17	3158.10	815.64	40400.29	362
	794	362	4558.19	3440.78	1155.09	47359.67	362
	795	362	13172.25	8305.13	2018.11	109637.18	362
	796	362	23841.29	14506.21	2632.22	202162.71	362
	797	362	3536.67	2568.17	374.85	31770.78	362
	798	362	12450.81	7576.33	2829.82	86953.01	362
	799	362	5696.46	4091.30	1094.76	50498.46	362
MO	301	362	10289.31	8372.72	2348.92	127070.23	362
	302	362	8465.35	6905.13	2107.21	99099.71	362
	305	362	4200.21	3368.28	999.96	46193.88	362
	309	362	3352.60	2448.71	889.27	29748.81	362
	311	362	3189.61	2750.74	719.66	39086.69	362
	312	362	5844.74	4048.55	1800.29	56012.46	362
	314	362	14632.26	9953.85	3720.89	133749.00	362
	317	362	3122.99	2396.76	472.15	33249.00	362
	323	362	5277.60	4054.38	1514.04	55282.23	362
	324	362	8677.55	6029.34	2848.97	79698.49	362
AR	402	362	5982.05	4241.80	361.76	55362.91	362
	403	362	10705.47	7764.41	570.04	111221.54	362
	404	362	4157.35	3691.64	348.03	47735.29	362
	405	362	28145.71	18570.28	31.67	218796.15	362
	406	362	11612.79	8588.58	129.47	112153.72	362
	407	362	11570.30	9114.99	659,76	120003.68	362
	408	362	12364.13	9131.88	1758.87	123137.46	362
	413	362	9115.09	7280.69	1395.17	95236.95	362
	698	362	219381 42	157179.31	111497 27	2473019 15	362
	896	257	54.98	191 41	6 75	3070.04	257
KS	322	362	15420.04	11033 22	4652 57	155416.00	362
	330	362	15363.06	10865.28	4481 20	148101 67	362
	331	362	7346.59	6285.01	1754.21	92717 29	362
	334	362	3497.85	2622.23	1039.52	35850.04	362
	334	362	4270 07	2830 13	900 27	33789.61	362
	336	362	7226.47	5023 47	2635.81	71813.67	362

Store Sales Differences

Generated by the SAS System ('SASApp',X64_ES08R2) on May 29, 2018 at 1:21:42 PM

A further investigation of store differences is accomplished with a two-tailed hypothesis test in which we will first determine if each store is significantly different than the average stores sales. The second step will be a determination if the store is significantly lower or significantly higher than that states average. The test is evaluated at the 95% confidence level. The model tested is written as follows for the 10 stores in Arkansas. NOTE: SAS Enterprise Guide tests each store individually with respect to the mean for the state.

 H_0 : \$ 32,242.55 = $\mu_{402} = \mu_{403} = \mu_{404} = \dots = \mu_{896}$

HA: At least two daily store revenue sales are not equal

The hypothesis test to answer our question of whether the individual stores are different than the state

mean is considered a Test of Location. Each store is compared to the State mean and the *Distribution analysis* task is used to determine the analysis results. We can start with the *Process flow* diagram that we ended with when creating our *Summary Statistics*.

1. Select the same filtered SAS Data set used for the *Summary Statistics*



- 2. Select <u>Tasks</u> \Rightarrow <u>Describe</u> \Rightarrow <u>Distribution Analysis...</u>.
- 3. Edit the Task Filter to only include AR (Arkansas)
- 4. Use the **amount** variable as the *analysis variable*
- 5. Use the **STORE** variable as the *Group analysis by*

- 6. Click Tables and...
 - a. Check the box for Basic confidence intervals
 - b. Check the box for Basic measures
 - c. Check the box for <u>Tests for</u> <u>location</u> and then type the value 32,242.55 in the field next to Null Hypothesis: Ho: Mu=



- 7. Click <u>Titles</u> and replace the default text with the following: Hypothesis test for Arkansas Stores Different from the State Average
- 8. Click Run
- 9. The 10 hypothesis tests ran result in significance for all of the stores. This translates to a rejection that any store is at the mean for the state average in Sales Revenue. The individual hypothesis tests will indicate whether each store is significantly below the state mean or significantly higher than the state mean. The following are an example of each and include potential questions that should be asked to address potential additional tests for decision-making.

Stores Performing above Average

Only Store 698 (1 out of 10 stores) exceeded the State Average daily Revenue in sales (\$32,243). Their mean sales were \$ 219,381 compared to the next closest store (405) at \$ 28,146. The positive Students' t statistic along with the p-value <.0001 indicates that Store 698 is significantly higher than the State average. The following questions may arise from this analysis

- Is Store 698 an outlier?
- What circumstances contribute to the performance of Store 698?
- Should Store 698 be compared to the other 9 stores in Arkansas?

Stores Performing below Average

Store 405, at \$ 28,146 was the best performing store that was below the State average. The negative Students' t statistic along with the p-value <.0001 indicates that Store 698 is significantly lower than the State average. Part of the purpose to understanding the differences among store performance is to learn from the top performing stores and improve the lower performing stores. With 9 stores below the State average, the following questions may arise from this analysis

• Should a new "State average" be computed eliminating Store 698?



Generated by the SAS System ('SASApp',X64_ES08R2) on May 29, 2018 at 1:47:19 PM



Generated by the SAS System ('SASApp',X64_ES08R2) on May 29, 2018 at 1:47:19 PM

- Which Stores should be differentiated as "best with this new analysis?
- Would Standards of performance hold from State to State?

In Conclusion

The final Process Flow should look similar to the screen shot below. Making use of the Enterprise Guide project to modify models tested allows the ability to keep prior analysis and finalize the results for distribution.

