Program Summary - Homework 4.sas

Execution Environment

Author:	chwang10
File:	/home/chwang10/Homework 4.sas
SAS Platform:	Linux LIN X64 3.10.0-1062.9.1.el7.x86_64
SAS Host:	ODAWS04-USW2.ODA.SAS.COM
SAS Version:	9.04.01M6P11072018
SAS Locale:	en_US
Submission Time:	10/21/2020, 9:00:31 PM
Browser Host:	ASTOUND-66-234-210-119.CA.ASTOUND.NET
User Agent:	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/86.0.4240.80 Safari/537.36
Application Server:	ODAMID00-USW2.ODA.SAS.COM

Code: Homework 4.sas

```
* Programmed by Charles Hwang *
* Coded in SAS OnDemand       *
* Wednesday, October 21, 2020 *
* Course: STAT 403            *
* Title: Homework 4           *;

/* 1a(i) */ * A type I error implies the water is more safe than the test implies. A consequence of this
is that safe water will potentially be rejected. A type II error implies the water is less safe than the
test implies. A consequence of this is that unsafe water will potentially be accepted. ;
/* 1a(ii) */ * A type I error implies the water is less safe than the test implies. A consequence of this
is that unsafe water will potentially be accepted. A type II error implies the water is more safe than
the test implies. A consequence of this is that safe water will potentially be rejected. ;
/* 1a(iii) */ * The alternative hypothesis with HA: μ < 10 ppb is more appropriate in this situation.
We want to ensure beyond a reasonable doubt that the arsenic level is below 10 ppb. The test with an
alternative hypothesis of HA: μ > 10 ppb would not be helpful in this situation because both the
null (H0: μ = 10 ppb) and alternative (HA: μ > 10 ppb) hypotheses are undesired outcomes. ;

/* 1b */ Data EPA;
Input Arsenic @@;
Datalines;
9.722 10.162 9.976 9.787 9.474 10.113 10.157 9.556 9.667 9.809 10.424 9.288
;

/* 1c */ Proc Ttest data=EPA h0=10 alpha=.01 sides=L; * One-sided test ;
Var Arsenic;
Title "1c. One-Sample T-Test on Arsenic Concentration in Water";
Run;
/* 1c(i) */ * t = -1.63, p = .06605 ;
/* 1c(ii) */ * We fail to reject H0 at the α = .01 level. There is insufficient evidence that the
mean arsenic level is below 10 ppb. ;
/* 1c(iii) */ * We are 99 percent confident that the mean arsenic level is 10.1043 ppb or less. ;

/* 2a */ Data Snoring;
Infile "/home/chwang10/Snoring.txt" dlm='09'x firstobs=2; * Skipping header row, starting on row 2 ;
Input Presurgery Postsurgery;
Difference=Presurgery-Postsurgery;
Run;

/* 2b */ * A paired t-test would be better in this situation because the data are clearly two separate
observations of a patient and thus not independent. ;

/* 2c(i) */ * H0: D = 0, HA: D ≠ 0 ;
/* 2c(ii) */ Proc Ttest data=Snoring;
Paired Presurgery*Postsurgery;
Title "2c(iii). Two-Sample Paired T-Test on Difference";
Run;
/* 2c(iii) */ * See output from Problem 2c(ii) ;
/* 2c(iv) */ * We are 95 percent confident the mean difference between pre-surgery snoring and
post-surgery snoring is between -0.0194 and 9.6860 decibels. ;
```

```
/* 2c(v) */ * t = 2.19, p = 0.0508, We fail to reject H0 at the α = .05 level. There is insufficient
evidence that the mean post-surgery snoring volume in decibels is lower than the mean pre-surgery
snoring volume in decibels. ;

/* 2d(i) */ * H0: The median post-surgery snoring volume in decibels is the same as the median
pre-surgery snoring volume in decibels. *
* HA: The median post-surgery snoring volume in decibels is not the same as the median pre-surgery
snoring volume in decibels. * ;
/* 2d(ii) */ Proc Univariate data=Snoring;
Var Difference;
Title "2d(iii). Wilcoxon Ranked Sign Test on Difference";
Run;
/* 2d(iii) */ * See output from Problem 2d(ii) ;
/* 2d(iv) */ * M = 4, p = 0.0386, We reject H0 at the α = .05 level. There is sufficient evidence that the
mean post-surgery snoring volume in decibels is lower than the mean pre-surgery snoring volume in decibels. ;

/* 2e(i) */ * The distribution is normal enough to use the t-test, but we should still exercise some
caution in doing so, as the number of observations in the data is not large. ;
/* 2e(ii) */ * The paired t-test is the better method. Because the data are paired, it makes sense to
use this test in this situation. ;
/* 2e(iii) */ * It is important to justify the test used because there can clearly be very different
conclusions and to avoid the appearance of data or analysis manipulation. ;

/* 3a */ Proc Import out=SS datafile="/home/chwang10/SubscriberSurvey.xlsx" dbms=xlsx;
Run;

/* 3b(i) */ * H0: Q_C = Q_S, HA: Q_C ≠ Q_S ;
/* 3b(ii) */ Proc Ttest data=SS;
Class Provider;
Var QOS;
Title "3b(iii). Two-Sample Independent T-Test on Quality of Service";
Run;
/* 3b(iii) */ * See output from Problem 3b(ii) ;
/* 3b(iv) */ * H0: σ_(Q_C) = σ_(Q_S), HA: σ_(Q_C) ≠ σ_(Q_S), F = 2.83, p = .0066 ;
* We reject H0 at the α = .05 level. There is sufficient evidence that the variance of the quality of
service of cable subscribers is different than the variance of the quality of service of satellite
subscribers. We should use the Satterthwaite approximation of degrees of freedom for unequal variances. ;
/* 3b(v) */ * (-1.7297,-0.0703), We are 95 percent confident that the difference between the
quality of service for cable subscribers and the quality of service for satellite subscribers
is -1.7297 and -0.0703. ;
/* 3b(vi) */ * t = -2.18, p = 0.0341, We reject H0 at the α = .05 level. There is sufficient evidence
that there is a difference between the quality of service for cable subscribers and the quality of
service for satellite subscribers. ;

/* 3c(i) */ * H0: The median quality of service for cable subscribers is the same as the median
quality of service for satellite subscribers. ;
* HA: The median quality of service for cable subscribers is different than the median quality of
service for satellite subscribers. ;
/* 3c(ii) */ Proc npar1way wilcoxon data=SS;
Class Provider;
Var QOS;
Exact wilcoxon;
Title "3c(iii). Wilcoxon Sum Rank Test on Quality of Service";
Run;
/* 3c(iii) */ * See output from Problem 3c(ii) ;
/* 3c(iv) */ * S = 1045.5, p = .0493, We reject H0 at the α = .05 level. There is sufficient evidence
that the median quality of service for cable subscribers is different than the median quality of
service for satellite subscribers. ;

/* 3d(i) */ * The distribution is normal enough to use the t-test. However, there are other problems
with doing so in this situation. ;
/* 3d(ii) */ * The Wilcoxon Rank Sum Test is the better method. Because quality of service is a
discrete variable (as demonstrated by the Q-Q plot in the output for problem 3b(ii)), a t-test
would be inappropriate in this situation. ;
```

Log: Homework 4.sas

Notes (16)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
```

```

       * Programmed by Charles Hwang *
       * Coded in SAS OnDemand       *
       * Wednesday, October 21, 2020 *
       * Course: STAT 403            *
       * Title: Homework 4           *;

       /* 1a(i) */ * A type I error implies the water is more safe than the test implies. A consequence of this
       is that safe water will potentially be rejected. A type II error implies the water is less safe than the
       test implies. A consequence of this is that unsafe water will potentially be accepted. ;
       /* 1a(ii) */ * A type I error implies the water is less safe than the test implies. A consequence of this
       is that unsafe water will potentially be accepted. A type II error implies the water is more safe than
       the test implies. A consequence of this is that safe water will potentially be rejected. ;
       /* 1a(iii) */ * The alternative hypothesis with HA: µ < 10 ppb is more appropriate in this situation.
       We want to ensure beyond a reasonable doubt that the arsenic level is below 10 ppb. The test with an
       alternative hypothesis of HA: µ > 10 ppb would not be helpful in this situation because both the
       null (H0: µ = 10 ppb) and alternative (HA: µ > 10 ppb) hypotheses are undesired outcomes. ;

       /* 1b */
     !          Data EPA;
       Input Arsenic @@;
       Datalines;

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.EPA has 12 observations and 1 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              666.06k
      OS Memory           37292.00k
      Timestamp           10/22/2020 04:00:30 AM
      Step Count                        398  Switch Count  2
      Page Faults                       0
      Page Reclaims                     92
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


       ;

       /* 1c */
     !          Proc Ttest data=EPA h0=10 alpha=.01 sides=L; * One-sided test ;
       Var Arsenic;
       Title "1c. One-Sample T-Test on Arsenic Concentration in Water";
       Run;

NOTE: PROCEDURE TTEST used (Total process time):
      real time           0.35 seconds
      user cpu time       0.18 seconds
      system cpu time     0.04 seconds
      memory              23902.40k
      OS Memory           53200.00k
      Timestamp           10/22/2020 04:00:30 AM
      Step Count                        399  Switch Count  25
      Page Faults                       0
      Page Reclaims                     15911
      Page Swaps                        0
      Voluntary Context Switches        732
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           1248


       /* 1c(i) */ * t = -1.63, p = .06605 ;
       /* 1c(ii) */ * We fail to reject H0 at the α = .01 level. There is insufficient evidence that the
       mean arsenic level is below 10 ppb. ;
       /* 1c(iii) */ * We are 99 percent confident that the mean arsenic level is 10.1043 ppb or less. ;

       /* 2a */
     !          Data Snoring;
       Infile "/home/chwang10/Snoring.txt" dlm='09'x firstobs=2; * Skipping header row, starting on row 2 ;
       Input Presurgery Postsurgery;
       Difference=Presurgery-Postsurgery;
       Run;

NOTE: The infile "/home/chwang10/Snoring.txt" is:
      Filename=/home/chwang10/Snoring.txt,
      Owner Name=chwang10,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=12Oct2020:02:04:47,
```

```
      File Size (bytes)=111

NOTE: 13 records were read from the infile "/home/chwang10/Snoring.txt".
      The minimum record length was 0.
      The maximum record length was 6.
NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.SNORING has 12 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              762.65k
      OS Memory           48044.00k
      Timestamp           10/22/2020 04:00:30 AM
      Step Count                        400  Switch Count  2
      Page Faults                       0
      Page Reclaims                     100
      Page Swaps                        0
      Voluntary Context Switches        18
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264



        /* 2b */ * A paired t-test would be better in this situation because the data are clearly two separate
        observations of a patient and thus not independent. ;

        /* 2c(i) */ * H0: D = 0, HA: D ≠ 0 ;
        /* 2c(ii) */
      !              Proc Ttest data=Snoring;
        Paired Presurgery*Postsurgery;
        Title "2c(iii). Two-Sample Paired T-Test on Difference";
        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              10709.93k
      OS Memory           56020.00k
      Timestamp           10/22/2020 04:00:30 AM
      Step Count                        401  Switch Count  31
      Page Faults                       0
      Page Reclaims                     16942
      Page Swaps                        0
      Voluntary Context Switches        1221
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1424


        /* 2c(iii) */ * See output from Problem 2c(ii) ;
        /* 2c(iv) */ * We are 95 percent confident the mean difference between pre-surgery snoring and
        post-surgery snoring is between -0.0194 and 9.6860 decibels. ;
        /* 2c(v) */ * t = 2.19, p = 0.0508, We fail to reject H0 at the α = .05 level. There is insufficient
        evidence that the mean post-surgery snoring volume in decibels is lower than the mean pre-surgery
        snoring volume in decibels. ;

        /* 2d(i) */ * H0: The median post-surgery snoring volume in decibels is the same as the median
        pre-surgery snoring volume in decibels. *
        * HA: The median post-surgery snoring volume in decibels is not the same as the median pre-surgery
        snoring volume in decibels. * ;
        /* 2d(ii) */
      !              Proc Univariate data=Snoring;
        Var Difference;
        Title "2d(iii). Wilcoxon Ranked Sign Test on Difference";
        Run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.05 seconds
      user cpu time       0.06 seconds
      system cpu time     0.00 seconds
      memory              819.62k
      OS Memory           49840.00k
      Timestamp           10/22/2020 04:00:31 AM
      Step Count                        402  Switch Count  1
      Page Faults                       0
      Page Reclaims                     56
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           16
```

```
      /* 2d(iii) */ * See output from Problem 2d(ii) ;
      /* 2d(iv) */ * M = 4, p = 0.0386, We reject H0 at the α = .05 level. There is sufficient evidence that the
      mean post-surgery snoring volume in decibels is lower than the mean pre-surgery snoring volume in decibels. ;

      /* 2e(i) */ * The distribution is normal enough to use the t-test, but we should still exercise some
      caution in doing so, as the number of observations in the data is not large. ;
      /* 2e(ii) */ * The paired t-test is the better method. Because the data are paired, it makes sense to
      use this test in this situation. ;
      /* 2e(iii) */ * It is important to justify the test used because there can clearly be very different
      conclusions and to avoid the appearance of data or analysis manipulation. ;

      /* 3a */
    !          Proc Import out=SS datafile="/home/chwang10/SubscriberSurvey.xlsx" dbms=xlsx;
      Run;

NOTE: Import cancelled.  Output dataset WORK.SS already exists.  Specify REPLACE option to overwrite it.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              687.46k
      OS Memory           50076.00k
      Timestamp           10/22/2020 04:00:31 AM
      Step Count                        403  Switch Count  0
      Page Faults                       0
      Page Reclaims                     142
      Page Swaps                        0
      Voluntary Context Switches        2
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


      /* 3b(i) */ * H0: Q_C = Q_S, HA: Q_C ≠ Q_S ;
      /* 3b(ii) */


    !              Proc Ttest data=SS;
      Class Provider;
      Var QOS;
      Title "3b(iii). Two-Sample Independent T-Test on Quality of Service";
      Run;

NOTE: PROCEDURE TTEST used (Total process time):
      real time           0.38 seconds
      user cpu time       0.20 seconds
      system cpu time     0.06 seconds
      memory              9979.34k
      OS Memory           56796.00k
      Timestamp           10/22/2020 04:00:31 AM
      Step Count                        404  Switch Count  49
      Page Faults                       0
      Page Reclaims                     25581
      Page Swaps                        0
      Voluntary Context Switches        1129
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1184


      /* 3b(iii) */ * See output from Problem 3b(ii) ;
      /* 3b(iv) */ * H0: σ_(Q_C) = σ_(Q_S), HA: σ_(Q_C) ≠ σ_(Q_S), F = 2.83, p = .0066 ;
      * We reject H0 at the α = .05 level. There is sufficient evidence that the variance of the quality of
      service of cable subscribers is different than the variance of the quality of service of satellite
      subscribers. We should use the Satterthwaite approximation of degrees of freedom for unequal variances. ;
      /* 3b(v) */ * (-1.7297,-0.0703), We are 95 percent confident that the difference between the
      quality of service for cable subscribers and the quality of service for satellite subscribers
      is -1.7297 and -0.0703. ;
      /* 3b(vi) */ * t = -2.18, p = 0.0341, We reject H0 at the α = .05 level. There is sufficient evidence
      that there is a difference between the quality of service for cable subscribers and the quality of
      service for satellite subscribers. ;

      /* 3c(i) */ * H0: The median quality of service for cable subscribers is the same as the median
      quality of service for satellite subscribers. ;
      * HA: The median quality of service for cable subscribers is different than the median quality of
      service for satellite subscribers. ;
      /* 3c(ii) */
    !              Proc npar1way wilcoxon data=SS;
      Class Provider;
      Var QOS;
      Exact wilcoxon;
```

```
        Title "3c(iii). Wilcoxon Sum Rank Test on Quality of Service";
        Run;

NOTE: PROCEDURE NPAR1WAY used (Total process time):
      real time           0.15 seconds
      user cpu time       0.09 seconds
      system cpu time     0.00 seconds
      memory              3165.59k
      OS Memory           51520.00k
      Timestamp           10/22/2020 04:00:31 AM
      Step Count                        405  Switch Count  1
      Page Faults                       0
      Page Reclaims                     372
      Page Swaps                        0
      Voluntary Context Switches        326
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           696


        /* 3c(iii) */ * See output from Problem 3c(ii) ;
        /* 3c(iv) */ * S = 1045.5, p = .0493, We reject H0 at the α = .05 level. There is sufficient evidence
        that the median quality of service for cable subscribers is different than the median quality of
        service for satellite subscribers. ;

        /* 3d(i) */ * The distribution is normal enough to use the t-test. However, there are other problems
        with doing so in this situation. ;
        /* 3d(ii) */ * The Wilcoxon Rank Sum Test is the better method. Because quality of service is a
        discrete variable (as demonstrated by the Q-Q plot in the output for problem 3b(ii)), a t-test
        would be inappropriate in this situation. ;

        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

```

Results: Homework 4.sas

1c. One-Sample T-Test on Arsenic Concentration in Water

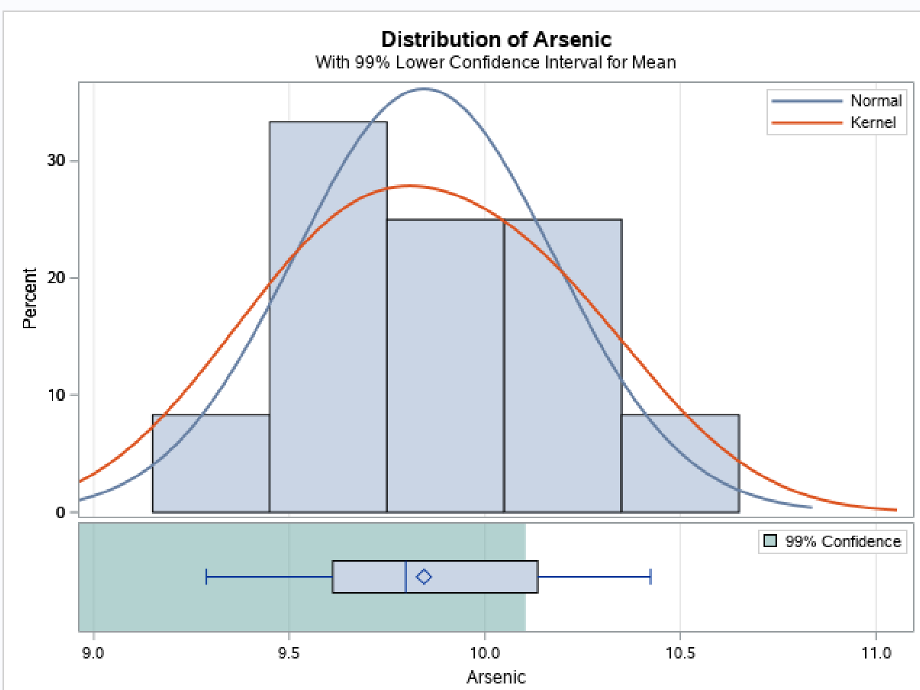
The TTEST Procedure

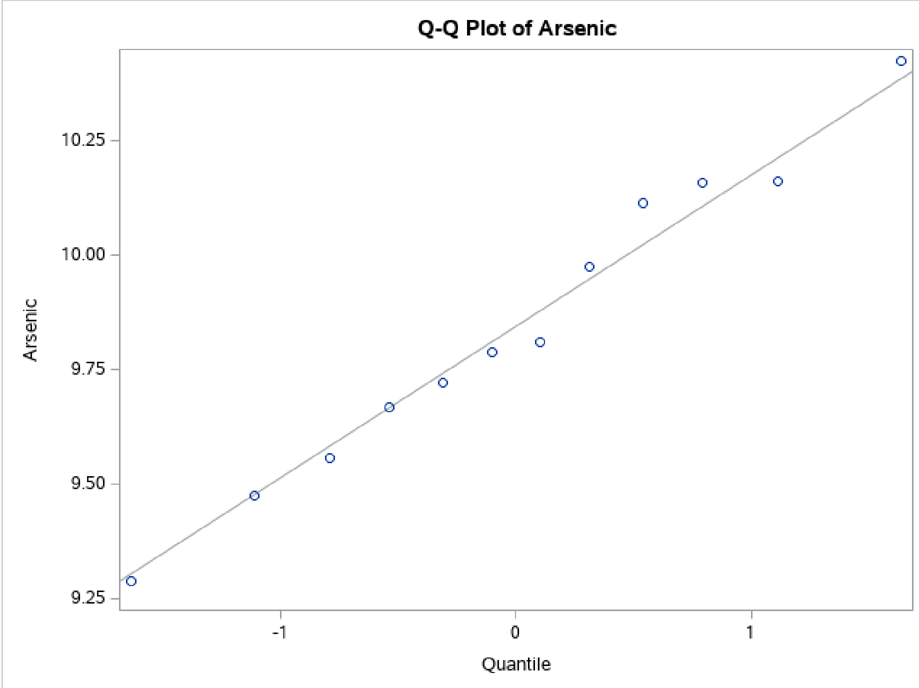
Variable: Arsenic

N	Mean	Std Dev	Std Err	Minimum	Maximum
12	9.8446	0.3310	0.0955	9.2880	10.4240

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
9.8446	-Infty	10.1043	0.3310	0.2122	0.6803

DF	t Value	Pr < t
11	-1.63	0.0660





2c(iii). Two-Sample Paired T-Test on Difference

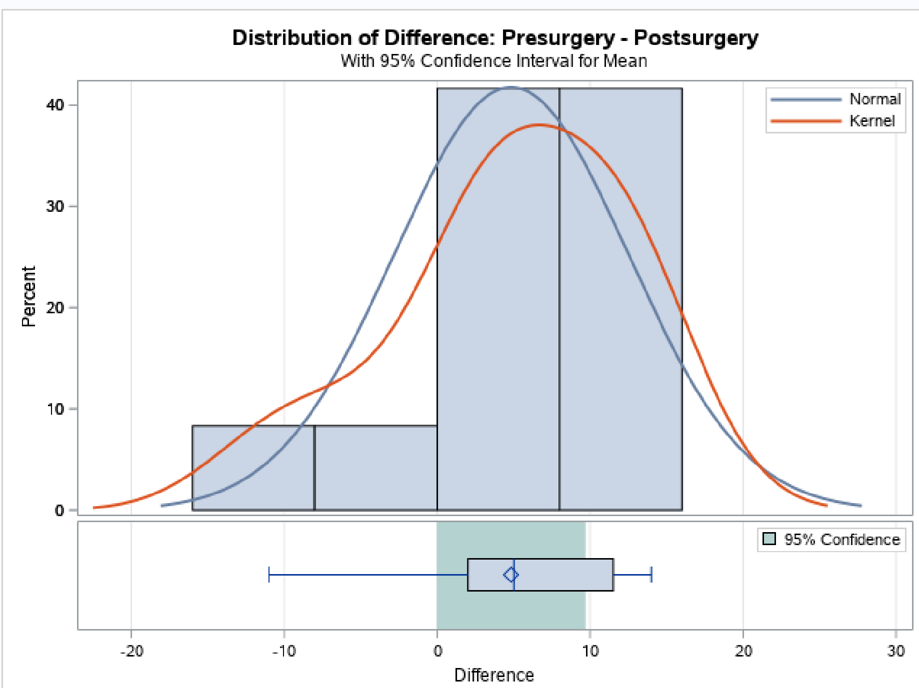
The TTEST Procedure

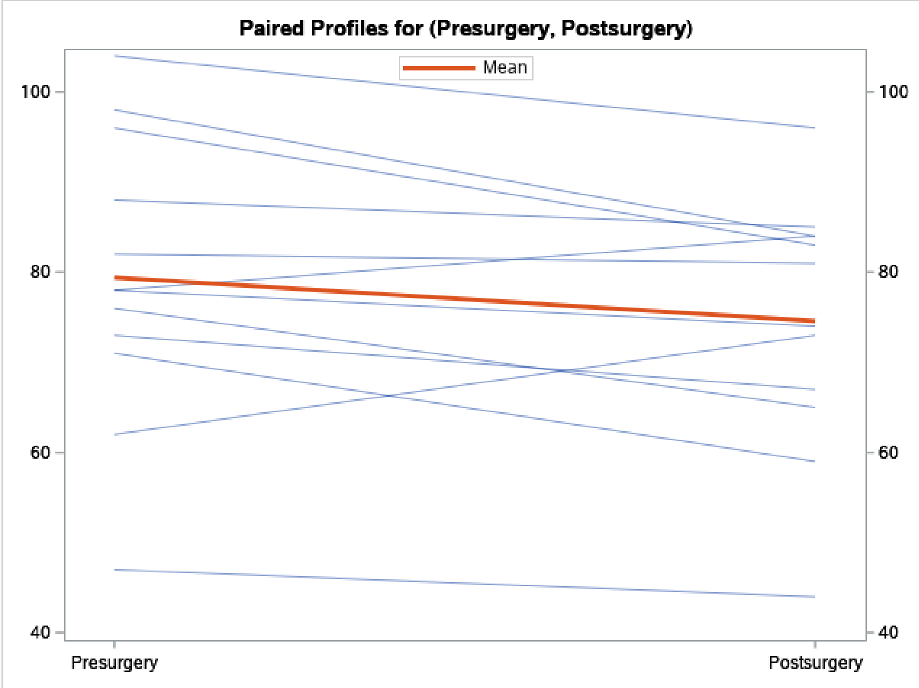
Difference: Presurgery - Postsurgery

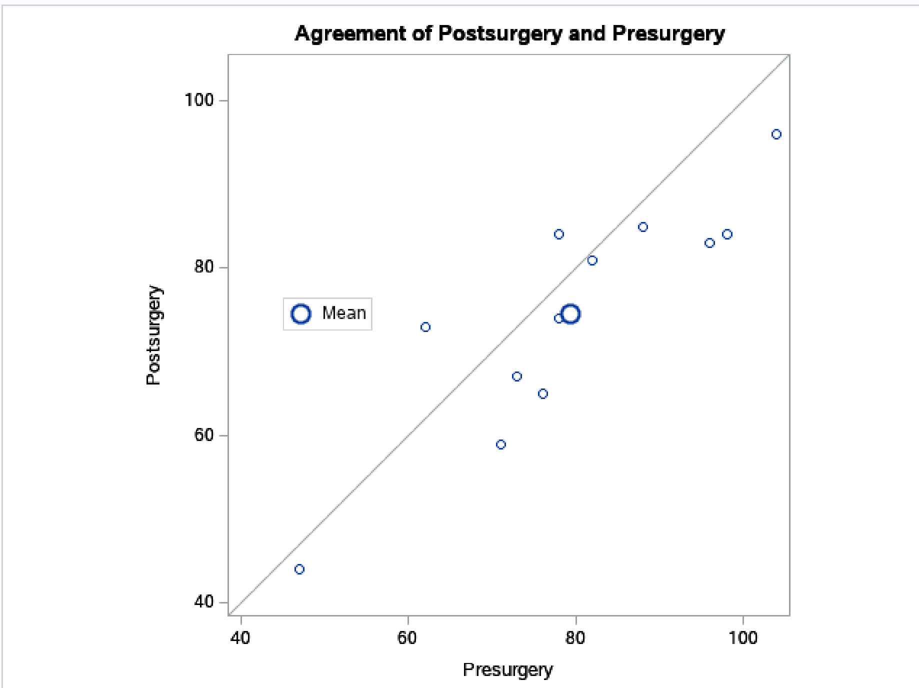
N	Mean	Std Dev	Std Err	Minimum	Maximum
12	4.8333	7.6376	2.2048	-11.0000	14.0000

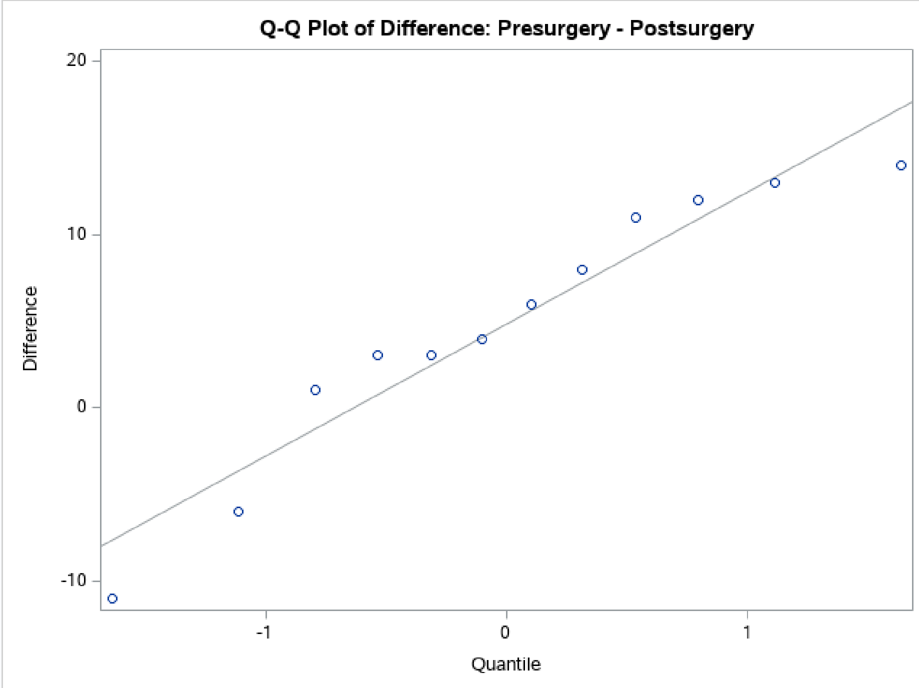
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
4.8333	-0.0194	9.6860	7.6376	5.4105	12.9678

DF	t Value	Pr > |t|
11	2.19	0.0508




Paired Profiles for (Presurgery, Postsurgery)
Mean
100
80
60
40
Presurgery
Postsurgery


Agreement of Postsurgery and Presurgery
Mean
Postsurgery
Presurgery
100
80
60
40



2d(iii). Wilcoxon Ranked Sign Test on Difference

The UNIVARIATE Procedure
Variable: Difference

Moments			
N	12	Sum Weights	12
Mean	4.83333333	Sum Observations	58
Std Deviation	7.63762616	Variance	58.3333333
Skewness	-0.8225035	Kurtosis	0.25273113
Uncorrected SS	922	Corrected SS	641.666667
Coeff Variation	158.019852	Std Error Mean	2.20479276

Basic Statistical Measures			
Location		Variability	
Mean	4.833333	Std Deviation	7.63763
Median	5.000000	Variance	58.33333
Mode	3.000000	Range	25.00000
		Interquartile Range	9.50000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	2.192194	Pr > |t|	0.0508
Sign	M	4	Pr >= |M|	0.0386
Signed Rank	S	25	Pr >= |S|	0.0483

Quantiles (Definition 5)	
Level	Quantile
100% Max	14.0
99%	14.0
95%	14.0
90%	13.0
75% Q3	11.5
50% Median	5.0
25% Q1	2.0
10%	-6.0
5%	-11.0
1%	-11.0
0% Min	-11.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-11	9	8	3
-6	4	11	10

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	7	12	2
3	8	13	1
3	5	14	11

3b(iii). Two-Sample Independent T-Test on Quality of Service

The TTEST Procedure

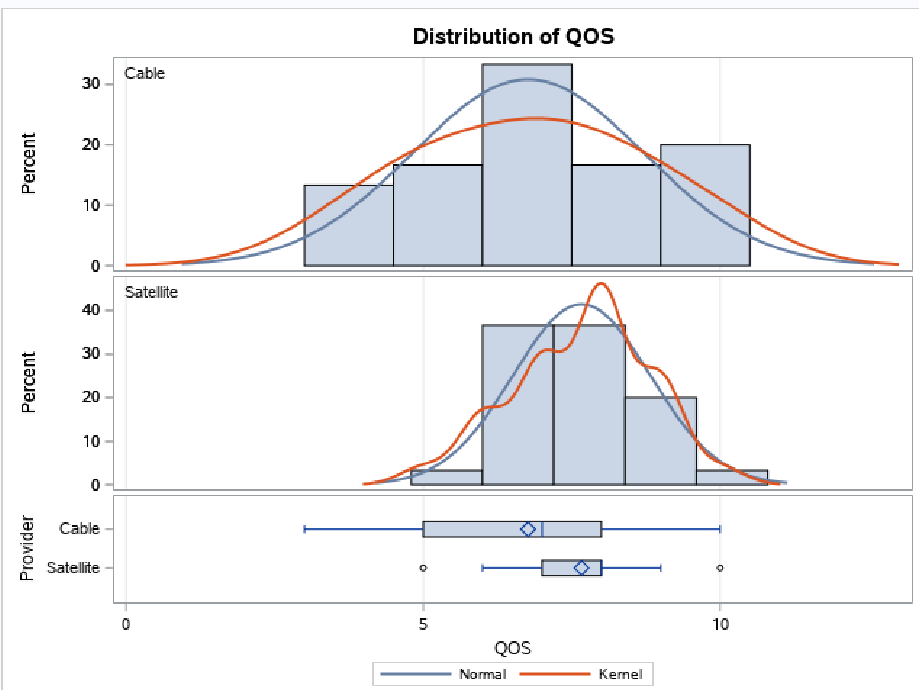
Variable: QOS (QOS)

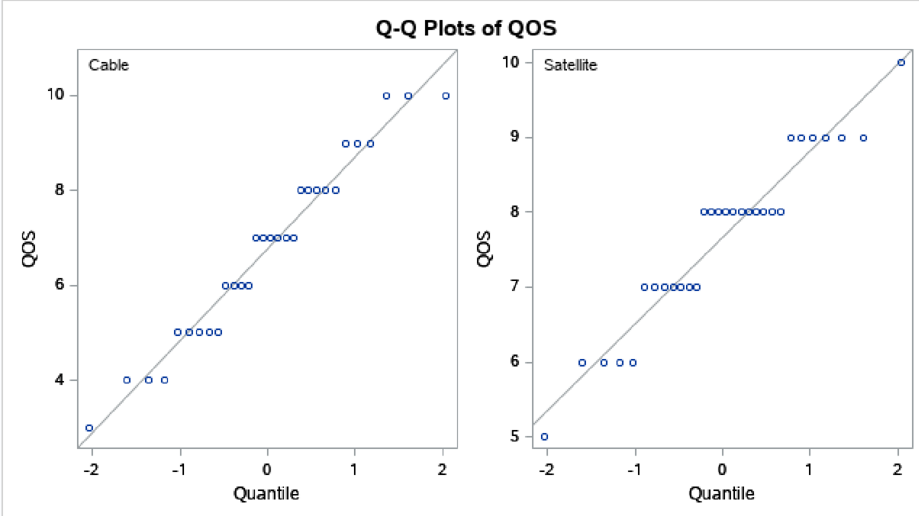
Provider	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Cable		30	6.7667	1.9420	0.3546	3.0000	10.0000
Satellite		30	7.6667	1.1547	0.2108	5.0000	10.0000
Diff (1-2)	Pooled		-0.9000	1.5976	0.4125		
Diff (1-2)	Satterthwaite		-0.9000		0.4125		

Provider	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Cable		6.7667	6.0415	7.4918	1.9420	1.5466	2.6106
Satellite		7.6667	7.2355	8.0978	1.1547	0.9196	1.5523
Diff (1-2)	Pooled	-0.9000	-1.7257	-0.0743	1.5976	1.3524	1.9522
Diff (1-2)	Satterthwaite	-0.9000	-1.7297	-0.0703			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	58	-2.18	0.0332
Satterthwaite	Unequal	47.228	-2.18	0.0341

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	29	29	2.83	0.0066





3c(iii). Wilcoxon Sum Rank Test on Quality of Service

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable QOS Classified by Variable Provider					
Provider	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Satellite	30	1045.50	915.0	66.408788	34.850
Cable	30	784.50	915.0	66.408788	26.150
Average scores were used for ties.					

Wilcoxon Two-Sample Test							
				t Approximation		Exact	
Statistic (S)	Z	Pr > Z	Pr > |Z|	Pr > Z	Pr > |Z|	Pr >= S	Pr >= |S-Mean|
1045.500	1.9576	0.0251	0.0503	0.0275	0.0550	0.0246	0.0493
Z includes a continuity correction of 0.5.							

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
3.8616	1	0.0494