

**STA311**  
**Quiz 04**

This quiz is based on a SAS dataset **Sashelp.cars** available in SAS permanent library SASHELP. If you can access SAS built-in library Sashelp, you can simply read that data directly from the library. For example,

**PROC MEANS DATA = Sashelp.cars;**

Otherwise, please download the SAS data file from the course web page and save it to a folder with an appropriate file name. The direct link to this datafile:

[https://sta311.s3.amazonaws.com/w05\\_quiz4\\_dataset\\_cars.sas7bdat](https://sta311.s3.amazonaws.com/w05_quiz4_dataset_cars.sas7bdat)

For example, I download the data file and save it in my folder **C:\STA311** and rename the SAS data file as **cars**. Assuming that Citrix SAS can access the folder **C:\STA311**. To use this data file in DATA/ PROC steps, I defined a permanent SAS library pointing to folder **C:\STA311**:

**LIBNAME my311 "C:\STA311";**  
**PROC MEANS DATA = my311.cars;**

More than half of the multiple-choice questions in this quiz require you to write some SAS procedure steps and generate output from which you can find the correct answer. The primary procedures you may consider are PROC MEANS, PROC FREQ, and PROC UNIVARIATE. Please read the block comments in my example SAS code for the list of available descriptive statistics you can request from each of the three procedures and their corresponding definitions.

Some of the questions will ask you to find some of the descriptive statistics such as mean (MEAN), standard deviation (STD), lower quartile (Q1), upper quartile (Q3), median (MEDIAN), minimum (MIN), and maximum (MAX) data values, etc. Some of the keywords and statements you need to know are TABLE, VAR, BY, and CLASS. You are expected to know which statistics are available in which procedures.

Throughout this quiz, I assume that the **SAS data set cars** is save in C:\STA311. The Citrix SAS can access this folder.

**Problem 1.**

**Assume that** the SAS data set **cars** is in folder **C:\STA311**.

This week, we studied three procedures PROC MEANS, PROC FREQ, and PROC UNIVARIATE. Let's assume that there is a numerical variable. We want to find the biggest 5 values of the variable. Which procedures can be used to find the top five values of the variable?

- A. PROC MEANS
- B. PROC FREQ
- C. PROC UNIVARIATE
- D. Both PROC MEANS and PROC FREQ
- E. Both PROC FREQ and PROC UNIVARIATE

**Correct Answer: C.**

**Problem 2.**

Assume that the SAS data set **cars** is in folder **C:\STA311**.

Which of the following code reads data correctly to SAS and produces an output that contains the following table describing the variables in it?

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
9	Cylinders	Num	8		
5	DriveTrain	Char	5		
8	EngineSize	Num	8		Engine Size (L)
10	Horsepower	Num	8		
7	Invoice	Num	8	DOLLAR8.	
15	Length	Num	8		Length (IN)
11	MPG_City	Num	8		MPG (City)
12	MPG_Highway	Num	8		MPG (Highway)
6	MSRP	Num	8	DOLLAR8.	
1	Make	Char	13		
2	Model	Char	40		
4	Origin	Char	6		
3	Type	Char	8		
13	Weight	Num	8		Weight (LBS)
14	Wheelbase	Num	8		Wheelbase (IN)

A.

```
LIBNAME 'my311 C:\STA311';
```

```
PROC CONTENT DATA = my311.cars;  
RUN;
```

B.

```
LIBNAME my311 "C:\STA311";
```

```
PROC CONTENTS DATA = my311.cars;  
RUN;
```

C.

```
LIBNAME 'my311 C:\STA311';
```

```
PROC CONTENTS DATA = cars;  
RUN;
```

D.

```
LIBNAME my311 'C:\STA311';
```

```
DATA CARS;  
SET cars;  
RUN;
```

```
PROC PRINT DATA = CARS;  
RUN;
```

E.

```
LIBNAME my311 'C:\STA311';
```

```
PROC PRINT DATA = my311.cars;;  
RUN;
```

**Correct Answer: B**

### Problem 3.

Which of these procedures produces output that is most useful for characterizing the distribution of a categorical variable?

- A. PROC PRINT
- B. PROC FREQ
- C. PROC MEANS
- D. PROC UNIVARIATE
- E. PROC CONTENTS

**Correct Answer: B**

PROC MEANS and PROC UNIVARIATE analyze numeric variables. PROC PRINT is more useful for producing a report of invalid values, with a WHERE statement to specify conditions. PROC FREQ identifies duplicate values by default in the Frequency column of the report.

### Problem 4.

We studied three statistical procedures this week: PROC MEANS, PROC FREQ, PROC UNIVARIATE. In PROC MEANS, we can use either BY or CLASS statements to split the statistics by a categorical variable. Which of the following statement about BY and CLASS is correct?

- A. BY statement creates multiple single-row tables to store the statistics while CLASS statement creates a single table with multiple rows to store the statistics.
- B. CLASS statement creates multiple single-row tables to store the statistics while BY statement creates a single table with multiple rows to store the statistics.
- C. Both BY and CLASS statements create a single table with multiple rows to store the statistics
- D. Both BY and CLASS statements create a single table with multiple single-row tables to store the statistics.
- E. All listed statements are not correct.

**Correct Answer: A**

### Problem 5.

Which of the following description of the TABLE statement is correct?

- A. Using TABLE statement in PROC MEANS will create a frequency table for a character variable.
- B. Using TABLE statement in PROC MEANS will create a table of statistics such as minimum, mean, variance, and maximum values of a variable.
- C. Using TABLE statement in PROC FREQ will create a frequency table of a character variable.
- D. Using TBABLE statement in PROC UNIVARIATE will produce a table with the top five biggest data values and the bottom five smallest values.
- E. Using the TABLE statement in the PROC FREQ will create a frequency table of a numerical variable.

**Correct Answer: C.**

#### Problem 6.

Assume that the SAS data set `cars` in folder `C:\STA311` and define the permanent library to be `LIBNAME my311 'C:\STA311';`

Write an appropriate procedure step to produce the frequency distribution of character variable `Type` (vehicle type). Based on the frequency table, what proportion of vehicles are **SUV**?

- A. 11.45%
- B. 7.01%
- C. 5.61%
- D. 14.02%
- E. 60.00%

**Correct Answer: D**

```
PROC FREQ DATA = my311.cars;  
TABLE Type;  
RUN;
```

Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Hybrid	3	0.70	3	0.70
SUV	60	14.02	63	14.72
Sedan	262	61.21	325	75.93
Sports	49	11.45	374	87.38
Truck	24	5.61	398	92.99
Wagon	30	7.01	428	100.00

#### Problem 7.

Assume that the SAS data set `cars` is in folder `C:\STA311` and define the permanent library to be `LIBNAME my311 'C:\STA311';`

Write an appropriate procedure step to produce the descriptive statistics of MSRP (manufacturer's suggested retail price). Based on the output of the procedure, what are the mean and standard deviation of the MSRP?

- A. mean = 19431.72, standard deviation = 32774.86
- B. mean = 32774.86, standard deviation = 181.0383
- C. mean = 3277.86, standard deviation = 1943.72
- D. mean = 181.0383, standard deviation = 19431.72
- E. mean = 32774.86, standard deviation = 19431.72

**Correct Answer: E**

```
PROC MEANS DATA = my311.cars;  
VAR MSRP;  
RUN;
```

Analysis Variable : MSRP				
N	Mean	Std Dev	Minimum	Maximum
428	32774.86	19431.72	10280.00	192465.00

#### Problem 8.

Assume that the SAS data set `cars` is in folder `C:\STA311` and define the permanent library to be `LIBNAME my311 'C:\STA311';`

Write an appropriate procedure step to produce the descriptive statistics of MPG\_City (mile per gallon in the city) that include the five-number summary (Min, Q1, Median, Q3, Max). Which of the following is the correct 5-number summary of MPG\_City?

A.

Min	Lower Quartile	Median	Upper Quartile	Max
11	17	19	31	60

B.

Min	Lower Quartile	Median	Upper Quartile	Max
10	14	19	29	60

C.

Min	Lower Quartile	Median	Upper Quartile	Max
10	17	19	21	60

D.

Min	Lower Quartile	Median	Upper Quartile	Max
10	17	23	28	60

E

Min	Lower Quartile	Median	Upper Quartile	Max
10	12	19	28	60

**Correct Answer: C**

```
PROC MEANS MIN Q1 MEDIAN Q3 MAX DATA = my311.cars;
VAR MPG_City;
RUN;
```

Analysis Variable : MPG_City MPG (City)				
Minimum	Lower Quartile	Median	Upper Quartile	Maximum
10.0000000	17.0000000	19.0000000	21.5000000	60.0000000

### Problem 9.

Assume that the SAS data set **cars** is in folder **C:\STA311** and define the permanent library to be  
`LIBNAME my311 'C:\STA311';`

Write an appropriate procedure step to produce the frequency distribution of **DriveTrain** (all = all-wheel drive, front = front-wheel drive, rear = rear-wheel drive). Based on the frequency table, which DriveTrain is most popular?

- A. Need more information to answer the questions.
- B. Rear-wheel and front-wheel are equally popular
- C. Rear-wheel drive
- D. Front-wheel drive
- E. All-wheel drive

**Correct Answer: D**

```
PROC FREQ DATA = my311.cars;
TABLE DriveTrain;
RUN;
```

DriveTrain	Frequency	Percent	Cumulative Frequency	Cumulative Percent
All	92	21.50	92	21.50
Front	226	52.80	318	74.30
Rear	110	25.70	428	100.00

**Problem 10.**

Assume that the SAS data set **cars** is in folder **C:\STA311** and define the permanent library to be **LIBNAME my311 'C:\STA311';**

Write an appropriate procedure step to produce the descriptive statistics that contain the mean and the standard deviation of **MSRP** split by **DriveTrain**. That is, we want to see the average **MSRP** and their corresponding standard deviation of all-wheel-drive, - front-wheel-drive, and rear-wheel-drive vehicles. Which of the following table gives the correct **DriveTrain**-specific **MSRP**? [Hint: using BY or CLASS statement, see SAS example code.]

A.

Analysis Variable: MSRP						
DriveTrain	N Obs	N	Mean	Std Dev	Minimum	Maximum
All	192	192	26483.49	14448.77	16497.00	86970.00
Front	126	126	34782.56	19697.04	10280.00	84600.00
Rear	110	110	46093.63	27306.18	12800.00	192465.00

B.

Analysis Variable: MSRP						
DriveTrain	N Obs	N	Mean	Std Dev	Minimum	Maximum
All	110	110	36483.49	14448.77	16497.00	86970.00
Front	226	226	44782.56	20697.04	10280.00	84600.00
Rear	92	92	26093.63	17306.18	12800.00	192465.00

C.

Analysis Variable: MSRP						
DriveTrain	N Obs	N	Mean	Std Dev	Minimum	Maximum
All	92	92	46483.49	24448.77	16497.00	86970.00
Front	226	226	34782.56	10697.04	10280.00	84600.00
Rear	110	110	26093.63	17306.18	12800.00	192465.00

D.

ANALYSIS VARIABLE: MSRP						
DRIVETRAIN	N Obs	N	Mean	Std Dev	Minimum	Maximum
ALL	92	92	36483.49	14448.77	16497.00	86970.00
FRONT	226	226	24782.56	10697.04	10280.00	84600.00
REAR	110	110	46093.63	27306.18	12800.00	192465.00

E.

Analysis Variable: MSRP						
DriveTrain	N Obs	N	Mean	Std Dev	Minimum	Maximum
All	226	226	36483.49	14448.77	16497.00	86970.00
Front	92	92	24782.56	10697.04	10280.00	84600.00
Rear	110	110	46093.63	27306.18	12800.00	192465.00

**Correct Answer: D**

```
PROC MEANS DATA = my311.cars;  
VAR MSRP;  
CLASS DriveTrain;  
RUN;
```

Analysis Variable : MSRP						
DriveTrain	N Obs	N	Mean	Std Dev	Minimum	Maximum
All	92	92	36483.49	14448.77	16497.00	86970.00
Front	226	226	24782.56	10697.04	10280.00	84600.00
Rear	110	110	46093.63	27306.18	12800.00	192465.00