MET MA 603: SAS Project Exam Help SAS Programming and https://powcoder.com Applications Add WeChat powcoder

Summarizing Datasets

Summarizing Datasets

When working with large datasets, it is not always practical or possible to try to understand the information in the dataset at a glance. Calculating summary statistics can server the following purposes:
Assignment Project Exam Help

1. Discover any data entry errors or anomalies contained

- within the datasets://powcoder.com
- 2. Understand the distribution of the variables in the Add WeChat powcoder dataset.
- 3. Perform a preliminary exploration of the dataset before conducting a more detailed analysis.

The Frequency Procedure

The **Frequency Procedure** produces statistics about the distributions of variables in a dataset.

The Tables statement is an optional statement which instructs SAS_Awhich variables to include in the distribution analysis. Single variables and/or combination of variables may be listed.

https://powcoder.com

```
proc freq data = data.occupancy;
tables smokers residents economic powcoder,
run;
```

The Frequency Procedure (cont.)

The Frequency Procedure produces statistics about the distributions of variables in a dataset.

Additional options must follow a backslash (/) symbol.

The **LIST** options prints two pways distribution appalyses in database form rather than in a two-way table.

NOPERCENT supptpssepondodgrofpercentages.

MISSING includes missing values in the distribution tables.

OUT=name creates a SAS dataset called name containing the distribution data, which can then be used in Data steps.

```
proc freq data = data1.occupancy ;
tables dogbreed * dogs / list nopercent missing
out=occupancy freq;
run;
```

Practice

Using the Golf.sas7bdat dataset, use Proc Freq to create an output that matches what is on the rightnment Project E

The SAS System

The FREQ Procedure

Cumulative

lgnment Project Exa	Hole1	Hele72	Frequency	Frequency
ignificati i roject Exa	11112		P 2	2
	3	3	4	6
https://powcoder.co)M³	4	7	13
• •	3	5	2	15
Add WeChat powc	od^3	6	1	16
ridd Weenat powe	Out	2	2	18
	4	3	6	24
	4	4	17	41
	4	5	8	49
	5	2	1	50
	5	4	2	52
	5	5	2	54

The Means Procedure

The **Means Procedure** produces summary statistics about the variables in a dataset.

The default statistics included are the number of observations (n); Mean Standard Peviation (stddev), and Min and Max values.

If specific statistica targe: listed in the analysis.

Options statistics include the Median, Mode, Range (Max – Min), and Sum.

Maxdec=d can be included in the Proc Means statement to specify the number of decimal places to print.

```
proc means data = data1.occupancy mean median mode range ;
run;
```

The Means Procedure (cont.)

The BY statement performs the analysis separately for the listed variables. The dataset must be sorted according to the BY variables.

The VAR statement analyzes only en the listed variables.

The **CLASS statement** performs the analysis separately for the listed variables and performs the analysis separately table. The data does not need to be sorted.

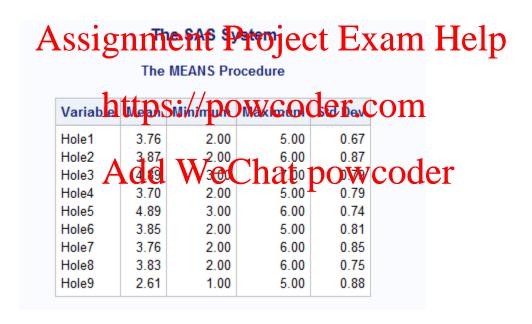
The BY and CLAS statements phouse dot both be used in the same Means Procedure.

The optional statement Output out=*name* creates a SAS dataset with the default means statistics.

```
proc means data = data1.occupancy maxdecs=2 n min max sum;
output out=occupancy_stats;
By dogbreed;
Var residents dogs;
run;
```

Practice

Using the Golf.sas7bdat dataset, use Proc Means to create an output that matches what is below.



Practice

Using the Scores1.sas7bdat dataset, use Proc Means to calculate the following statistics for the score of each school: average, standard deviation, minimum, and maximum. Assignment Project Exam Help

Do not display any decimal places in the result.

Add WeChat powcoder

Readings

Textbook sections 4.10, 4.12

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder