

## Level 2 Practice: Using Procedures to Validate Data

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

Reg	Description	Type	Description
A	Alaska	NM	National Monument
IM	Intermountain	NP	National Park
MW	Midwest	NS	National Seashore
NC	National Capital	PRE	National Preserve
NE	Northeast	RVR	National River
PW	Pacific West		
SE	Southeast		

1. 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.

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

2. What invalid values exist for **Reg**?

No invalid values exist for **Reg**.

3. What invalid values exist for **Type**?

*NPRES*, *PRESERVE*, and *RIVERWAYS* are invalid values for **Type**.

4. Write a PROC UNIVARIATE step to generate statistics for the **Acres** column in the **pg1.np\_summary** table. Submit the step.

```
proc univariate data=pg1.np_summary;  
    var Acres;  
run;
```

5. What are the observation numbers for the smallest park and the largest park?

Smallest: Observation 78  
Largest: Observation 6

6. View the **pg1.np\_summary** table to identify the name and size of the smallest and largest parks.

Smallest: African Burial Ground Monument, .35 acres  
Largest: Noatak National Preserve, 6,587,071.39 acres