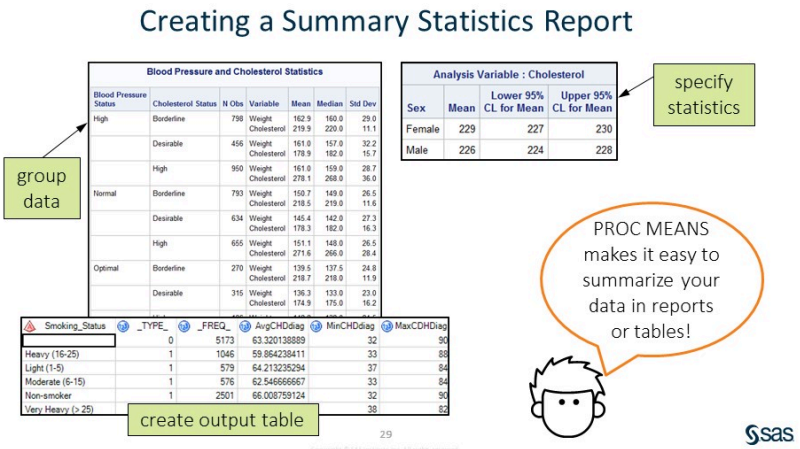
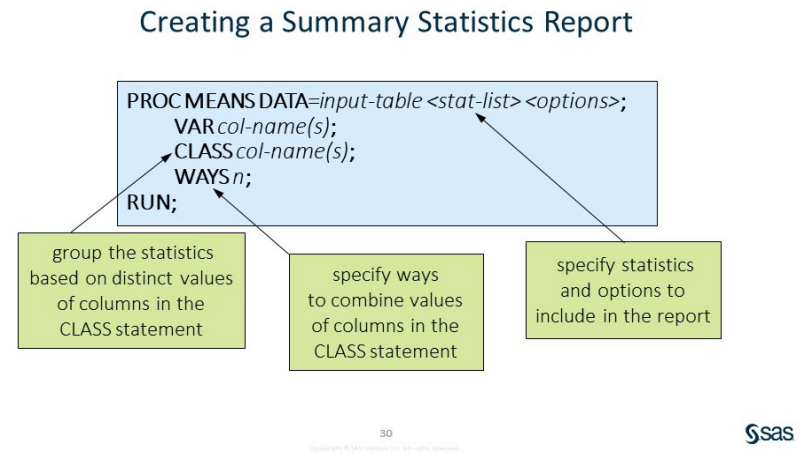
B5.3 - Creating Summary Statistics Reports

Summary Statistics Reports



The MEANS procedure is helpful for calculating basic summary statistics and looking for numeric values that might be outside of an expected range. But now that we're beyond validation, we can use PROC MEANS to generate complex reports that include various statistics and groupings within the data.



We've used the VAR statement to identify the numeric columns to analyze. Now we can add more to PROC MEANS to customize how data is summarized. In the PROC MEANS statement, you can specify the statistics you want calculated and how they should be displayed. The CLASS statement enables you to name one or more columns to group the data. Statistics are calculated for each unique value of the CLASS columns, and when you have more than one CLASS column, you can use the WAYS statement to control the combination of values of the CLASS columns. Let's go into a demo to see these new options and statements in action.

Demo: Creating Summary Statistics Reports

5_3 - Demo - Creating Summary Statistics Reports.pdf
(https://clemson.instructure.com/courses/237270/files/23074705/download?wrap=1)
(https://clemson.instructure.com/courses/237270/files/23074705/download?download_frd=1)

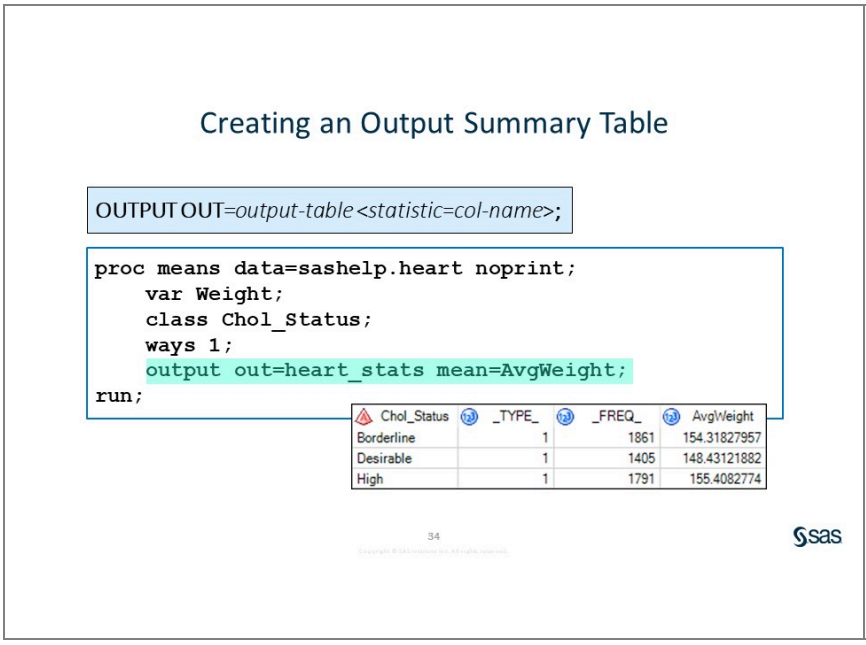
Activity 5.06

Open **p105a06.sas** from the **activities** folder and perform the following tasks:

1. Add options to include N (count), MEAN, and MIN statistics. Round each statistic to the nearest integer.
2. Add a CLASS statement to group the data by **Season** and **Ocean**. Run the program.
3. Modify the program to add the WAYS statement so that separate reports are created for **Season** and **Ocean** statistics. Run the program.
   Which ocean had the lowest mean for minimum pressure?
   Which season had the lowest mean for minimum pressure?

Click here for Solution

Output Summary Table



When you analyze detailed data, you might want to create a SAS table that summarizes the data for further analysis. PROC MEANS is a great way to create summary tables. The OUTPUT statement offers several options to customize the table that is generated. You use the OUT= option to name the output table. The OUTPUT statement also enables you to generate output statistics and name a column to store them in.

Activity 5.07

Open **p105a07.sas** from the **activities** folder and perform the following tasks:

1. Run the PROC MEANS step and compare the report and the **wind_stats** table. Are the same statistics in the report and table? What do the first five rows in the table represent?
2. Uncomment the WAYS statement. Delete the statistics listed in the PROC MEANS statement and add the NOPRINT option. Run the program. Notice that a report is not generated and the first five rows from the previous table are excluded.
3. Add the following options in the OUTPUT statement and run the program again. How many rows are in the output table?

```
output out=wind_stats mean=AvgWind max=MaxWind;
```

Click here for Solution.

Activity 5.08

Open **p105a08.sas** from the **activities** folder. On line 33 of the program add an 's' after 'http' to the website address so that the it is now "https://services.arcgisonline.com/arcgis/rest/services/World_Physical_Map".

Run the program and examine the results to see examples of other procedures that analyze and report on the data.