Chapter 1 - Minitab Details

Case 1.1.1. Motivation and Creativity - 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, case0101.csv. The data will appear as two columns in Minitab with titles SCORE and TREATMENT.

Step 2: Manipulate Data:

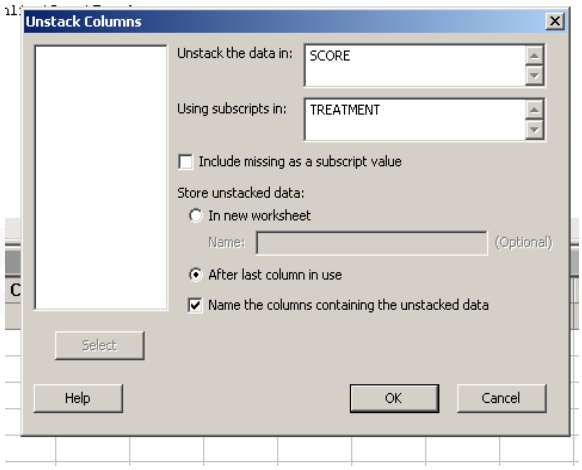
The data consists of EXTRINSIC and INTRINSIC treatment groups stacked one on top of the other. To see the data in the form shown in ***Display 1.1 in R&S p. 2***, you can un-stack the data.

To do this, go to Data → Unstack Columns →select variable C1 Score into the window/box labeled Unstack the Data in;

Select C2 TREATMEN 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 untitled.

You can go into the grey box in C3 and C4 above Row 1 and insert titles: Extrinsic group and Intrinsic group, respectively.



Step 3: Display Data similar to ***Display 1.1 in R&S p. 2***:

Go to Data →Display Data; Select C4 Intrinsic group first, and then select C3 Extrinsic group, and click OK to produce a display on the output portion of your Minitab session as shown below:

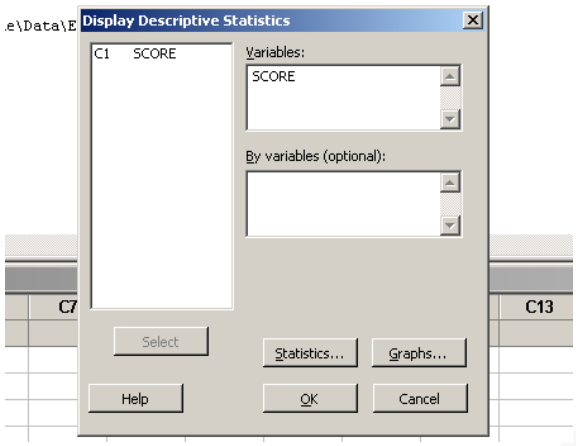
Data Display

Row	SCORE	TREATMENT	SCORE_EXTRINSIC	SCORE_INTRINSIC
1	5.00	EXTRINSIC	5.00	12.00
2	5.40	EXTRINSIC	5.40	12.00
3	6.10	EXTRINSIC	6.10	12.90
4	10.90	EXTRINSIC	10.90	13.60
5	11.80	EXTRINSIC	11.80	16.60
6	12.00	EXTRINSIC	12.00	17.20
7	12.30	EXTRINSIC	12.30	17.50
8	14.80	EXTRINSIC	14.80	18.20
9	15.00	EXTRINSIC	15.00	19.10
10	16.80	EXTRINSIC	16.80	19.30
11	17.20	EXTRINSIC	17.20	19.80
12	17.20	EXTRINSIC	17.20	20.30
13	17.40	EXTRINSIC	17.40	20.50
14	17.50	EXTRINSIC	17.50	20.60
15	18.50	EXTRINSIC	18.50	21.30
16	18.70	EXTRINSIC	18.70	21.60
17	18.70	EXTRINSIC	18.70	22.10
18	19.20	EXTRINSIC	19.20	22.20
19	19.50	EXTRINSIC	19.50	22.60
20	20.70	EXTRINSIC	20.70	23.10
21	21.20	EXTRINSIC	21.20	24.00
22	22.10	EXTRINSIC	22.10	24.30
23	24.00	EXTRINSIC	24.00	26.70
24	12.00	INTRINSIC		29.70
25	12.00	INTRINSIC		
26	12.90	INTRINSIC		
27	13.60	INTRINSIC		
28	16.60	INTRINSIC		
29	17.20	INTRINSIC		
30	17.50	INTRINSIC		
31	18.20	INTRINSIC		
32	19.10	INTRINSIC		
33	19.30	INTRINSIC		
34	19.80	INTRINSIC		
35	20.30	INTRINSIC		
36	20.50	INTRINSIC		
37	20.60	INTRINSIC		
38	21.30	INTRINSIC		
39	21.60	INTRINSIC		
40	22.10	INTRINSIC		
41	22.20	INTRINSIC		
42	22.60	INTRINSIC		
43	23.10	INTRINSIC		
44	24.00	INTRINSIC		
45	24.30	INTRINSIC		
46	26.70	INTRINSIC		
47	29.70	INTRINSIC		

Step 4: Summary Statistics for both groups:

Go to Stat →Basic Statistics →Display Descriptive Statistics;

Select C4 Intrinsic group; click OK, to see this display on the output portion:



Descriptive Statistics: SCORE_INTRINSIC

```
Variable          N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median
SCORE_INTRINSIC  24   0  19.883    0.906  4.440   12.000  17.275  20.400

Variable             Q3  Maximum
SCORE_INTRINSIC  22.500   29.700
```

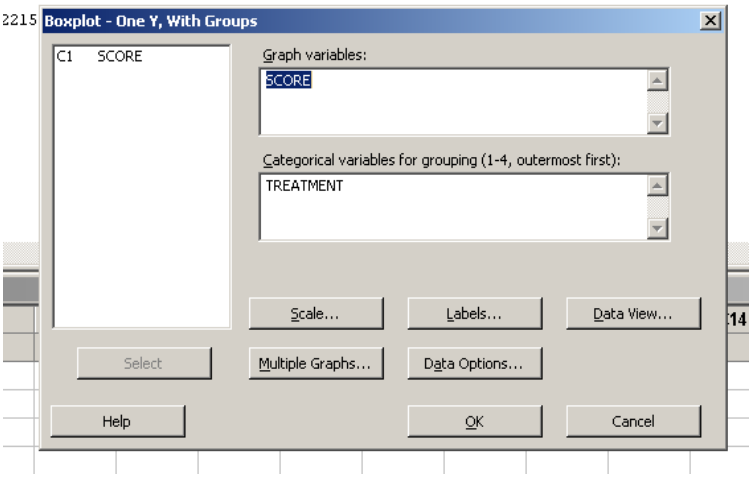
Next, repeat for the second group: Go to Stat → Basic Statistics →Display Descriptive Statistics; Select C3 Extrinsic group; click OK, to see this display on the output portion:

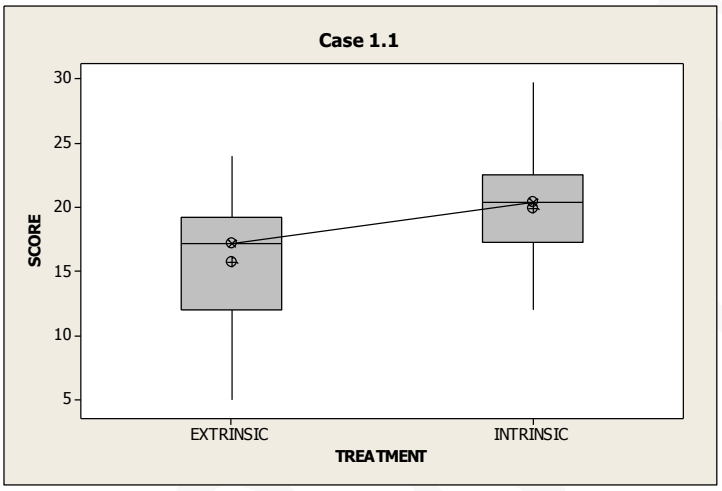
Descriptive Statistics: SCORE_EXTRINSIC

```
Variable          N  N*   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3
SCORE_EXTRINSIC  23   0  15.74     1.10   5.25     5.00  12.00   17.20  19.20

Variable         Maximum
SCORE_EXTRINSIC    24.00
```

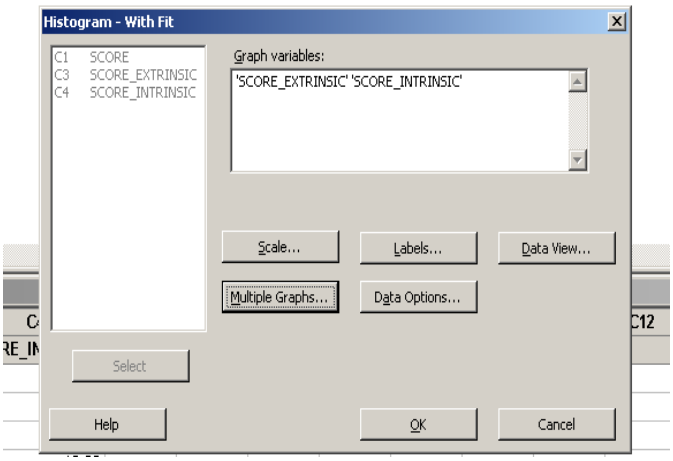
Step 5: Some graphs: Side-by-side Box-plots.

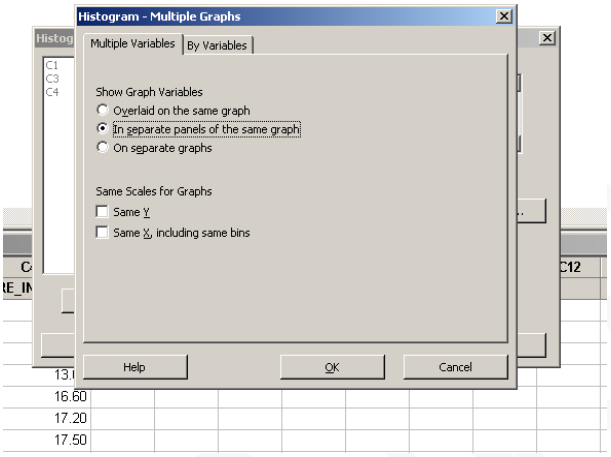
Click on Graph→Boxplot→Select One Y with Groups; Select C1 SCORE in the left window/box, and see it appear in the box Graph Variables on the right. Next, click on the box on the right, Categorical Variables for Grouping; select C2 TREATMENT in the left box, and see it appear in the box Categorical Variables 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:

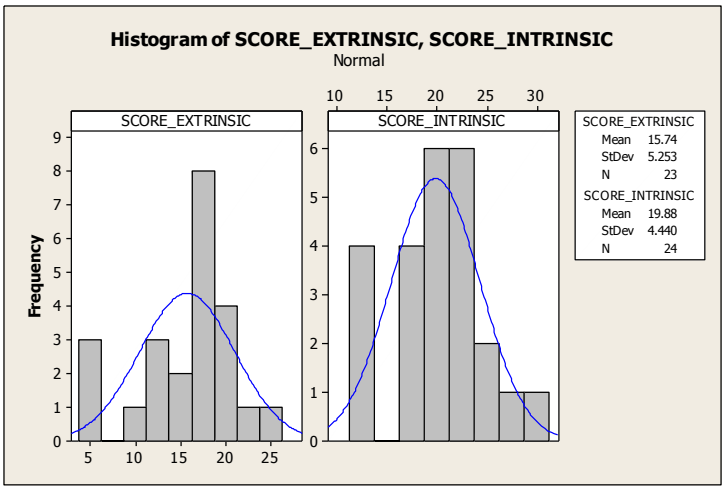




Histograms: Go to Graph→Histogram; select with fit; Select SCORE_EXTRINSIC and SCORE_INTRINSIC in the box on the left into the Graph variables box on the right; select Multiple Graphs in separate panels on the same graph; and click OK, to get this plot.


Histogram - With Fit
C1 SCORE
C3 SCORE_EXTRINSIC
C4 SCORE_INTRINSIC
Graph variables:
'SCORE_EXTRINSIC' 'SCORE_INTRINSIC'
Scale...
Labels...
Data View...
Multiple Graphs...
Data Options...
Select
Help
OK
Cancel


Histogram - Multiple Graphs
Multiple Variables
By Variables
Show Graph Variables
Overlaid on the same graph
In separate panels of the same graph
On separate graphs
Same Scales for Graphs
Same Y
Same X, including same bins
Help
OK
Cancel
16.60
17.20
17.50



Stem-and-leaf plots: Go to Graph→Stem-and-leaf ; select SCORE_EXTRINSIC and SCORE_INTRINSIC in the box on the left into the Graph variables box on the right; and click Ok to see the following:

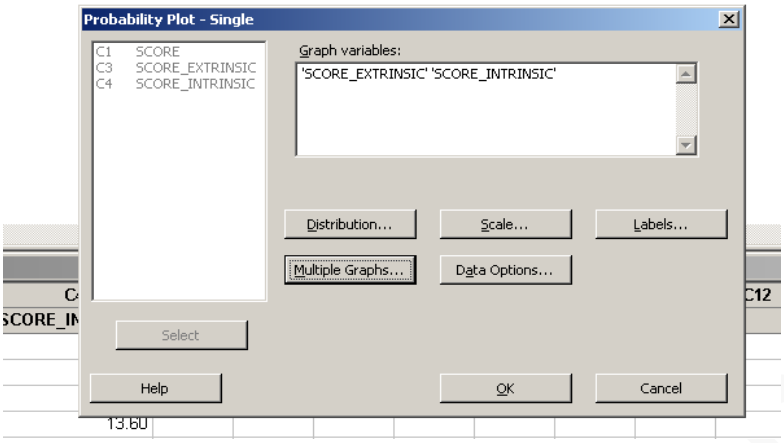
Stem-and-Leaf Display: SCORE_EXTRINSIC, SCORE_INTRINSIC

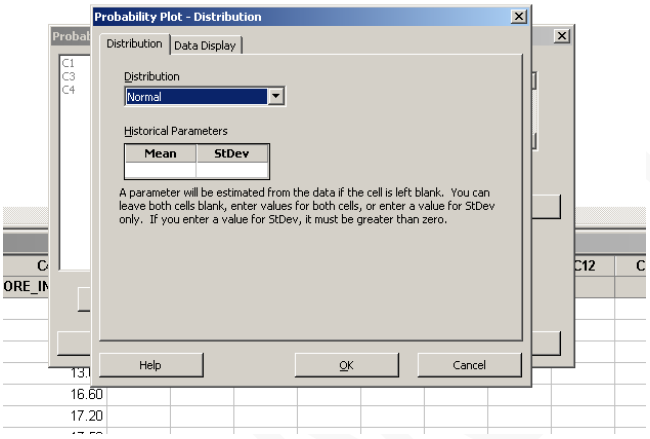
```
Stem-and-leaf of SCORE_EXTRINSIC  N  = 23
Leaf Unit = 1.0
 2   0  55
 3   0  6
 3   0
 5   1  01
 7   1  22
 9   1  45
(5)  1  67777
 9   1  88899
 4   2  01
 2   2  2
 1   2  4
```

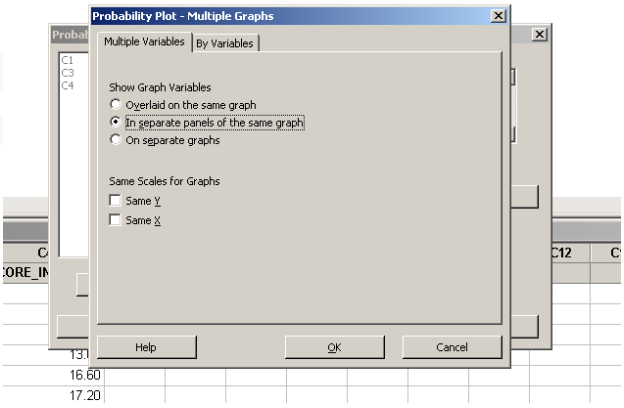
```
Stem-and-leaf of SCORE_INTRINSIC  N  = 24
Leaf Unit = 1.0
 4   1  2223
 4   1
 7   1  677
 11  1  8999
(5)  2  00011
 8   2  2223
 4   2  44
 2   2  6
 1   2  9
```

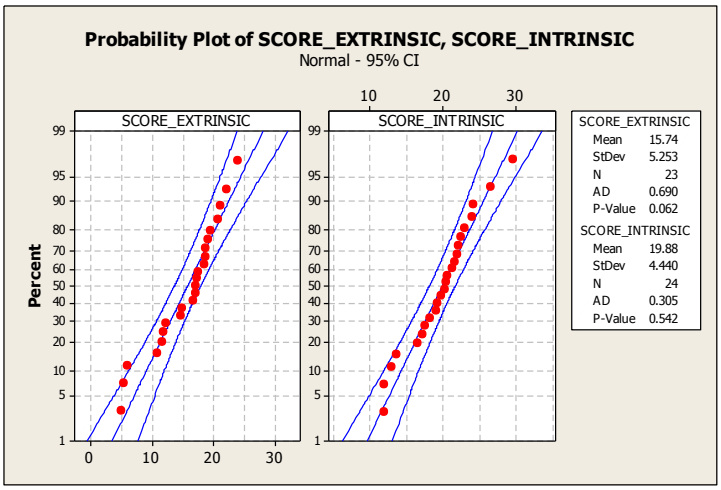
Step 6: Normal Probability Plots

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


Probability Plot - Single
C1 SCORE
C3 SCORE_EXTRINSIC
C4 SCORE_INTRINSIC
Graph variables:
'SCORE_EXTRINSIC' 'SCORE_INTRINSIC'
Distribution...
Scale...
Labels...
Multiple Graphs...
Data Options...
Select
Help
OK
Cancel

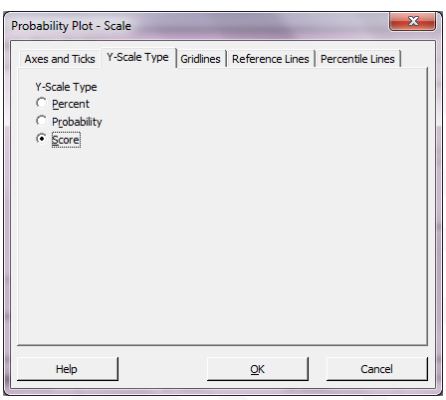

Probability Plot - Distribution
Distribution
Data Display
Distribution
Normal
Historical Parameters
Mean
StDev
A parameter will be estimated from the data if the cell is left blank. You can leave both cells blank, enter values for both cells, or enter a value for StDev only. If you enter a value for StDev, it must be greater than zero.
Help
OK
Cancel

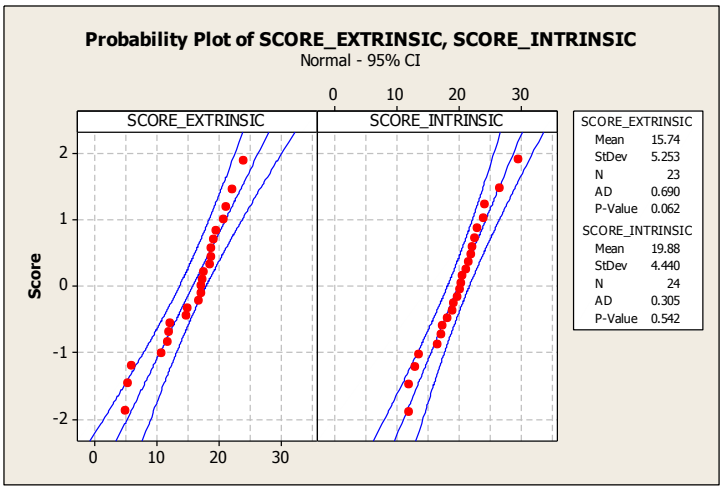

Probability Plot - Multiple Graphs
Multiple Variables
By Variables
Show Graph Variables
Overlaid on the same graph
In separate panels of the same graph
On separate graphs
Same Scales for Graphs
Same Y
Same X
Help
OK
Cancel



Note: The y-axis shows percent and values go from 0 to 1. In Minitab, this is the default option under Graphs → Probability Plot. This is a Normal Probability Plot. It is not a Quantile-Quantile or Q-Q plot.

To create a **Normal Q-Q plot**, do the following steps.
Go to Graphs → Probability Plot → Single; Select the two variables from the left into the right hand side box; select Multiple Graphs – in separate panels of the same graph and click OK.
Click on Scale, and **click on Y-Scale Type and click on Score**.



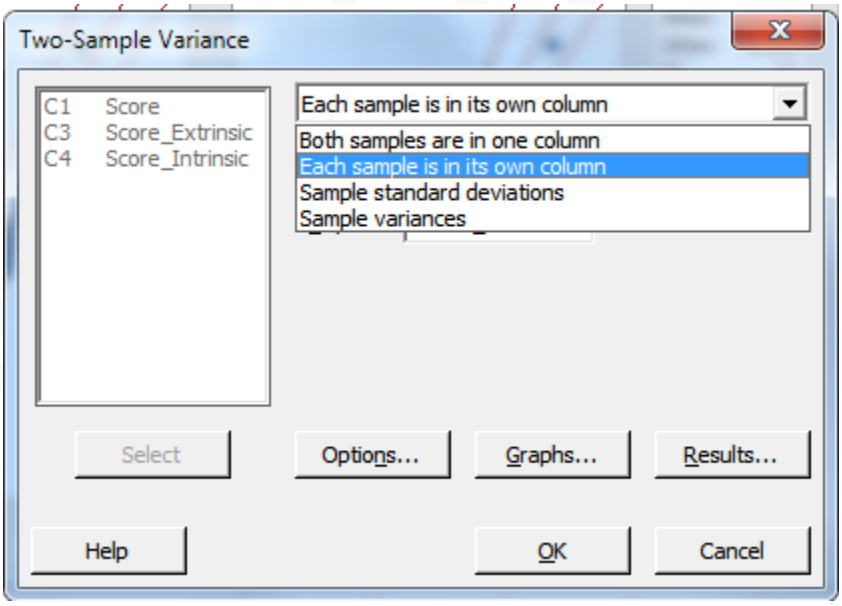


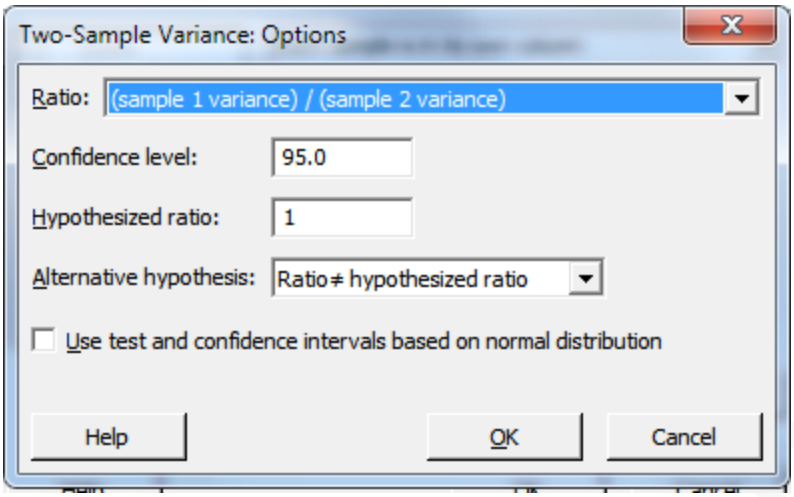
Note that the y-axis shows z-scores and not percents in this Q-Q plot.

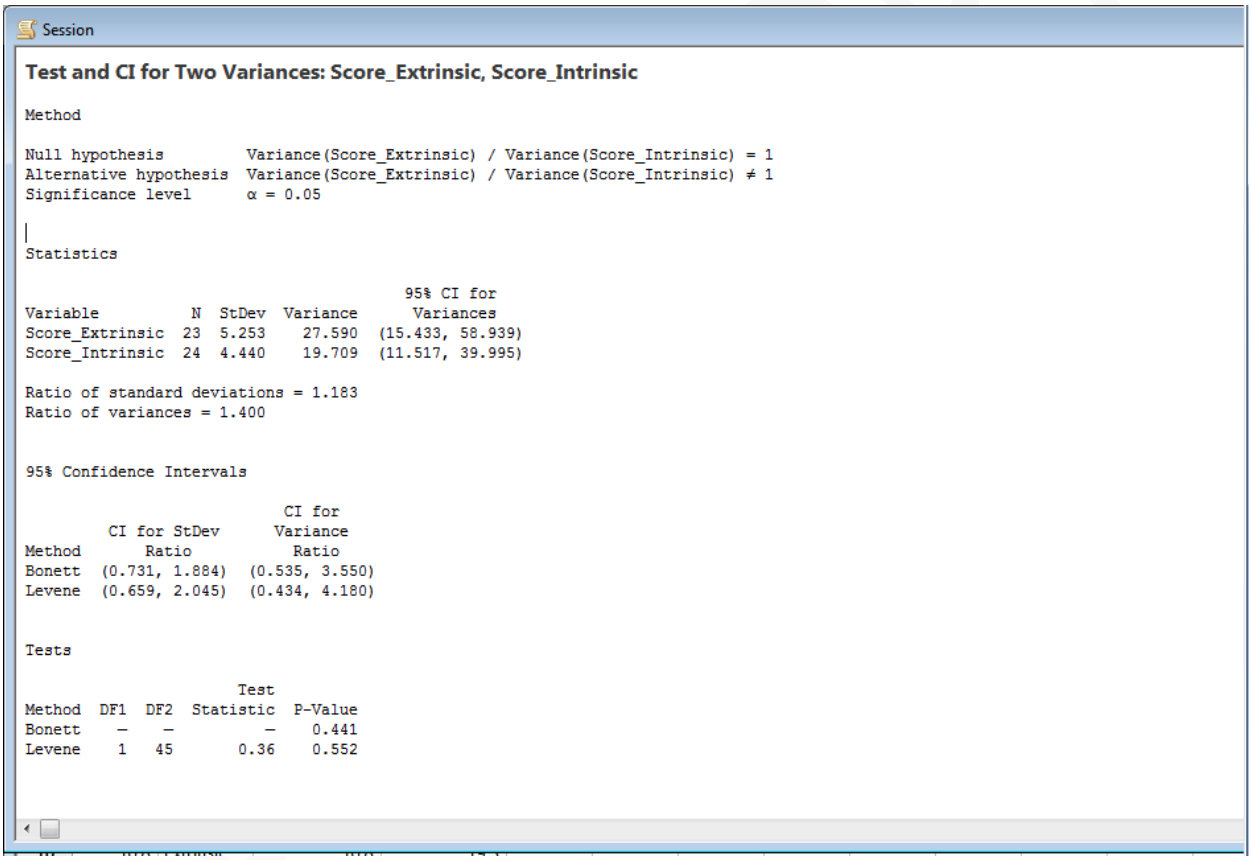
Note: The information in the Probability plot and the Q-Q plot, about whether points lie along a straight line, is similar. You can use either one to assess normality, as long as you label/understand the plot/axes correctly.

Test for equal variances (needs normality assumption on both samples to be valid)

Go to Stat →Basic Statistics →2 Variances; select two variables into the box on the right; click on Options and Hypothesized Ratio, select Variance 1/Variance 2 and click OK.







Session

Test and CI for Two Variances: Score_Extrinsic, Score_Intrinsic

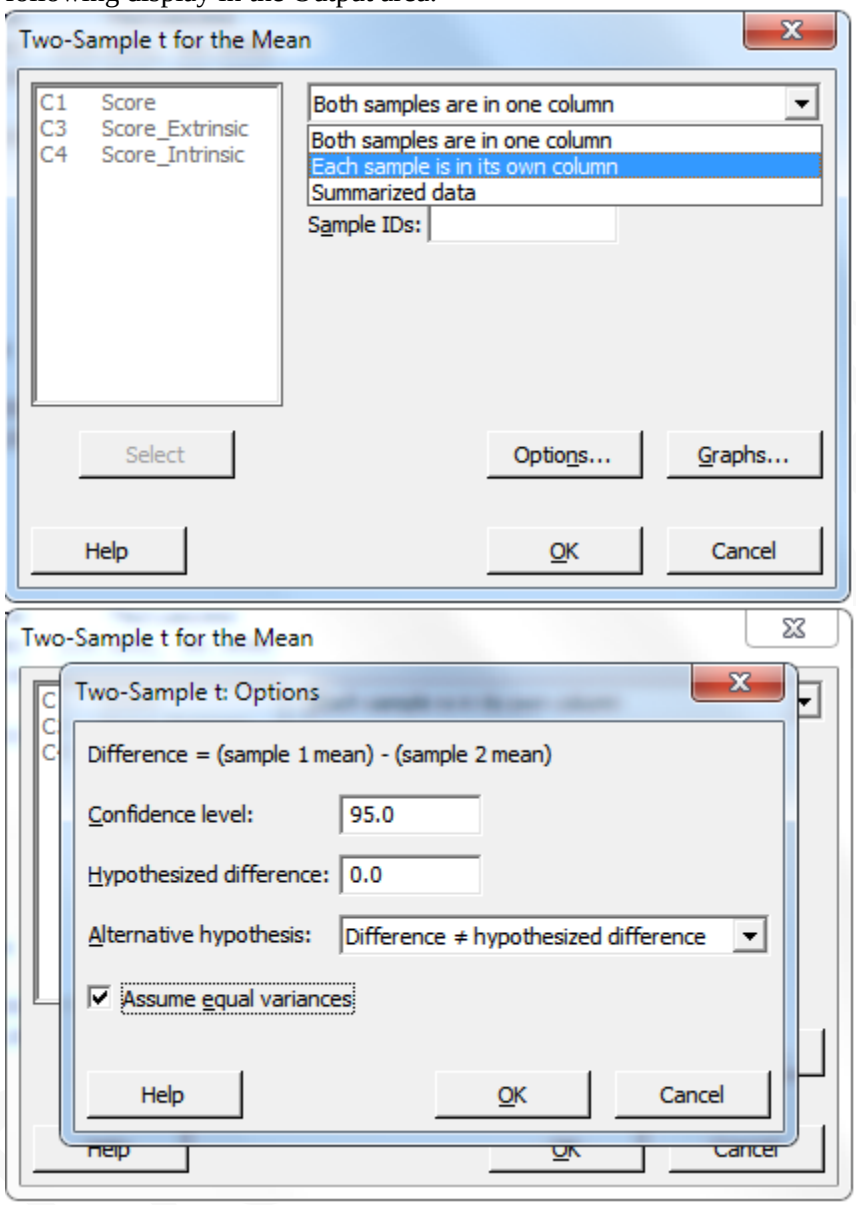
```
Method

Null hypothesis         Variance(Score_Extrinsic) / Variance(Score_Intrinsic) = 1
Alternative hypothesis  Variance(Score_Extrinsic) / Variance(Score_Intrinsic) ≠ 1
Significance level      α = 0.05

Statistics

                                                95% CI for
Variable          N  StDev  Variance            Variances
Score_Extrinsic  23  5.253    27.590  (15.433, 58.939)
Score_Intrinsic  24  4.440    19.709  (11.517, 39.995)

Ratio of standard deviations = 1.183
Ratio of variances = 1.400


95% Confidence Intervals

                                CI for
          CI for StDev        Variance
Method       Ratio             Ratio
Bonett  (0.731, 1.884)  (0.535, 3.550)
Levene  (0.659, 2.045)  (0.434, 4.180)


Tests

                          Test
Method  DF1  DF2  Statistic  P-Value
Bonett    –    –          –    0.441
Levene    1   45       0.36    0.552
```

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 C4 Intrinsic group into the window/box for the First group, and select C3 Extrinsic group 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 SCORE_INTRINSIC vs SCORE_EXTRINSIC

                  N   Mean  StDev  SE Mean
SCORE_INTRINSIC  24  19.88   4.44     0.91
SCORE_EXTRINSIC  23  15.74   5.25      1.1


Difference = mu (SCORE_INTRINSIC) - mu (SCORE_EXTRINSIC)
Estimate for difference:  4.14
95% CI for difference:  (1.29, 7.00)
T-Test of difference = 0 (vs not =): T-Value = 2.93  P-Value = 0.005  DF = 45
Both use Pooled StDev = 4.8541
```

Note: You should compare the results you obtained here with the Statistical Conclusion in ***R&S p. 3***.

Step 7: Discussion of Results and Scope of Inference:

The observed p-value of the test is 0.005, which is less than the pre-chosen level of significance $\alpha = 0.05$. Therefore, we reject the null hypothesis at the 5% level of significance.

Case 1.1.2. Sex Discrimination in Employment - 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 case0102.csv. The data consists of SALARY for FEMALES and those for MALES stacked one on top of the other.

Step 2: you can un-stack the data. To do this, go to Data → Unstack Columns →select variable C1 Salary into the window/box labeled Unstack the Data in;

Select C2 SEX 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 SALARY_FEMALE and SALARY_MALE

Step 3: Display Data

Row	SALARY_FEMALE	SALARY_MALE
1	3900	4620
2	4020	5040
3	4290	5100
4	4380	5100
5	4380	5220
6	4380	5400
7	4380	5400
8	4380	5400
9	4440	5400
10	4500	5400
11	4500	5700
12	4620	6000
13	4800	6000
14	4800	6000
15	4800	6000
16	4800	6000
17	4800	6000
18	4800	6000
19	4800	6000
20	4800	6000
21	4800	6000
22	4800	6000
23	4980	6000
24	5100	6000
25	5100	6300
26	5100	6600
27	5100	6600
28	5100	6600
29	5100	6840
30	5160	6900
31	5220	6900
32	5220	8100
33	5280	
34	5280	
35	5280	
36	5400	
37	5400	
38	5400	
39	5400	
40	5400	

41	5400
42	5400
43	5400
44	5400
45	5400
46	5400
47	5400
48	5520
49	5520
50	5580
51	5640
52	5700
53	5700
54	5700
55	5700
56	5700
57	6000
58	6000
59	6120
60	6300
61	6300

Step 4: Summary Statistics for both groups:

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

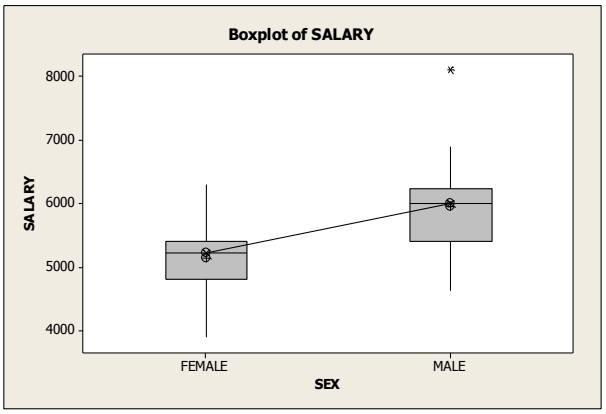
Descriptive Statistics: SALARY_FEMALE, SALARY_MALE

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
SALARY_FEMALE	61	0	5138.9	69.1	539.9	3900.0	4800.0	5220.0	5400.0
SALARY_MALE	32	0	5957	122	691	4620	5400	6000	6225

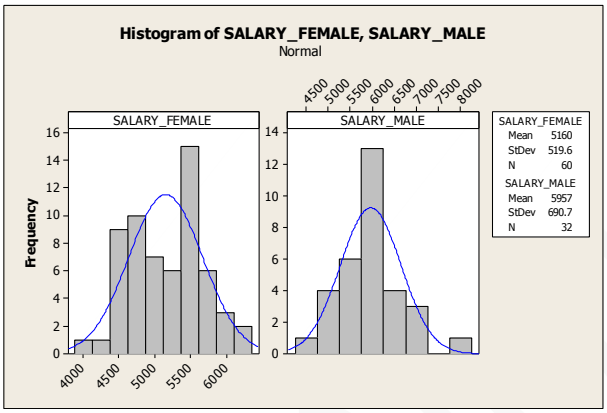
Variable	Maximum
SALARY_FEMALE	6300.0
SALARY_MALE	8100

Step 5: Some graphs: Side-by-side Box-plots.

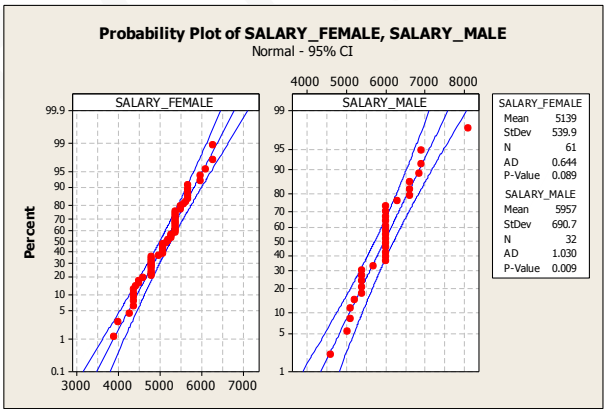
Click on Graph→Boxplot→Select One Y with Groups; Select C1 SALARY in the left window/box, and see it appear in the box Graph Variables on the right. Next, click on the box on the right, Categorical Variables for Grouping; select C2 SEX in the left box, and see it appear in the box Categorical Variables 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:



Histograms: Go to Graph→Histogram; select With fit; Select SALARY_FEMALE and SALARY_MALE in the box on the left into the Graph variables box on the right; select Multiple Graphs in separate panels on the same graph; and click OK, to get this plot.



Step 6: Two-sample t-test: we are doing the pooled t-test here. We first verify whether assumptions for this procedure are valid.



Test and CI for Two Variances: Salary_Female, Salary_Male

```
Method

Null hypothesis         σ(Salary_Female) / σ(Salary_Male) = 1
Alternative hypothesis  σ(Salary_Female) / σ(Salary_Male) ≠ 1
Significance level      α = 0.05


Statistics

Variable        N    StDev    Variance   95% CI for StDevs
Salary_Female  61  539.871  291460.328  (463.067,  650.308)
Salary_Male    32  690.733  477112.500  (499.792, 1016.906)

Ratio of standard deviations = 0.782
Ratio of variances = 0.611


95% Confidence Intervals

                            CI for
         CI for StDev      Variance
Method       Ratio           Ratio
Bonett  (0.560, 1.167)  (0.313, 1.363)
Levene  (0.624, 1.361)  (0.389, 1.853)


Tests

                       Test
Method  DF1  DF2  Statistic  P-Value
Bonett    –    –          –    0.180
Levene    1   91       0.19    0.666
```

Go to Stat →Basic Statistics →2-Sample t; Select Samples in two different columns; Select Salary_Female in the window/box for the First group, and select Salary_Male group 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-Test and CI: SALARY, SEX
Two-sample T for SALARY

SEX	N	Mean	StDev	SE Mean
FEMALE	61	5139	540	69
MALE	32	5957	691	122

Difference = mu (FEMALE) - mu (MALE)
Estimate for difference: -818
95% CI for difference: (-1076, -560)
T-Test of difference = 0 (vs not =): T-Value = -6.29 P-Value = 0.000 DF = 91
Both use Pooled StDev = 595.5707

Try this:
Go to Stat →Basic Statistics →2-Sample t; Select Samples in two different columns; Select Salary_Male in the window/box for the First group, and select Salary_Female group into the window/box for the Second group; select Assume equal variances, and click OK to see the numbers given in ***R&S***.