

EPI 5143 Winter 2024 QUIZ 1

Due Tuesday February 6th, 11:59pm via Github

Answer all the questions by writing the necessary SAS code and producing any output from SAS procedures used. Provide the SAS code, the SAS log, and SAS output with your quiz solutions.

You are expected to do your own coding and submit your own independent work—no copying/collaboration is permitted for quizzes and exams, we will be monitoring this closely.

1. Save the quiz1 data (quiz1_data.sas7bdat) on your computer in your course data folder. Make it read only so you don't accidentally change it.
(this is a suggestion not for marks).

2. Create a new permanent SAS library called quiz1 that points to the folder on your computer where you saved your quiz1 dataset (you could also write protect the SAS library to prevent changing the original quiz data).
(ie so to reference this dataset in SAS it would be quiz1.quiz1_data).

```
LIBNAME Quiz1 '/home/u63577751/Quiz1/';
```

```
data Quiz1.quiz1;  
    set Quiz1.Quiz1_data;  
run;
```

3. Use PROC CONTENTS to find out some information about this dataset.

```
proc contents data=Quiz1.quiz1_data;  
run;
```

- a) How many observations does the dataset have?
The dataset contains 1400 observations
- b) How many variables does the dataset have?
The dataset has 9 variables

4. Use PROC FREQ to provide information about the variable diabetes. If this variable represents those individuals in the dataset with diabetes, what proportion of people in the dataset have diabetes?

```
proc freq;  
    table mentalhealth * (sex community health income DHHGAGE) / nopercnt norow cmh;  
run;
```

```
proportion= individuals with diabetes / total sample  
proportion= 190/1400  
proportion= 0.13
```

5. Use PROC UNIVARIATE to provide information about the variable X1.

```
proc univariate data= Quiz1.quiz1_data;  
    variable X1;  
    histogram X1;
```

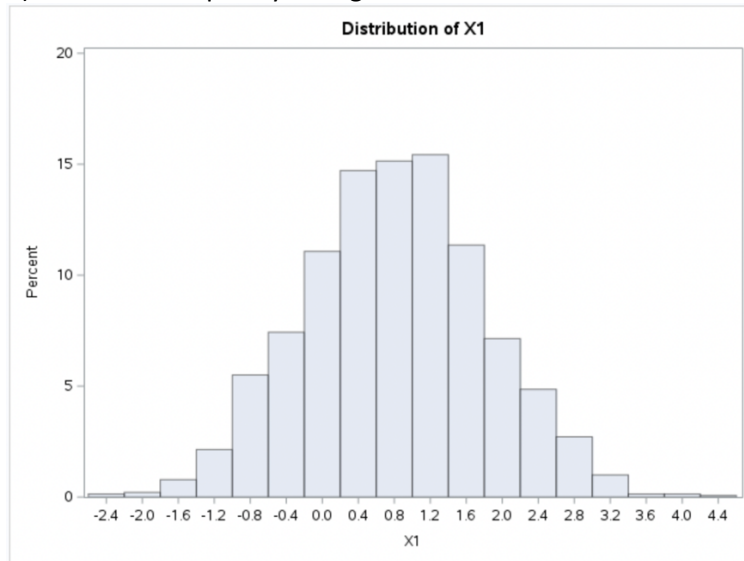
run;

a) What are the mean and standard deviation of X1?

Mean= 0.81

Standard Deviation= 1.01

b) Produce a frequency histogram of X1.



6. Create a temporary copy of the quiz1 dataset called work.quiz1.

```
data work.quiz1;  
    set Quiz1.quiz1_data;  
    sum_V1=X1+X2+X3;  
    sum_V2=sum(X1,X2,X3);  
run;
```

The remainder of the questions involve working with the work.quiz1 dataset.

7. a) Create a new variable called sum_V1 that is the sum of X1, X2 and X3 using mathematical operators.

```
data work.quiz1;  
    set Quiz1.quiz1_data;  
    sum_V1=X1+X2+X3;  
run;
```

b) Create a new variable called sum_V2 that is the sum of X1, X2 and X3 using a SAS function.

```
data work.quiz1;  
    sum_V1=X1+X2+X3;  
    sum_V2=sum(X1,X2,X3);
```

```
run;
```

8. Consult_dt and Surgery_dt are SAS dates. Create a new variable called wait_time that calculates the time in days between consult and surgery.

```
data work.quiz1;  
    wait_time= Surgery_dt-Consult_dt;  
run;
```

9. Create a new variable called X2_high which has a value of 1 if X2 is greater than or equal to the overall mean of X2 and 0 otherwise (you can find the mean of X2 using PROC UNIVARIATE).

```
data work.quiz1;  
    set work.  
    X2_high=0;  
    if X2>=11.97 then X2_high=1;  
run;
```

10. a) Use PROC UNIVARIATE to determine the mean and standard deviation of the variable sum_V2, and the median, minimum and maximum values for wait_time.

Sum_V2

Mean= 36.80

SD= 1.72

Wait_time

Median= 48

Minimum= 0

Maximum=99

c) Use PROC FREQ to create a 2x2 frequency table for X2_high vs. diabetes.

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of X2_high by diabetes			
	X2_high	diabetes		Total
		0	1	
0	602	88	690	49.29
	43.00	6.29		
	87.25	12.75		
	49.75	46.32		
1	608	102	710	50.71
	43.43	7.29		
	85.63	14.37		
	50.25	53.68		
Total	1210	190	1400	
	86.43	13.57	100.00	