

MET MA 603:
Assignment 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:

1. Discover any data entry errors or anomalies contained within the dataset.
2. Understand the distribution of the variables in the dataset.
3. Perform a preliminary exploration of the dataset before conducting a more detailed analysis.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

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 which 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 dogbreed * dogs ;  
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** option prints two-way distribution analyses in database form rather than in a two-way table.

NOPERCENT suppresses printing of percentages.

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 right.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

The SAS System

The FREQ Procedure

| Hole1 | Hole72 | Frequency | Cumulative Frequency |
|-------|--------|-----------|----------------------|
| 2 | 4 | 2 | 2 |
| 3 | 3 | 4 | 6 |
| 3 | 4 | 7 | 13 |
| 3 | 5 | 2 | 15 |
| 3 | 6 | 1 | 16 |
| 4 | 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 Deviation (**stddev**), and **Min** and **Max** values.

If specific statistics are listed in the Proc Means statement, only those statistics will be included 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 on the listed variables.

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

The BY and CLASS statements should not 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.

Assignment Project Exam Help

The SAS System

The MEANS Procedure

| Variable | Mean | Minimum | Maximum | Std. Dev. |
|----------|------|---------|---------|-----------|
| Hole1 | 3.76 | 2.00 | 5.00 | 0.67 |
| Hole2 | 3.87 | 2.00 | 6.00 | 0.87 |
| Hole3 | 4.49 | 3.00 | 6.00 | 0.77 |
| Hole4 | 3.70 | 2.00 | 5.00 | 0.79 |
| Hole5 | 4.89 | 3.00 | 6.00 | 0.74 |
| Hole6 | 3.85 | 2.00 | 5.00 | 0.81 |
| Hole7 | 3.76 | 2.00 | 6.00 | 0.85 |
| Hole8 | 3.83 | 2.00 | 6.00 | 0.75 |
| Hole9 | 2.61 | 1.00 | 5.00 | 0.88 |

<https://powcoder.com>

Add WeChat powcoder

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. <https://powcoder.com>

[Add WeChat powcoder](#)

Readings

- Textbook sections 4.10, 4.12

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder