**EPI 5143 Winter 2021 QUIZ 1**

**Due Friday February 12th, 5pm via Github**

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

**1. Save the quiz1 data 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 quiz and specify the physical directory on your computer where you saved your quiz1 dataset. (ie so to reference this dataset in SAS it would be quiz.quiz1\_data).**

LIBNAME Quiz01 "C:\Users\Amber\Documents\Waterloo PhD Courses\EPI 5143 - Epidemiological Research Using Large Databases\Quiz Files\Quiz #1";

**RUN**;

**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=Quiz01.quiz1\_data;

**run**;

Number of observations: 1150

Number of variables: 6

*(\*\*I didn’t want to create a library that was simply ‘quiz’ as was suggested so I could tell this quiz apart from future ones\*\*)*

**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** data=Quiz01.quiz1\_data;

tables diabetes;

**run**;

Based on the results of this table, the proportion of individuals with diabetes is 3.3%

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

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

**PROC** **UNIVARIATE** data=Quiz01.quiz1\_data;

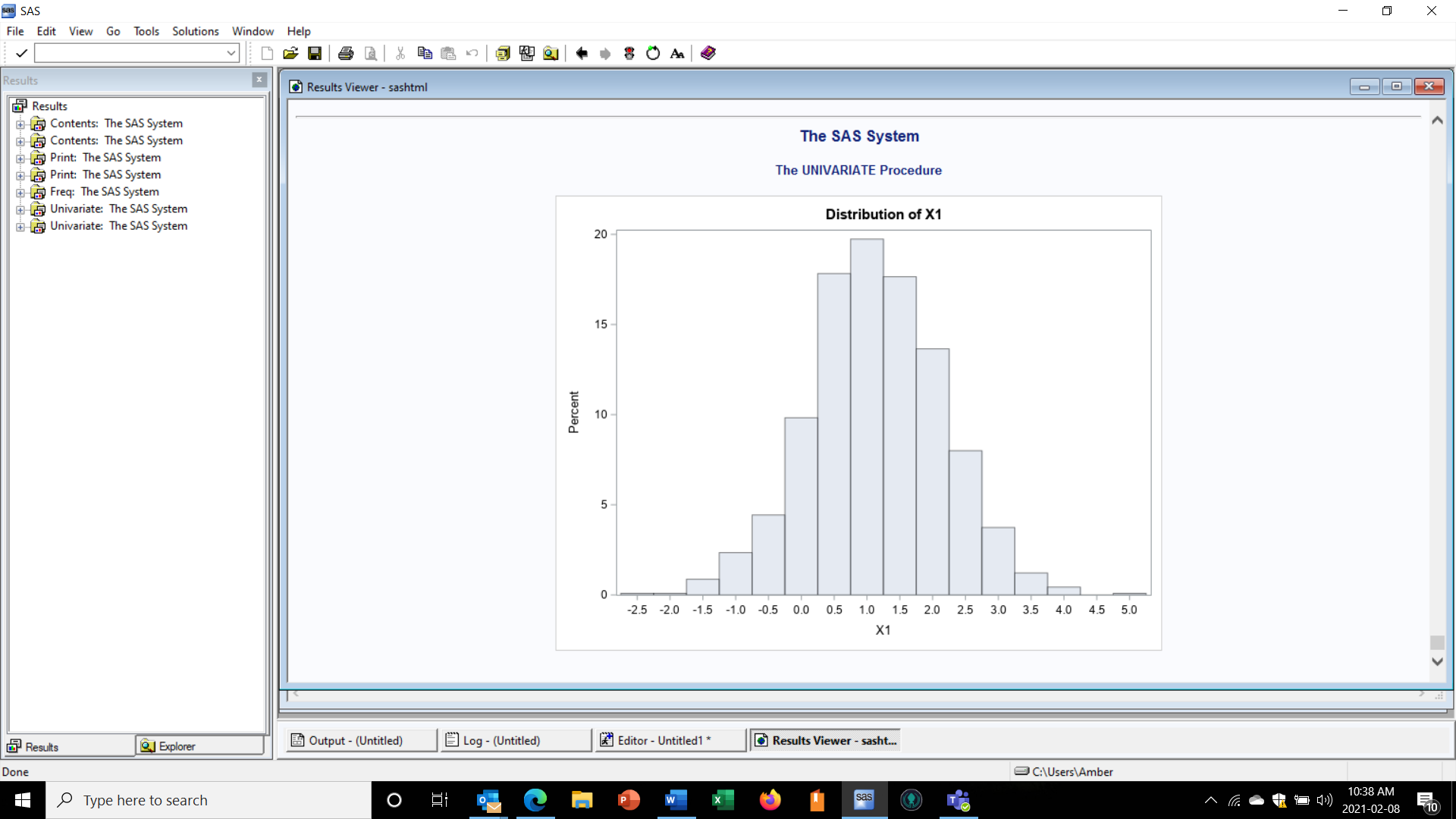
var X1;

**Run**;

Mean = 1.14

Standard deviation = 1.004

**b) Produce a frequency histogram of X1.**



**PROC** **UNIVARIATE** data=Quiz01.quiz1\_data;

var X1;

histogram;

**Run**;

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

**data** work.quiz1;

set quiz01.quiz1\_data;

**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;

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;

set quiz1;

sum\_V2 = sum (of X1--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;

set quiz1;

wait\_time = surgery\_dt-consult\_dt;

**run**;

**9. Create a new variable called X1\_level which has a value of 1 if X1 is greater than or equal to 1.25 and 0 otherwise.**

**data** work.quiz1;

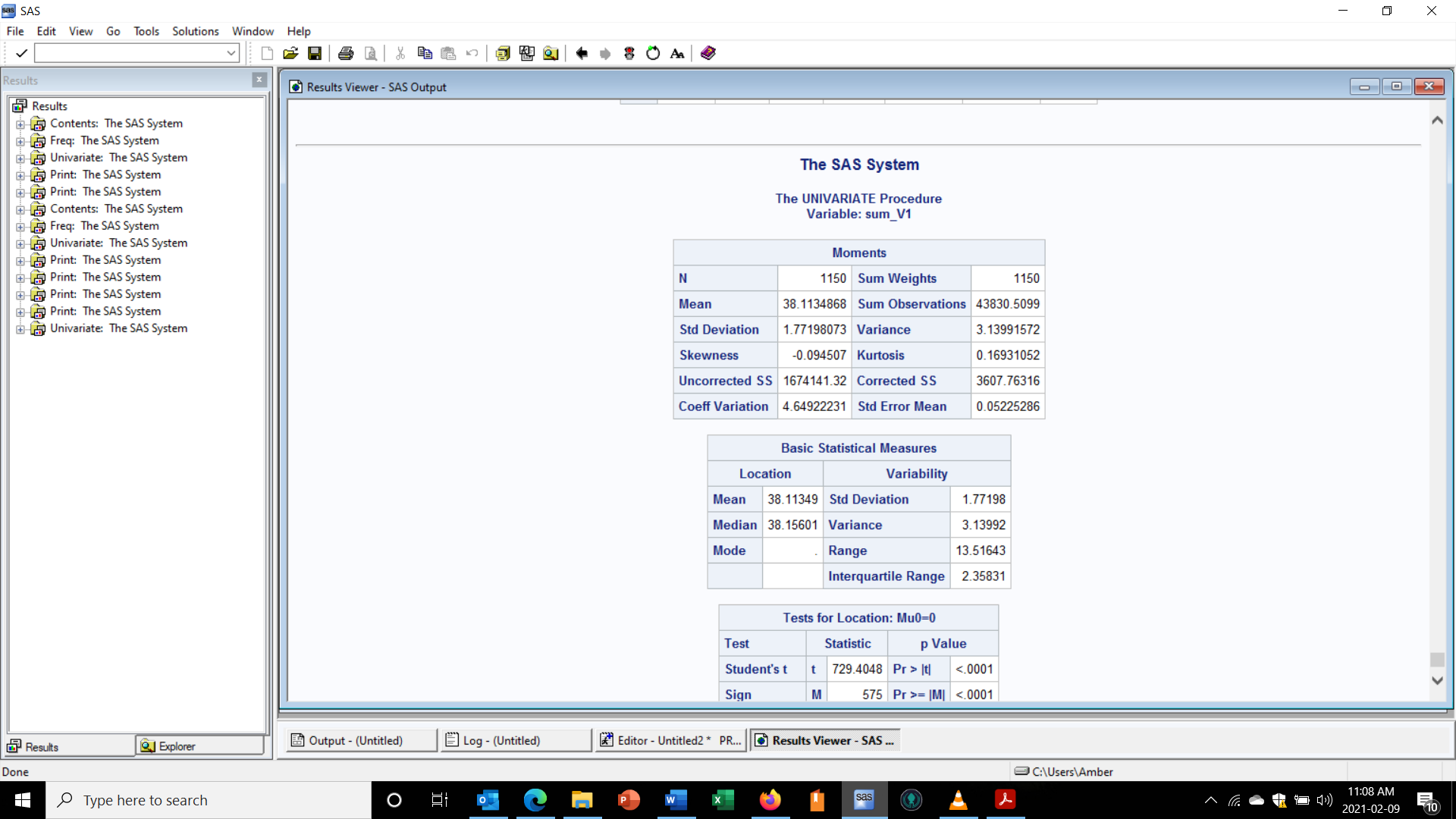
set quiz1;

X1\_level=**0**;

if X1>=**1** then X1\_level=**1**;

**run**;

**10. a) Use PROC UNIVARIATE to find out the mean of sum\_V1, sum\_V2, and wait\_time for the work.quiz1 dataset.**



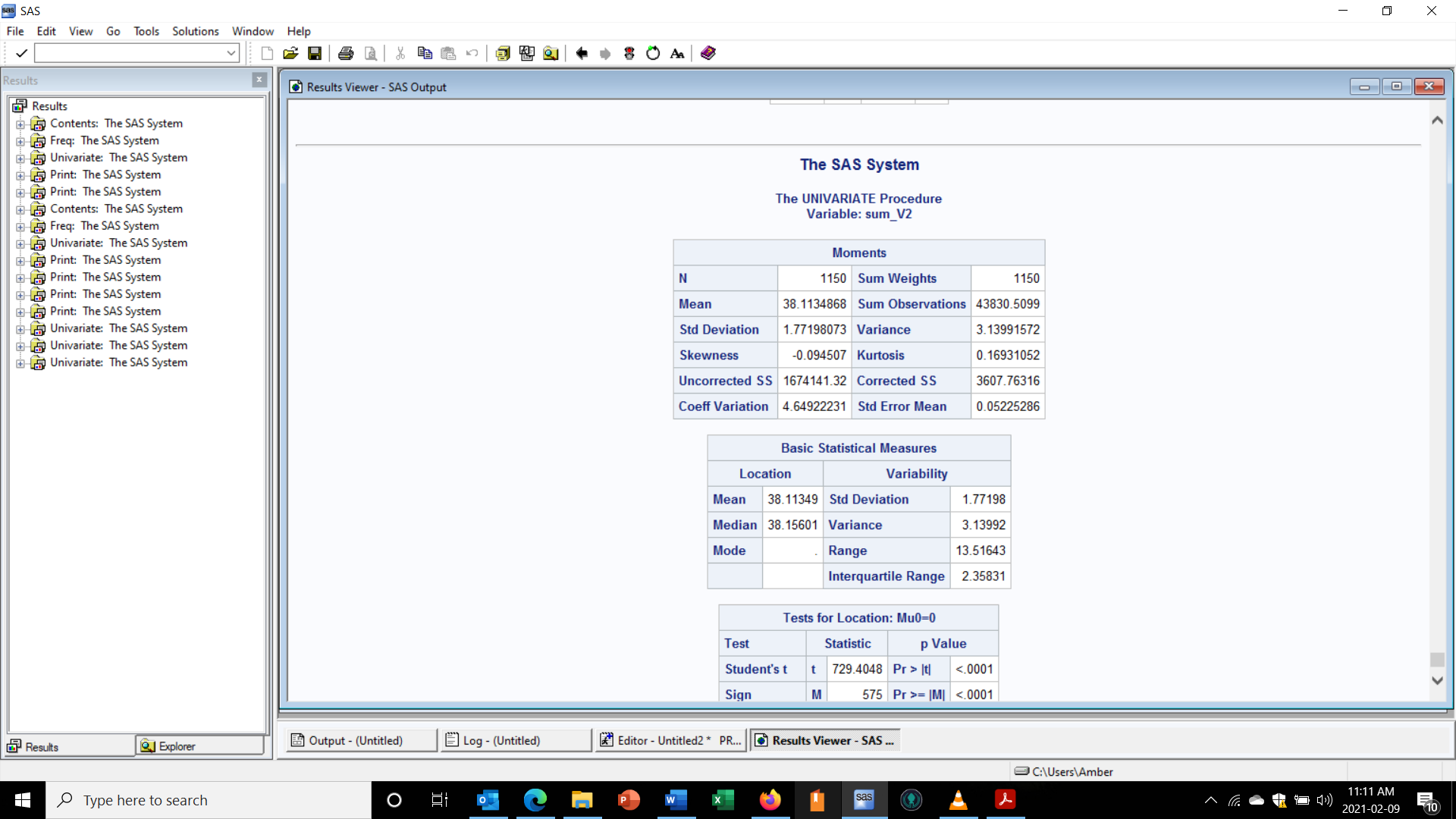
1. **sum\_V1**

**Proc** **Univariate** data=quiz1;

var sum\_v1;

**run**;

sum\_V1 mean = 38.11334868



1. **sum\_V2**

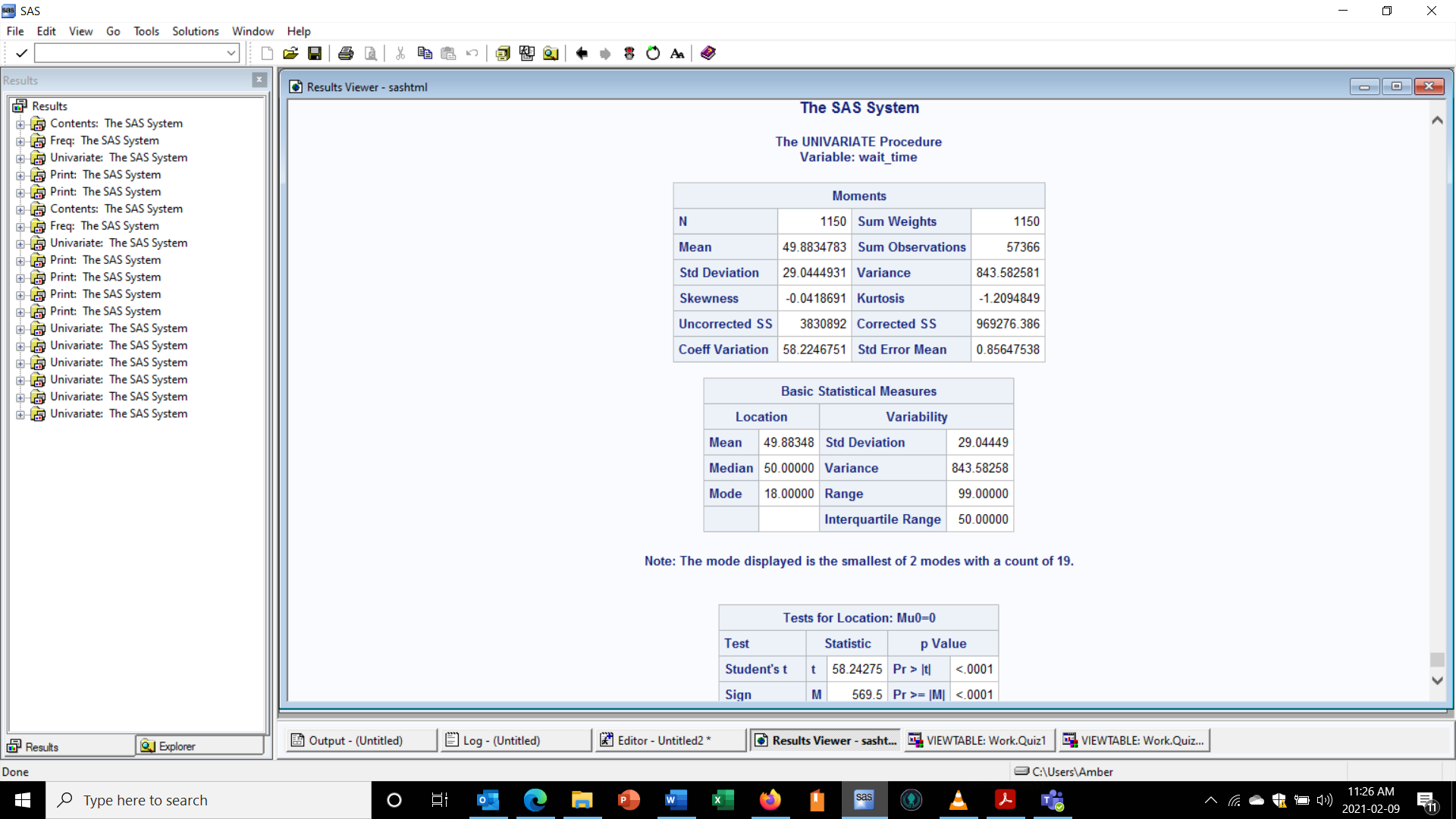
**Proc** **Univariate** data=quiz1;

var sum\_v2;

run;

sum\_V2 mean = 38.11334868

1. **wait\_time**



**Proc** **Univariate** data=quiz1;

var wait\_time;

**run**;

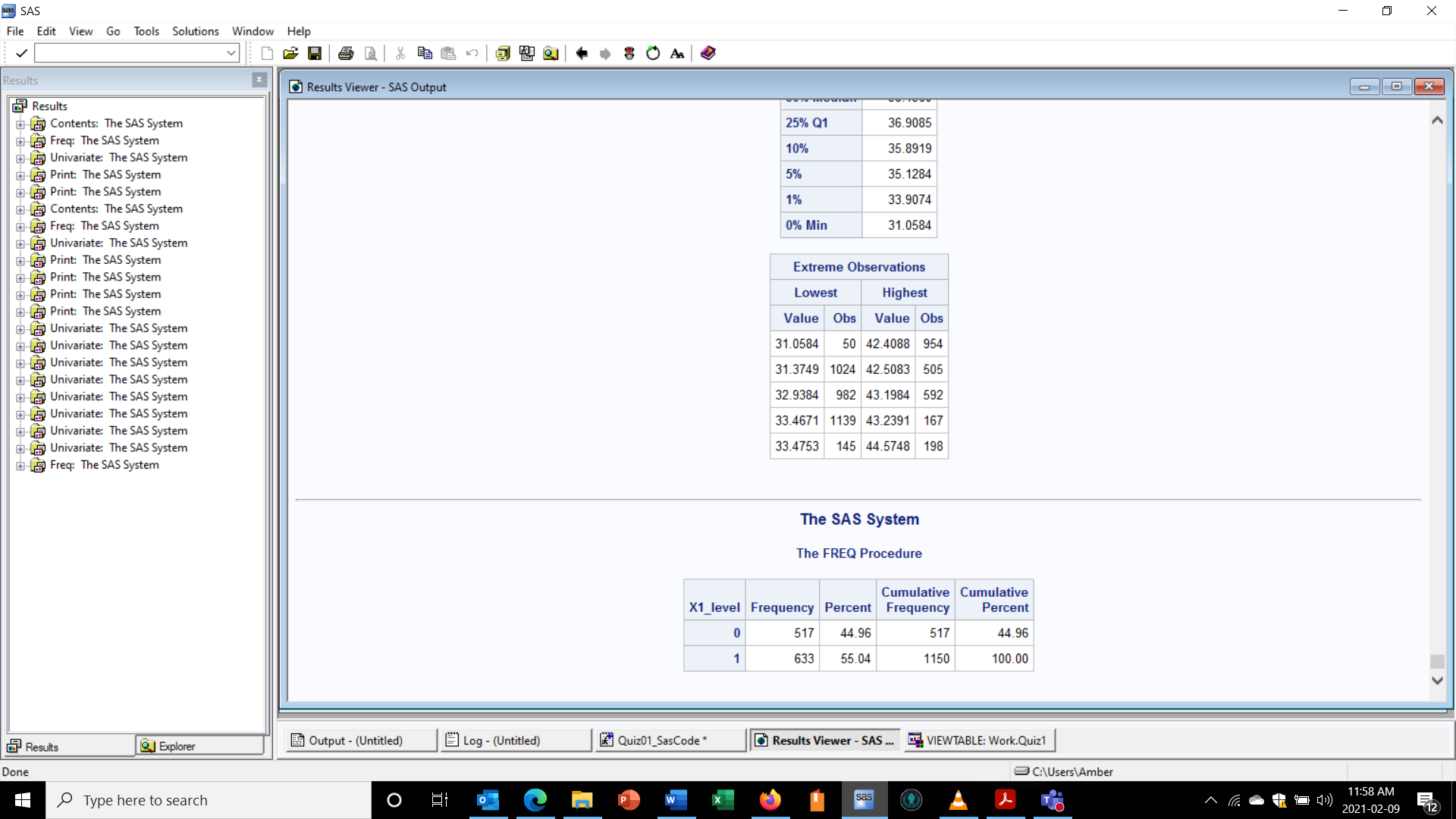
wait\_time mean = 41.8834783 days

**b) Use PROC FREQ to find out the frequency count for the X1\_level variable.**

**proc** **freq** data=work.quiz1;

tables X1\_level;

**run**;



Based on the above, the count frequency for the X1\_level variable is 44.96% for level 0 and 55.04% for level 1.