Unit 1-12 Exercises

1. Counting Levels of a Variable with PROC FREQ

a. Retrieve the starter program **p112e01**.

b. Modify the program to produce two separate reports:

1) Display the number of distinct levels of **`Customer_ID`** and **`Employee_ID`** for retail orders.

a) Use a WHERE statement to limit the report to retail sales by specifying the condition **`Order_Type=1`**.

b) Display this report title: **`Unique Customers and Salespersons for Retail Sales`**.

If you do not want to see the counts for individual levels of **`Customer_ID`** and **`Employee_ID`**, add the NOPRINT option to the TABLES statement after a forward slash.

2) Display the number of distinct levels for **`Customer_ID`** for catalog and Internet orders.

a) Use a WHERE statement to limit the report to catalog and Internet sales by specifying the condition corresponding to **`Order_Type`** values other than `1`.

b) Display this report title: **`Unique Customers for Catalog and Internet`**.

If you do not want to see the counts for individual levels of **`Customer_ID`**, add the NOPRINT option to the TABLES statement after a forward slash.

c. Submit the program to produce the following reports:

PROC FREQ Output

Unique Customers and Salespersons for Retail Sales

The FREQ Procedure

Number of Variable Levels

Variable	Label	Levels
Customer_ID	Customer ID	31
Employee_ID	Employee ID	100

Unique Customers for Catalog and Internet Sales

The FREQ Procedure

Number of Variable Levels

Variable	Label	Levels
Customer_ID	Customer ID	63

2. Producing Frequency Reports with PROC FREQ

a. Retrieve the starter program **p112e02**.

b. Add TABLES statements to the PROC FREQ step to produce three frequency reports:

1) Number of orders in each year: Apply the YEAR4. format to the **`Order_Date`** variable to combine all orders within the same year.

2) Number of orders of each order type: Apply the **`ordertypes.`** format defined in the starter program to the **`Order_Type`** variable. Suppress the cumulative frequency and percentages.

3) Number of orders for each combination of year and order type: Suppress all percentages that normally appear in each cell of an *n*-way table.

c. Submit the program to produce the following output:

PROC FREQ Output

Order Summary by Year and Type

The FREQ Procedure

Date Order was placed by Customer

Order_Date	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2003	104	21.22	104	21.22
2004	87	17.76	191	38.98
2005	70	14.29	261	53.27
2006	113	23.06	374	76.33
2007	116	23.67	490	100.00

Order Type

Order_ Type	Frequency	Percent
Retail	260	53.06
Catalog	132	26.94
Internet	98	20.00

Table of Order_Date by Order_Type

Order_Date(Date Order was placed by Customer)
Order_Type(Order Type)

Frequency	Retail	Catalog	Internet	Total
2003	45	41	18	104
2004	51	20	16	87
2005	27	23	20	70
2006	67	33	13	113
2007	70	15	31	116
Total	260	132	98	490

3. Displaying PROC FREQ Output in Descending Frequency Order

a. Retrieve the starter program **p112e03**.

b. Submit the program to produce the following report:

PROC FREQ Output

Customer Demographics

(Top two levels for each variable?)

The FREQ Procedure

Customer Country

Customer_ Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AU	8	10.39	8	10.39
CA	15	19.48	23	29.87
DE	10	12.99	33	42.86
IL	5	6.49	38	49.35
TR	7	9.09	45	58.44
US	28	36.36	73	94.81
ZA	4	5.19	77	100.00

Customer Type Name

Customer_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Internet/Catalog Customers	8	10.39	8	10.39
Orion Club members high activity	11	14.29	19	24.68
Orion Club members medium activity	20	25.97	39	50.65
Orion Club Gold members high activity	10	12.99	49	63.64
Orion Club Gold members low activity	5	6.49	54	70.13
Orion Club Gold members medium activity	6	7.79	60	77.92
Orion Club members low activity	17	22.08	77	100.00

Customer Age Group

Customer_ Age_Group	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15-30 years	22	28.57	22	28.57
31-45 years	27	35.06	49	63.64
46-60 years	14	18.18	63	81.82
61-75 years	14	18.18	77	100.00

c. What are the two most common values for each variable?

1) **`Country`** _____ _____

2) **`Customer Type`** _____ _____

3) **`Customer Age Group`** _____ _____

d. Modify the program to display the frequency counts in descending order.

Documentation about the FREQ procedure can be found in the SAS Help and Documentation from the Contents tab (**SAS Products** ⇨ **Base SAS** ⇨ **Base SAS Procedures Guide: Statistical Procedures** ⇨ **The FREQ Procedure**).
Look for an option in the PROC FREQ statement that can perform the requested action.

e. Submit the modified program.

f. What are the two most common values for each variable?

1) **`Country`** _____ _____

2) **`Customer Type`** _____ _____

3) **`Customer Age Group`** _____ _____

Do these answers match the previous set of answers?

Which report was easier to use to answer the questions correctly?

4. Creating an Output Data Set with PROC FREQ

a. Retrieve the starter program **p112e04**.

b. Create an output data set containing the frequency counts based on **`Product_ID`**.

c. Combine the output data set with **`orion.product_list`** to obtain the **`Product_Name`** value for each **`Product_ID`** code.

d. Sort the merged data so that the most frequently ordered products appear at the top of the resulting data set. Print the first 10 observations, that is, those that represent the 10 products ordered most often.

.

e. Submit the program to produce the following report:

PROC PRINT Output

Top Ten Products by Number of Orders

Obs	Orders	Product Number	Product
1	6	230100500056	Knife
2	6	230100600030	Outback Sleeping Bag, Large,Left,Blue/Black
3	5	230100600022	Expedition10,Medium,Right,Blue Ribbon
4	5	240400300035	Smasher Shorts
5	4	230100500082	Lucky Tech Intergal Wp/B Rain Pants
6	4	230100600005	Basic 10, Left , Yellow/Black
7	4	230100600016	Expedition Zero,Medium,Right,Charcoal
8	4	230100600028	Expedition 20,Medium,Right,Forestgreen
9	4	230100700008	Family Holiday 4
10	4	230100700011	Hurricane 4

5. Creating a Summary Report with PROC MEANS

a. Retrieve the starter program **p112e05**.

b. Display only the SUM statistic for the **`Total_Retail_Price`** variable.

c. Display separate statistics for the combination of **`Order_Date`** and **`Order_Type`**. Apply the ORDERTYPES. format so that the order types are displayed as text descriptions, not numbers. Apply the YEAR4. format so that order dates are displayed as years, not individual dates.

d. Submit the program to produce the following report:

Partial PROC MEANS Output

Revenue (in U.S. Dollars) Earned from All Orders

The MEANS Procedure

Analysis Variable : Total_Retail_Price Total Retail Price for This Product

Date Order was placed by Customer	Order Type	N Obs	Sum
2003	Retail	53	7938.80
	Catalog	52	10668.08
	Internet	23	4124.05
2004	Retail	63	9012.22
	Catalog	23	3494.60
	Internet	22	3275.70
2005	Retail	34	5651.29
	Catalog	33	6569.98
	Internet	23	4626.40

6. Analyzing Missing Numeric Values with PROC MEANS

a. Retrieve the starter program **p112e06**.

b. Display the number of missing values and the number of nonmissing values present in the **`Birth_Date`**, **`Emp_Hire_Date`**, and **`Emp_Term_Date`** variables.

c. Suppress any decimal places in the displayed statistics.

d. Display separate statistics for each value of **`Gender`**.

e. Suppress the output column that displays the total number of observations in each classification group.

f. Submit the program to produce the following report:

PROC MEANS Output

Number of Missing and Non-Missing Date Values

The MEANS Procedure

Employee Gender	Variable	Label	N Miss	N
F	Birth_Date	Employee Birth Date	0	191
	Emp_Hire_Date	Employee Hire Date	0	191
	Emp_Term_Date	Employee Termination Date	139	52
M	Birth_Date	Employee Birth Date	0	233
	Emp_Hire_Date	Employee Hire Date	0	233
	Emp_Term_Date	Employee Termination Date	169	64

7. Analyzing All Possible Classification Levels with PROC MEANS

a. Retrieve the starter program **p112e07**.

b. Display the following statistics in the report:

1) Lower Confidence Limit for the Mean
2) Mean
3) Upper Confidence Limit for the Mean

c. Change the α value for the confidence limits to `0.10`, resulting in a 90% confidence limit.

d. Display all countries stored in the `Work.countries` data set in the report, even if there are no customers from that country.

Documentation about the MEANS procedure can be found in the SAS Help and Documentation from the Contents tab (**SAS Products** ⇨ **Base SAS** ⇨ **Base SAS 9.3 Procedures Guide** ⇨ **Procedures** ⇨ **The MEANS Procedure**). Look for options in the PROC MEANS statement that can perform the requested actions.

e. Submit the program to produce the following report:

PROC MEANS Output

Average Age of Customers in Each Country

The MEANS Procedure

Analysis Variable : Customer_Age Customer Age

Customer Country	N Obs	Lower 90% CL for Mean	Mean	Upper 90% CL for Mean
AU	8	42.4983854	52.3750000	62.2516146
BE	0	.	.	.
CA	15	31.2270622	40.0000000	48.7729378
DE	10	35.2564025	46.6000000	57.9435975
DK	0	.	.	.
ES	0	.	.	.
FR	0	.	.	.
GB	0	.	.	.
IL	5	30.1150331	40.0000000	49.8849669
NL	0	.	.	.
NO	0	.	.	.
PT	0	.	.	.
SE	0	.	.	.
TR	7	30.5050705	39.4285714	48.3520724
US	28	35.6505942	40.4285714	45.2065486
ZA	4	12.1696649	34.7500000	57.3303351

8. Creating an Output Data Set with PROC MEANS

a. Retrieve the starter program **p112e08**.

b. Create an output data set containing the sum of **`Total_Retail_Price`** values for each **`Product_ID`**.

c. Combine the output data set with **`orion.product_list`** to obtain the **`Product_Name`** value for each **`Product_ID`** code.

d. Sort the merged data so that the products with higher revenues appear at the top of the resulting data set. Print the first 10 observations, that is, those that represent the ten products with the most revenue.

e. Display the revenue values with a leading euro symbol (€), a period that separates every three digits, and a comma that separates the decimal fraction.

f. Submit the program to produce the following report:

PROC MEANS Output

Top Ten Products by Revenue

Obs	Revenue	Product Number	Product
1	€3.391,80	230100700009	Family Holiday 6
2	€3.080,30	230100700008	Family Holiday 4
3	€2.250,00	230100700011	Hurricane 4
4	€1.937,20	240200100173	Proplay Executive Bi-Metal Graphite
5	€1.796,00	240200100076	Expert Men's Firesole Driver
6	€1.561,80	240300300090	Top R&D Long Jacket
7	€1.514,40	240300300070	Top Men's R&D Ultimate Jacket
8	€1.510,80	240100400098	Rollerskate Roller Skates Ex9 76mm/78a Biofl
9	€1.424,40	240100400129	Rollerskate Roller Skates Sq9 80-76mm/78a
10	€1.343,30	240100400043	Perfect Fit Men's Roller Skates

9. Creating a Simple Tabular Report with PROC TABULATE

a. Retrieve the starter program **p112e09**.

b. Add a CLASS statement to enable **`Customer_Group`** and **`Customer_Gender`** as classification variables.

c. Add a VAR statement to enable **`Customer_Age`** as an analysis variable

d. Add a TABLE statement to create a report with the following characteristics:

1) **`Customer_Group`** defines the rows.

2) An extra row that combines all groups appears at the bottom of the table.

3) **`Customer_Gender`** defines the columns.

4) The N and MEAN statistics based on **`Customer_Age`** are displayed for each combination of **`Customer_Group`** and **`Customer_Gender`**.

e. Submit the program to produce the following report:

PROC TABULATE Output

Ages of Customers by Group and Gender

	Customer Gender			
	F		M	
	Customer Age		Customer Age	
	N	Mean	N	Mean
Customer Group Name				
Internet/Catalog Customers	4.00	49.35	4.00	54.25
Orion Club Gold members	11.00	35.36	10.00	38.90
Orion Club members	15.00	32.53	33.00	47.03
All	30.00	35.80	47.00	45.91

10. Creating a Three-Dimensional Tabular Report with PROC TABULATE

a. Retrieve the starter program **p112e10**.

b. Define a tabular report with the following characteristics:

1) **`Customer_Gender`** defines the page dimension.

2) **`Customer_Group`** defines the row dimension.

3) The column dimension should display the number of customers and the percentage of customers in each category (COLPCTN).

Change the headers for the statistic columns with a KEYLABEL statement. Documentation about the KEYLABEL statement can be found in the SAS Help
and Documentation from the Contents tab (**SAS Products** ⇨ **Base SAS** ⇨ **Base SAS 9.3 Procedures Guide** ⇨ **Procedures** ⇨ **The TABULATE Procedure**).

c. Submit the program to produce the following two-page report:

PROC TABULATE Output

Customers by Group and Gender

Customer Gender F

	Number	Percentage
Customer Group Name		
Internet/Catalog Customers	4.00	13.33
Orion Club Gold members	11.00	36.67
Orion Club members	15.00	50.00

Customers by Group and Gender

Customer Gender M

	Number	Percentage
Customer Group Name		
Internet/Catalog Customers	4.00	8.51
Orion Club Gold members	10.00	21.28
Orion Club members	33.00	70.21

11. Creating a Customized Tabular Report with PROC TABULATE

a. Retrieve the starter program **p112e11**.

b. Modify the label for the **`Total_Retail_Price`** variable.

c. Suppress the labels for the **`Order_Date`** and **`Product_ID`** variables.

d. Suppress the label for the SUM keyword.

e. Insert this text into the box above the row titles: **`High Cost Products (Unit Cost > $250)`**. Suppress all titles.

f. Display all calculated cell values with the DOLLAR12. format.

g. Display **`$0`** in all cells that have no calculated value.

Documentation about the TABULATE procedure can be found in the SAS Help and Documentation from the Contents tab (**SAS Products** ⇨ **Base SAS** ⇨ **Base SAS 9.3 Procedures Guide** ⇨ **Procedures** ⇨ **The TABULATE Procedure**).
Look for features of the PROC TABULATE statement, the TABLE statement, and
the KEYLABEL statement that can perform the requested actions.

h. Submit the program to produce the following report:
PROC TABULATE Output

High Cost Products (Unit Cost > $250)	Revenue for Each Product			
	230100700008	230100700009	240300100028	240300100032
2003	$0	$0	$0	$1,200
2005	$2,057	$2,256	$0	$0
2006	$0	$1,136	$0	$0
2007	$519	$0	$1,066	$0

12. Creating an Output Data Set with PROC TABULATE

a. Retrieve the starter program **p112e12**.

b. Create an output data set from the PROC TABULATE results. The output data set should contain average salaries for each combination of **`Company`** and **`Employee_Gender`**, plus overall averages for each **`Company`**.

Creating an output data set from PROC TABULATE results is discussed in the self-study content at the end of this section.

c. Sort the data set by **`average salary`**.

d. Print the sorted data set. Assign a format and column header to the **`average salary`** column.

e. Submit the program to produce the following report:

PROC PRINT Output

Average Employee Salaries

Obs	Company	Employee Gender	Average Salary
1	Orion Australia	F	$27,760
2	Orion USA	F	$29,167
3	Orion Australia		$30,574
4	Orion USA		$31,226
5	Orion USA	M	$32,534
6	Orion Australia	M	$32,963
7	Concession	F	$33,375
8	Purchasing	M	$33,462
9	Concession		$33,839
10	Concession	M	$34,650
11	Purchasing		$38,408
12	Logistics	F	$39,055
13	Purchasing	F	$41,556
14	Marketing	M	$42,645
15	Logistics		$43,128
16	Shared Functions	M	$43,428
17	Marketing		$44,390
18	Shared Functions		$44,631
19	Shared Functions	F	$46,016
20	Marketing	F	$47,132
21	Logistics	M	$47,630
22	Board of Directors	F	$68,370
23	Board of Directors		$134,034
24	Board of Directors	M	$212,831