Sumaiya Ahmed
6341568

EPI 5143 Winter 2025 QUIZ 1

1. Save the quiz1 data (quiz1_data.sas7bdat) on your computer in your course data folder. Make sure it is 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.
(ie. so to reference this dataset in SAS it would be quiz1.quiz1_data).

```
LIBNAME quiz1 "/home/u64004207/EPI5143 - Large Databases";

PROC DATASETS LIB=quiz1;
RUN;
QUIT;
```

3. Use PROC CONTENTS to find out some information about this dataset. How many observations does the dataset have? How many variables does the dataset have?

```
proc contents data =quiz1.quiz1_data;
run;
```

900 observations
8 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? (provide the frequency table from your SAS results with your answers).

```
proc freq data =quiz1.quiz1_data;
  tables diabetes;
run;
```

The FREQ Procedure

diabetes	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	803	89.22	803	89.22
1	97	10.78	900	100.00

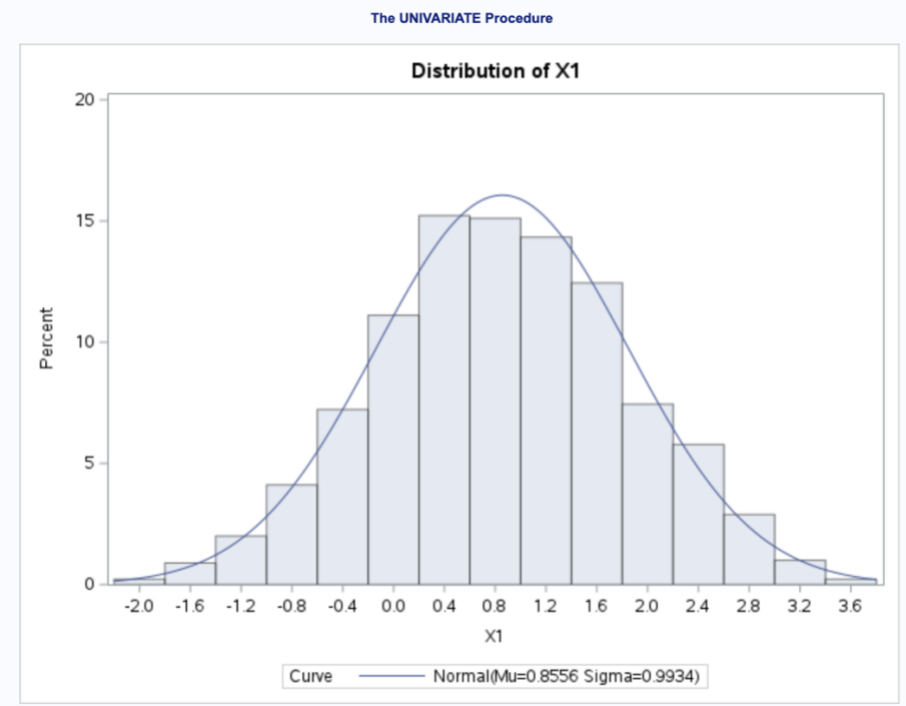
Proportion of people in the dataset have diabetes = 10.78%

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

```
proc univariate data=quiz1.quiz1_data;
```

```
  var X1;
  histogram X1 / normal;
run;
```

a) What are the mean and standard deviation of X1? Mean = 0.86, SD = 0.99
b) Produce a frequency histogram of X1 (provide with your answers).



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

```
data work.quiz1;
  set quiz1.quiz1_data;
run;
```

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

7. a) Create a new variable called mean_V1 that is the mean of X1, X2 and X3 using mathematical operators.
b) Create a new variable called mean_V2 that is the mean of X1, X2 and X3 using a SAS function.

```
data work.quiz1;
  set work.quiz1;
  mean_V1 = (X1 + X2 + X3) / 3;
  mean_V2 = mean(X1, X2, X3);
run;
```

Sumaiya Ahmed
6341568

EPI 5143 Winter 2025 QUIZ 1

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

9. Create a new variable called X1_high which has a value of 1 if X1 is greater than or equal to the mean of X1 in the dataset and 0 otherwise (you can find the mean of X1 using PROC UNIVARIATE or PROC MEANS).

```
proc means data =work.quiz1 mean;
   var X1;
   output out =mean_X1_data mean=X1_mean;
run;

data work.quiz1;
   set work.quiz1;
   if _N_ = 1 then set mean_X1_data;
   X1_high = (X1 >= X1_mean);
run;
```

10. a) Use PROC UNIVARIATE to find out the mean values of the variables mean_V1, and mean_V2, and the median, minimum and maximum values for wait_time.

```
proc univariate data =work.quiz1;
   var mean_V1 mean_V2 wait_time;
   output out=univ_output mean=mean_V1_mean mean_V2_mean wait_time_mean
      median=wait_time_median min=wait_time_min max=wait_time_max;
run;
```

Sumaiya Ahmed
6341568

EPI 5143 Winter 2025 QUIZ 1

The UNIVARIATE Procedure
Variable: mean_V1

Moments			
N	900	Sum Weights	900
Mean	12.3058822	Sum Observations	11075.294
Std Deviation	0.59283406	Variance	0.35145222
Skewness	0.04521235	Kurtosis	-0.138687
Uncorrected SS	136607.22	Corrected SS	315.955543
Coeff Variation	4.8174852	Std Error Mean	0.01976114

Basic Statistical Measures			
Location		Variability	
Mean	12.30588	Std Deviation	0.59283
Median	12.30553	Variance	0.35145
Mode	.	Range	3.63810
		Interquartile Range	0.84021

The UNIVARIATE Procedure
Variable: mean_V2

Moments			
N	900	Sum Weights	900
Mean	12.3058822	Sum Observations	11075.294
Std Deviation	0.59283406	Variance	0.35145222
Skewness	0.04521235	Kurtosis	-0.138687
Uncorrected SS	136607.22	Corrected SS	315.955543
Coeff Variation	4.8174852	Std Error Mean	0.01976114

Basic Statistical Measures			
Location		Variability	
Mean	12.30588	Std Deviation	0.59283
Median	12.30553	Variance	0.35145
Mode	.	Range	3.63810
		Interquartile Range	0.84021

Sumaiya Ahmed
6341568

EPI 5143 Winter 2025 QUIZ 1

The UNIVARIATE Procedure
Variable: wait_time

Moments			
N	900	Sum Weights	900
Mean	48.9555556	Sum Observations	44060
Std Deviation	28.7475341	Variance	826.420714
Skewness	0.01031414	Kurtosis	-1.1831833
Uncorrected SS	2899934	Corrected SS	742952.222
Coeff Variation	58.7216992	Std Error Mean	0.95825114

Basic Statistical Measures			
Location		Variability	
Mean	48.95556	Std Deviation	28.74753
Median	49.00000	Variance	826.42071
Mode	16.00000	Range	99.00000
		Interquartile Range	49.00000

b) Use PROC FREQ to create a 2x2 frequency table for X1_high vs. diabetes. (provide SAS frequency table with your answers).

```
proc freq data=work.quiz1;
   tables X1_high * diabetes / chisq norow nocol nopercent;
run;
```

The FREQ Procedure

Frequency

Table of X1_high by diabetes			
	diabetes		
X1_high	0	1	Total
0	410	50	460
1	393	47	440
Total	803	97	900