Chapter 3 - Minitab Details

In Minitab, under each procedure, go to Help, click on <u>see also</u> and <u>Methods and Formulas</u> for details about the procedure.

Case 3.1.1. Cloud Seeding to Increase Rainfall – A Randomized Experiment

Step 1: Copy the data into a Minitab Worksheet: use these steps:
File → Open Worksheet →Browse your local directory and upload the csv file case0301.csv

The data will appear as two columns in Minitab with C1 RAINFALL and C2-T TREATMENT . So Column 2 is a Text column with entries UNSEEDED or SEEDED.

Step 2: Display Data,
Go to Data →Display Data; Select C1 and then select C2, and click OK to produce a display on the output portion of your Minitab session as shown below:

Data Display

Row	RAINFALL	TREATMENT
1	1202.60	UNSEEDED
2	830.10	UNSEEDED
3	372.40	UNSEEDED
4	345.50	UNSEEDED
5	321.20	UNSEEDED
6	244.30	UNSEEDED
7	163.00	UNSEEDED
8	147.80	UNSEEDED
9	95.00	UNSEEDED
10	87.00	UNSEEDED
11	81.20	UNSEEDED
12	68.50	UNSEEDED
13	47.30	UNSEEDED
14	41.10	UNSEEDED
15	36.60	UNSEEDED
16	29.00	UNSEEDED
17	28.60	UNSEEDED
18	26.30	UNSEEDED
19	26.10	UNSEEDED
20	24.40	UNSEEDED
21	21.70	UNSEEDED
22	17.30	UNSEEDED
23	11.50	UNSEEDED
24	4.90	UNSEEDED
25	4.90	UNSEEDED
26	1.00	UNSEEDED
27	2745.60	SEEDED
28	1697.80	SEEDED
29	1656.00	SEEDED
30	978.00	SEEDED
31	703.40	SEEDED
32	489.10	SEEDED
33	430.00	SEEDED
34	334.10	SEEDED
35	302.80	SEEDED

36	274.70	SEEDED
37	274.70	SEEDED
38	255.00	SEEDED
39	242.50	SEEDED
40	200.70	SEEDED
41	198.60	SEEDED
42	129.60	SEEDED
43	119.00	SEEDED
44	118.30	SEEDED
45	115.30	SEEDED
46	92.40	SEEDED
47	40.60	SEEDED
48	32.70	SEEDED
49	31.40	SEEDED
50	17.50	SEEDED
51	7.70	SEEDED
52	4.10	SEEDED

Step 3: you can un-stack the data. To do this, go to Data → Unstack Columns →select variable C1 into the window/box labeled Unstack the Data in;

Select C2 into the window/box labeled Using Subscripts in; Store un-stacked data: click on After last column in use;

Un-tick Name the Columns containing the un-stacked data; click OK.

This will create 2 new columns in your worksheet, C3 and C4, which are titled RAINFALL_SEEDED and RAINFALL_UNSEEDED

Data Display

Row	RAINFALL_SEEDED	RAINFALL_UNSEEDED
1	2745.60	1202.60
2	1697.80	830.10
3	1656.00	372.40
4	978.00	345.50
5	703.40	321.20
6	489.10	244.30
7	430.00	163.00
8	334.10	147.80
9	302.80	95.00
10	274.70	87.00
11	274.70	81.20
12	255.00	68.50
13	242.50	47.30
14	200.70	41.10
15	198.60	36.60
16	129.60	29.00
17	119.00	28.60
18	118.30	26.30
19	115.30	26.10
20	92.40	24.40
21	40.60	21.70
22	32.70	17.30
23	31.40	11.50
24	17.50	4.90
25	7.70	4.90
26	4.10	1.00

Step 4: Summary Statistics

Go to Stat →Basic Statistics →Display Descriptive Statistics;

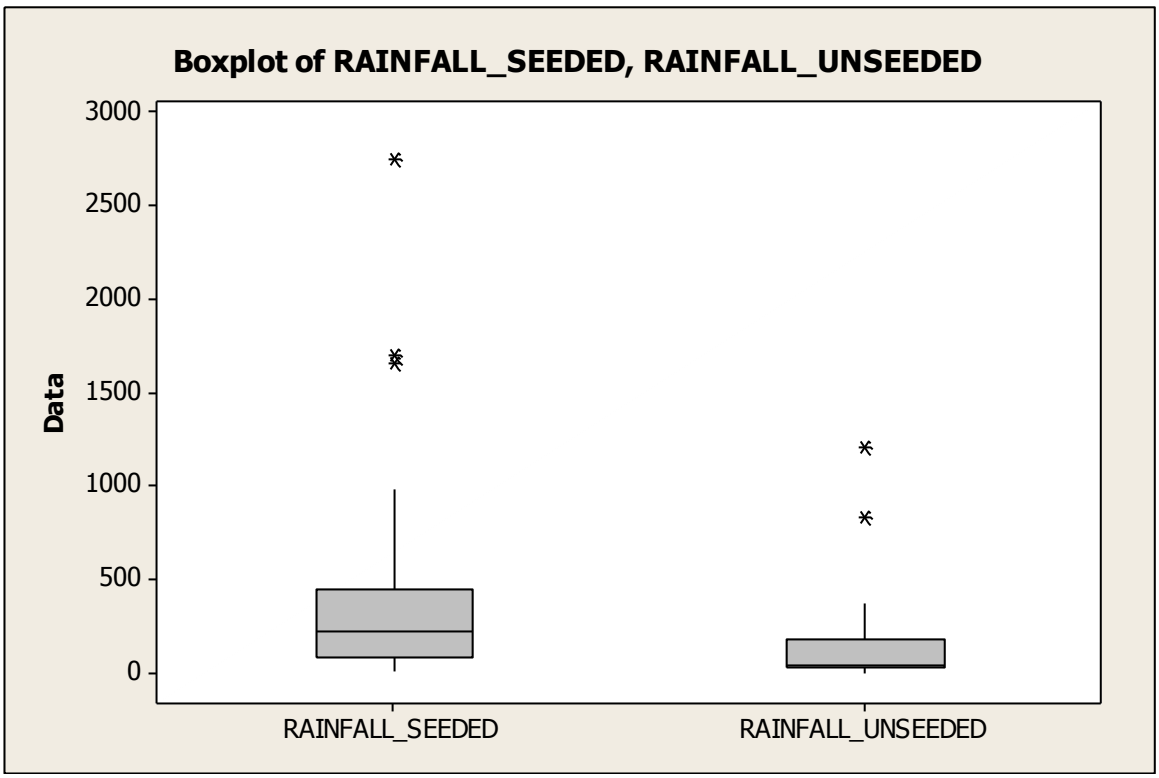
Select C1 RAINFALL into Variables; select C2 TREATMENT into By Variables and click OK, to see this display on the output portion:

Descriptive Statistics: RAINFALL

Variable	TREATMENT	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median
RAINFALL	SEEDED	26	0	442	128	651	4	79	222
	UNSEEDED	26	0	164.6	54.6	278.4	1.0	23.7	44.2

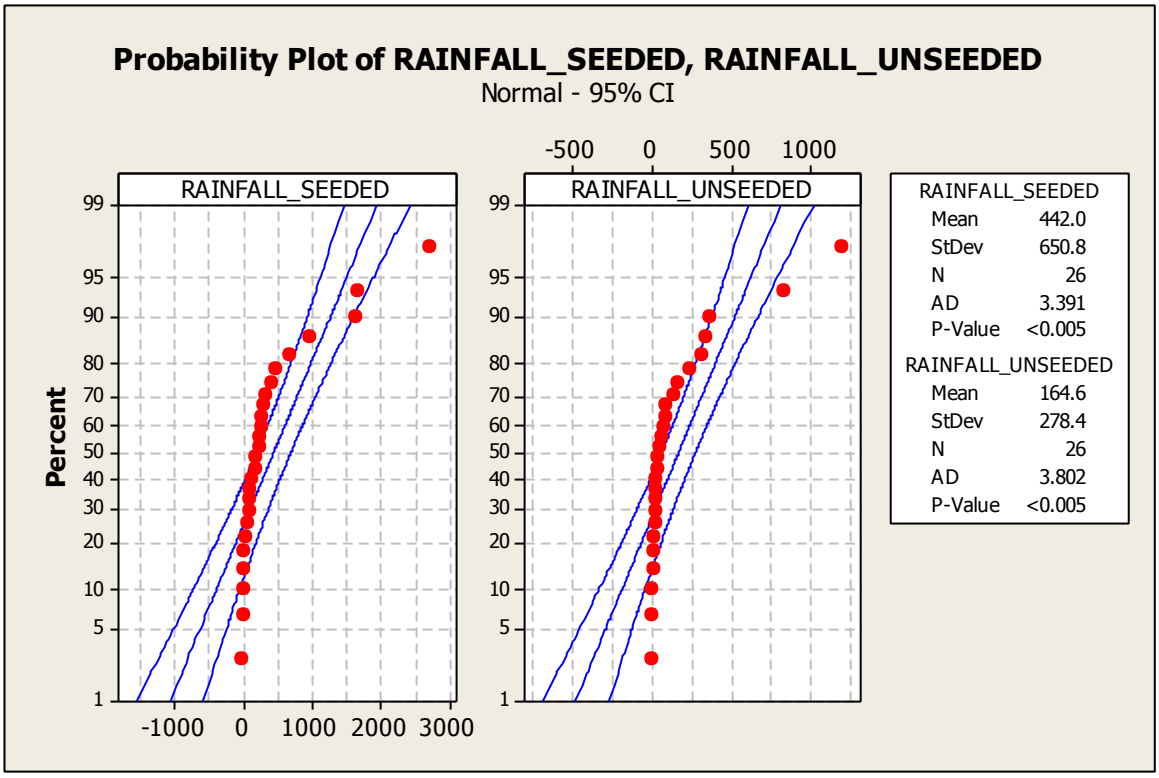
Variable	TREATMENT	Q3	Maximum
RAINFALL	SEEDED	445	2746
	UNSEEDED	183.3	1202.6

Step 4: Some graphs:
Click on Graph→Boxplot→Select Multiple Y's Simple option; click on C3 and C4 to select them as variables; click OK, and see the following plot:



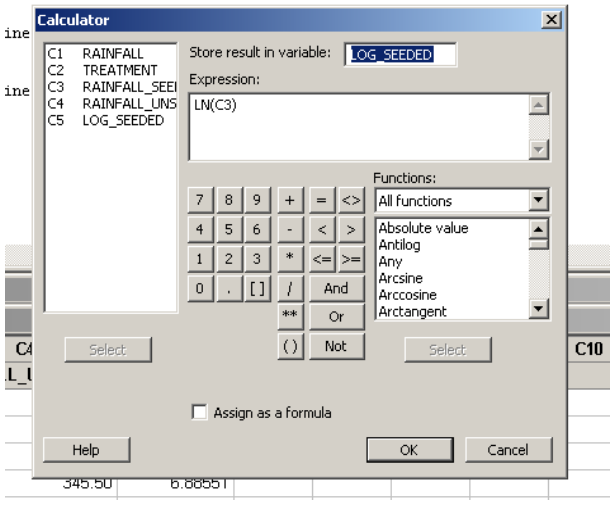
Step 5: Normal Probability Plot

Go to Graphs → Probability Plot → Single→Multiple Graphs→In separate panels of the same graph; click OK. Select the variables C3 and C4 from the left into the right hand side box; and click OK.



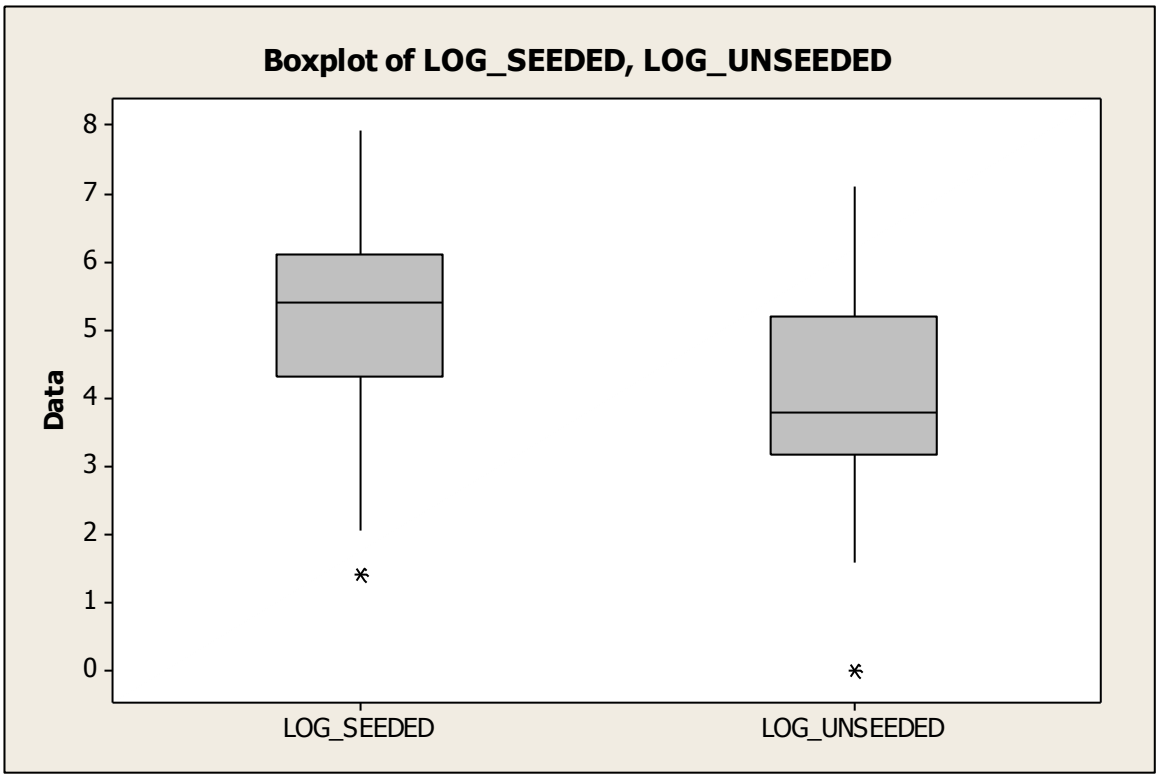
Step 6: Natural Log transform of the two samples.

Go to Calc→ Calculator→ Store Result in Variable C5; Under Expression, select Natural log (log base e); it will bring up LN( ) in the Expression box. Insert C3 within parenthesis, so we have LN(C3) and click ok. This will create a new column C5; you can name it – I called it LOG_SEEDED. Repeat these steps to create a new Column 6: LOG_UNSEEDED.

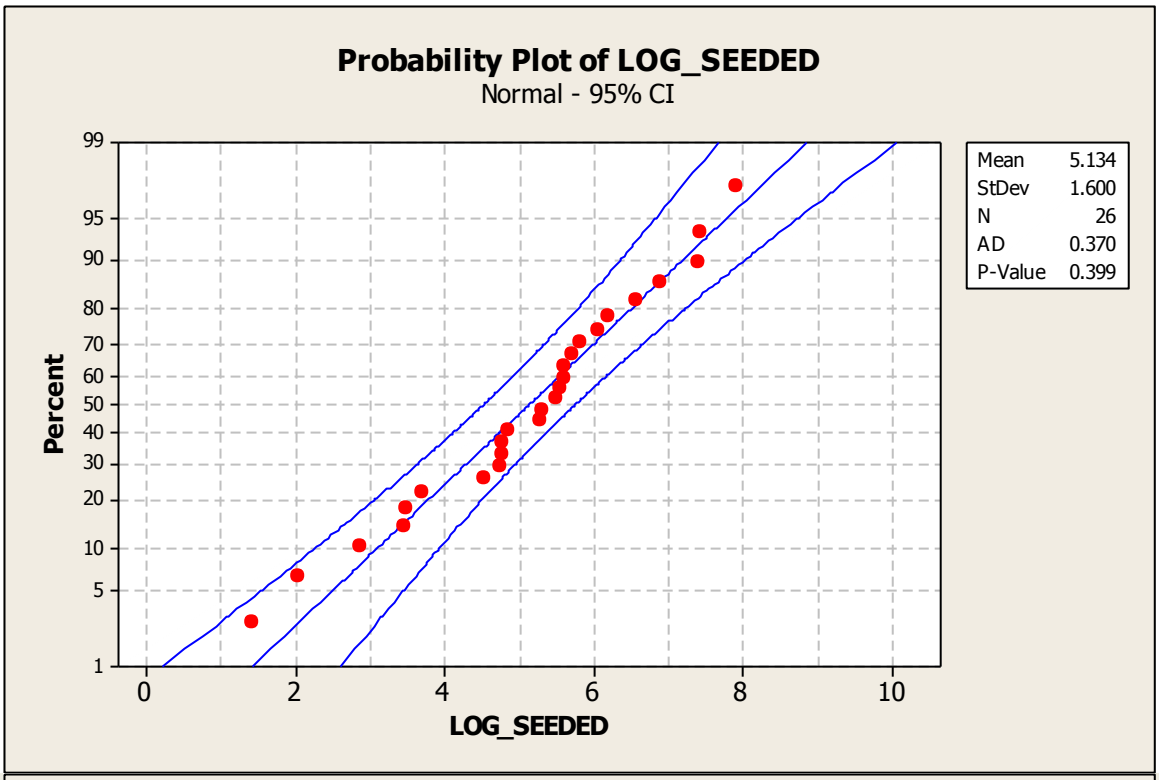


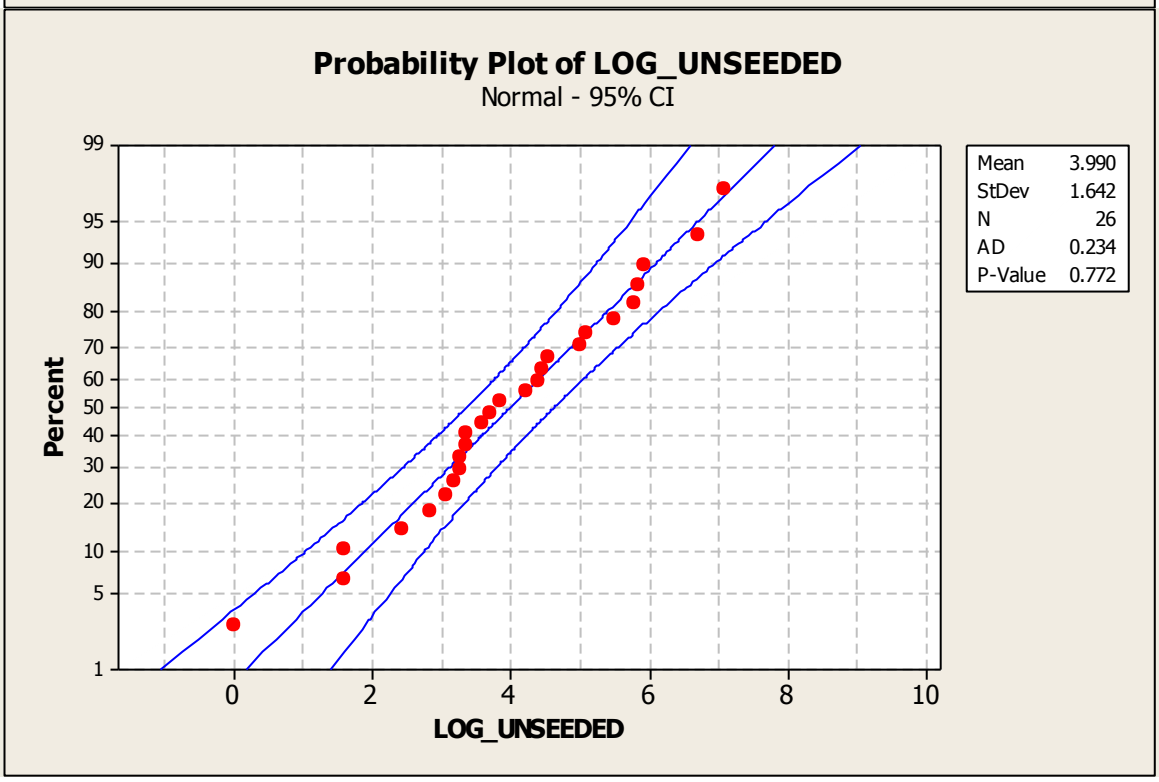
Row	RAINFALL_SEEDED	LOG_SEEDED	RAINFALL_UNSEEDED	LOG_UNSEEDED
1	2745.60	7.91775	1202.60	7.09224
2	1697.80	7.43709	830.10	6.72155
3	1656.00	7.41216	372.40	5.91997
4	978.00	6.88551	345.50	5.84499
5	703.40	6.55593	321.20	5.77206
6	489.10	6.19257	244.30	5.49840
7	430.00	6.06379	163.00	5.09375
8	334.10	5.81144	147.80	4.99586
9	302.80	5.71307	95.00	4.55388
10	274.70	5.61568	87.00	4.46591
11	274.70	5.61568	81.20	4.39692
12	255.00	5.54126	68.50	4.22683
13	242.50	5.49100	47.30	3.85651
14	200.70	5.30181	41.10	3.71601
15	198.60	5.29129	36.60	3.60005
16	129.60	4.86445	29.00	3.36730
17	119.00	4.77912	28.60	3.35341
18	118.30	4.77322	26.30	3.26957
19	115.30	4.74754	26.10	3.26194
20	92.40	4.52613	24.40	3.19458
21	40.60	3.70377	21.70	3.07731
22	32.70	3.48738	17.30	2.85071
23	31.40	3.44681	11.50	2.44235
24	17.50	2.86220	4.90	1.58924
25	7.70	2.04122	4.90	1.58924
26	4.10	1.41099	1.00	0.00000

Click on Graph→Boxplot→Select Multiple Y's Simple option; click on C5 and C6 to select them as variables; click OK, and see the following plot:

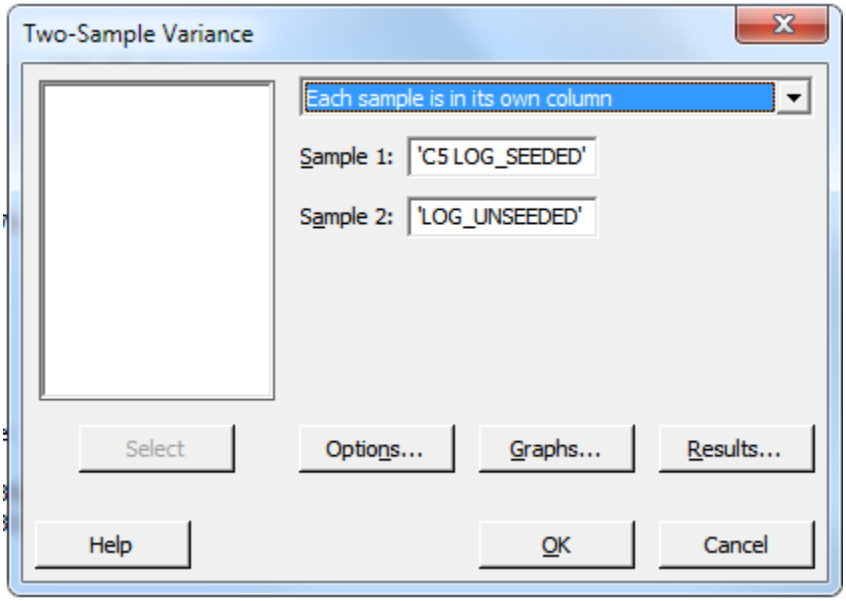


Go to Graphs → Probability Plot → Single→Multiple Graphs→In separate panels of the same graph; click OK. Select the variables C5 and C6 from the left into the right hand side box; and click OK.


Probability Plot of LOG_SEEDED
Normal - 95% CI
Percent
LOG_SEEDED
Mean 5.134
StDev 1.600
N 26
AD 0.370
P-Value 0.399


Probability Plot of LOG_UNSEEDED
Normal - 95% CI
Percent
LOG_UNSEEDED
Mean 3.990
StDev 1.642
N 26
AD 0.234
P-Value 0.772

Go to Stat →Basic Statistics →2 Variances; select two variables into the box on the right and click OK.



Test and CI for Two Variances: C5 LOG_SEEDED, LOG_UNSEEDED

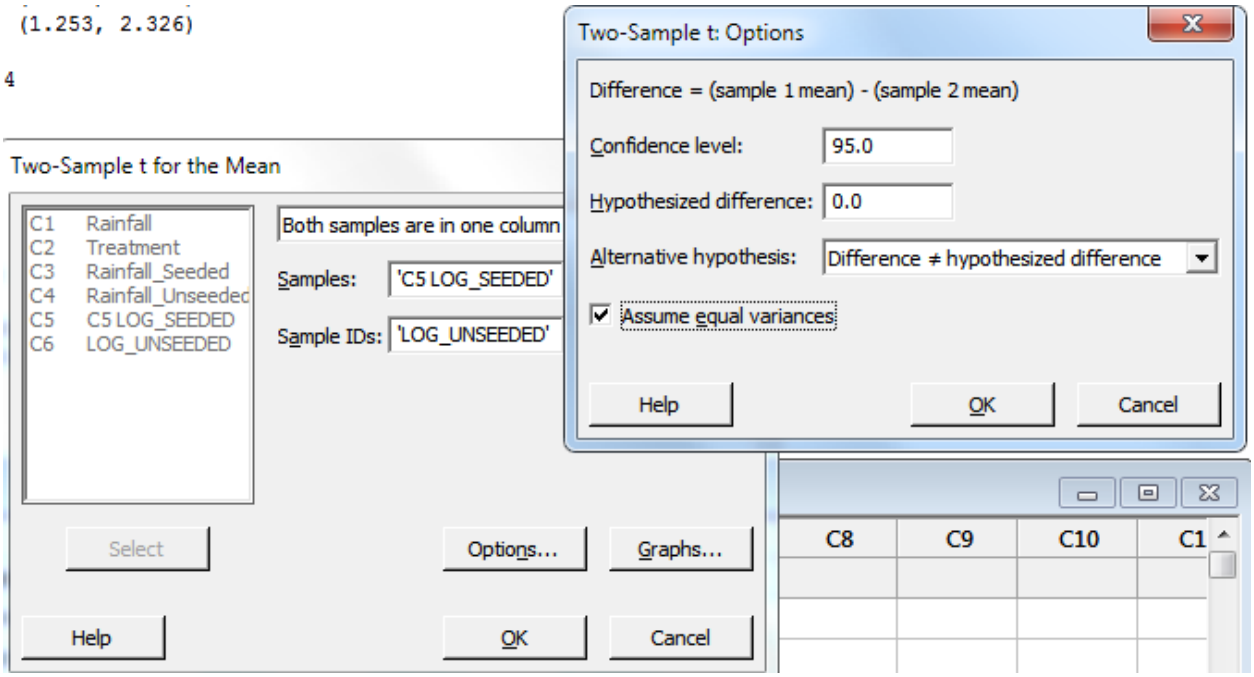
```
Method

Null hypothesis         σ(C5 LOG_SEEDED) / σ(LOG_UNSEEDED) = 1
Alternative hypothesis  σ(C5 LOG_SEEDED) / σ(LOG_UNSEEDED) ≠ 1
Significance level      α = 0.05


Statistics

                                          95% CI for
Variable        N  StDev  Variance        StDevs
C5 LOG_SEEDED  26  1.600     2.558  (1.217, 2.273)
LOG_UNSEEDED   26  1.642     2.696  (1.253, 2.326)

Ratio of standard deviations = 0.974
Ratio of variances = 0.949


95% Confidence Intervals

                              CI for
         CI for StDev       Variance
Method       Ratio            Ratio
Bonett  (0.617, 1.534)  (0.380, 2.352)
Levene  (0.560, 1.521)  (0.314, 2.315)


Tests

                        Test
Method  DF1  DF2  Statistic  P-Value
Bonett    1    -       0.02    0.900
Levene    1   50       0.08    0.781
```

Two-sample t-test : we are doing the pooled t-test here **(needs normality and equal variances assumptions to be valid)**

Go to Stat →Basic Statistics →2-Sample t;

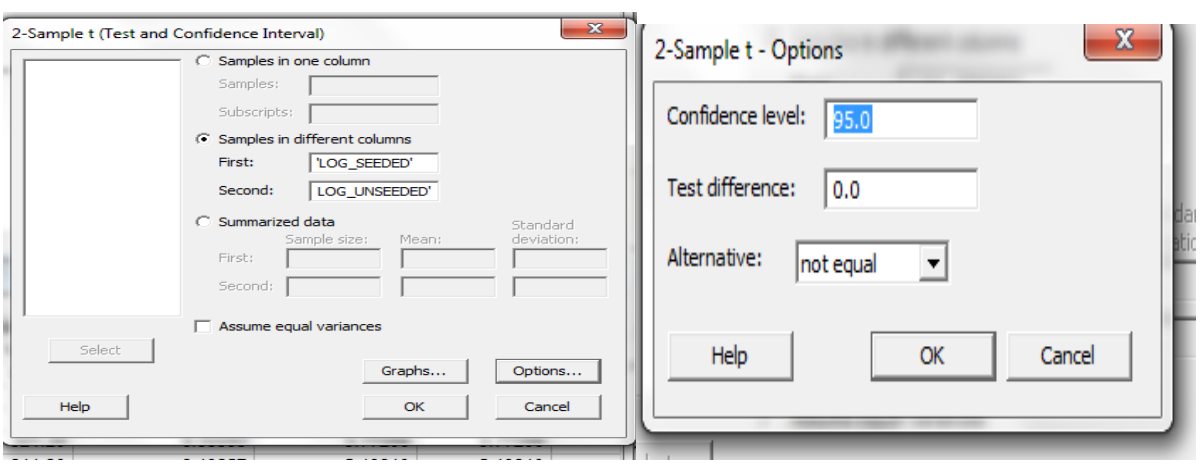
Select Samples in two different columns;

Select C5 into the window/box for the First group, and select C6 into the window/box for the Second group; select Assume equal variances, and click OK to see the following display in the Output area:




```
Two-sample T for LOG_SEEDED vs LOG_UNSEEDED

                N  Mean  StDev  SE Mean
LOG_SEEDED     26  5.13   1.60     0.31
LOG_UNSEEDED   26  3.99   1.64     0.32


Difference = mu (LOG_SEEDED) - mu (LOG_UNSEEDED)
Estimate for difference:  1.144
95% CI for difference:  (0.241, 2.047)
T-Test of difference = 0 (vs not =): T-Value = 2.54  P-Value = 0.014  DF = 50
Both use Pooled StDev = 1.6208
```

Welch's t-test



This does not use the Pooled estimate of the standard deviation. These results will differ from results from the 2-sample pooled t-procedure.

Two-Sample T-Test and CI: LOG_SEEDED, LOG_UNSEEDED

```
Two-sample T for LOG_SEEDED vs LOG_UNSEEDED

               N  Mean  StDev  SE Mean
LOG_SEEDED    26  5.13   1.60     0.31
LOG_UNSEEDED  26  3.99   1.64     0.32


Difference = mu (LOG_SEEDED) - mu (LOG_UNSEEDED)
Estimate for difference:  1.144
95% CI for difference:  (0.240, 2.047)
T-Test of difference = 0 (vs not =): T-Value = 2.54  P-Value = 0.014  DF = 49
```

Case 3.1.2. Effects of Agent Orange on Troops in Vietnam – An Observational Study.

Step 1: Copy the data into a Minitab Worksheet: use these steps: File → Open Worksheet →Browse your local directory and upload the csv file case0302.csv. The data consists of 2 columns; C1 has Dioxin levels and C2 has VETERAN (Vietnam or Other)

Step 2: you can un-stack the data. To do this, go to Data → Unstack Columns →select variable C1 into the window/box labeled Unstack the Data in; select C2 into the window/box labeled Using Subscripts in; Store un-stacked data: click on After last column in use; un-tick Name the Columns containing the un-stacked data; click OK. This will create 2 new columns in your worksheet, C3 and C4, which are titled DIOXIN_OTHER and DIOXIN_VIETNAM.

Step 3: Summary Statistics for both groups:

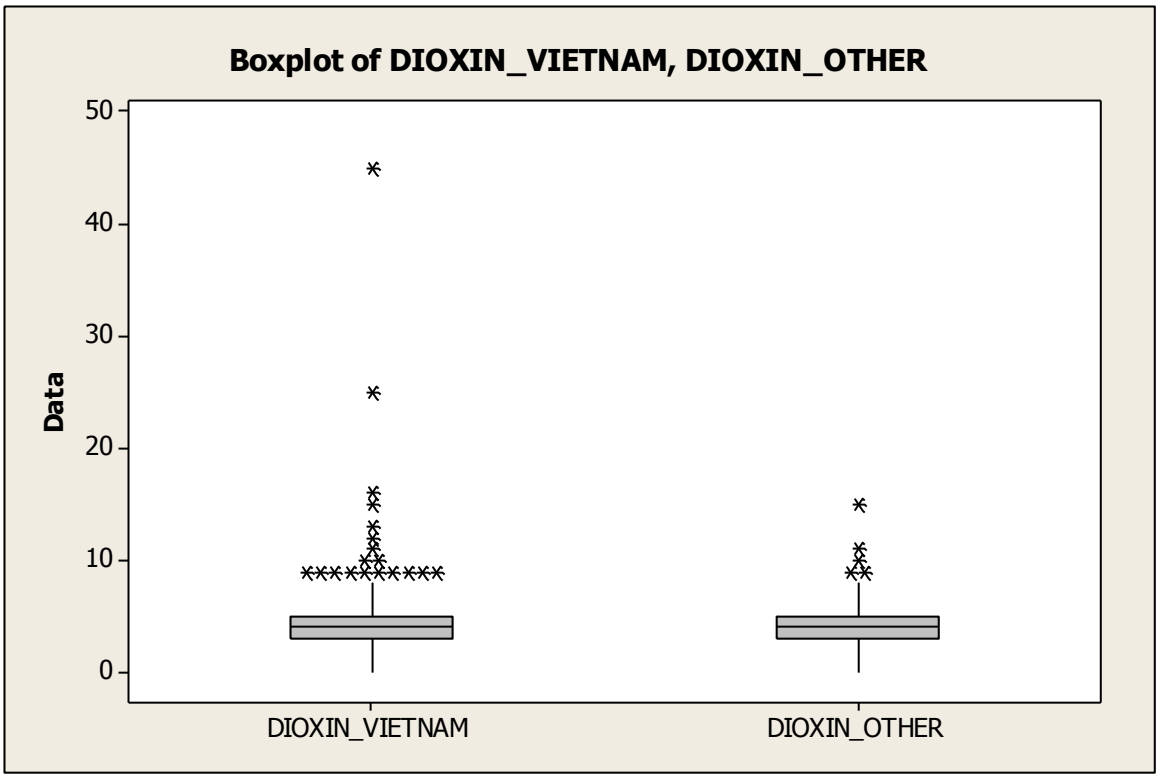
Go to Stat →Basic Statistics →Display Descriptive Statistics;
Select C3 and C4; click OK, to see this display on the output portion:

Descriptive Statistics: DIOXIN

Variable	VETERAN	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median
DIOXIN	OTHER	97	0	4.186	0.234	2.302	0.000	3.000	4.000
	VIETNAM	646	0	4.260	0.104	2.643	0.000	3.000	4.000

Variable	VETERAN	Q3	Maximum
DIOXIN	OTHER	5.000	15.000
	VIETNAM	5.000	45.000

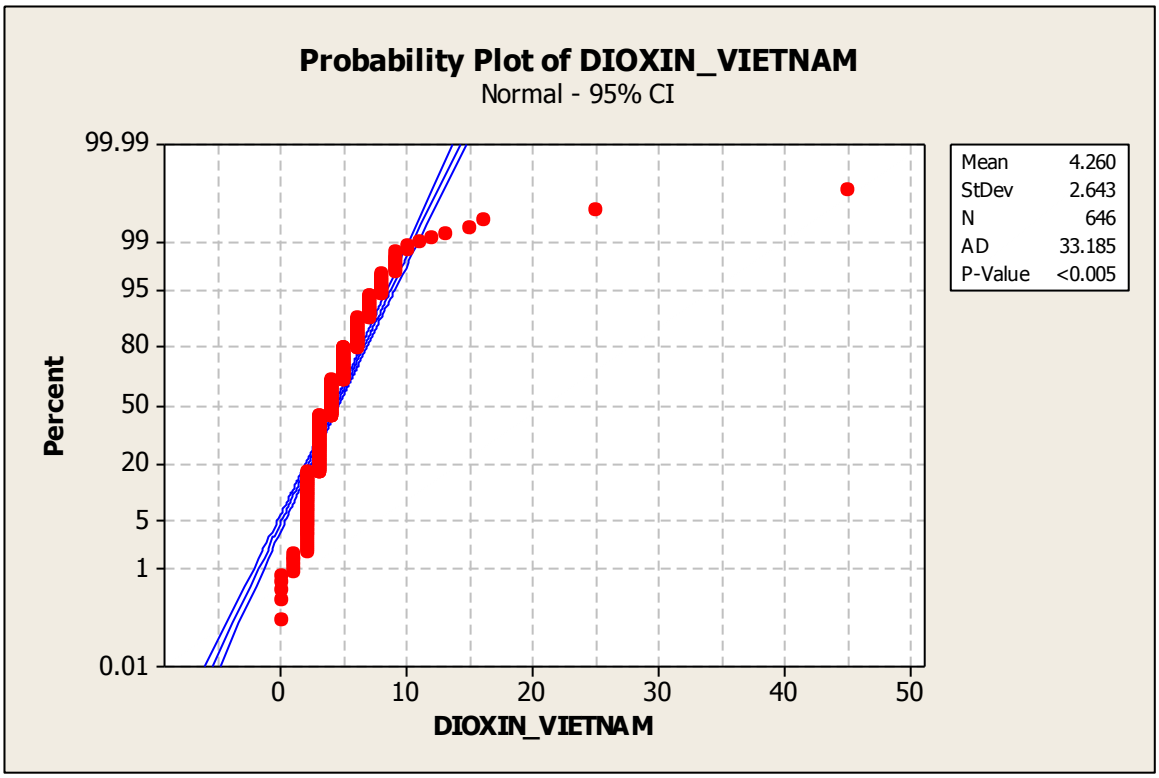
Step 5: Some graphs: Side-by-side Box-plots.
Click on Graph→Boxplot→Select One Y with Groups; Select C4 in the left window/box, and see it appear in the box Graph Variables on the right. Next, click on the box on the right, Categorical Variables for Grouping; select C3 in the left box, and see it appear in the box Categorical Variables for Grouping on the right. Click on Data View and select the following: Interquartile Range box, Outlier symbols, Median symbol, Mean symbol, Median connect line. Click OK, and see the following plot:

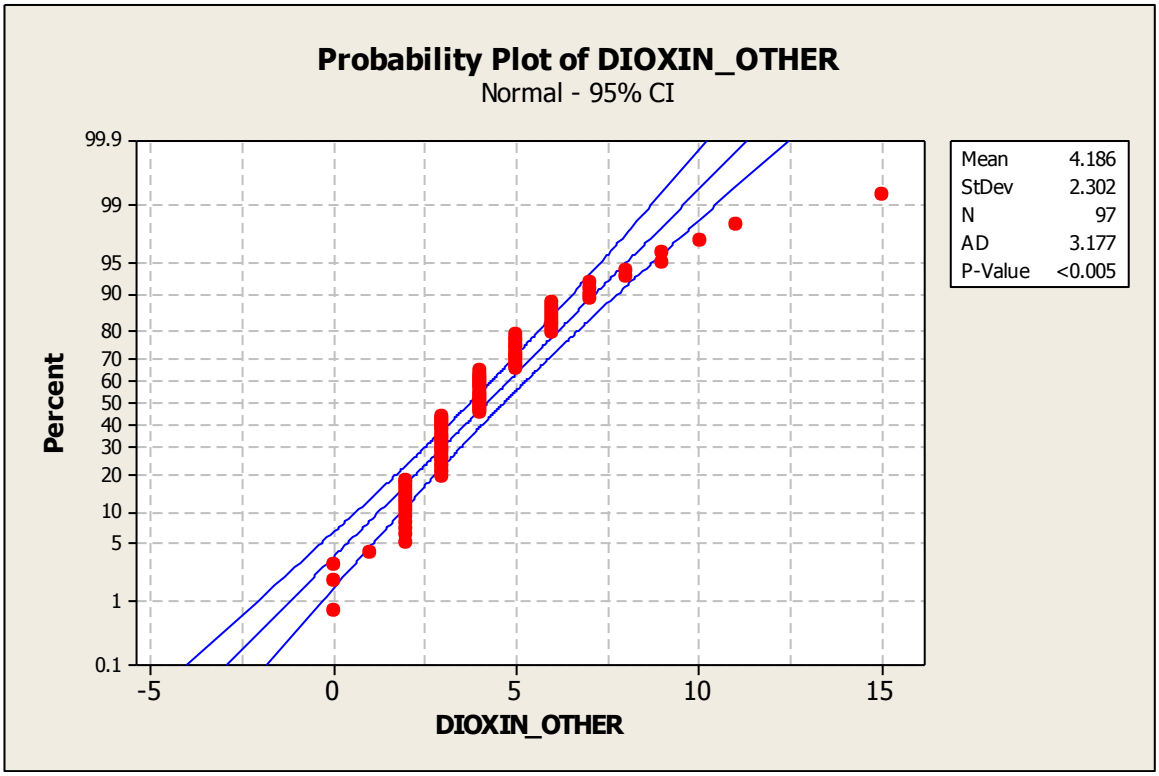


Using all the data, carry out a 2-sample pooled t-test for comparing the mean dioxin levels in the two populations, after verifying assumptions. Steps are same as shown under Case 2.1.2.

Test and CI for Two Variances: Dioxin_Other, Dioxin_Vietnam

```
Method

Null hypothesis         σ(Dioxin_Other) / σ(Dioxin_Vietnam) = 1
Alternative hypothesis  σ(Dioxin_Other) / σ(Dioxin_Vietnam) ≠ 1
Significance level      α = 0.05


Statistics

                                          95% CI for
Variable          N  StDev  Variance      StDevs
Dioxin_Other     97  2.302     5.299  (1.771, 3.054)
Dioxin_Vietnam  646  2.643     6.983  (1.811, 3.868)

Ratio of standard deviations = 0.871
Ratio of variances = 0.759


95% Confidence Intervals

                            CI for
         CI for StDev      Variance
Method       Ratio           Ratio
Bonett  (0.414, 2.668)  (0.171, 7.117)
Levene  (0.803, 1.483)  (0.645, 2.200)


Tests

                       Test
Method  DF1  DF2  Statistic  P-Value
Bonett    —    —          —    0.792
Levene    1  741       0.10    0.750
```





Two-sample Pooled t-test – Upper-tailed test

Two-Sample T-Test and CI: DIOXIN_VIETNAM, DIOXIN_OTHER

```
Two-sample T for DIOXIN_VIETNAM vs DIOXIN_OTHER

                  N  Mean  StDev  SE Mean
DIOXIN_VIETNAM  646  4.26   2.64     0.10
DIOXIN_OTHER     97  4.19   2.30     0.23


Difference = mu (DIOXIN_VIETNAM) - mu (DIOXIN_OTHER)
Estimate for difference:  0.074
95% lower bound for difference:  -0.392
T-Test of difference = 0 (vs >): T-Value = 0.26  P-Value = 0.396  DF = 741
Both use Pooled StDev = 2.6010
```

Let us repeat the analyses after removing the most extreme outlier, the observation for Vietnam veteran # 646. To do this, go to the worksheet, and make sure you wish to delete Obs 646 under C4.
Go to Data →Delete Rows; insert 646 in the window for Rows to Delete, and insert/select C4 in the window Columns from which to delete, and click OK. You will see that the worksheet now does not have the obs valued 45 for Vietnam vet #646. Repeat all the previous analyses and see if there is a marked difference in the summary statistics, plots and conclusions from the pooled t-test.

Descriptive Statistics: DIOXIN_VIETNAM, DIOXIN_OTHER

```
Variable          N  N*    Mean  SE Mean   StDev  Minimum      Q1  Median
DIOXIN_VIETNAM  645   0  4.1969   0.0827  2.1007   0.0000  3.0000  4.0000
DIOXIN_OTHER     97   0   4.186    0.234   2.302    0.000   3.000   4.000

Variable            Q3  Maximum
DIOXIN_VIETNAM  5.0000  25.0000
DIOXIN_OTHER     5.000   15.000
```

Let us repeat the analyses after removing the most extreme outlier, the observation for Vietnam veteran # 645 in addition to #646 we already deleted. To do this, go to the worksheet, and make sure you wish to delete Obs 645 under C4. Go to Data →Delete Rows; insert 646 in the window for Rows to Delete, and insert/select C4 in the window Columns from which to delete, and click OK. You will see that the worksheet now does not have the obs valued 25 for Vietnam vet #645. Repeat all the previous analyses and see if there is a marked difference in the summary statistics, plots and conclusions from the pooled t-test.

Descriptive Statistics: DIOXIN_VIETNAM, DIOXIN_OTHER

```
Variable          N  N*    Mean  SE Mean   StDev  Minimum      Q1  Median
DIOXIN_VIETNAM  644   0  4.1646   0.0763  1.9354   0.0000  3.0000  4.0000
DIOXIN_OTHER     97   0   4.186    0.234   2.302    0.000   3.000   4.000

Variable            Q3  Maximum
DIOXIN_VIETNAM  5.0000  16.0000
DIOXIN_OTHER     5.000   15.000
```