Practice

If you restarted your SAS session, open and submit the **libname.sas** program in the course files.

Level 1

1. Exploring Data with Procedures

The **pg1.np_summary** table contains public use statistics from the National Park Service. Use the PRINT, MEANS, UNIVARIATE, and FREQ procedures to explore the data for possible inconsistencies.

a. Open **p103p01.sas** from the **practices** folder. Complete the PROC PRINT statement to list the first 20 observations in **pg1.np_summary**.

b. Add a VAR statement to include only the following variables: **Reg**, **Type**, **ParkName**, **DayVisits**, **TentCampers**, and **RVCampers**. Highlight the step and run the selected code.

Do you observe any possible inconsistencies in the data?

Obs	Reg	Type	ParkName	DayVisits	TentCampers	RVCampers
1	A	NM	Cape Krusenstern National Monument	15,000	0	0
2	A	NP	Kenai Fjords National Park	346,534	1,514	0
3	A	NP	Kobuk Valley National Park	15,500	0	0
4	A	PRE	Yukon-Charley Rivers National Preserve	1,146	0	0
5	A	PRE	Bering Land Bridge National Preserve	2,642	0	0
6	A	PRESERVE	Noatak National Preserve	17,000	0	0

c. Copy the PROC PRINT step and paste it at the end of the program. Change **PRINT** to **MEANS** and remove the OBS= data set option. Modify the VAR statement to calculate summary statistics for **DayVisits**, **TentCampers**, and **RVCampers**. Highlight the step and run the selected code.

What is the minimum value for tent campers? Is that value unexpected?

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
DayVisits	Recreational Day Visitors	135	966022.48	1568838.29	1146.00	11312786.00
TentCampers	Tent Campers	135	23870.81	60590.83	0	490431.00
RVCampers	RV Campers	135	14761.33	40977.10	0	376744.00

d. Copy the PROC MEANS step and paste it at the end of the program. Change **MEANS** to **UNIVARIATE**. Highlight the step and run the selected code.

Are there negative values for any of the columns?

e. Copy the PROC UNIVARIATE step and paste it at the end of the program. Change **UNIVARIATE** to **FREQ**. Change the VAR statement to a TABLES statement to produce frequency tables for **Reg** and **Type**. Highlight the step and run the selected code.

Are there any lowercase codes? Are there any codes that occur only once in the table?

Region Code				
Reg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	6	4.44	6	4.44
IM	52	38.52	58	42.96
MW	18	13.33	76	56.30
NC	1	0.74	77	57.04
NE	13	9.63	90	66.67
PW	23	17.04	113	83.70
SE	22	16.30	135	100.00

Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NM	63	46.67	63	46.67
NP	51	37.78	114	84.44
NPRE	1	0.74	115	85.19
NS	10	7.41	125	92.59
PRE	3	2.22	128	94.81
PRESERVE	4	2.96	132	97.78
RIVERWAYS	1	0.74	133	98.52
RVR	2	1.48	135	100.00

f. Add comments before each step to document the program. Save the program as **np_validate.sas** in the output folder.

Level 2

2. Using Procedures to Validate Data

The **pg1.np_summary** table contains information about US national parks, monuments, preserves, rivers, and seashores. Valid values for the columns **Reg** and **Type** are as follows:

Reg	Description
A	Alaska
IM	Intermountain
MW	Midwest
NC	National Capital
NE	Northeast
PW	Pacific West
SE	Southeast

Type	Description
NM	National Monument
NP	National Park
NS	National Seashore
PRE	National Preserve
RVR	National River

a. Create a new program. Write a PROC FREQ step to produce frequency tables for the **Reg** and **Type** columns in the **pg1.np_summary** table. Submit the step and look for invalid values.

b. Write a PROC UNIVARIATE step to generate statistics for the **Acres** column in the **pg1.np_summary** table. Notice the observation numbers for the smallest park and the largest park.

c. View the **pg1.np_summary** table to identify the name of the smallest and largest parks.

Challenge

3. Generating Extreme Observations Output

The **pg1.eu_occ** table includes monthly occupancy counts for European countries between January 2004 and September 2017.

The SAS Output Delivery System (ODS) gives you options for controlling the type and format of the output that is generated by SAS code. The ODS SELECT statement is used to specify output objects for results. The ODS SELECT statement can be used to generate a report containing only the Extreme Observations output.

Note: To specify an output object, you need to know which output objects your SAS program produces. The ODS TRACE statement writes to the SAS log a trace record that includes the path, the label, and other information about each output object that your SAS program produces. You can find documentation about the ODS TRACE and ODC SELECT statements in the SAS Help Facility and in the online documentation.

a. Create a new program. Write a PROC UNIVARIATE step to examine **Camp** in the **pg1.eu_occ** table.

b. Add the ODS TRACE statements before and after PROC UNIVARIATE as follows.

```
ods trace on;
proc univariate data=pg1.eu_occ;
    var camp;
run;
ods trace off;
```

c. Submit the program and notice the trace information in the SAS log. Determine the name of the Extreme Observations output object.

d. Delete the ODS TRACE statements. Add an ODS SELECT statement immediately before the PROC UNIVARIATE step and provide the name of the Extreme Observation output object.

Note: This method can be used with other procedures that create multiple tables (such as PROC CONTENTS) to select a portion of the output.

e. Using the SAS documentation or the syntax Help in the editor, identify the option that specifies the number of extreme observations that are listed in the table. Use the option to change the number of extreme observations from five to 10. Submit the program.

The UNIVARIATE Procedure
Variable: Camp (Nights Spent at Camp Grounds or RV Parks)

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	3204	36451329	1701
0	3203	37544738	1665
0	3202	38075450	1677
0	3201	38118264	1653
0	3200	44896837	1712
0	3199	45472015	1676
0	3198	45863306	1688
0	3197	46045595	1664
0	3196	46513414	1700
0	3195	47550694	1652

End of Practices