

Level 1 Practice: 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.

1. Open **p103p01.sas** from the **practices** folder.

- Complete the PROC PRINT statement to list the first 20 observations in **pg1.np_summary**.
- 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.

```
proc print data=pg1.np_summary(obs=20);  
    var Reg Type ParkName DayVisits TentCampers RVCampers;  
run;
```

2. Do you observe any possible inconsistencies in the data?

Yes. The **Type** column has inconsistencies. Notice that national preserve locations have the code **PRE** and **PRESERVE**.

3. Copy the PROC PRINT step and paste it at the end of the program.

- Replace PRINT with 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.

```
proc means data=pg1.np_summary;  
    var DayVisits TentCampers RVCampers;  
run;
```

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

The minimum value is zero. That value is not unexpected, because it's possible that a park had zero tent campers.

5. Copy the PROC MEANS step and paste it at the end of the program.

- Replace MEANS with UNIVARIATE.
- Highlight the step and run the selected code.

```
proc univariate data=pg1.np_summary;  
    var DayVisits TentCampers RVCampers;  
run;
```

6. Are there negative values for any of the columns?

No.

7. Copy the PROC UNIVARIATE step and paste it at the end of the program.

- Replace UNIVARIATE with FREQ.
- Replace the VAR statement with a TABLES statement to produce frequency tables for **Reg** and **Type**.
- Highlight the step and run the selected code.

```
proc freq data=pg1.np_summary;  
    tables Reg Type;  
run;
```

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

There are no lowercase codes. **NC**, **NPRE**, and **RIVERWAYS** occur once in the table.

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

```
/*list first 20 rows*/  
proc print data=pg1.np_summary(obs=20);  
    var Reg Type ParkName DayVisits TentCampers RVCampers;  
run;  
  
/*calculate summary statistics*/  
proc means data=pg1.np_summary;  
    var DayVisits TentCampers RVCampers;  
run;  
  
/*examine extreme values*/  
proc univariate data=pg1.np_summary;  
    var DayVisits TentCampers RVCampers;  
run;  
  
/*list unique values and frequency counts*/  
proc freq data=pg1.np_summary;  
    tables Reg Type;  
run;
```