

# **Introduction To SAS Final Take Home Exam**

**Name** Ian Liu      **Grade** 93.04

# Grading Instruction

5 pts for each multiple-choice    13 Questions with a total of 65 points    -20

**10 pts for short answer**      **3 questions with 30 points , -8**

5 pts for each sub-question in the programming problem 1 Problem (a-d) 20, -5  
1 extra problem (a-e) 25

## **Section 1. Multiple Choice**

The final Grade  $(115-20-8-5+25)/115=93.04$

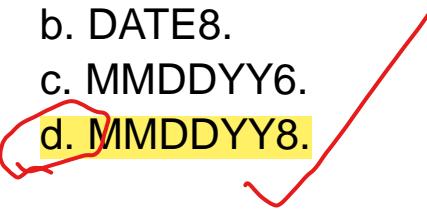
1. A great advantage of storing dates and times as SAS numeric date and time values is that

  - a. they can easily be edited.
  - b. they can easily be read and understood.
  - c. they can be used in text strings like other character values.
  - d. they can be used in calculations like other numeric values.

2. An input data file has date expressions in the form 10222001. Which SAS informat should you use to read these dates?

5

- a. DATE6.
- b. DATE8.
- c. MMDDYY6.
- d. MMDDYY8.



3. Which SAS statement reads the raw data values in order and assigns them to the variables shown below?

Variables: FirstName (character), LastName (character), Age (numeric), School (character), Class (numeric)

- a. input FirstName \$ LastName \$ Age School \$ Class;
- b. input FirstName LastName Age School Class;
- c. input FirstName \$ 1-4 LastName \$ 6-12 Age 14-15 School \$ 17-19 Class 21; —5
- d. input FirstName 1-4 LastName 6-12 Age 14-15 School 17-19 Class 21;

Correct answer is a (Because the data is free format, list input is used to read the values. With list input, you simply name each variable and identify its type.)

4. Which SAS program correctly reads the data in the raw data file that is referenced by the fileref Volunteer?

- a. data perm.contest;  
  infile volunteer;  
  input FirstName \$ LastName \$ Age School \$ Class;  
  run;
- b. data perm.contest;  
  infile volunteer;  
  length LastName \$ 11;  
  input FirstName \$ lastname \$ Age School \$ Class;  
  run;
- c. data perm.contest;  
  infile volunteer;  
  input FirstName \$ lastname \$ Age School \$ Class;  
  length LastName \$ 11;  
  run;
- d. data perm.contest;  
  infile volunteer;  
  input FirstName \$ LastName \$ 11. Age School \$ Class;  
  run;

5 Which SAS statement can be used to read a SAS data set?

- a. SET
- b. INFILE
- c. INPUT
- d. All of the above

6 Which of the following is considered a sum statement in the DATA step?

- a. X = A + B;
- b. X = SUM(A, B);
- c. A + B;
- d. All of the above

7 Which of the following is a valid name for a user-defined format?

- a. Body\_Mass\_Index\_Categories
- b. Description2
- c. Age(yrs)
- d. Varchar\$

8. Which PROC uses a TABLES statement?

- a. PRINT
- b. SORT
- c. FREQ
- d. All of the above

9.. To change the appearance of a data cell, which of the following PROC TABULATE statements would you use?

- a. CLASS
- b. CLASSLEV
- c. TABLE
- d. VAR

10 Suppose you are given one SAS data set that contains the GPAs of students in a certain college with seven majors of study and a second SAS data set with just one observation containing a variable for the overall average GPA in the college (CollegeGPA). How can you combine the two data sets so that the value for the variable CollegeGPA will be repeated for all observations in the data set of students?

- a. In a DATA step using a MERGE and a BY statement
- b. In a DATA step using a MERGE without a BY statement
- c. In a DATA step using two SET statements
- d. In a DATA step using one SET statement and a BY statement

→ 5

Correct Answer is C

11 If the following code is submitted and then followed by a PROC PRINT,  
what will be the resulting title in the output?

```
%LET title = The Amazing Race;
```

```
TITLE "&title";
```

- a. The Amazing Race
- b. &title
- c. title
- d. '&title'

12 What is true about conditional macro logic statements such as %IF  
-%THEN?

- a. They appear in the program after macro resolution
- b. They can be used to apply logic to sections of code
- c. They are only used to create variables in a data set
- d. None of the above

Correct answer is b

-5

13 What will happen if you specify TRANSPARENCY = 1 in a plot statement  
in PROC SGPlot?

- a. Transparency will be turned on
- b. Transparency will be turned off
- c. The plot feature will be dark
- d. The plot feature will disappear

Correct answer is d

-5

14 Chi-square tests can be computed using which procedure?

- a. PREQ
- b. UNIVARIATE
- c. CORR
- d. None of the above

15 A MODEL statement is required by which procedure?

- a. REG
- b. ANOVA
- c. Both
- d. Neither

## Section 2. Short Answer

1. Describe an advantage of using the CLASS statement rather than the BY statement with PROC MEANS.

If your data are not sorted and do not need to be sorted, then stacking requires fewer steps than interleaving. You would also want to stack if the goal is to maintain the original order of the observations in the data sets. If your data need to be sorted, then it may be more efficient to stack the data sets and then sort the combined data. Interleaving would require sorting each data set separately, which requires longer code and may waste computer resources for large data sets

**One advantage of using the CLASS statement rather than the BY statement with PROC MEANS is that the data do not have to be sorted first and the output is also more compact.**

2. Describe a situation where you might want to stack two SAS data sets rather than interleave them.

**One scenario where one might want to stack 2 SAS datasets instead of interleaving datasets is in question B where we want to have a single Plan variable for ANOVA class grouping. In general, when one wants to group data, it is better to stack variables then interleaving it.**

I don't understand your answer here. What is question B that you referred? The correct answer is provided on the top right corner. Since i did not cover this in the class, i will eliminate this question from the final.

3. What would you do if you found the following message in your SAS log when performing a one-to-many merge?

NOTE: MERGE statement has more than one data set with repeats of BY values.

If you get this message in the log, then you need to remove the duplicate observations from one of the data sets before doing the merge. You can check for duplicates in a data set using PROC SORT with the NODUPKEY option. This will remove all observations from the data set that have duplicate values of the BY variables

We can use the IN= option to specify which dataset we want to keep in its entirety. If the other dataset doesn't match the 1st dataset, then that row isn't merged.



- 4 You would like to carry out a hypothesis test to compare the average GPA of males and females for students in the College of Science and Math at your university. Your data set includes one observation for each student with GPA stored as one variable and gender stored as another variable. Besides the PROC and RUN statements, which two additional statements would be required to accomplish this with one PROC TTEST? Be specific about which variable goes with each statement.

You would need the CLASS and VAR statements.

```
PROC TTEST DATA=COLLEGE_GPA;  
  CLASS Gender;  
  VAR GPA;  
RUN;
```

### Section 3. Programming

You can only complete one problem (A or B) to get full points and Problem C is bonus points (30)

- A** The SAS data set called CATS contains information about breeds of cats. The variables are the name of the breed, the place where the breed originated, how the breed was derived, the type of hair, and a description of the breed's appearance. The following program produces counts of breed by place of origin, and then prints a report listing all the data for breeds that originated in Thailand. Modify the program using macro variables and a macro to increase its flexibility.

```
LIBNAME sasdata 'c:\MySASLib';
** Original code without macro variables;
PROC FREQ DATA = sasdata.cats;
TABLES Origin;
TITLE 'Cat Breeds by Origin';
RUN;

PROC PRINT DATA = sasdata.cats;
WHERE Origin = 'Thailand';
TITLE 'Cat Breeds with Origin = Thailand';
RUN;
```

- a. Examine this SAS data set including the variable labels and attributes.

Use %LET statements to create two macro variables: one to replace the variable name Origin, and another to replace the data value 'Thailand'.

Use an option that will enable you to see the standard SAS statements generated by the macro processor.

- b. Use the macro variables to produce counts for the variable Derivation, and to list data for breeds that were derived by mutation.

- c. Convert the code to a macro. Pass the values for the two macro variables into the macro as parameters. Call the macro to produce counts for the variable Hair, and to list data for breeds with long hair.
- d. Add programming to your macro from part c) that will save your output in a PDF file. Name this file CatRpt, and append the filename with a suffix that is the name of the variable used in the TABLES statement.

**B**) The local public transportation company would like to conduct a study of the impact of three possible plans on a certain popular bus route in town. Plan one consists of traffic lights timed in sync via a special remote transmitter on the bus that interacts with the traffic signal. Plan two consists of shutting down a regular traffic lane to become an express lane for buses only. Plan three is the current normal bus route with no changes. To assess the impact of a change, the company implemented plan one for a month, and then implemented plan two for a different month. All three plans were measured during summer months to ensure consistency in weather. The time (in minutes) for a bus to complete its route was measured for randomly selected weekdays and weekends, and these data are stored in the SAS data set called BUS.

- a. Examine this SAS data set including the variable labels and attributes. Calculate the mean, median, and standard deviation travel time for each plan. Present these descriptive statistics to one decimal place.
- b. Test to see whether there is any difference between the mean times for any of the plans, and compare the resulting p-value to alpha = 0.05.
- c. Add pairwise comparisons of group means to the analysis from part b).
- d. In a comment in your program, discuss your findings. Which plan should the city adopt?

## Extra Points

© Researchers at a local medical center have just completed enrollment for a clinical trial of a new cholesterol-lowering medication for use in subjects with borderline high total cholesterol. They keep their enrollment data in two SAS data sets so as to not bias the clinicians. The VISITS data set contains basic information about each subject at their baseline visit (Visit = 0). The TXGROUP data set contains information about whether the subject received the treatment or a placebo.

- a. Examine these SAS data sets including the variable labels and attributes. Combine these two data sources to create a data set that identifies the treatment group for each subject.
- b. The data entry clerk informs you that there may be duplicate records in the TXGROUP data set. Write code that will identify any duplicate entries and modify the programming for part a), if necessary.
- c. Using the combined data set from part b), calculate the median baseline cholesterol measurement across all subjects, and use this information to create a variable that groups subjects as less than or equal to the median, or more than the median. Do this without typing the calculated median value by hand into your code.
- d. Using the data set from part c), create another data set that schedules subjects for their next three visits, one every 30 days starting at the baseline visit date, for a total of four visits per subject. Each of the visits should appear as a new observation with the original subject information, the median grouping, the corresponding visit date, and a visit number (0, 1, 2, or 3).
- e. Add a comment to your program that states the number of observations and variables in the final data sets created in parts c) and d).

**B** The local public transportation company would like to conduct a study of the impact of three possible plans on a certain popular bus route in town. Plan one consists of traffic lights timed in sync via a special remote transmitter on the bus that interacts with the traffic signal. Plan two consists of shutting down a regular traffic lane to become an express lane for buses only. Plan three is the current normal bus route with no changes. To assess the impact of a change, the company implemented plan one for a month, and then implemented plan two for a different month. All three plans were measured during summer months to ensure consistency in weather. The time (in minutes) for a bus to complete its route was measured for randomly selected weekdays and weekends, and these data are stored in the SAS data set called BUS.

- a. Examine this SAS data set including the variable labels and attributes. Calculate the mean, median, and standard deviation travel time for each plan. Present these descriptive statistics to one decimal place.
- b. Test to see whether there is any difference between the mean times for any of the plans, and compare the resulting p-value to alpha = 0.05.
- c. Add pairwise comparisons of group means to the analysis from part b).
- d. In a comment in your program, discuss your findings. Which plan should the city adopt?

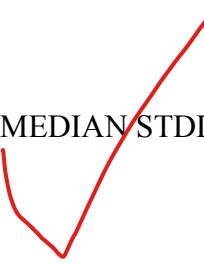
~~A~~ B

## CODE

```
/* B.a */  
DATA BUS;  
set '/home/u62223361/Intro to SAS/FinalExam/bus.sas7bdat';  
RUN;
```

```
PROC CONTENTS DATA = BUS;  
RUN;
```

```
PROC PRINT DATA = BUS;  
RUN;  
  
PROC MEANS DATA = BUS MEAN MEDIAN STDDEV MAXDEC=1;  
    VAR Plan1 Plan2 Plan3;  
RUN;
```



## LOG

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
68  
69          /* B.a */  
70          DATA BUS;  
71          set '/home/u62223361/Intro to  
SAS/FinalExam/bus.sas7bdat';  
NOTE: Data file WC000001.BUS.DATA is in a format that is native  
to another host, or the file encoding does not match the session  
encoding. Cross Environment Data Access will be used,  
which might require additional CPU resources and might reduce  
performance.  
72          RUN;  
NOTE: There were 10 observations read from the data set  
/home/u62223361/Intro to SAS/FinalExam/bus.sas7bdat.  
NOTE: The data set WORK.BUS has 10 observations and 4 variables.  
NOTE: DATA statement used (Total process time):  
      real time        0.00 seconds  
      user cpu time    0.00 seconds  
      system cpu time  0.00 seconds  
      memory           1188.09k  
      OS Memory        28216.00k
```

Timestamp	05/03/2024 03:46:55 AM		
Step Count	310	Switch Count	3
Page Faults	0		
Page Reclaims	155		
Page Swaps	0		
Voluntary Context Switches	22		
Involuntary Context Switches	0		
Block Input Operations	0		
Block Output Operations	264		

73

74 PROC CONTENTS DATA = BUS;  
75 RUN;

NOTE: PROCEDURE CONTENTS used (Total process time) :

real time	0.02 seconds		
user cpu time	0.02 seconds		
system cpu time	0.00 seconds		
memory	1963.15k		
OS Memory	28476.00k		
Timestamp	05/03/2024 03:46:55 AM		
Step Count	311	Switch Count	0
Page Faults	0		
Page Reclaims	108		
Page Swaps	0		
Voluntary Context Switches	2		
Involuntary Context Switches	0		
Block Input Operations	0		
Block Output Operations	16		

76

77 PROC PRINT DATA = BUS;  
78 RUN;

NOTE: There were 10 observations read from the data set  
WORK.BUS.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.01 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	775.12k
OS Memory	28216.00k
Timestamp	05/03/2024 03:46:55 AM
Step Count	312 Switch Count 0
Page Faults	0
Page Reclaims	68
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	0

79

80       PROC MEANS DATA = BUS MEAN MEDIAN STDDEV MAXDEC=1;  
81       VAR Plan1 Plan2 Plan3;  
82       RUN;

NOTE: There were 10 observations read from the data set  
WORK.BUS.

NOTE: PROCEDURE MEANS used (Total process time):

real time	0.01 seconds
user cpu time	0.01 seconds
system cpu time	0.01 seconds
memory	6652.31k
OS Memory	33868.00k
Timestamp	05/03/2024 03:46:55 AM
Step Count	313 Switch Count 1

Page Faults	0
Page Reclaims	1461
Page Swaps	0
Voluntary Context Switches	20
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

83

84           OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

94

User: u62223361

Messages: 30

## RESULTS

The CONTENTS Procedure

<b>Data Set Name</b>	WORK.BUS	<b>Observations</b>	10
<b>Member Type</b>	DATA	<b>Variables</b>	4
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	05/02/2024 23:46:55	<b>Observation Length</b>	32
<b>Last Modified</b>	05/02/2024 23:46:55	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	4078
<b>Obs in First Data Page</b>	10
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/saswork/SAS_work5358000083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/bus.sas7bdat
<b>Release Created</b>	9.0401M7
<b>Host Created</b>	Linux
<b>Inode Number</b>	1744845476
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	u62223361
<b>File Size</b>	256KB
<b>File Size (bytes)</b>	262144

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Label
4	Day	Char	8	Day (WE = weekend, WD = weekday)
1	Plan1	Num	8	Plan 1 - remote sensored lights
2	Plan2	Num	8	Plan 2 - express bus lane
3	Plan3	Num	8	Plan 3 - current route

Obs	Plan1	Plan2	Plan3	Day
1	17	23	24	WE
2	19	28	27	WD
3	15	30	30	WE
4	24	31	31	WD
5	22	24	35	WE
6	20	32	33	WD
7	22	28	30	WE
8	22	32	34	WD
9	16	26	28	WE
10	26	33	33	WD

### The MEANS Procedure

Variable	Label	Mean	Median	Std Dev
Plan1	Plan 1 - remote sensored lights	20.3	21.0	3.6
Plan2	Plan 2 - express bus lane	28.7	29.0	3.5
Plan3	Plan 3 - current route	30.5	30.5	3.4

# B

## CODE

```
/* B.b */

DATA BUS_PLAN1 (KEEP=Plan1 Plan Day RENAME= (Plan1 = Time));
    SET BUS;
    Plan = "Plan1";
RUN;

DATA BUS_PLAN2 (KEEP=Plan2 Plan Day RENAME= (Plan2 = Time));
    SET BUS;
    Plan = "Plan2";
RUN;

DATA BUS_PLAN3 (KEEP=Plan3 Plan Day RENAME= (Plan3 = Time));
    SET BUS;
    Plan = "Plan3";
RUN;

DATA BUSRESHAPED;
    SET BUS_PLAN1 BUS_PLAN2 BUS_PLAN3;
RUN;

PROC ANOVA DATA = BUSRESHAPED;
    CLASS Plan;
    MODEL Time = Plan;
    MEANS Plan / BON DUNCAN SCHEFFE SIDAK T TUKEY WALLER WELCH
    ALPHA=0.05 ;
RUN;
```

In this code, statement of  
MEANS Plan/SCHEFFE has already answered question C

## LOG

```
1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69      /* B.b */
70      DATA BUS_PLAN1 (KEEP=Plan1 Plan Day RENAME= (Plan1 =
Time));
71      SET BUS;
72      Plan = "Plan1";
73      RUN;
```

NOTE: There were 10 observations read from the data set  
WORK.BUS.

NOTE: The data set WORK.BUS\_PLAN1 has 10 observations and 3  
variables.

NOTE: DATA statement used (Total process time):

real time	0.00	seconds
user cpu time	0.00	seconds
system cpu time	0.00	seconds
memory	943.40k	
OS Memory	28732.00k	
Timestamp	05/03/2024 03:47:40	AM
Step Count	319	Switch Count 2
Page Faults	0	
Page Reclaims	127	
Page Swaps	0	
Voluntary Context Switches	11	
Involuntary Context Switches	0	
Block Input Operations	0	
Block Output Operations	264	

```
75      DATA BUS_PLAN2 (KEEP=Plan2 Plan Day RENAME= (Plan2 =  
Time));
```

```
76      SET BUS;
```

```
77      Plan = "Plan2";
```

```
78      RUN;
```

NOTE: There were 10 observations read from the data set  
WORK.BUS.

NOTE: The data set WORK.BUS\_PLAN2 has 10 observations and 3  
variables.

NOTE: DATA statement used (Total process time):

real time	0.00	seconds
user cpu time	0.00	seconds
system cpu time	0.00	seconds
memory	943.40k	
OS Memory	28732.00k	
Timestamp	05/03/2024 03:47:40	AM
Step Count	320	Switch Count 2
Page Faults	0	
Page Reclaims	129	
Page Swaps	0	
Voluntary Context Switches	11	
Involuntary Context Switches	0	
Block Input Operations	0	
Block Output Operations	264	

```
79
```

```
80      DATA BUS_PLAN3 (KEEP=Plan3 Plan Day RENAME= (Plan3 =  
Time));
```

```
81      SET BUS;
```

```
82      Plan = "Plan3";
```

```
83      RUN;
```

NOTE: There were 10 observations read from the data set WORK.BUS.

NOTE: The data set WORK.BUS\_PLAN3 has 10 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1055.53k
OS Memory	28732.00k
Timestamp	05/03/2024 03:47:40 AM
Step Count	321 Switch Count 2
Page Faults	0
Page Reclaims	130
Page Swaps	0
Voluntary Context Switches	10
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

84

85       DATA BUSRESHAPED;  
86       SET BUSPLAN1 BUSPLAN2 BUSPLAN3;  
87       RUN;

NOTE: There were 10 observations read from the data set WORK.BUS\_PLAN1.

NOTE: There were 10 observations read from the data set WORK.BUS\_PLAN2.

NOTE: There were 10 observations read from the data set WORK.BUS\_PLAN3.

NOTE: The data set WORK.BUSRESHAPED has 30 observations and 3 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.01 seconds
memory	1665.87k
OS Memory	29252.00k
Timestamp	05/03/2024 03:47:40 AM
Step Count	322    Switch Count  2
Page Faults	0
Page Reclaims	201
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

88  
89        PROC ANOVA DATA = BUSRESHAPED;  
90        CLASS Plan;  
91        MODEL Time = Plan;  
92        MEANS Plan / BON DUNCAN SCHEFFE SIDAK T TUKEY WALLER  
WELCH ALPHA=0.05 ;  
93        RUN;  
94  
95        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
105

## RESULTS

### The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Plan	3	Plan1 Plan2 Plan3

Number of Observations Read	30
Number of Observations Used	30

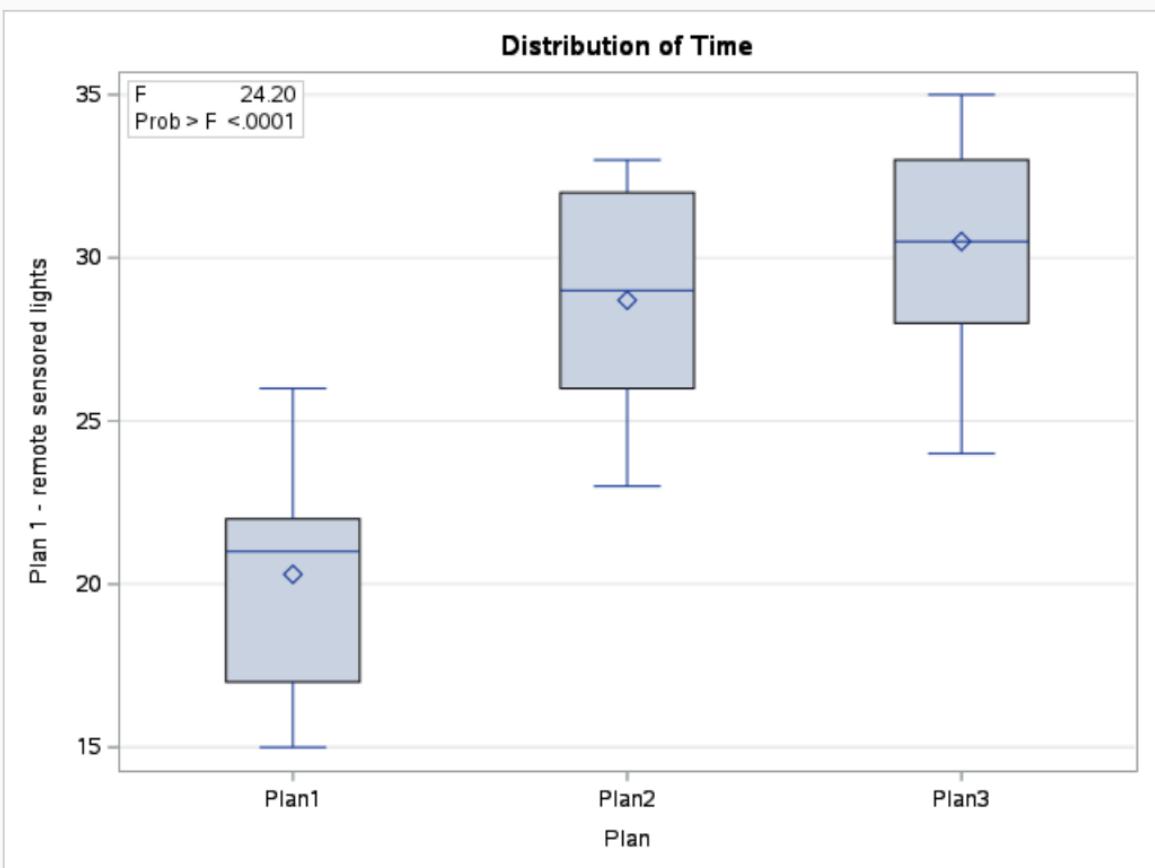
### The ANOVA Procedure

Dependent Variable: Time Plan 1 - remote sensed lights

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	592.8000000	296.4000000	24.20	<.0001
Error	27	330.7000000	12.2481481		
Corrected Total	29	923.5000000			

R-Square	Coeff Var	Root MSE	Time Mean
0.641906	13.20655	3.499735	26.50000

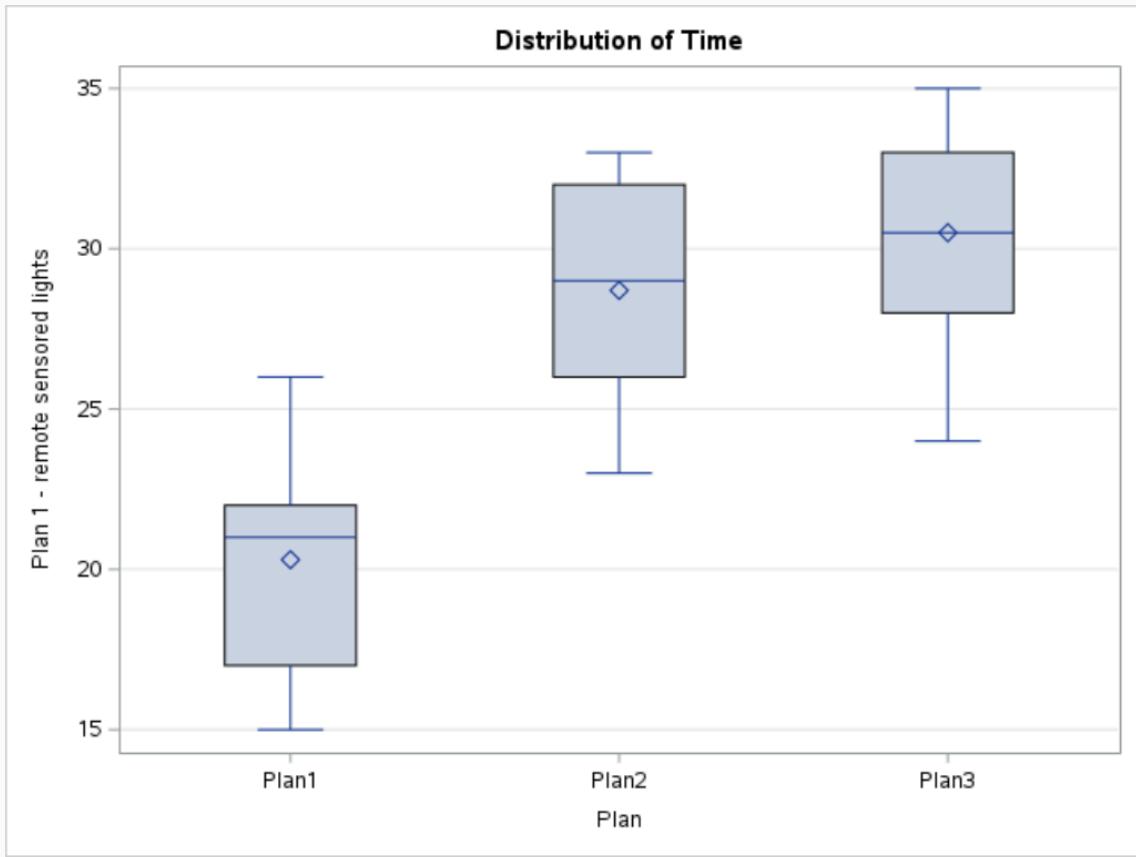
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Plan	2	592.8000000	296.4000000	24.20	<.0001



**The ANOVA Procedure**

Welch's ANOVA for Time			
Source	DF	F Value	Pr > F
Plan	2.0000	23.03	<.0001
Error	17.9964		

The ANOVA Procedure



### The ANOVA Procedure

#### Waller-Duncan K-ratio t Test for Time

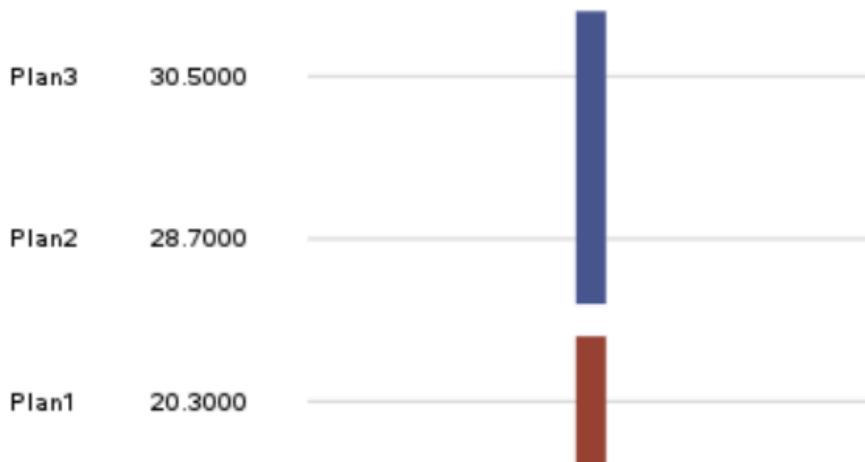
**Note:** This test minimizes the Bayes risk under additive loss and certain other assumptions.

Kratio	100
Error Degrees of Freedom	27
Error Mean Square	12.24815
F Value	24.20
Critical Value of t	1.86919
Minimum Significant Difference	2.9255

#### Time Waller Grouping for Means of Plan (Alpha = 0.05)

Means covered by the same bar are not significantly different.

##### Plan Estimate



### The ANOVA Procedure

#### t Tests (LSD) for Time

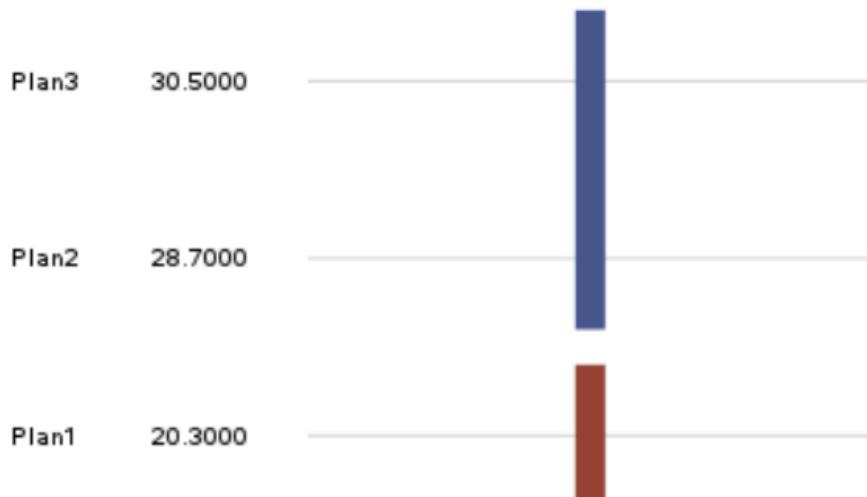
**Note:** This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

<b>Alpha</b>	0.05
<b>Error Degrees of Freedom</b>	27
<b>Error Mean Square</b>	12.24815
<b>Critical Value of t</b>	2.05183
<b>Least Significant Difference</b>	3.2114

#### Time t Grouping for Means of Plan (Alpha = 0.05)

Means covered by the same bar are not significantly different.

##### Plan Estimate



### The ANOVA Procedure

#### Duncan's Multiple Range Test for Time

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

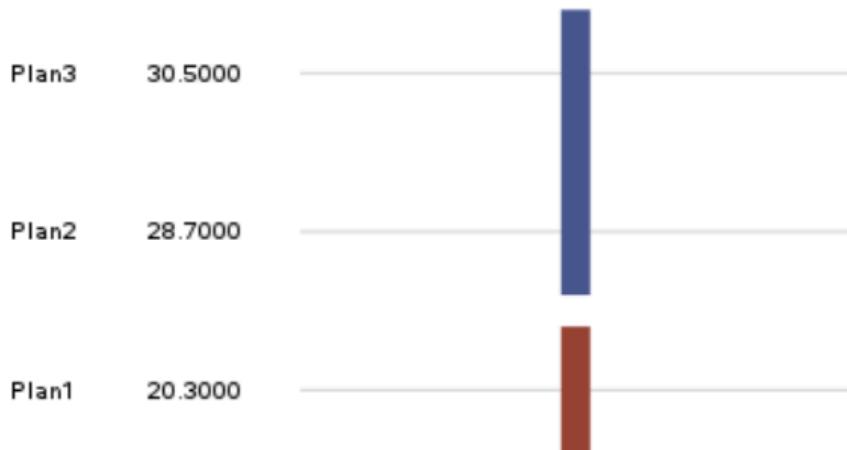
Alpha	0.05
Error Degrees of Freedom	27
Error Mean Square	12.24815

Number of Means	2	3
Critical Range	3.211	3.374

#### Time Duncan Grouping for Means of Plan (Alpha = 0.05)

Means covered by the same bar are not significantly different.

##### Plan Estimate



**The ANOVA Procedure**

**Tukey's Studentized Range (HSD) Test for Time**

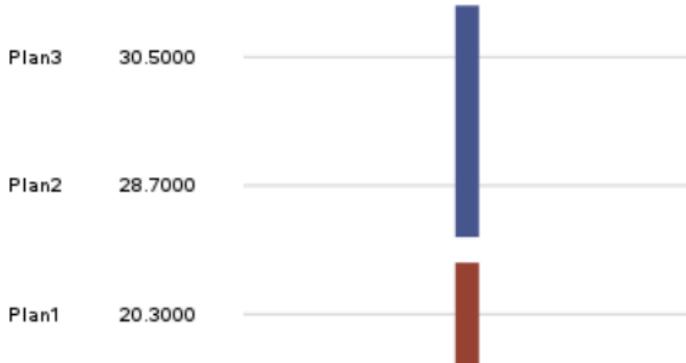
**Note:** This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

<b>Alpha</b>	0.05
<b>Error Degrees of Freedom</b>	27
<b>Error Mean Square</b>	12.24815
<b>Critical Value of Studentized Range</b>	3.50633
<b>Minimum Significant Difference</b>	3.8805

**Time Tukey Grouping for Means of Plan (Alpha = 0.05)**

Means covered by the same bar are not significantly different.

**Plan Estimate**



### The ANOVA Procedure

#### Sidak t Tests for Time

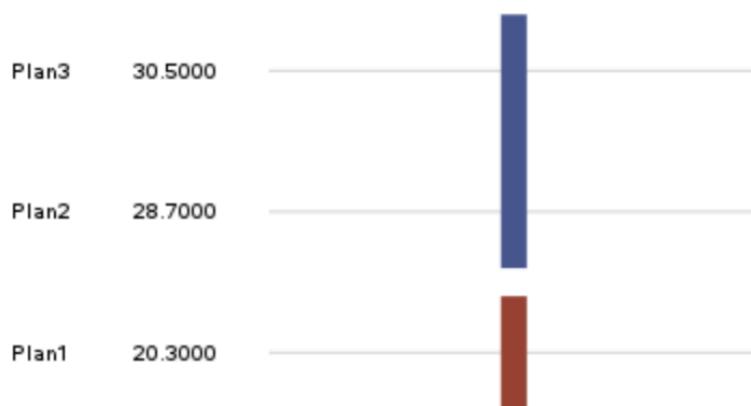
**Note:** This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

<b>Alpha</b>	0.05
<b>Error Degrees of Freedom</b>	27
<b>Error Mean Square</b>	12.24815
<b>Critical Value of t</b>	2.54506
<b>Minimum Significant Difference</b>	3.9834

#### Time Sidak Grouping for Means of Plan (Alpha = 0.05)

Means covered by the same bar are not significantly different.

##### Plan Estimate



**The ANOVA Procedure**

**Bonferroni (Dunn) t Tests for Time**

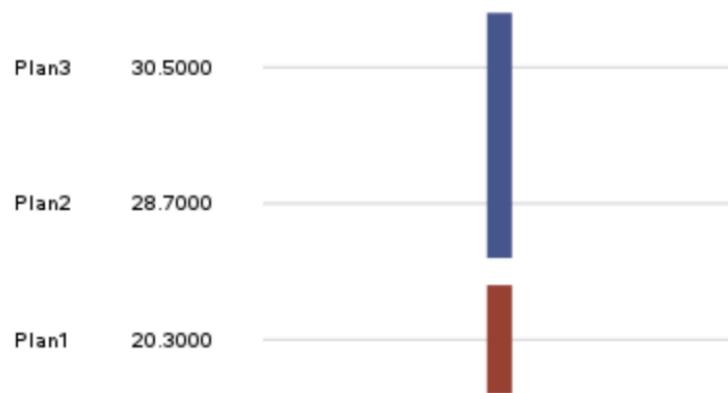
**Note:** This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

<b>Alpha</b>	0.05
<b>Error Degrees of Freedom</b>	27
<b>Error Mean Square</b>	12.24815
<b>Critical Value of t</b>	2.55246
<b>Minimum Significant Difference</b>	3.9949

**Time Bonferroni Grouping for Means of Plan  
(Alpha = 0.05)**

Means covered by the same bar are not significantly different.

**Plan Estimate**



## The ANOVA Procedure

### Scheffe's Test for Time

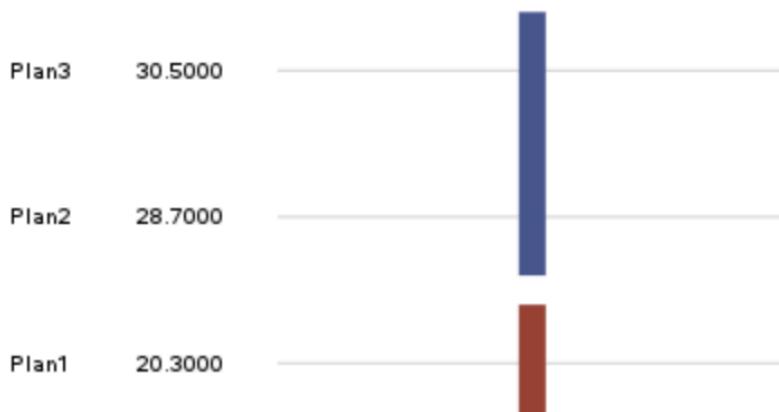
**Note:** This test controls the Type I experimentwise error rate.

<b>Alpha</b>	0.05
<b>Error Degrees of Freedom</b>	27
<b>Error Mean Square</b>	12.24815
<b>Critical Value of F</b>	3.35413
<b>Minimum Significant Difference</b>	4.0537

### Time Scheffe Grouping for Means of Plan (Alpha = 0.05)

Means covered by the same bar are not significantly different.

#### Plan Estimate



C

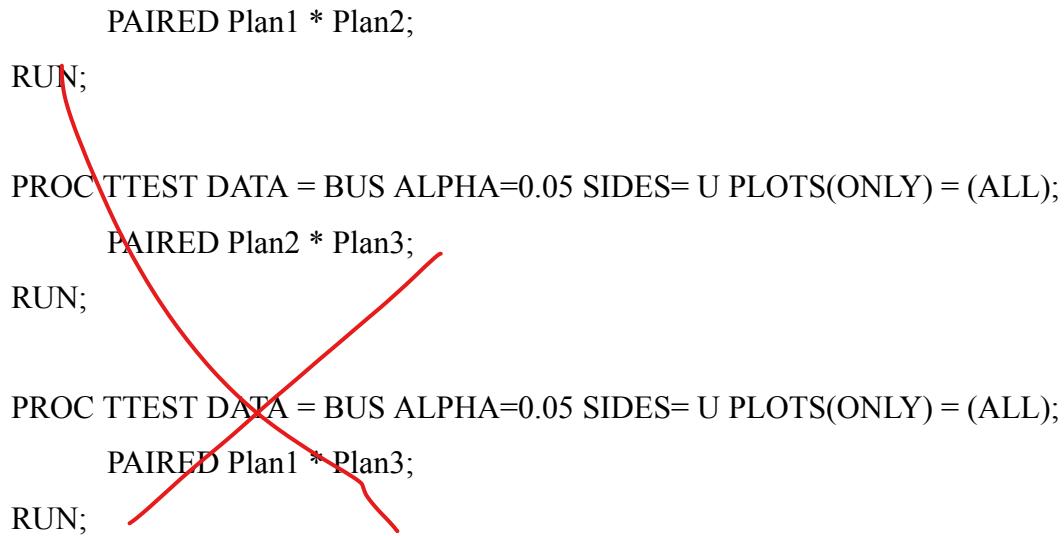
CODE

Pairwise comparison doesn't mean to be paired test. The paired test is only applied to two measurements observed from each observation

/\* B.c \*/

PROC TTEST DATA = BUS ALPHA=0.05 SIDES= U PLOTS(ONLY) = (ALL);

```
PAIRED Plan1 * Plan2;  
RUN;  
  
PROC TTEST DATA = BUS ALPHA=0.05 SIDES= U PLOTS(ONLY) = (ALL);  
PAIRED Plan2 * Plan3;  
RUN;  
  
PROC TTEST DATA = BUS ALPHA=0.05 SIDES= U PLOTS(ONLY) = (ALL);  
PAIRED Plan1 * Plan3;  
RUN;
```



## LOG

```
1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
68  
69      /* B.c */  
70      PROC TTEST DATA = BUS ALPHA=0.05 SIDES= U PLOTS(ONLY)  
= (ALL);  
71      PAIRED Plan1 * Plan2;  
72      RUN;  
  
NOTE: PROCEDURE TTEST used (Total process time):  
      real time          0.45 seconds  
      user cpu time      0.22 seconds  
      system cpu time    0.05 seconds  
      memory            21245.68k  
      OS Memory          46432.00k  
      Timestamp          05/03/2024 03:49:20 AM  
      Step Count          329   Switch Count  30  
      Page Faults         0  
      Page Reclaims       20529  
      Page Swaps          0
```

Voluntary Context Switches	1461
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	2016

73

```
74      PROC TTEST DATA = BUS ALPHA=0.05 SIDES= U PLOTS(ONLY)
= (ALL);
75      PAIRED Plan2 * Plan3;
76      RUN;
```

NOTE: PROCEDURE TTEST used (Total process time):

real time	0.42 seconds		
user cpu time	0.21 seconds		
system cpu time	0.04 seconds		
memory	13037.21k		
OS Memory	45664.00k		
Timestamp	05/03/2024 03:49:21 AM		
Step Count	330	Switch Count	30
Page Faults	0		
Page Reclaims	18188		
Page Swaps	0		
Voluntary Context Switches	1482		
Involuntary Context Switches	2		
Block Input Operations	0		
Block Output Operations	1832		

77

```
78      PROC TTEST DATA = BUS ALPHA=0.05 SIDES= U PLOTS(ONLY)
= (ALL);
79      PAIRED Plan1 * Plan3;
80      RUN;
```

NOTE: PROCEDURE TTEST used (Total process time):

real time	0.42 seconds		
user cpu time	0.21 seconds		
system cpu time	0.04 seconds		
memory	12949.50k		
OS Memory	46176.00k		
Timestamp	05/03/2024 03:49:21 AM		
Step Count	331	Switch Count	30
Page Faults	0		
Page Reclaims	18209		
Page Swaps	0		
Voluntary Context Switches	1456		
Involuntary Context Switches	0		
Block Input Operations	0		
Block Output Operations	1952		

81

82           OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

92

## RESULTS

The TTEST Procedure

Difference: Plan1 - Plan2

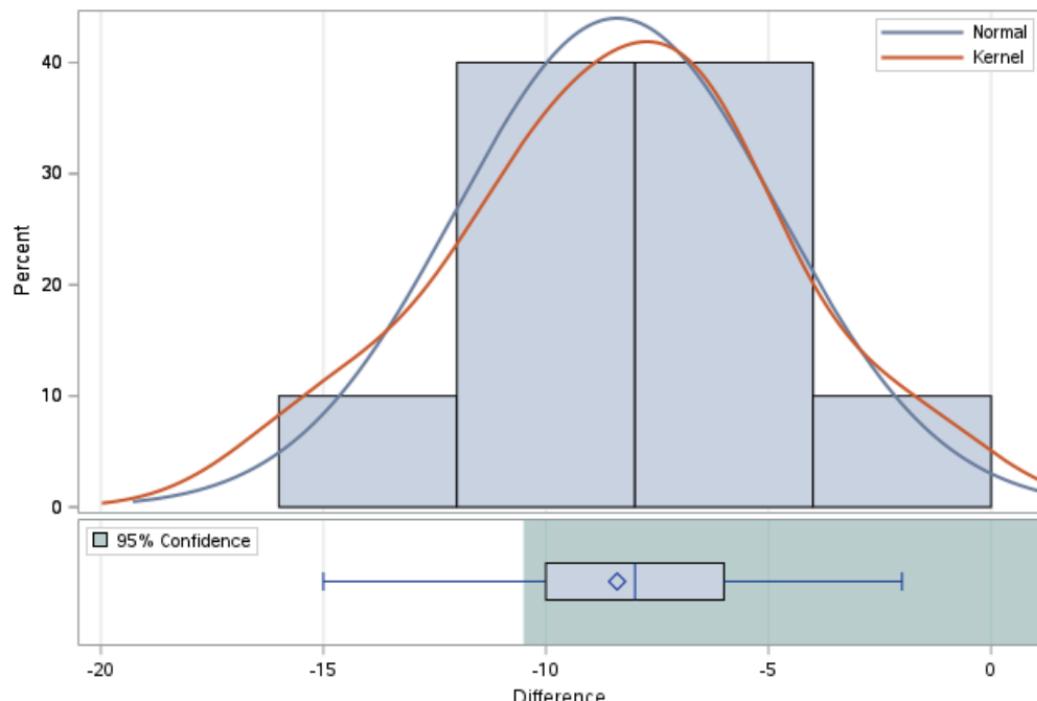
N	Mean	Std Dev	Std Err	Minimum	Maximum
10	-8.4000	3.6271	1.1470	-15.0000	-2.0000

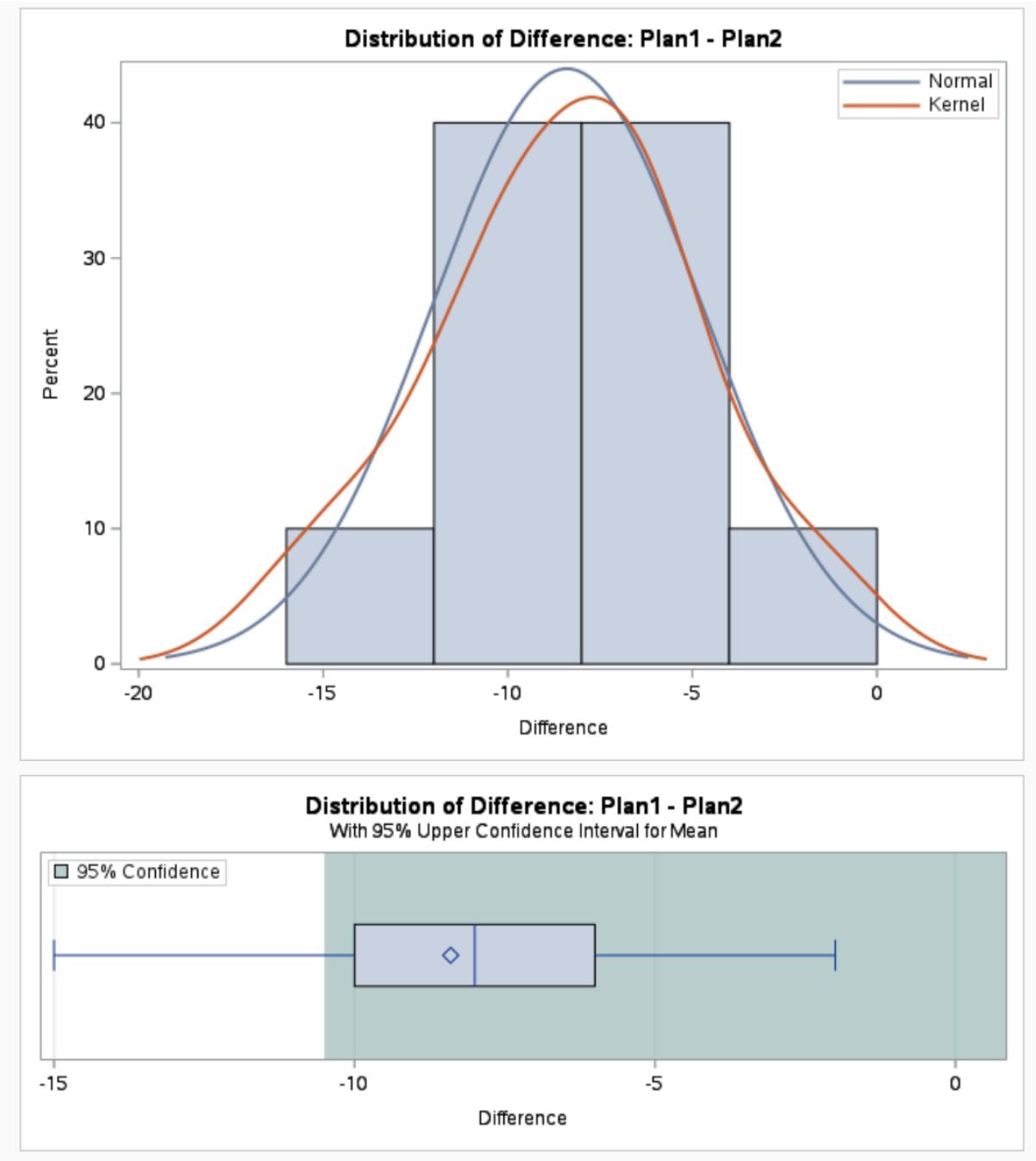
Mean	95% CL Mean	Std Dev	95% CL Std Dev
-8.4000	-10.5025	Infy	3.6271 2.4948 6.6216

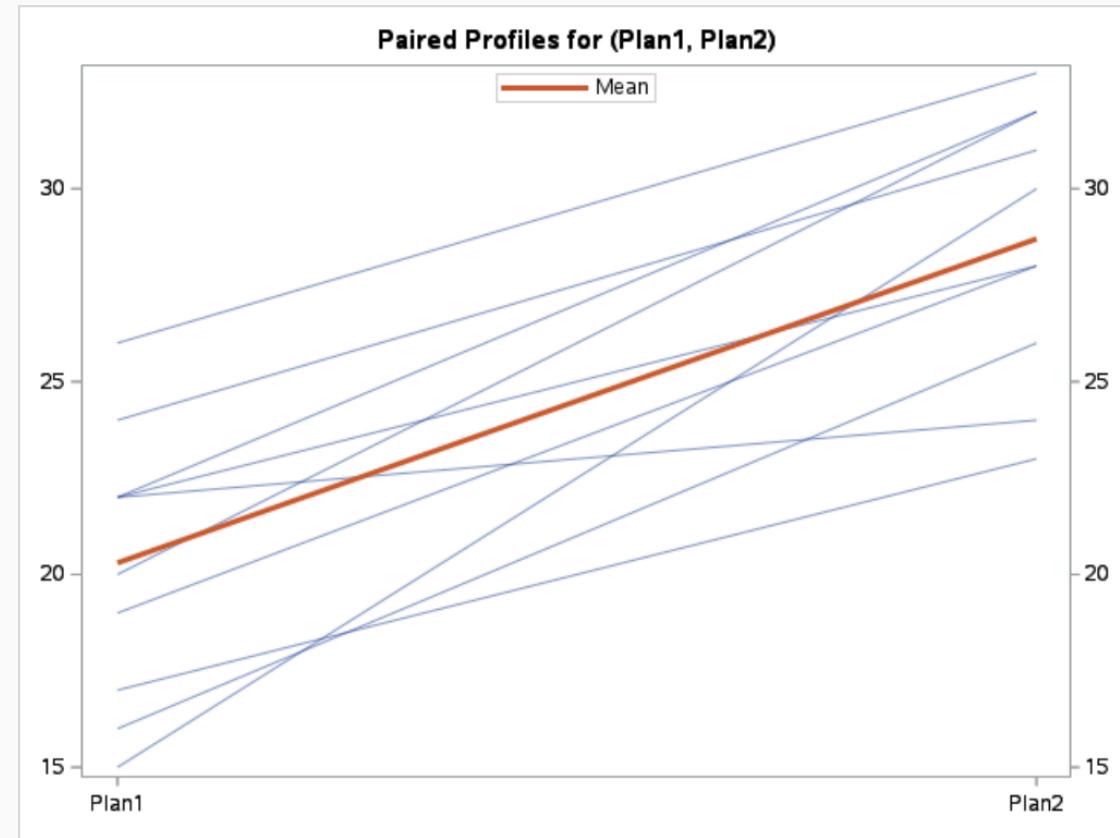
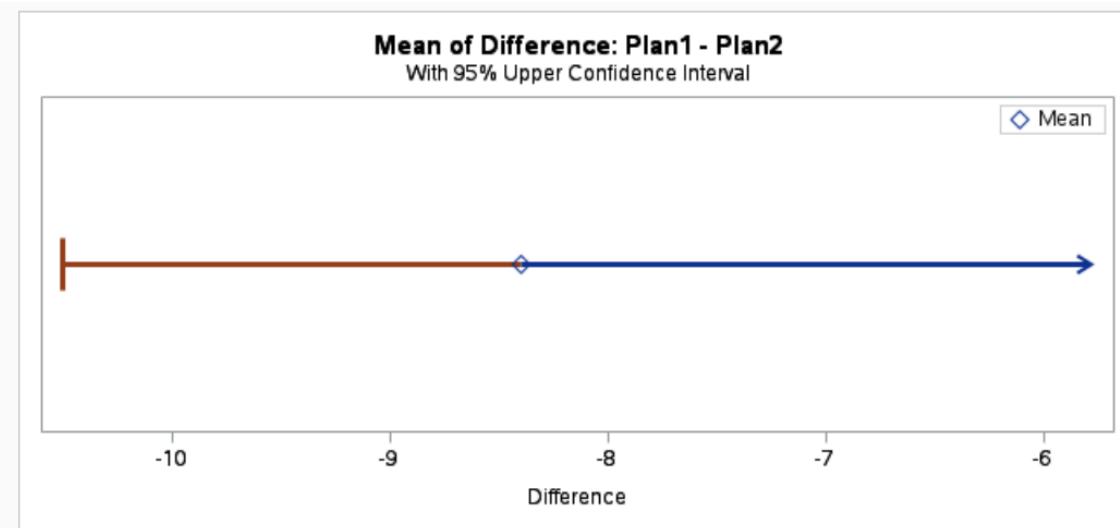
DF	t Value	Pr > t
9	-7.32	1.0000

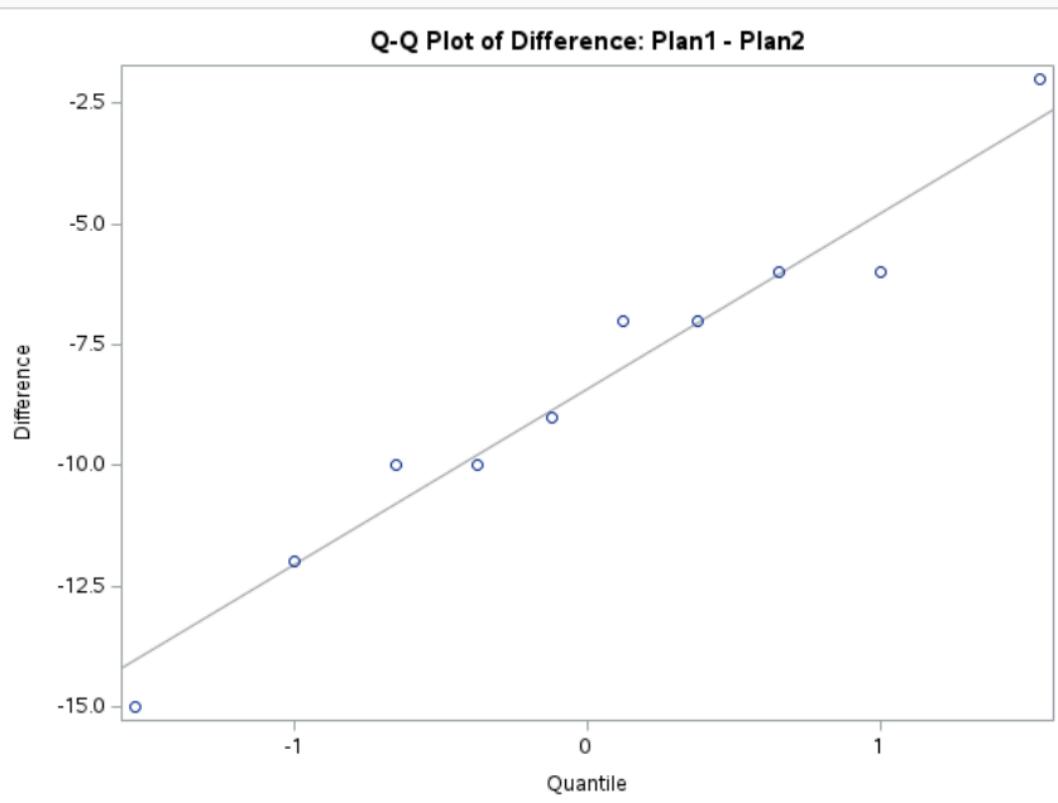
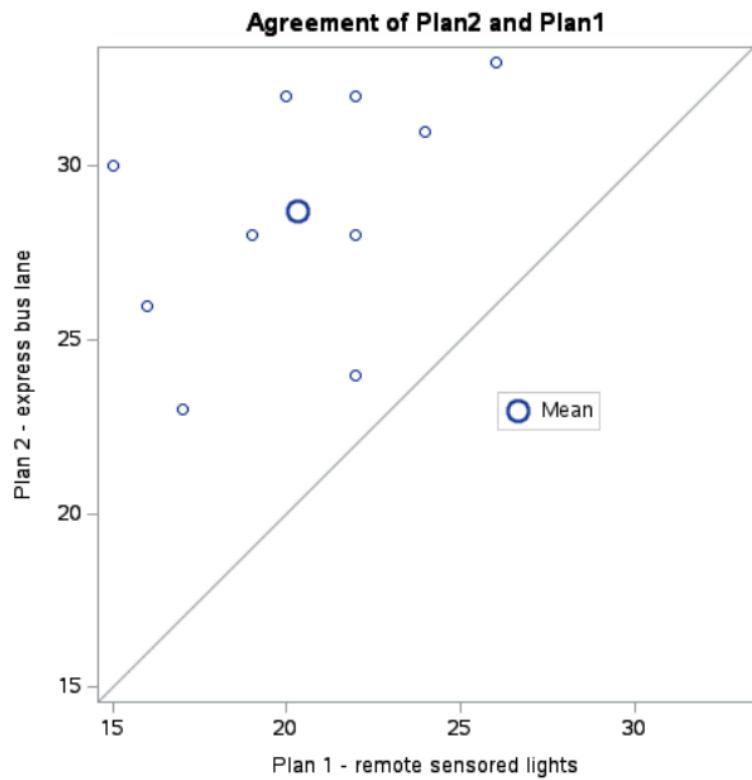
**Distribution of Difference: Plan1 - Plan2**

With 95% Upper Confidence Interval for Mean









The TTEST Procedure

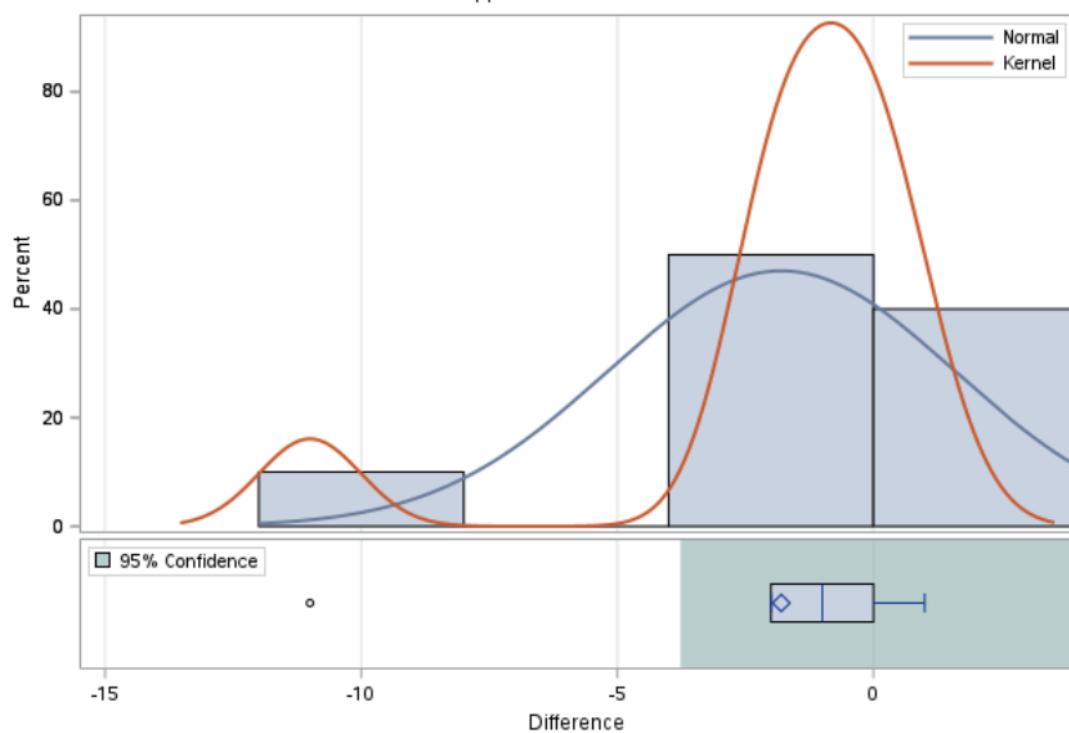
Difference: Plan2 - Plan3

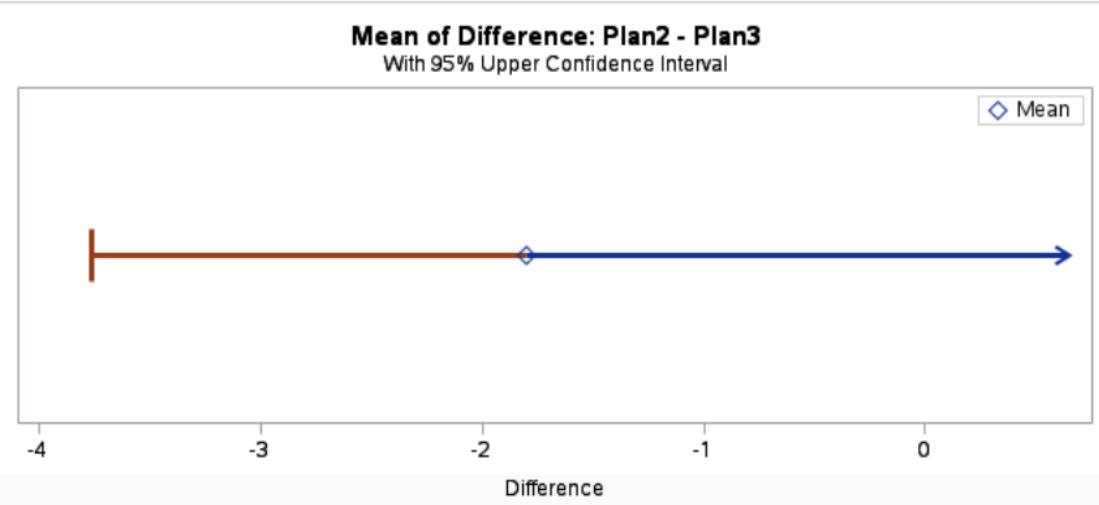
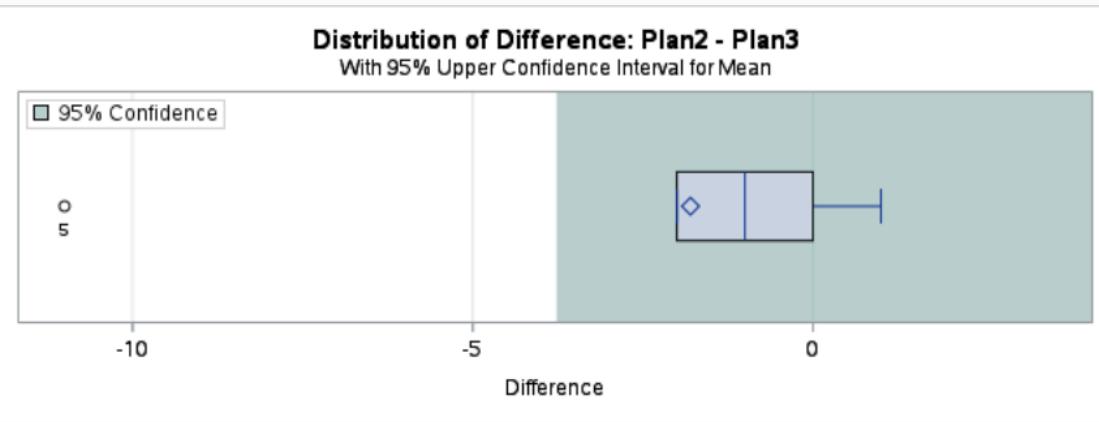
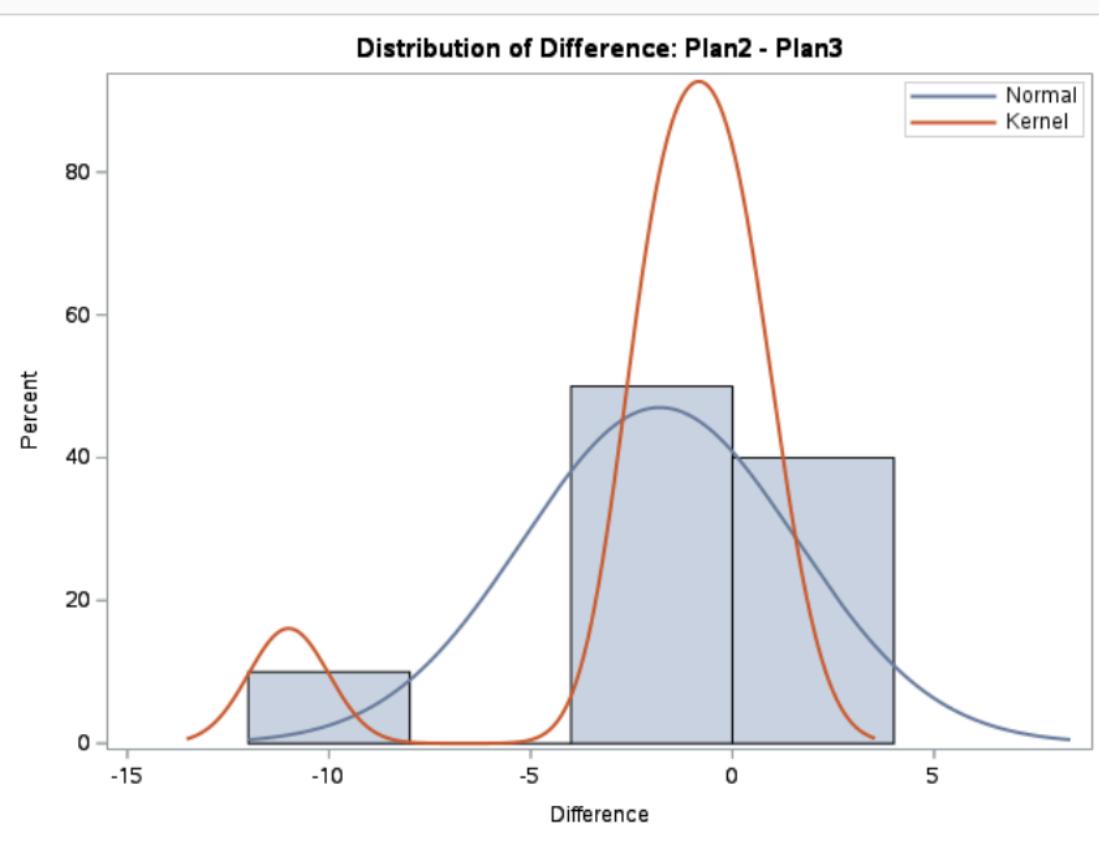
N	Mean	Std Dev	Std Err	Minimum	Maximum
10	-1.8000	3.3928	1.0729	-11.0000	1.0000

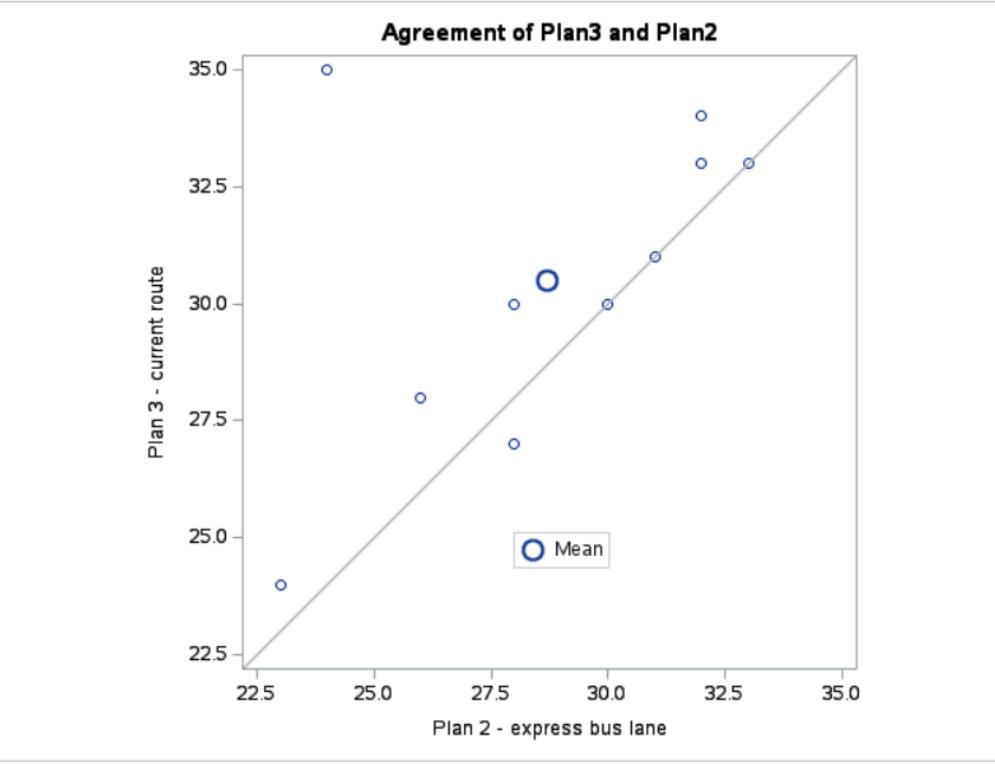
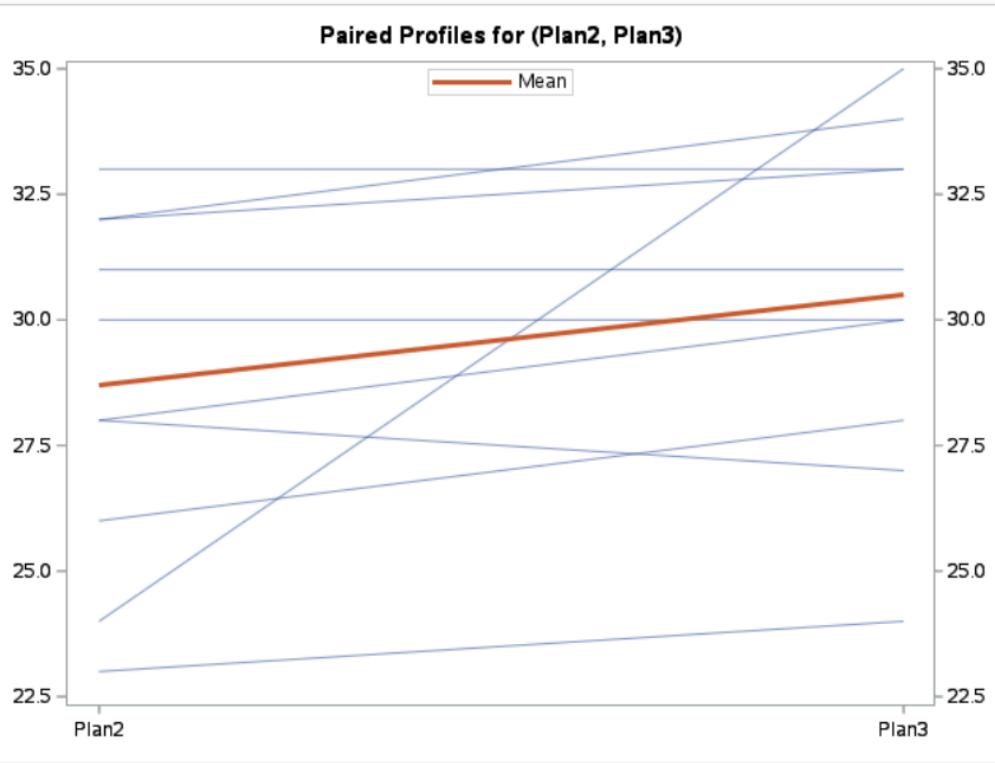
Mean	95% CL Mean	Std Dev	95% CL Std Dev
-1.8000	-3.7667	Infty	3.3928

DF	t Value	Pr > t
9	-1.68	0.9361

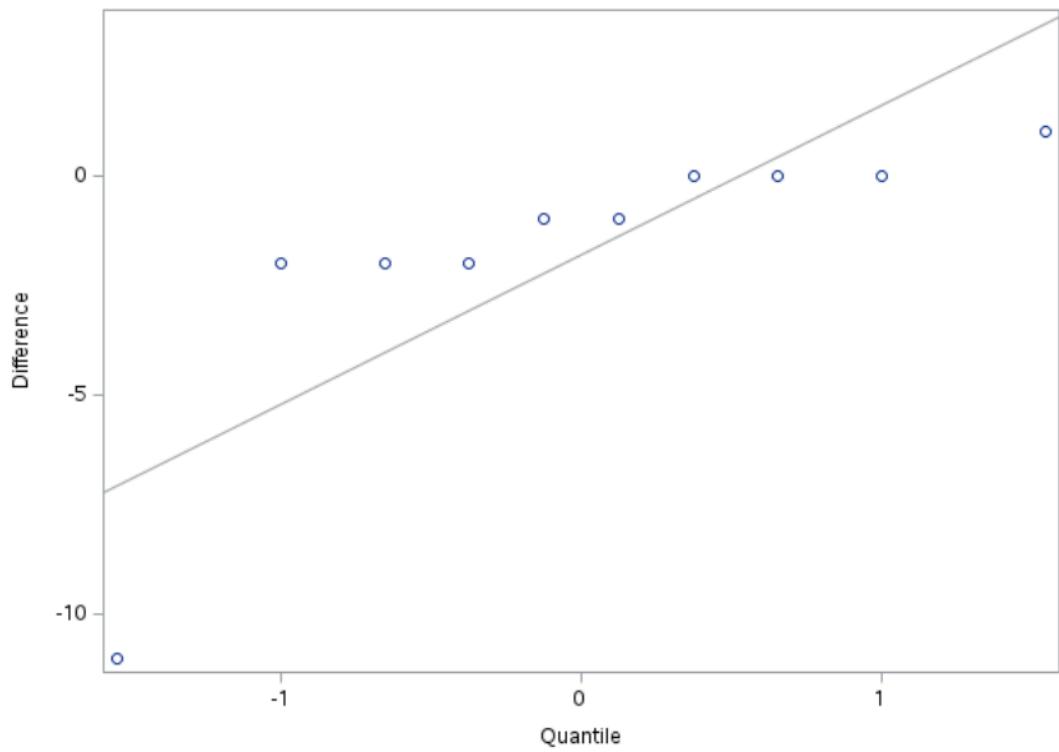
**Distribution of Difference: Plan2 - Plan3**  
With 95% Upper Confidence Interval for Mean







**Q-Q Plot of Difference: Plan2 - Plan3**



The TTEST Procedure

Difference: Plan1 - Plan3

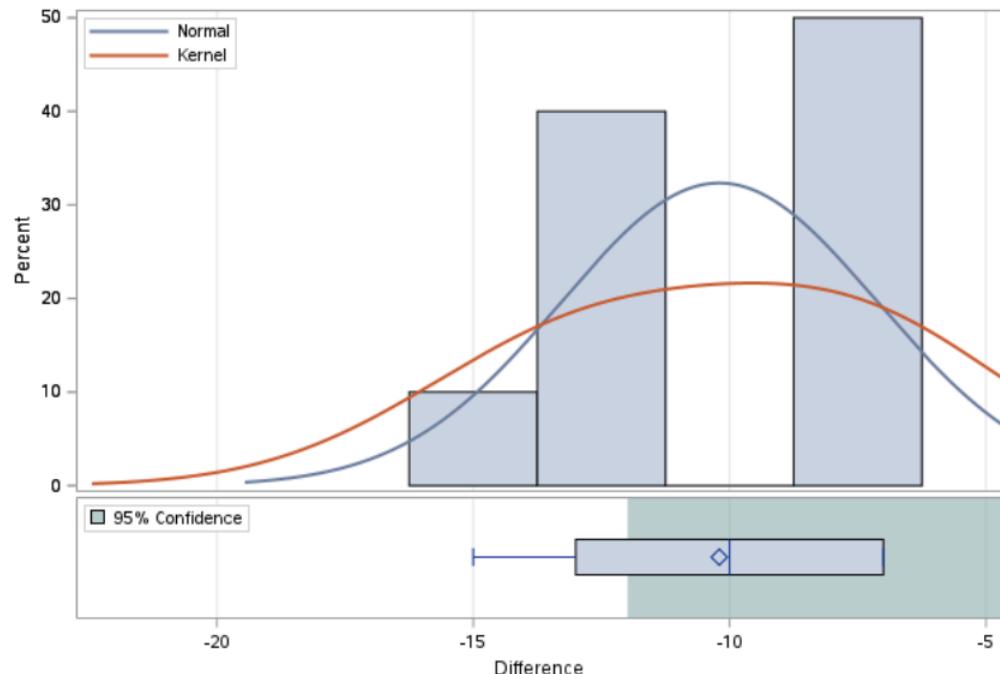
N	Mean	Std Dev	Std Err	Minimum	Maximum
10	-10.2000	3.0840	0.9752	-15.0000	-7.0000

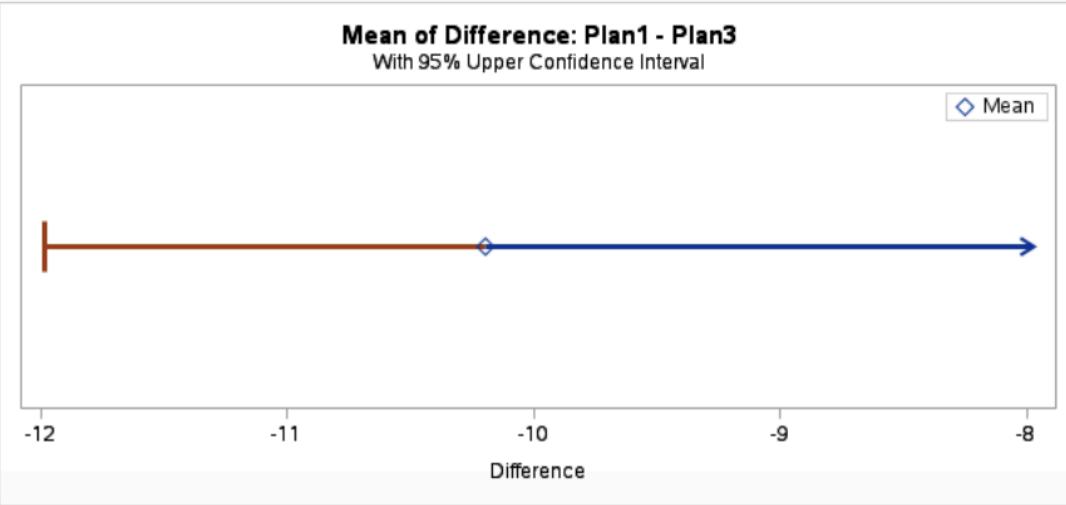
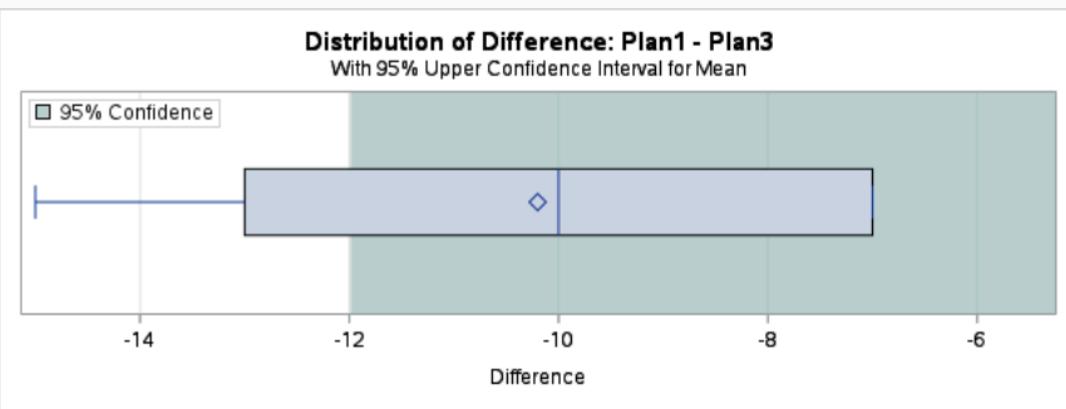
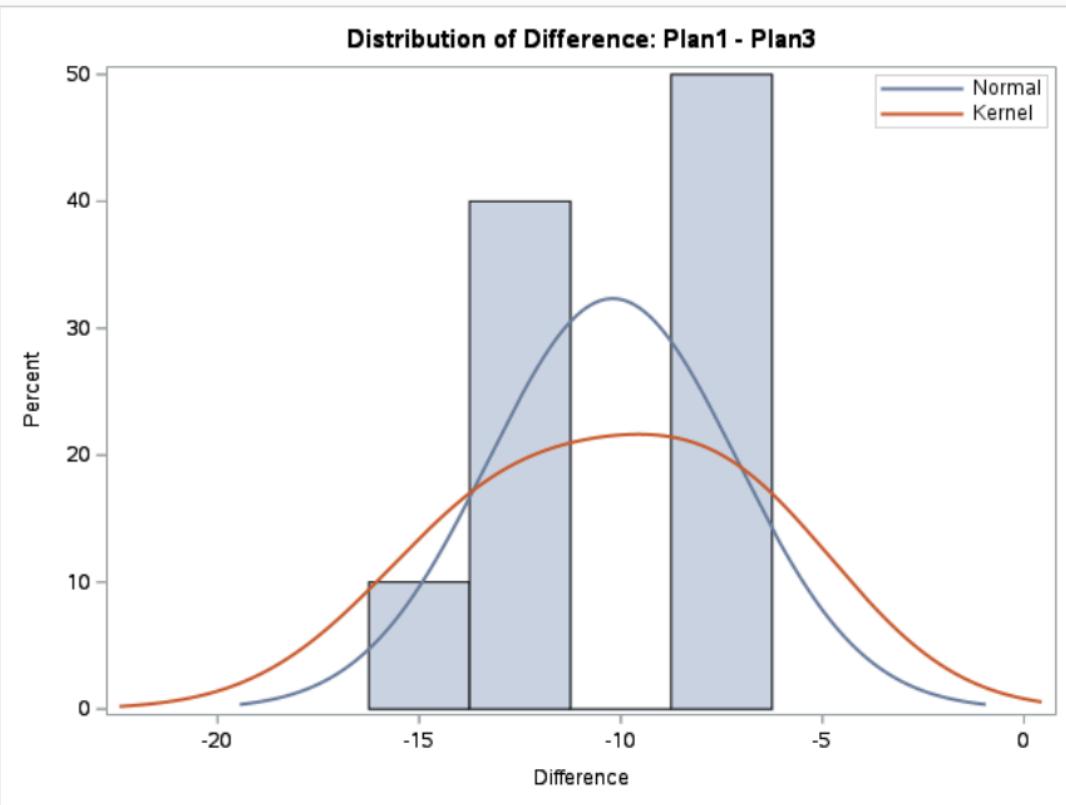
Mean	95% CL Mean	Std Dev	95% CL Std Dev
-10.2000	-11.9877	Infty	3.0840

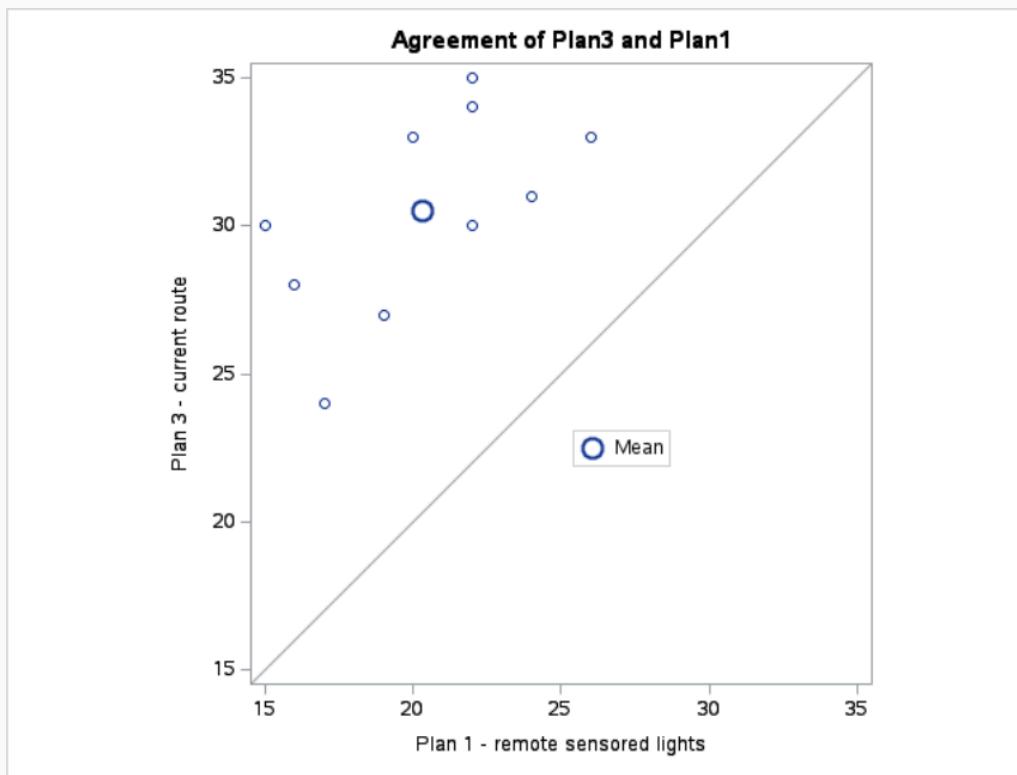
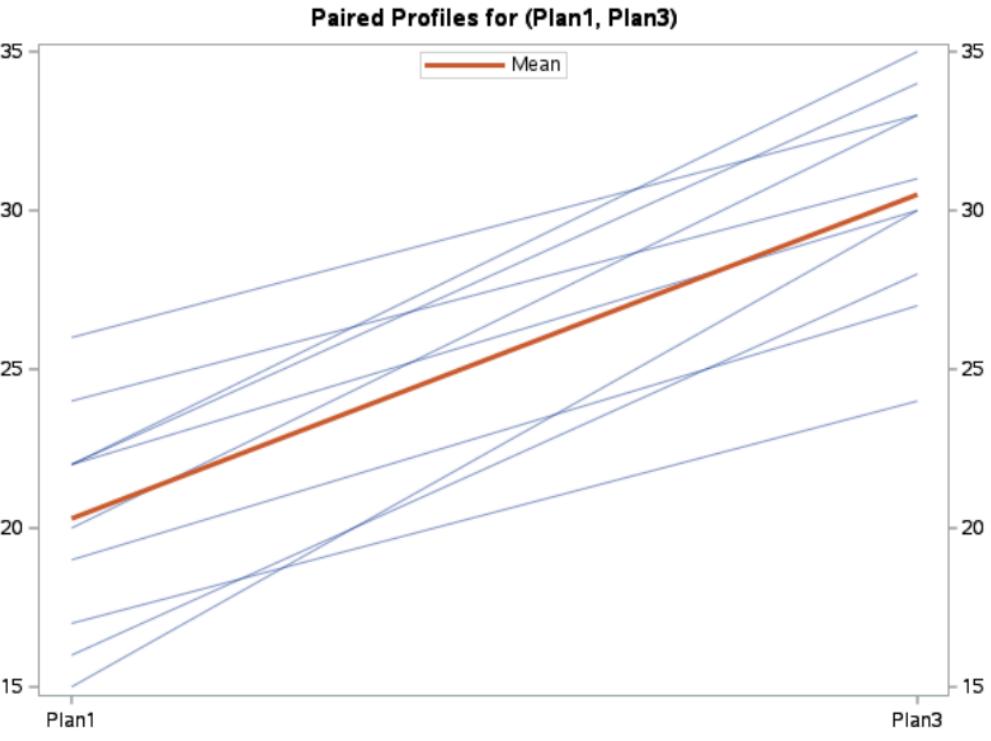
DF	t Value	Pr > t
9	-10.46	1.0000

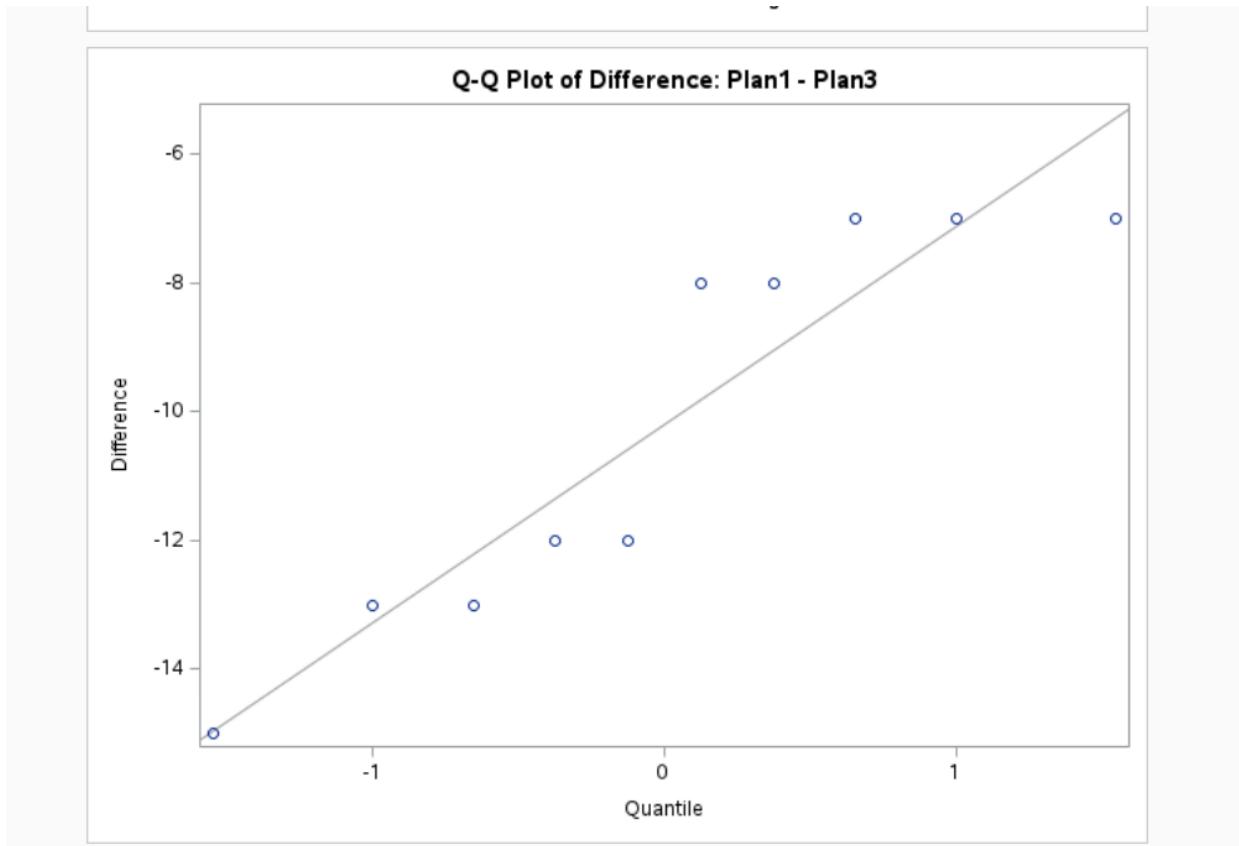
**Distribution of Difference: Plan1 - Plan3**

With 95% Upper Confidence Interval for Mean









D

/\* B.d \*/

/\* The pairwise t-tests between each plan show p-values that are close to 1.  
This means that we fail to reject the null hypothesis that there is no difference  
between the two means.

→ 5

In particular, the boxplots show that plan 2 and plan 3 overlap quite a lot.

However, if I had to choose a plan, I would choose plan 1.

Based on the boxplot, it shows that the time is in general faster than other plans. \*/

The overall F test suggests that there is a true mean difference among three plans ( $p < 0.0001$ ).

Based on the multiple comparisons we would recommend plan one as it has the lowest mean (20.3 minutes) route time and is significantly different from the other two plans;

## Extra Points

- © Researchers at a local medical center have just completed enrollment for a clinical trial of a new cholesterol-lowering medication for use in subjects with borderline high total cholesterol. They keep their enrollment data in two SAS data sets so as to not bias the clinicians. The VISITS data set contains basic information about each subject at their baseline visit (Visit = 0). The TXGROUP data set contains information about whether the subject received the treatment or a placebo.
- a. Examine these SAS data sets including the variable labels and attributes. Combine these two data sources to create a data set that identifies the treatment group for each subject.
  - b. The data entry clerk informs you that there may be duplicate records in the TXGROUP data set. Write code that will identify any duplicate entries and modify the programming for part a), if necessary.
  - c. Using the combined data set from part b), calculate the median baseline cholesterol measurement across all subjects, and use this information to create a variable that groups subjects as less than or equal to the median, or more than the median. Do this without typing the calculated median value by hand into your code.
  - d. Using the data set from part c), create another data set that schedules subjects for their next three visits, one every 30 days starting at the baseline visit date, for a total of four visits per subject. Each of the visits should appear as a new observation with the original subject information, the median grouping, the corresponding visit date, and a visit number (0, 1, 2, or 3).
  - e. Add a comment to your program that states the number of observations and variables in the final data sets created in parts c) and d).

A

CODE

```
/* C.a */
```

```
DATA VISITS;  
set '/home/u62223361/Intro to SAS/FinalExam/visits.sas7bdat';  
FORMAT VisitDt DATE.;  
RUN;
```

```
DATA TXGROUP;  
set '/home/u62223361/Intro to SAS/FinalExam/txgroup.sas7bdat';  
RUN;
```

```
PROC CONTENTS DATA = VISITS;  
RUN;  
PROC CONTENTS DATA = TXGROUP;  
RUN;
```

```
PROC PRINT DATA = VISITS;  
RUN;  
PROC PRINT DATA = TXGROUP;  
RUN;
```

Those codes are not necessary

```
PROC SORT DATA = TXGROUP;  
BY ID;  
RUN;
```

```
PROC SORT DATA = VISITS;  
BY ID;  
RUN;
```

```
DATA VISITS_TXGROUP;  
MERGE VISITS (IN=V) TXGROUP;  
BY ID;  
IF V;
```

```
RUN;  
  
PROC CONTENTS DATA = VISITS_TXGROUP;  
RUN;
```



## LOG

```
1           OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
68  
69           /* C.a */  
70           DATA VISITS;  
71           set '/home/u62223361/Intro to  
SAS/FinalExam/visits.sas7bdat';  
NOTE: Data file WC000001.VISITS.DATA is in a format that is  
native to another host, or the file encoding does not match the  
session  
encoding. Cross Environment Data Access will be used,  
which might require additional CPU resources and might reduce  
performance.  
72           FORMAT VisitDt DATE.;  
73           RUN;  
NOTE: There were 2363 observations read from the data set  
/home/u62223361/Intro to SAS/FinalExam/visits.sas7bdat.  
NOTE: The data set WORK.VISITS has 2363 observations and 5  
variables.  
NOTE: DATA statement used (Total process time):  
      real time          0.00 seconds  
      user cpu time      0.00 seconds  
      system cpu time    0.00 seconds  
      memory             1265.31k  
      OS Memory          27960.00k
```

Timestamp	05/03/2024 03:51:23 AM		
Step Count	337	Switch Count	3
Page Faults	0		
Page Reclaims	182		
Page Swaps	0		
Voluntary Context Switches	23		
Involuntary Context Switches	0		
Block Input Operations	0		
Block Output Operations	264		

74

75       DATA TXGROUP;  
76           set '/home/u62223361/Intro to  
SAS/FinalExam/txgroup.sas7bdat';

NOTE: Data file WC000001.TXGROUP.DATA is in a format that is  
native to another host, or the file encoding does not match the  
session encoding. Cross Environment Data Access will be  
used, which might require additional CPU resources and might  
reduce

performance.

77       RUN;

NOTE: There were 2455 observations read from the data set  
/home/u62223361/Intro to SAS/FinalExam/txgroup.sas7bdat.

NOTE: The data set WORK.TXGROUP has 2455 observations and 2  
variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1066.59k
OS Memory	27960.00k
Timestamp	05/03/2024 03:51:23 AM

Step Count	338	Switch Count	3
Page Faults	0		
Page Reclaims	116		
Page Swaps	0		
Voluntary Context Switches	23		
Involuntary Context Switches	0		
Block Input Operations	0		
Block Output Operations	264		

78

79       PROC CONTENTS DATA = VISITS;  
80       RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.02 seconds		
user cpu time	0.02 seconds		
system cpu time	0.00 seconds		
memory	1842.00k		
OS Memory	28220.00k		
Timestamp	05/03/2024 03:51:23 AM		
Step Count	339	Switch Count	0
Page Faults	0		
Page Reclaims	95		
Page Swaps	0		
Voluntary Context Switches	2		
Involuntary Context Switches	0		
Block Input Operations	0		
Block Output Operations	24		

81       PROC CONTENTS DATA = TXGROUP;  
82       RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time           0.02 seconds

user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	919.53k
OS Memory	28220.00k
Timestamp	05/03/2024 03:51:23 AM
Step Count	340 Switch Count 0
Page Faults	0
Page Reclaims	95
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	24

83

84 PROC PRINT DATA = VISITS;  
85 RUN;

NOTE: There were 2363 observations read from the data set  
WORK.VISITS.

NOTE: PROCEDURE PRINT used (Total process time):

real time	1.51 seconds
user cpu time	1.52 seconds
system cpu time	0.00 seconds
memory	649.46k
OS Memory	27960.00k
Timestamp	05/03/2024 03:51:24 AM
Step Count	341 Switch Count 0
Page Faults	0
Page Reclaims	63
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	1

Block Input Operations	0
Block Output Operations	928

86 PROC PRINT DATA = TXGROUP;  
87 RUN;

NOTE: There were 2455 observations read from the data set  
WORK.TXGROUP.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.85 seconds
user cpu time	0.86 seconds
system cpu time	0.01 seconds
memory	661.81k
OS Memory	27960.00k
Timestamp	05/03/2024 03:51:25 AM
Step Count	342 Switch Count 0
Page Faults	0
Page Reclaims	63
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	552

88  
89 PROC SORT DATA = TXGROUP;  
90 BY ID;  
91 RUN;

NOTE: There were 2455 observations read from the data set  
WORK.TXGROUP.

NOTE: The data set WORK.TXGROUP has 2455 observations and 2  
variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1055.62k
OS Memory	28220.00k
Timestamp	05/03/2024 03:51:25 AM
Step Count	343 Switch Count 2
Page Faults	0
Page Reclaims	107
Page Swaps	0
Voluntary Context Switches	16
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

92

93 PROC SORT DATA = VISITS;  
94 BY ID;  
95 RUN;

NOTE: There were 2363 observations read from the data set  
WORK.VISITS.

NOTE: The data set WORK.VISITS has 2363 observations and 5  
variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	1600.43k
OS Memory	28740.00k
Timestamp	05/03/2024 03:51:25 AM
Step Count	344 Switch Count 2
Page Faults	0

Page Reclaims	98
Page Swaps	0
Voluntary Context Switches	13
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

96

```

97      DATA VISITS_TXGROUP;
98      MERGE VISITS (IN=V) TXGROUP;
99      BY ID;
100     IF V;
101     RUN;
```

NOTE: There were 2363 observations read from the data set WORK.VISITS.

NOTE: There were 2455 observations read from the data set WORK.TXGROUP.

NOTE: The data set WORK.VISITS\_TXGROUP has 2455 observations and 6 variables.

NOTE: DATA statement used (Total process time):

real time	0.00	seconds
user cpu time	0.00	seconds
system cpu time	0.00	seconds
memory	1390.93k	
OS Memory	28480.00k	
Timestamp	05/03/2024 03:51:25 AM	
Step Count	345	Switch Count 2
Page Faults	0	
Page Reclaims	149	
Page Swaps	0	
Voluntary Context Switches	16	
Involuntary Context Switches	0	

Block Input Operations	0
Block Output Operations	264

102

103        PROC CONTENTS DATA = VISITS\_TXGROUP;

104        RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.02 seconds
user cpu time	0.03 seconds
system cpu time	0.00 seconds
memory	928.50k
OS Memory	28220.00k
Timestamp	05/03/2024 03:51:25 AM
Step Count	346    Switch Count 0
Page Faults	0
Page Reclaims	96
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	24

105

106        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

116

## RESULTS

The CONTENTS Procedure

<b>Data Set Name</b>	WORK.VISITS	<b>Observations</b>	2363
<b>Member Type</b>	DATA	<b>Variables</b>	5
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	05/02/2024 23:51:23	<b>Observation Length</b>	40
<b>Last Modified</b>	05/02/2024 23:51:23	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	3265
<b>Obs in First Data Page</b>	2363
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/saswork/SAS_work5358000083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/visits.sas7bdat
<b>Release Created</b>	9.0401M7
<b>Host Created</b>	Linux
<b>Inode Number</b>	1744845489
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	u62223361
<b>File Size</b>	256KB
<b>File Size (bytes)</b>	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Label
5	B_Cholesterol	Num	8		Baseline Cholesterol measurement
3	Gender	Char	6		Subject gender
1	ID	Num	8		Subject ID
4	Visit	Num	8		Visit number
2	VisitDt	Num	8	DATE.	Baseline visit date

The CONTENTS Procedure

<b>Data Set Name</b>	WORK.TXGROUP	<b>Observations</b>	2455
<b>Member Type</b>	DATA	<b>Variables</b>	2
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	05/02/2024 23:51:23	<b>Observation Length</b>	16
<b>Last Modified</b>	05/02/2024 23:51:23	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

<b>Data Set Page Size</b>	131072
<b>Number of Data Set Pages</b>	1
<b>First Data Page</b>	1
<b>Max Obs per Page</b>	8126
<b>Obs in First Data Page</b>	2455
<b>Number of Data Set Repairs</b>	0
<b>Filename</b>	/saswork/SAS_work535800083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/txgroup.sas7bdat
<b>Release Created</b>	9.0401M7
<b>Host Created</b>	Linux
<b>Inode Number</b>	1744845473
<b>Access Permission</b>	rw-r--r--
<b>Owner Name</b>	u62223361
<b>File Size</b>	256KB
<b>File Size (bytes)</b>	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
1	ID	Num	8	Subject ID
2	TX	Num	8	Treatment group (0=placebo, 1=treatment)

<b>Obs</b>	<b>ID</b>	<b>VisitDt</b>	<b>Gender</b>	<b>Visit</b>	<b>B_Cholesterol</b>
<b>1</b>	178772	30DEC17	Male	0	228
<b>2</b>	300732	30DEC17	Female	0	236
<b>3</b>	372409	30DEC17	Male	0	239
<b>4</b>	418589	30DEC17	Female	0	232
<b>5</b>	472192	30DEC17	Female	0	223
<b>6</b>	652213	30DEC17	Male	0	223
<b>7</b>	742277	30DEC17	Male	0	222
<b>8</b>	788007	30DEC17	Male	0	216
<b>9</b>	795908	30DEC17	Female	0	221
<b>10</b>	801436	30DEC17	Male	0	211
<b>11</b>	827486	30DEC17	Male	0	222
<b>12</b>	920988	30DEC17	Female	0	226
<b>13</b>	997739	30DEC17	Male	0	239
<b>14</b>	110839	31DEC17	Male	0	235
<b>15</b>	486674	31DEC17	Female	0	224
<b>16</b>	627763	31DEC17	Female	0	232
<b>17</b>	693700	31DEC17	Male	0	240
<b>18</b>	747920	31DEC17	Male	0	200
<b>19</b>	775998	31DEC17	Female	0	231
<b>20</b>	360291	01JAN18	Male	0	226
<b>21</b>	405071	01JAN18	Female	0	207
<b>22</b>	436144	01JAN18	Male	0	212
<b>23</b>	440775	01JAN18	Female	0	206
<b>24</b>	708278	01JAN18	Female	0	201
<b>25</b>	712869	01JAN18	Male	0	234
<b>26</b>	975230	01JAN18	Female	0	228
<b>27</b>	102430	02JAN18	Male	0	229
<b>28</b>	131762	02JAN18	Male	0	240
<b>29</b>	147834	02JAN18	Male	0	219
<b>30</b>	440081	02JAN18	Male	0	223
<b>31</b>	594307	02JAN18	Female	0	236
<b>32</b>	788782	02JAN18	Male	0	226
<b>33</b>	794427	02JAN18	Male	0	205
<b>34</b>	871928	02JAN18	Male	0	231
<b>35</b>	107395	05JAN18	Female	0	240
<b>36</b>	189171	05JAN18	Male	0	232
<b>37</b>	273495	05JAN18	Female	0	212
<b>38</b>	699910	05JAN18	Male	0	237

1363	742210	28DEC18	Male	0	234
------	--------	---------	------	---	-----

Obs	ID	TX
1	43100	0
2	100153	1
3	100405	0
4	100597	0
5	100732	1
6	101927	1
7	102430	1
8	102669	0
9	102700	0
10	103181	1
11	103631	1
12	104050	1
13	104300	0
14	104344	1
15	104409	1
16	104649	0
17	105080	1
18	105164	0
19	105587	1
20	105894	1
21	106303	0
22	106559	1
23	107216	1
24	107395	0
25	107579	0
26	107986	0

2453	999453	0	
2454	999679	0	
2455	999880	1	

---

The CONTENTS Procedure			
<b>Data Set Name</b>	WORK.VISITS_TXGROUP	<b>Observations</b>	2455
<b>Member Type</b>	DATA	<b>Variables</b>	6
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	05/02/2024 23:51:26	<b>Observation Length</b>	48
<b>Last Modified</b>	05/02/2024 23:51:26	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information			
<b>Data Set Page Size</b>	131072		
<b>Number of Data Set Pages</b>	1		
<b>First Data Page</b>	1		
<b>Max Obs per Page</b>	2722		
<b>Obs in First Data Page</b>	2455		
<b>Number of Data Set Repairs</b>	0		
<b>Filename</b>	/saswork/SAS_work5358000083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/visits_txgroup.sas7bdat		
<b>Release Created</b>	9.0401M7		
<b>Host Created</b>	Linux		
<b>Inode Number</b>	1744845489		
<b>Access Permission</b>	rw-r--r--		
<b>Owner Name</b>	u62223361		
<b>File Size</b>	256KB		
<b>File Size (bytes)</b>	262144		

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format
5	B_Cholesterol	Num	8	Baseline Cholesterol measurement
3	Gender	Char	6	Subject gender
1	ID	Num	8	Subject ID
6	TX	Num	8	Treatment group (0=placebo, 1=treatment)
4	Visit	Num	8	Visit number
2	VisitDt	Num	8	DATE. Baseline visit date

B

## CODE

```
/* C.b */

PROC SORT DATA = TXGROUP OUT = TXGROUP_NEAT NODUPKEY DUPOUT=
TXGROUP_EXTRA;
  BY ID TX;
RUN;
```

```
DATA VISITS_TXGROUP;  
  MERGE VISITS (IN=V) TXGROUP_NEAT;  
  BY ID;  
  IF V;  
RUN;
```

```
PROC CONTENTS DATA = VISITS_TXGROUP;  
RUN;
```

It would be better to list of all those duplications using the code below

```
PROC Print data=TXgroupextra;  
run;
```

## LOG

```
1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
68  
69      /* C.b */  
70      PROC SORT DATA = TXGROUP OUT = TXGROUP_NEAT NODUPKEY  
DUPOUT= TXGROUP_EXTRA;  
71      BY ID TX;  
72      RUN;  
  
NOTE: There were 2455 observations read from the data set  
WORK.TXGROUP.  
  
NOTE: 92 observations with duplicate key values were deleted.  
  
NOTE: The data set WORK.TXGROUP_NEAT has 2363 observations and 2  
variables.  
  
NOTE: The data set WORK.TXGROUP_EXTRA has 92 observations and 2  
variables.  
  
NOTE: PROCEDURE SORT used (Total process time):  
      real time          0.00 seconds  
      user cpu time      0.01 seconds  
      system cpu time    0.00 seconds  
      memory            1590.37k
```

Those are duplication records

OS Memory	28484.00k
Timestamp	05/03/2024 03:52:48 AM
Step Count	352 Switch Count 4
Page Faults	0
Page Reclaims	208
Page Swaps	0
Voluntary Context Switches	21
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	536

```

73
74      DATA VISITS_TXGROUP;
75      MERGE VISITS (IN=V) TXGROUP_NEAT;
76      BY ID;
77      IF V;
78      RUN;

```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS.

NOTE: There were 2363 observations read from the data set  
WORK.TXGROUP\_NEAT.

NOTE: The data set WORK.VISITS\_TXGROUP has 2363 observations and  
6 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1497.09k
OS Memory	28224.00k
Timestamp	05/03/2024 03:52:48 AM
Step Count	353 Switch Count 2
Page Faults	0

Page Reclaims	143
Page Swaps	0
Voluntary Context Switches	10
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

79

80       PROC CONTENTS DATA = VISITS\_TXGROUP;  
81       RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.02 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	1987.31k
OS Memory	27964.00k
Timestamp	05/03/2024 03:52:48 AM
Step Count	354   Switch Count  0
Page Faults	0
Page Reclaims	95
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	16

82

83       OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
93

User: u62223361

Messages: 34

## RESULTS

The CONTENTS Procedure			
<b>Data Set Name</b>	WORK.VISITS_TXGROUP	<b>Observations</b>	2363
<b>Member Type</b>	DATA	<b>Variables</b>	6
<b>Engine</b>	V9	<b>Indexes</b>	0
<b>Created</b>	05/02/2024 23:52:49	<b>Observation Length</b>	48
<b>Last Modified</b>	05/02/2024 23:52:49	<b>Deleted Observations</b>	0
<b>Protection</b>		<b>Compressed</b>	NO
<b>Data Set Type</b>		<b>Sorted</b>	NO
<b>Label</b>			
<b>Data Representation</b>	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
<b>Encoding</b>	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information			
<b>Data Set Page Size</b>	131072		
<b>Number of Data Set Pages</b>	1		
<b>First Data Page</b>	1		
<b>Max Obs per Page</b>	2722		
<b>Obs in First Data Page</b>	2363		
<b>Number of Data Set Repairs</b>	0		
<b>Filename</b>	/saswork/SAS_work5358000083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/visits_txgroup.sas7bdat		
<b>Release Created</b>	9.0401M7		
<b>Host Created</b>	Linux		
<b>Inode Number</b>	1744845472		
<b>Access Permission</b>	rw-r--r--		
<b>Owner Name</b>	u62223361		
<b>File Size</b>	256KB		
<b>File Size (bytes)</b>	262144		

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
5	B_Cholesterol	Num	8		Baseline Cholesterol measurement
3	Gender	Char	6		Subject gender
1	ID	Num	8		Subject ID
6	TX	Num	8		Treatment group (0=placebo, 1=treatment)
4	Visit	Num	8		Visit number
2	VisitDt	Num	8	DATE.	Baseline visit date

## C

## CODE

```
/* C.c */
```

```
PROC MEANS DATA= VISITS_TXGROUP MEDIAN;
VAR B_Cholesterol;
```

```
OUTPUT OUT = VISITS_TXGROUP_MEDIAN MEDIAN(B_Cholesterol) =  
Median_B_Choles;  
RUN;
```

```
DATA VISITS_TXGROUP_MEDIAN;  
SET VISITS_TXGROUP_MEDIAN;  
TEMP = "TEMP";  
RUN;
```

```
DATA VISITS_TXGROUP;  
SET VISITS_TXGROUP;  
TEMP = "TEMP";  
RUN;
```

```
DATA VISITS_TXGROUP (DROP= TEMP _FREQ__TYPE_);  
MERGE VISITS_TXGROUP (IN=I) VISITS_TXGROUP_MEDIAN;  
BY TEMP;  
IF I;  
IF B_Cholesterol <= Median_B_Choles THEN B_Chol_Strata = "LESS THAN";  
ELSE B_Chol_Strata = "GREATER THAN";  
LABEL Median_B_Choles = "Median of Baseline Cholesterol measurement";  
LABEL B_Chol_Strata = "Greater than or less than equal to the median of  
B_Cholesterol";  
RUN;
```

PROC CONTENTS DATA=VISITS\_TXGROUP;  
RUN;

LOG

1

OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
DATA compare (DROP = MedChol);  
IF \_N\_ = 1 THEN SET med (KEEP = MedChol);  
SET identify;  
IF B\_Cholesterol <= MedChol THEN B\_Group  
= '<=Median';  
ELSE B\_Group = '>Median';  
RUN;

You can simplify you code as the following

PROC MEANS DATA = ??? NOPRINT;  
VAR B\_Cholesterol;  
OUTPUT OUT = med MEDIAN(B\_Cholesterol) =  
MedChol;  
RUN;

DATA compare (DROP = MedChol);  
IF \_N\_ = 1 THEN SET med (KEEP = MedChol);  
SET identify;  
IF B\_Cholesterol <= MedChol THEN B\_Group  
= '<=Median';  
ELSE B\_Group = '>Median';  
RUN;

```
68
69      /* C.c */
70      PROC MEANS DATA= VISITS_TXGROUP MEDIAN;
71      VAR B_Cholesterol;
72      OUTPUT OUT = VISITS_TXGROUP_MEDIAN
MEDIAN(B_Cholesterol) = Median_B_Choles;
73      RUN;

NOTE: There were 2363 observations read from the data set
WORK.VISITS_TXGROUP.

NOTE: The data set WORK.VISITS_TXGROUP_MEDIAN has 1 observations
and 3 variables.

NOTE: PROCEDURE MEANS used (Total process time):
      real time          0.01 seconds
      user cpu time      0.01 seconds
      system cpu time    0.01 seconds
      memory             8019.09k
      OS Memory          34144.00k
      Timestamp           05/03/2024 03:53:14 AM
      Step Count          360   Switch Count   3
      Page Faults         0
      Page Reclaims       1690
      Page Swaps          0
      Voluntary Context Switches  31
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 264

74
75      DATA VISITS_TXGROUP_MEDIAN;
76      SET VISITS_TXGROUP_MEDIAN;
77      TEMP = "TEMP";
78      RUN;
```

NOTE: There were 1 observations read from the data set  
WORK.VISITS\_TXGROUP\_MEDIAN.

NOTE: The data set WORK.VISITS\_TXGROUP\_MEDIAN has 1 observations  
and 4 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	942.00k
OS Memory	27964.00k
Timestamp	05/03/2024 03:53:14 AM
Step Count	361 Switch Count 2
Page Faults	0
Page Reclaims	127
Page Swaps	0
Voluntary Context Switches	15
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

79

```
80      DATA VISITS_TXGROUP;
81      SET VISITS_TXGROUP;
82      TEMP = "TEMP";
83      RUN;
```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_TXGROUP.

NOTE: The data set WORK.VISITS\_TXGROUP has 2363 observations and  
7 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds

system cpu time	0.00	seconds
memory	1165.90k	
OS Memory	27964.00k	
Timestamp	05/03/2024 03:53:14	AM
Step Count	362	Switch Count 2
Page Faults	0	
Page Reclaims	97	
Page Swaps	0	
Voluntary Context Switches	11	
Involuntary Context Switches	0	
Block Input Operations	0	
Block Output Operations	520	

```

84
85      DATA VISITS_TXGROUP (DROP= TEMP _FREQ_ _TYPE_);
86      MERGE VISITS_TXGROUP (IN=I) VISITS_TXGROUP_MEDIAN;
87      BY TEMP;
88      IF I;
89      IF B_Cholesterol <= Median_B_Choles THEN
B_Chol_Strata = "LESS THAN";
90      ELSE B_Chol_Strata = "GREATER THAN";
91      LABEL Median_B_Choles = "Median of Baseline
Cholesterol measurement";
92      LABEL B_Chol_Strata = "Greater than or less than
equal to the median of B_Cholesterol";
93      RUN;

```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_TXGROUP.

NOTE: There were 1 observations read from the data set  
WORK.VISITS\_TXGROUP\_MEDIAN.

NOTE: The data set WORK.VISITS\_TXGROUP has 2363 observations and  
8 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	1868.53k
OS Memory	28224.00k
Timestamp	05/03/2024 03:53:14 AM
Step Count	363 Switch Count 2
Page Faults	0
Page Reclaims	144
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	520

94

95 PROC CONTENTS DATA=VISITS\_TXGROUP;  
96 RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.02 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	1496.84k
OS Memory	27964.00k
Timestamp	05/03/2024 03:53:14 AM
Step Count	364 Switch Count 0
Page Faults	0
Page Reclaims	98
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0

Block Input Operations	0
Block Output Operations	8

97

98           OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

108

# RESULTS

The MEANS Procedure					
Analysis Variable : B_Cholesterol Baseline Cholesterol measurement					
					Median
					220.0000000

The CONTENTS Procedure					
Data Set Name	WORK.VISITS_TXGROUP		Observations	2363	
Member Type	DATA		Variables	8	
Engine	V9		Indexes	0	
Created	05/02/2024 23:53:15		Observation Length	64	
Last Modified	05/02/2024 23:53:15		Deleted Observations	0	
Protection			Compressed	NO	
Data Set Type			Sorted	NO	
Label					
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64				
Encoding	utf-8 Unicode (UTF-8)				

Engine/Host Dependent Information					
Data Set Page Size	131072				
Number of Data Set Pages	2				
First Data Page	1				
Max Obs per Page	2043				
Obs in First Data Page	1996				
Number of Data Set Repairs	0				
Filename	/saswork/SAS_work5358000083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/visits_txgroup.sas7bdat				
Release Created	9.0401M7				
Host Created	Linux				
Inode Number	1744845472				
Access Permission	rw-r--r--				
Owner Name	u62223361				
File Size	384KB				
File Size (bytes)	393216				

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
8	B_Chol_Strata	Char	9		Greater than or less than equal to the median of B_Cholesterol
5	B_Cholesterol	Num	8		Baseline Cholesterol measurement
3	Gender	Char	6		Subject gender
1	ID	Num	8		Subject ID
7	Median_B_Choles	Num	8		Median of Baseline Cholesterol measurement
6	TX	Num	8		Treatment group (0=placebo, 1=treatment)
4	Visit	Num	8		Visit number
2	VisitDt	Num	8	DATE.	Baseline visit date

# D

## CODE

```
/* C.d */  
DATA VISITS_0;  
    SET VISITS_TXGROUP;  
RUN;
```

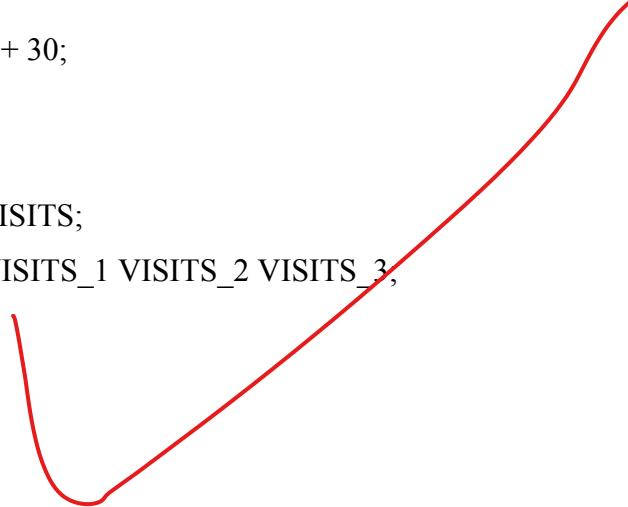
```
DATA VISITS_1;  
    SET VISITS_0;  
    Visit = 1;  
    VisitDt = VisitDt + 30;  
RUN;
```

```
DATA VISITS_2;  
    SET VISITS_1;  
    Visit = 2;  
    VisitDt = VisitDt + 30;  
RUN;
```

You can simplify your code using DO LOOP statement in the DATA STEP

```
DATA VISITS_3;  
    SET VISITS_2;  
    Visit = 3;  
    VisitDt = VisitDt + 30;  
RUN;
```

```
DATA SCHEDULED_VISITS;  
    SET VISITS_0 VISITS_1 VISITS_2 VISITS_3,  
RUN;
```



```
PROC CONTENTS DATA=SCHEDULED_VISITS;  
RUN;
```

## LOG

```
1           OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;  
68  
69           /* C.d */  
70           DATA VISITS_0;  
71           SET VISITS_TXGROUP;  
72           RUN;
```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_TXGROUP.

NOTE: The data set WORK.VISITS\_0 has 2363 observations and 8  
variables.

NOTE: DATA statement used (Total process time):

real time	0.00	seconds
user cpu time	0.00	seconds
system cpu time	0.00	seconds
memory	1311.78k	
OS Memory	27708.00k	
Timestamp	05/03/2024 03:54:24	AM
Step Count	381	Switch Count 2
Page Faults	0	
Page Reclaims	97	
Page Swaps	0	
Voluntary Context Switches	11	
Involuntary Context Switches	0	
Block Input Operations	0	
Block Output Operations	520	

```
73  
74      DATA VISITS_1;  
75      SET VISITS_0;  
76      Visit = 1;  
77      VisitDt = VisitDt + 30;  
78      RUN;  
  
NOTE: There were 2363 observations read from the data set  
WORK.VISITS_0.  
  
NOTE: The data set WORK.VISITS_1 has 2363 observations and 8  
variables.  
  
NOTE: DATA statement used (Total process time):  
      real time          0.00 seconds  
      user cpu time      0.01 seconds  
      system cpu time    0.00 seconds  
      memory            1313.09k  
      OS Memory         27708.00k  
      Timestamp          05/03/2024 03:54:24 AM  
      Step Count          382   Switch Count   2  
      Page Faults        0  
      Page Reclaims       98  
      Page Swaps          0  
      Voluntary Context Switches  12  
      Involuntary Context Switches 0  
      Block Input Operations 0  
      Block Output Operations 520  
  
79  
80      DATA VISITS_2;  
81      SET VISITS_1;  
82      Visit = 2;  
83      VisitDt = VisitDt + 30;  
84      RUN;
```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_1.

NOTE: The data set WORK.VISITS\_2 has 2363 observations and 8  
variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1425.21k
OS Memory	27708.00k
Timestamp	05/03/2024 03:54:24 AM
Step Count	383    Switch Count    2
Page Faults	0
Page Reclaims	97
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	528

85

```
86      DATA VISITS_3;
87      SET VISITS_2;
88      Visit = 3;
89      VisitDt = VisitDt + 30;
90      RUN;
```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_2.

NOTE: The data set WORK.VISITS\_3 has 2363 observations and 8  
variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
-----------	--------------

user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1313.09k
OS Memory	27708.00k
Timestamp	05/03/2024 03:54:24 AM
Step Count	384    Switch Count  2
Page Faults	0
Page Reclaims	99
Page Swaps	0
Voluntary Context Switches	10
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	520

91

```
92      DATA SCHEDULED_VISITS;
93      SET VISITS_0 VISITS_1 VISITS_2 VISITS_3;
94      RUN;
```

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_0.

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_1.

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_2.

NOTE: There were 2363 observations read from the data set  
WORK.VISITS\_3.

NOTE: The data set WORK.SCHEDULED\_VISITS has 9452 observations  
and 8 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.01 seconds

memory	2962.90k
OS Memory	29256.00k
Timestamp	05/03/2024 03:54:24 AM
Step Count	385 Switch Count 2
Page Faults	0
Page Reclaims	319
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	1288

95

96 PROC CONTENTS DATA=SCHEDULED\_VISITS;  
97 RUN;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.02 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	2406.40k
OS Memory	28092.00k
Timestamp	05/03/2024 03:54:24 AM
Step Count	386 Switch Count 0
Page Faults	0
Page Reclaims	192
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	8

98

99  
109

OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

## RESULTS

The CONTENTS Procedure			
Data Set Name	WORK.SCHEDULED_VISITS	Observations	9452
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	05/02/2024 23:54:25	Observation Length	64
Last Modified	05/02/2024 23:54:25	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information			
Data Set Page Size	131072		
Number of Data Set Pages	5		
First Data Page	1		
Max Obs per Page	2043		
Obs in First Data Page	1996		
Number of Data Set Repairs	0		
Filename	/saswork/SAS_work5358000083ED_odaws02-usw2-2.oda.sas.com/SAS_workCBF4000083ED_odaws02-usw2-2.oda.sas.com/scheduled_visits.sas7bdat		
Release Created	9.0401M7		
Host Created	Linux		
Inode Number	1744845474		
Access Permission	rW-r--r--		
Owner Name	u62223361		
File Size	768KB		
File Size (bytes)	786432		

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
8	B_Chol_Strata	Char	9		Greater than or less than equal to the median of B_Cholesterol
5	B_Cholesterol	Num	8		Baseline Cholesterol measurement
3	Gender	Char	6		Subject gender
1	ID	Num	8		Subject ID
7	Median_B_Choles	Num	8		Median of Baseline Cholesterol measurement
6	TX	Num	8		Treatment group (0=placebo, 1=treatment)
4	Visit	Num	8		Visit number
2	VisitDt	Num	8	DATE.	Baseline visit date

E

/\* C.e \*/  
/\* SCHEDULED\_VISITS has 9452 observations and 8 variables \*/  
/\* VISITS\_TXGROUP has 2363 observations and 8 variables \*/