



5.3 - Practice

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

Level 1

1. Producing a Descriptive Statistic Report

The **pg1.np_weather** table contains weather-related information for four national parks: Death Valley National Park, Grand Canyon National Park, Yellowstone National Park, and Zion National Park. Use the MEANS procedure to analyze the data in this table.

- a. Create a new program. Write a PROC MEANS step to analyze rows from **pg1.np_weather** with the following specifications:
 - 1) Generate the mean, minimum, and maximum statistics for the **Precip**, **Snow**, **TempMin**, and **TempMax** columns.
 - 2) Use the MAXDEC= option to display the values with a maximum of two decimal positions.
 - 3) Use the CLASS statement to group the data by **Year** and **Name**.
 - 4) Use **Weather Statistics by Year and Park** as the report title. Run the program and review the results.

Weather Statistics by Year and Park						
Year	NAME	N Obs	Variable	Mean	Minimum	Maximum
2015	DEATH VALLEY, CA US	365	PRECIP	0.01	0.00	0.55
			SNOW	0.00	0.00	0.00
			TEMPMIN	64.44	26.00	97.00
			TEMPMAX	93.29	53.00	125.00
	GRAND CANYON VISITOR CENTER, AZ US	365	PRECIP	0.07	0.00	2.20
			SNOW	0.13	0.00	5.70
			TEMPMIN	40.99	8.00	69.00
			TEMPMAX	61.60	17.00	93.00
	YELLOWSTONE NATIONAL PARK EAST ENTRANCE, WY US	363	PRECIP	0.06	0.00	0.85
			SNOW	0.33	0.00	10.00
			TEMPMIN	23.05	-22.00	48.00
			TEMPMAX	53.11	9.00	89.00
	ZION NATIONAL PARK, UT US	362	PRECIP	0.05	0.00	1.23
			SNOW	0.01	0.00	4.00
			TEMPMIN	49.63	7.00	78.00
			TEMPMAX	75.60	35.00	110.00

Level 2

2. Creating an Output Table with Custom Columns

The **pg1.np_weather** table contains weather-related information for four national parks: Death Valley National Park, Grand Canyon National Park, Yellowstone National Park, and Zion National Park. Use the MEANS procedure to analyze the data in this table.

- Create a new program. Write a PROC MEANS step to analyze rows from **pg1.np_westweather** where values for **Precip** are **not** equal to zero. Analyze precipitation amounts grouped by **Name** and **Year**. Create only an output table, named **rainstats**, with columns for the N and SUM statistics. Name the columns **RainDays** and **TotalRain** respectively. Keep only those rows that are the combination of **Year** and **Name**.
- Write a PROC PRINT step to print the **rainstats** table. Suppress the printing of observation numbers, and display column labels. Display the columns in the following order: **Name**, **Year**, **RainDays**, and **TotalRain**. Label **Name** as **Park Name**, **RainDays** as **Number of Days Raining**, and **TotalRain** as **Total Rain Amount (inches)**. Use **Rain Statistics by Year and Park** as the report title.
- Run the program and review the results.

Rain Statistics by Year and Park			
Park Name	Year	Number of Days Raining	Total Rain Amount (inches)
DEATH VALLEY, CA US	2015	15	2.45
DEATH VALLEY, CA US	2016	16	1.42
DEATH VALLEY, CA US	2017	11	1.46
GRAND CANYON VISITOR CENTER, AZ US	2015	97	25.9
GRAND CANYON VISITOR CENTER, AZ US	2016	82	21.1
GRAND CANYON VISITOR CENTER, AZ US	2017	65	11
YELLOWSTONE NATIONAL PARK EAST ENTRANCE, WY US	2015	150	22.2
YELLOWSTONE NATIONAL PARK EAST ENTRANCE, WY US	2016	149	23.4
YELLOWSTONE NATIONAL PARK EAST ENTRANCE, WY US	2017	143	25.7
ZION NATIONAL PARK, UT US	2015	77	16.9
ZION NATIONAL PARK, UT US	2016	68	21.7
ZION NATIONAL PARK, UT US	2017	56	14.5

Challenge

3. Identifying the Top Three Extreme Values with the Output Statistics

- Create a new program. Write a PROC MEANS step to analyze rows from **pg1.np_multiyr** and create a table named **top3parks** with the following attributes:
 - Suppress the display of the PROC MEANS report.
 - Analyze **Visitors** grouped by **Region** and **Year**.
 - Drop the **_FREQ_** and **_TYPE_** columns from **top3parks** and keep only rows that are a result of a combination of **Region** and **Year**.
 - Create a column for **TotalVisitors** in the output table.
 - Include in the output table the top three parks in terms of the number of visitors. Automatically resolve conflicts in the column names when names are assigned to the new columns in the output table.

Note: Use SAS Help to learn about the IDGROUP option in the OUTPUT statement.

- Run the program and review the output.

	Region	Year	TotalVisitors	Visitors_1	Visitors_2	Visitors_3	ParkName_1	ParkName_2	ParkName_3
1	Alaska	2010	2,274,843	193,116	191,495	188,594	Klondike Go...	Klondike Go...	Klondike Go...
2	Alaska	2011	2,333,919	208,958	189,427	187,383	Klondike Go...	Klondike Go...	Klondike Go...
3	Alaska	2012	2,412,524	201,814	187,285	183,204	Klondike Go...	Klondike Go...	Klondike Go...
4	Alaska	2013	2,585,980	260,494	235,738	229,747	Klondike Go...	Klondike Go...	Klondike Go...
5	Alaska	2014	2,684,693	278,870	259,349	237,976	Klondike Go...	Klondike Go...	Klondike Go...
6	Alaska	2015	2,664,293	239,023	213,899	209,604	Klondike Go...	Klondike Go...	Klondike Go...
7	Alaska	2016	2,783,011	224,793	221,231	219,057	Klondike Go...	Klondike Go...	Klondike Go...
8	Intermountain	2010	42,652,924	957,785	854,837	694,841	Yellowstone...	Yellowstone...	Yellowstone...
9	Intermountain	2011	40,543,746	906,935	805,173	743,741	Yellowstone...	Yellowstone...	Rocky Mount...
10	Intermountain	2012	41,274,295	922,225	790,286	671,925	Yellowstone...	Yellowstone...	Yellowstone...