Seneca

SENECA COLLEGE OF APPLIED ARTS AND TECHNOLOGY

SENECA BUSINESS

BAN100 - Statistics for Analytics Other

Version NA

DATE: 5/30/2023 TIME ALLOWED: 14 days

PROFESSOR(S): Samaneh Gholami

Allowable Examination Aids: (check applicable boxes)

☒ Calculators (non-programmable only)	☒ Math Tables (normal distribution table)	☒ Periodic Tables
☐ Formula Sheets (attached)	☒ Textbooks	☒ Probability Tables
☒ Dictionary	☒ Notes	☐ Other

Answers to be completed on:

☐ Exam Booklet	☐ GradeMaster Card	☐ Exam Paper

TOTAL MARKS: 20 WEIGHTED VALUE: 20

INSTRUCTIONS:

Academic Integrity Policy. Seneca upholds a learning community that values academic integrity, honesty, fairness, trust, respect, responsibility and courage. These values enhance Seneca's commitment to students by delivering high-quality education and teaching excellence, while supporting a positive learning environment. The AI policy is always in effect. Note Sections 2.3 and 2.4:

"…2.3 Should there be a suspected violation of this policy (e.g.…cheating, falsification, impersonation or plagiarism), the academic integrity sanctions will be applied according to the severity of the offence committed. Refer to Appendix B for the academic integrity sanctions. 2.4 Should a suspected violation of this policy be a result of, or in combination with, a suspected violation of Seneca's Student Code of Conduct and/or another non-academic-related Seneca policy, the matter will be investigated and adjudicated through the process found in the Student Code of Conduct."

TO BE COMPLETED BY STUDENT
SUBJECT SECTION NUMBER (e.g. QNM223 AA):
STUDENT NAME: Ugonna Okengwu
STUDENT NUMBER: 114939192
STUDENT SIGNATURE:

APPROVED BY: _____

Cristina Italia, Interim Chair

School of Management and Entrepreneurship

DATE: _____

In my investigative analysis, 2 sample hypothesis were used. This means the variables had only 2 levels of 0 and 1. The following variables were used to investigate the relationship with baby weight:

- Married
- Boy
- MomSmoke
- Black

Below is the code used to derive the investigative results.

6/11/23, 9:28 PM Code: ASSIGNMENT1.sas

```
libname Relation "/home/u63417899/BAN100ZBB";

proc import datafile="/home/u63417899/BAN100ZBB/File BIRTH(1).xlsx"
    out=FileBirth
    dbms=xlsx
    replace;
    getnames=YES;
run;

data Relation.FileBirth;
    set FileBirth;
run;

Title "Relationship between Weight and Married mother";
proc ttest data = Relation.filebirth;
     class Married;
     var Weight;
run;

Title "Relationship between Weight and Gender variable";
proc ttest data = Relation.filebirth;
     class Boy;
     var Weight;
run;

Title "Relationship between Weight and Smoking mother";
proc ttest data = Relation.filebirth;
     class Momsmoke;
     var Weight;
run;

Title "Relationship between Weight and black variable";
proc ttest data = Relation.filebirth;
     class Black;
     var Weight;
run;
```

FIRST INVESTIGATION – RELATIONSHIP BETWEEN WEIGHT AND MARRIED MOTHERS

This investigation was carried out in 3 steps.

1. **Hypothesis**

H0: $\mu 1 = \mu 2$: The average weight of babies from married mothers is the same with the average weight of unmarried mothers. This is the Null hypothesis.

Ha: $\mu 1 \neq \mu 2$: The average weight of babies from married mothers is not the same with average weight of unmarried mothers. This is the alternative hypothesis.

Based on the SAS unmarried mother is represented by μ1=0

While married mother is represented by μ2=1.

2. **Equal/Unequal variance.**

Based on the result shown in SAS, P value was less than 0.0001.

P VALUE <0.0001.

Since it is less than 0.05, we reject the null hypothesis because we have an unequal variance. Since we have an unequal variance, we will make use of the scatterthwaite variance to base our decisions on.

3. **Conclusion**

The mean weight of married mothers' babies is not the same with the mean weight of unmarried mother's babies for both the Null hypothesis and the alternative hypothesis. This means that the marital status of married mothers has an influence on their baby's weight.

The TTEST Procedure

Variable: Weight (Weight)

Married	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		14369	3234.4	579.0	4.8302	284.0	6350.0
1		35631	3425.7	551.8	2.9231	240.0	5970.0
Diff (1-2)	Pooled		-191.3	559.7	5.5315		
Diff (1-2)	Satterthwaite		-191.3		5.6459		

Married	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3234.4	3225.0	3243.9	579.0	572.4	585.8
1		3425.7	3420.0	3431.5	551.8	547.8	555.9
Diff (1-2)	Pooled	-191.3	-202.1	-180.5	559.7	556.3	563.2
Diff (1-2)	Satterthwaite	-191.3	-202.4	-180.2			

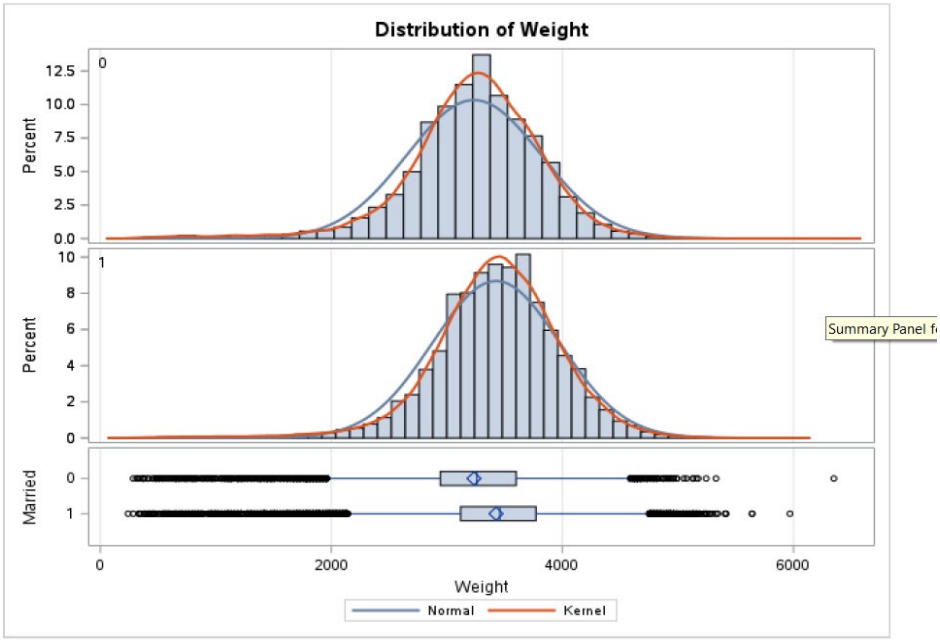
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	-34.58	<.0001
Satterthwaite	Unequal	25443	-33.88	<.0001

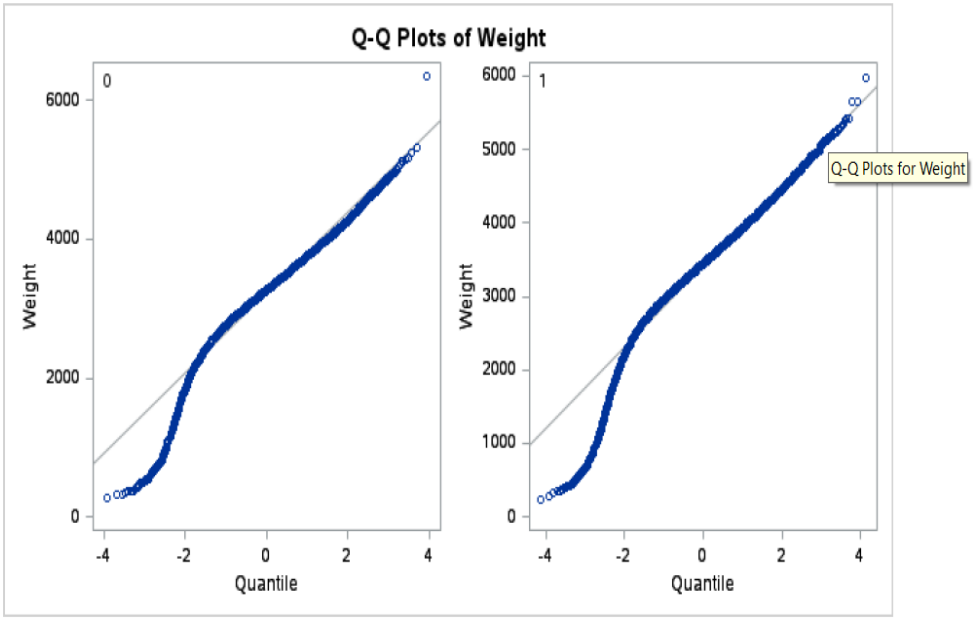
Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	14368	35630	1.10	<.0001

The analysis shows that the P value <0.0001 and the t value is -33.88. We rejected the null hypothesis, and the coding of the variable is why the t value is negative. This is because SAS studio arranges the levels of 0 and 1 in groups with higher mean as the 0 level and codes the small mean as the 1 level. Small sample size and data distributions can also be responsible for the negative t value.

The Q-Q plot depicts the quantiles of the sample data. It shows the normality for the weight of both married and unmarried mothers. The quantile usually falls along a straight line, this is because the

sample data is normally distributed. Based on the graph, the sample data has a heavier tail than a normal distribution.





SECOND INVESTIGATION – RELATIONSHIP BETWEEN WEIGHT AND GENDER

This investigation was carried out in 3 steps.

1. **Hypothesis**

H0: μ1= μ2: The average weight of male and female babies is equal. This is the Null hypothesis.

Ha: μ1 ≠ μ2: The average weight of male and female babies is equal. This is the alternative hypothesis.

Based on the SAS, not a Boy child is represented by μ1=0

While a Boy child is represented by μ2=1.

2. **Equal/Unequal variance.**

H0: $\sigma_1^2 = \sigma_2^2$ (we reject)

H0: $\sigma_1^2 \neq \sigma_2^2$

Based on the result shown in SAS, P value is less than 0.0001, making it less than α=0.05.

P VALUE <0.0001.

Since it is less than 0.05, we reject the null hypothesis because we have an unequal variance.

3. **Conclusion**

From the SAS result, it can be said that the weight of the baby is being influenced by their gender. This means there is a relationship between the baby weight and the baby's gender.

Relationship between Weight and Gender variable

The TTEST Procedure

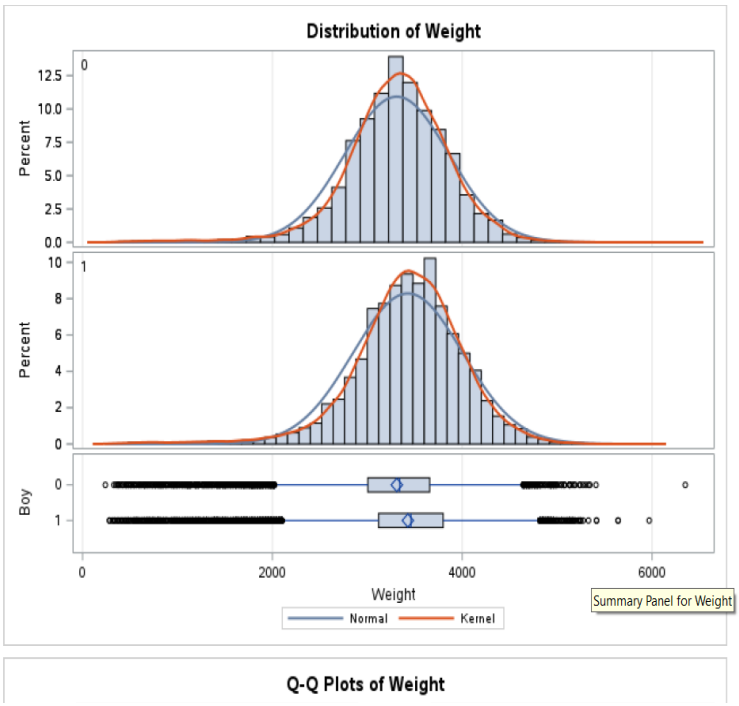
Variable: Weight (Weight)

Boy	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		24208	3310.6	547.7	3.5204	240.0	6350.0
1		25792	3427.3	577.7	3.5970	284.0	5970.0
Diff (1-2)	Pooled		-116.7	563.4	5.0416		
Diff (1-2)	Satterthwaite		-116.7		5.0331		

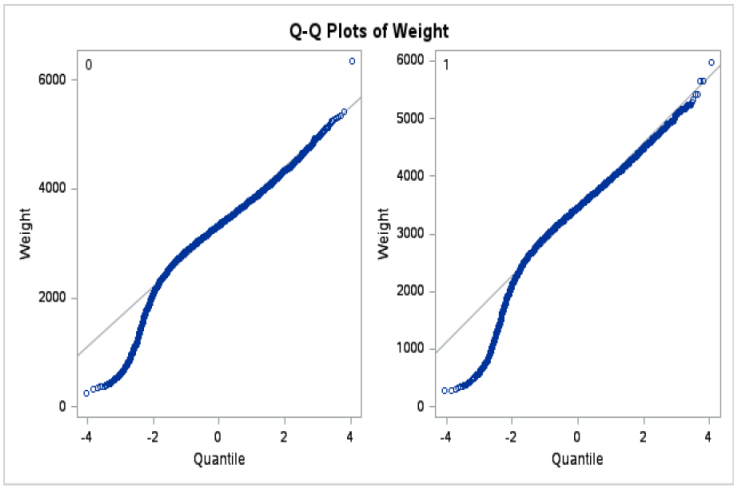
Boy	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3310.6	3303.7	3317.5	547.7	542.9	552.7
1		3427.3	3420.2	3434.3	577.7	572.7	582.7
Diff (1-2)	Pooled	-116.7	-126.6	-106.8	563.4	559.9	566.9
Diff (1-2)	Satterthwaite	-116.7	-126.6	-106.8			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	-23.15	<.0001
Satterthwaite	Unequal	49993	-23.18	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	25791	24207	1.11	<.0001



The analysis shows that the P value <0.0001 and the t value is -23.18. We rejected the null hypothesis, and the coding of the variable is why the t value is negative. The median of the 2 samples has a slight difference and the 2 data sets have similar distributions.



THIRD INVESTIGATION – RELATIONSHIP BETWEEN WEIGHT AND SMOKING MOTHERS

This investigation was carried out in 3 steps.

1. **Hypothesis**

H0: μ1= μ2: The average weight of babies from smoking mother's babies and non-smoking mothers are equal. This is the Null hypothesis.

Ha: μ1 ≠ μ2: The average weight of babies from smoking mother's babies and non-smoking mothers are not equal. This is the alternative hypothesis.

Based on the SAS, non-smoking mother is represented by μ1=0

While smoking mother is represented by μ2=1.

2. Equal/Unequal variance.

H0: $\sigma_1^2 = \sigma_2^2$ (we reject)

H0: $\sigma_1^2 \neq \sigma_2^2$

Based on the result shown in SAS, P value is less than 0.0004, making it less than α=0.05.

P VALUE <0.0004.

Since it is less than 0.05, we reject the null hypothesis because we have an unequal variance.

1. Conclusion

From the SAS result, it can be said that the weight of the baby is being influenced by the smoking habits of the mother. This means there is a relationship between the baby weight and the mother's smoking habits.

Relationship between Weight and Smoking mother

The TTEST Procedure

Variable: Weight (Weight)

-usw2.oda.sas.com/SASStudio/sasexec/submissions/e5d64b54-ff22-46f5-b253-b62716a4a59a/results

---

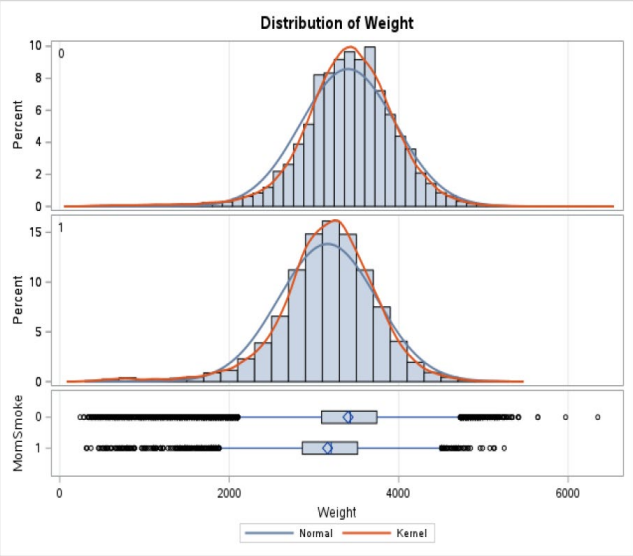
PM Results: ASSIGNMENT1.sas

MomSmoke	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		43467	3402.3	558.0	2.6766	240.0	6350.0
1		6533	3160.9	576.8	7.1358	312.0	5245.0
Diff (1-2)	Pooled		241.5	560.5	7.4376		
Diff (1-2)	Satterthwaite		241.5		7.6213		

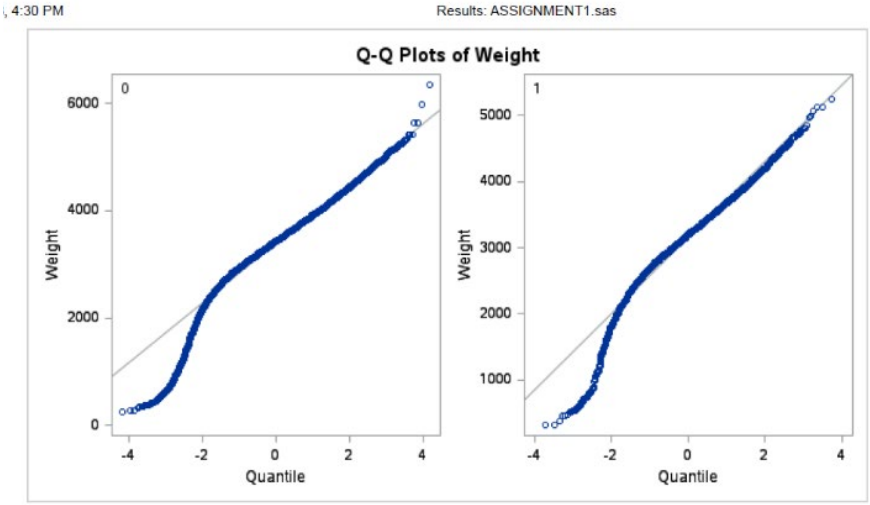
MomSmoke	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3402.3	3397.1	3407.6	558.0	554.3	561.8
1		3160.9	3146.9	3174.8	576.8	567.0	586.8
Diff (1-2)	Pooled	241.5	226.9	256.0	560.5	557.1	564.0
Diff (1-2)	Satterthwaite	241.5	226.5	256.4			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	32.46	<.0001
Satterthwaite	Unequal	8474.1	31.68	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6532	43466	1.07	0.0004



The analysis shows that the P value <0.0004 and the t value is 31.68. We rejected the null hypothesis, and the coding of the variable is why the t value is negative. The mean of the first level (3402.3) is greater than the mean of the second level (3160.9). This means the variables were coded accurately.



From the graph, it is shown identical shapes as the median has a slight difference. The plot of the quantile falls along a straight line. It depicts a normal distribution.

FOURTH INVESTIGATION – RELATIONSHIP BETWEEN WEIGHT AND BLACK VARIABLE

This investigation was carried out in 3 steps.

1. **Hypothesis**

H0: μ1= μ2: The average weight of black babies and nonblack babies are equal. This is the Null hypothesis.

Ha: $\mu 1 \neq \mu 2$: The average weight of black babies and nonblack babies are not equal. This is the alternative hypothesis.

Based on the SAS, non-black babies represented by $\mu 1=0$

While black babies is represented by $\mu 2=1$.

2. Equal/Unequal variance.

H0: $\sigma_1^2 = \sigma_2^2$ (we reject)

H0: $\sigma_1^2 \neq \sigma_2^2$

Based on the result shown in SAS, P value is less than 0.0001, making it less than $\alpha=0.05$.

P VALUE <0.0001.

Since it is less than 0.05, we reject the null hypothesis because we have an unequal variance.

3. Conclusion

From the SAS result, it can be said that the baby being black influences the weight of the baby. This means there is a relationship between the baby being black and the baby weight.

Relationship between Weight and black variable

The TTEST Procedure

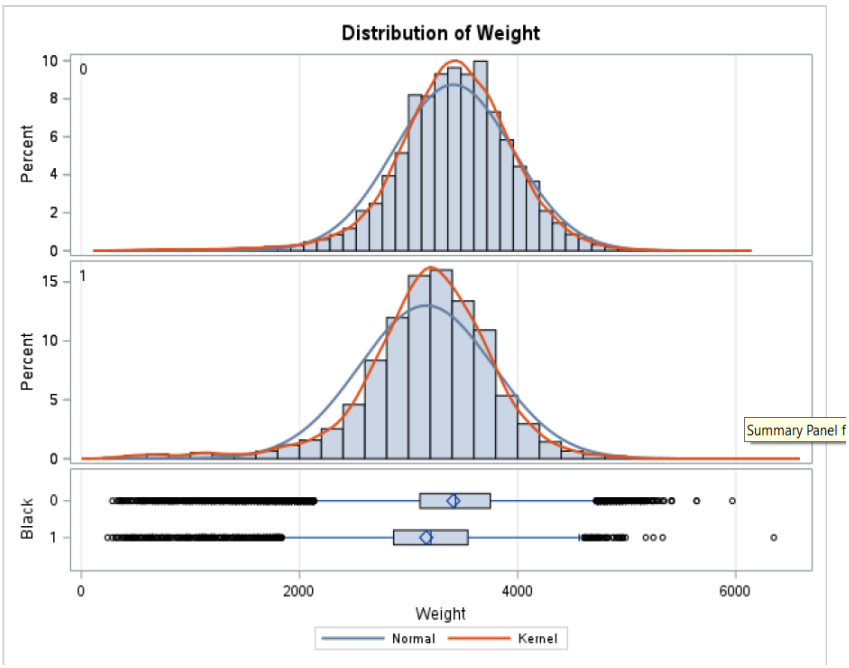
Variable: Weight (Weight)

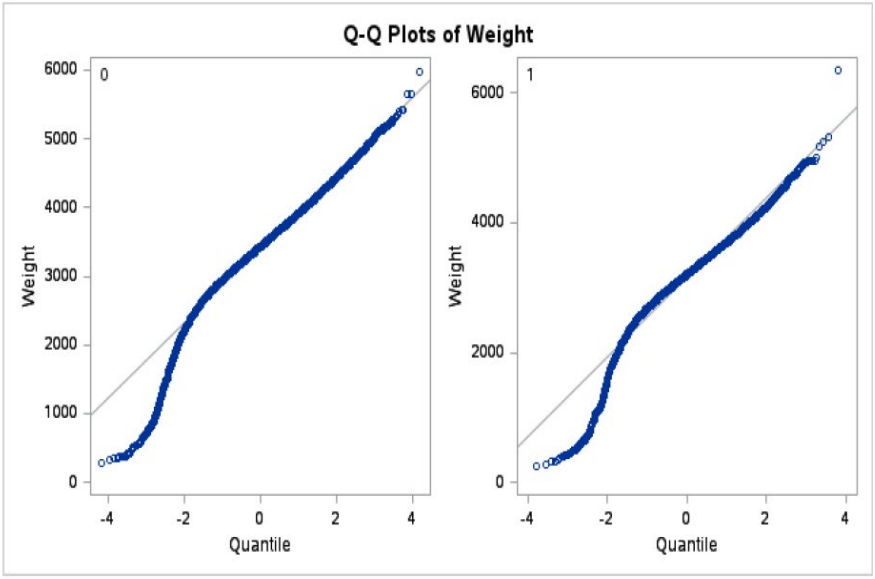
Black	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		41858	3411.2	547.6	2.6766	284.0	5970.0
1		8142	3162.7	613.7	6.8011	240.0	6350.0
Diff (1-2)	Pooled		248.6	558.9	6.7697		
Diff (1-2)	Satterthwaite		248.6		7.3088		

Black	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3411.2	3406.0	3416.5	547.6	543.9	551.4
1		3162.7	3149.3	3176.0	613.7	604.4	623.3
Diff (1-2)	Pooled	248.6	235.3	261.8	558.9	555.5	562.4
Diff (1-2)	Satterthwaite	248.6	234.2	262.9			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	36.72	<.0001
Satterthwaite	Unequal	10808	34.01	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	8141	41857	1.26	<.0001





The analysis shows that the P value <0.0001 and the t value is 34.01. We rejected the null hypothesis, and the coding of the variable is why the t value is negative.