FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF DESIGN AND MANUFACTURING TECHNOLOGY

EXAMINATION PAPER 2016

MODULE CODE: MS5431

SEMESTER: Autumn 2016

MODULE TITLE: Quality Science 1

DURATION OF EXAM: 2.5 hours

LECTURER: Mr. Kevin O'Brien

GRADING SCHEME: 100 marks
50% of module grade

EXTERNAL EXAMINER: Prof. John Davies

INSTRUCTIONS TO CANDIDATES

Scientific calculators approved by the University of Limerick can be used.
Students must attempt 4 questions from 5.

- Question 1 is worth 40%. Each other question is worth 20%.
- Question 1 and Question 2 are compulsory.
- You must attempt any two questions from Questions 3, 4 and 5.

Question 1 - Short Questions (Compulsory)

Answer any ten of the following twelve questions. Do not attempt more than ten.

(i) (4 marks) What is meant by the sampling distribution of the mean? Provide a hypothetical example in your explanation.

(ii) (4 marks) What is a trimmed mean? In what circumstances would you use this measure in preference to the arithmetic mean?

(iii) (4 marks) What information does a 95% confidence interval for the mean give us?

(iv) (4 marks) What is a Type I error and a Type II error?

(v) (4 marks) What are the tests that you can perform when comparing two populations?

(vi) (4 marks) What are the key components that need to be identified when designing an experiment?

(vii) (4 marks) What distinguishes a factorial experiment from a completely randomised experiment or a randomised block experiment?

(viii) (4 marks) In the context of Experimental Design, what is the difference between a "between treatments" estimate and a "within treatments" estimate?

(ix) (4 marks) What is meant by multicollinearity?

(x) (4 marks) How does stepwise regression work and why would you use it?

(xi) (4 marks) In the context of analyzing categorical data, what is a Goodness of Fit test?

(xii) (4 marks) State the purpose of the Mann-Whitney U Test and the Kruskall Wallis Test. Include in your answer comparisons to the parametric counterpart to these tests.

Question 2 - Experimental Design (Compulsory)

Part A - One Way ANOVA (10 Marks)

The R&D team from a manufacturer of advanced composite materials is testing the Stiffness of a new composite material. Stiffness is a function of the percentage concentration of a particular compound within the composite material.

The R&D team investigated the Stiffness for 5 different percentage levels of the compound, A=5%, B=7.5%, C=10%, D =12.5% and E = 15%.
The data from the test was collected and is given below. Higher scores indicate higher Stiffness.

5% (A)	7.5% (B)	10% (C)	12.5% (D)	15% (E)
17.02	17.72	18.87	18.27	17.82
17.17	17.41	17.74	17.75	18.68
17.65	17.81	17.70	18.05	18.53
18.56	17.55	17.70	17.91	18.68
17.35	17.94	17.20	19.42	18.37
18.34	17.35	18.40	18.42	18.16
17.25	17.33	18.18	19.22	19.45

This data was entered into Minitab and the output on the next page was generated. Write a short report about what this analysis says about the relationship between the percentage level of the compound and the composite Stiffness. Explain your conclusions.

```
One-way ANOVA: X1, X2, X3, X4, X5

Method

Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      alpha = 0.05

Equal variances were assumed for the analysis.


Factor Information

Factor  Levels  Values
Factor       5  X1, X2, X3, X4, X5
```

Minitab Output for Question 2 Part A (Continued)

```
Analysis of Variance

Source  DF  Adj SS  Adj MS  F-Value  P-Value
Factor   4   5.449  1.3622     4.85    0.004
Error   30   8.418  0.2806
Total   34  13.866


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.529706  39.29%     31.20%      17.37%


Means

Factor  N     Mean   StDev         95% CI
X1      7   17.620   0.602  ( 17.211,  18.029)
X2      7  17.5871  0.2405  (17.1783, 17.9960)
X3      7   17.970   0.552  ( 17.561,  18.379)
X4      7   18.434   0.646  ( 18.025,  18.843)
X5      7   18.527   0.510  ( 18.118,  18.936)

Pooled StDev = 0.529706


Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Factor  N     Mean  Grouping
X5      7   18.527  A
X4      7   18.434  A B
X3      7   17.970  A B C
X1      7   17.620    B C
X2      7  17.5871      C

Means that do not share a letter are significantly different.


..............
```

Part B - Two Way ANOVA (10 Marks)

Suppose you want to determine whether the brand of cleaning product used and the temperature affects the amount of dirt removed from your machinery.
You are also interested in determining if there is an interaction between the two variables.
You buy two different brand of detergent ("`Super`" and "`Best`") and choose three different temperature levels ("`Cold`", "`Warm`", and "`Hot`"). There are four measurements per treatment group.

	Cold	Warm	Hot
Super	4,5,6,5	7,9,8,12	10,12,11,9
Best	6,6,4,4	13,15,12,12	12,13,10,13

This data was entered into Minitab and the output on this page and the next page was generated. Write a short report about what this analysis says about the relationship between the amount of dirt removed and both of the factors. Explain your conclusions.

Notation

- Detergent is Factor A.
- Temperature is Factor B.

```
General Linear Model: x versus FactA, FactB

Method

Factor coding  (-1, 0, +1)

Factor Information

Factor  Type   Levels  Values
FactA   Fixed       2  Best, Super
FactB   Fixed       3  Cold, Hot, Warm


Analysis of Variance

Source         DF  Adj SS   Adj MS  F-Value  P-Value
  FactA         1   20.17   20.167     9.81    0.006
  FactB         2  200.33  100.167    48.73    0.000
  FactA*FactB   2   16.33    8.167     3.97    0.037
Error          18   37.00    2.056
Total          23  273.83
```

Minitab Output for Question 2 Part B

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.43372  86.49%     82.73%      75.98%


Coefficients

Term            Coef  SE Coef  T-Value  P-Value   VIF
Constant       9.083    0.293    31.04    0.000
FactA
  Best         0.917    0.293     3.13    0.006  1.00
FactB
  Cold        -4.083    0.414    -9.87    0.000  1.33
  Hot          2.167    0.414     5.24    0.000  1.33
FactA*FactB
  Best Cold   -0.917    0.414    -2.21    0.040  1.33
  Best Hot    -0.167    0.414    -0.40    0.692  1.33



..............
```

Question 3 - Data Analysis

Part A - Boxplots (6 Marks)

Construct a pair of box plots for the data in the below. Construct one box plot for the data related to Material A, the other for Material B.
Comment on the features of the box plots and what conclusions, if any, you can derive from the two box plots.

Material	Sample size	Bonding Strength (Newton Metres)
A (80% pure)	12	2.0 2.1 2.1 2.1 2.2 2.3 2.3 2.3 2.4 2.4 2.5 2.6
B (60% pure)	10	1.9 1.9 2.0 2.1 2.1 2.2 2.2 2.5 2.7 2.8

Part B - Testing Distributional Assumptions (8 Marks)

Consider the results of a statistical analysis carried on both of the sample data sets X and Y. These results are presented as Minitab output on subsequent pages.

(a) (1 Mark) What sort of analysis are we carrying out?

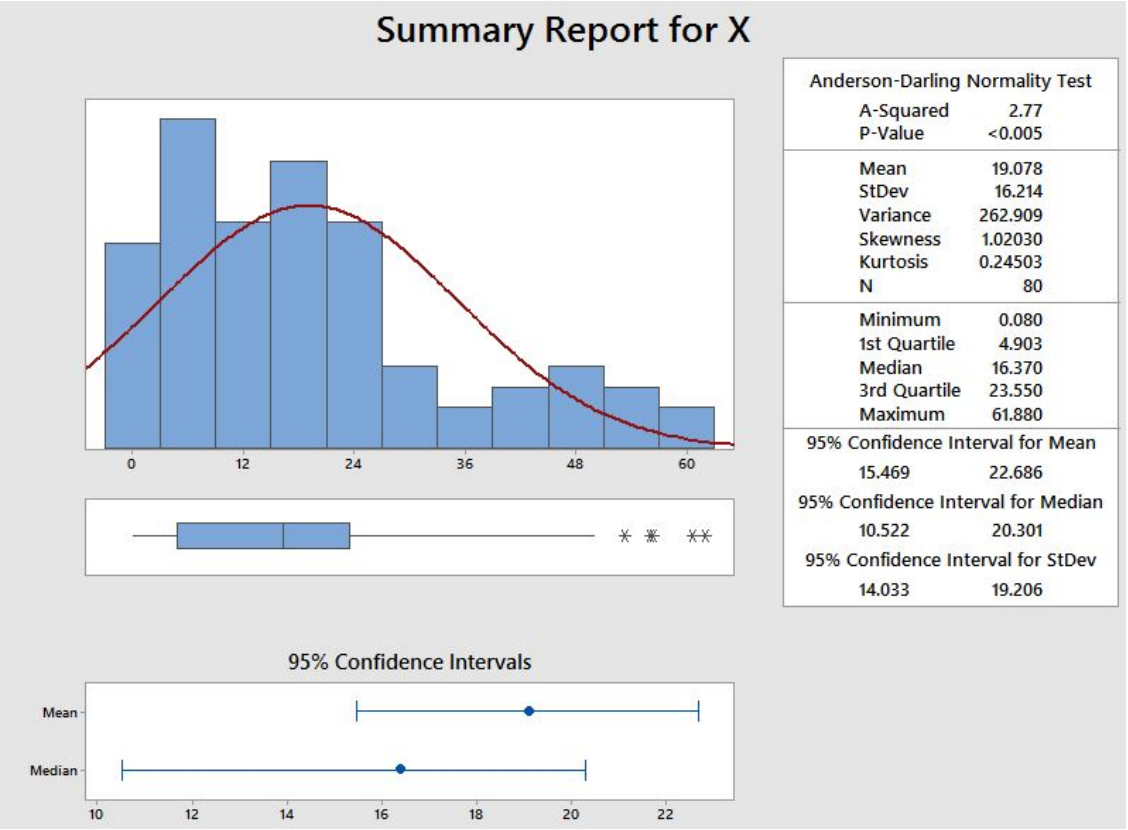
(b) (1 Mark) What is the relevance of this analysis as part of an overall statistical study.

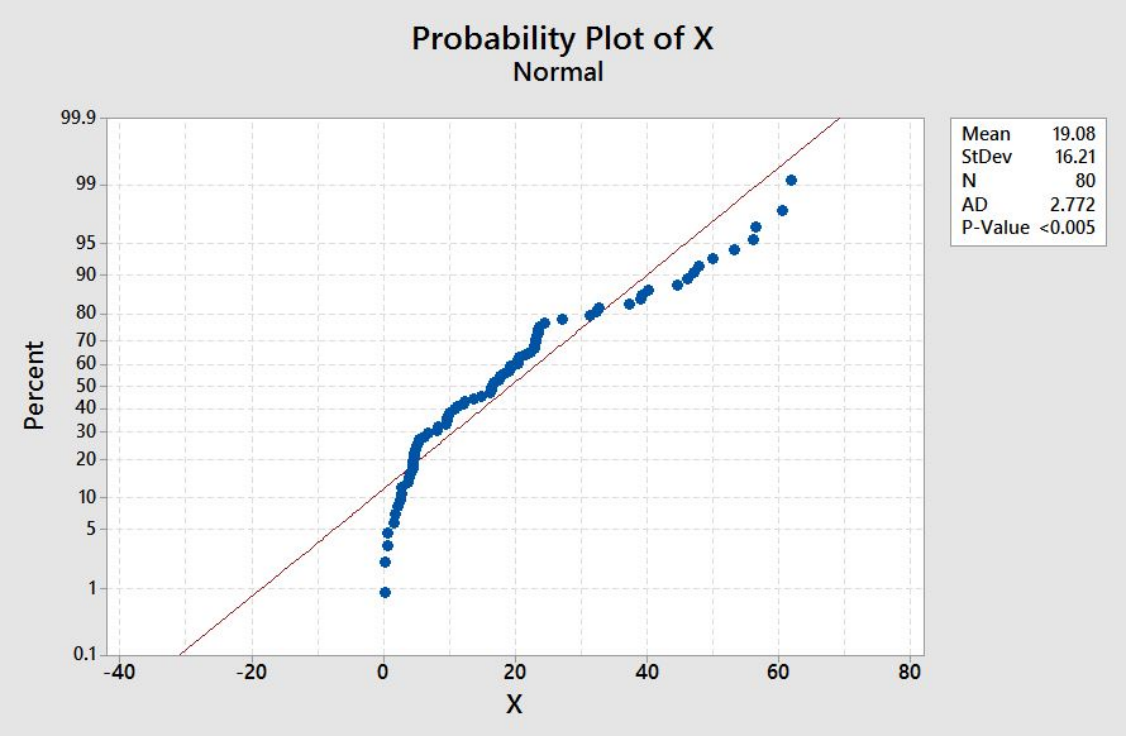
(c) (3 Marks) What is the conclusion of this analysis for the Variable X? Justify your answer with reference to 3 separate indications.

(d) (3 Marks) What is the conclusion of this analysis for the Variable Y? Justify your answer with reference to 3 separate indications.

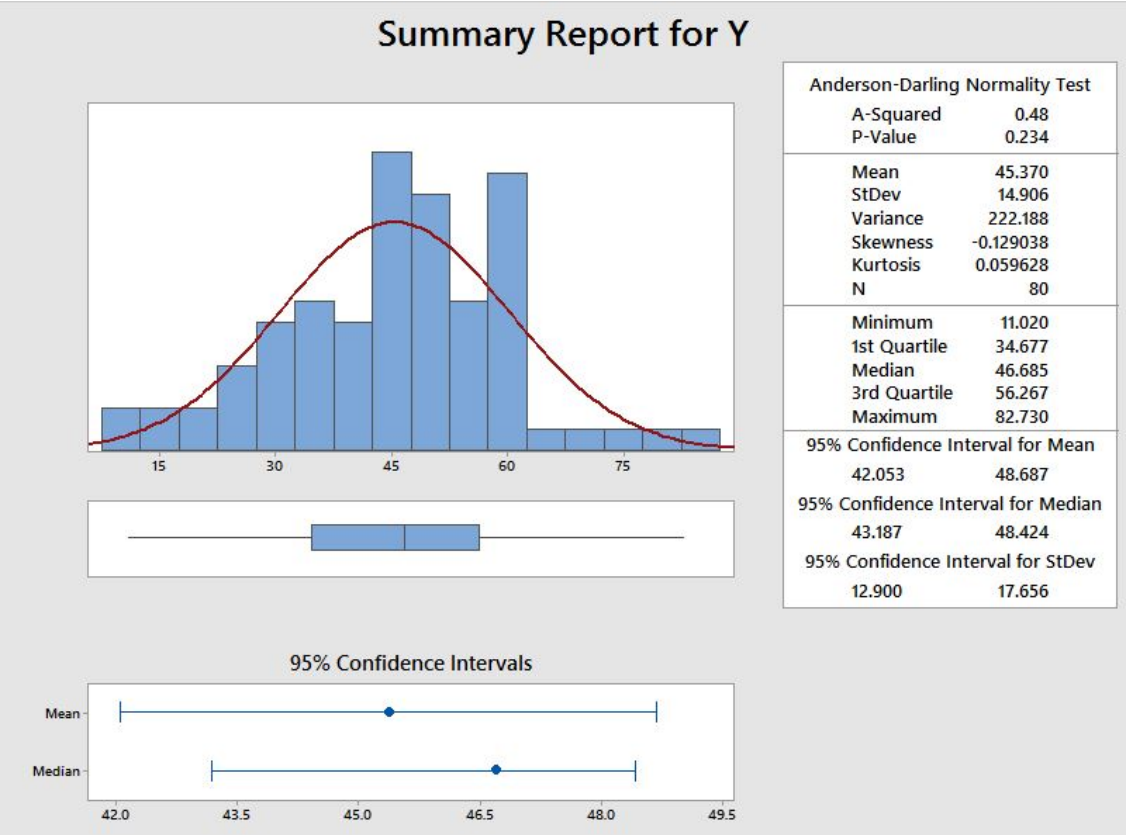
Important: *Question 3 comprises a third part: Part C. This part is presented in subsequent pages.*

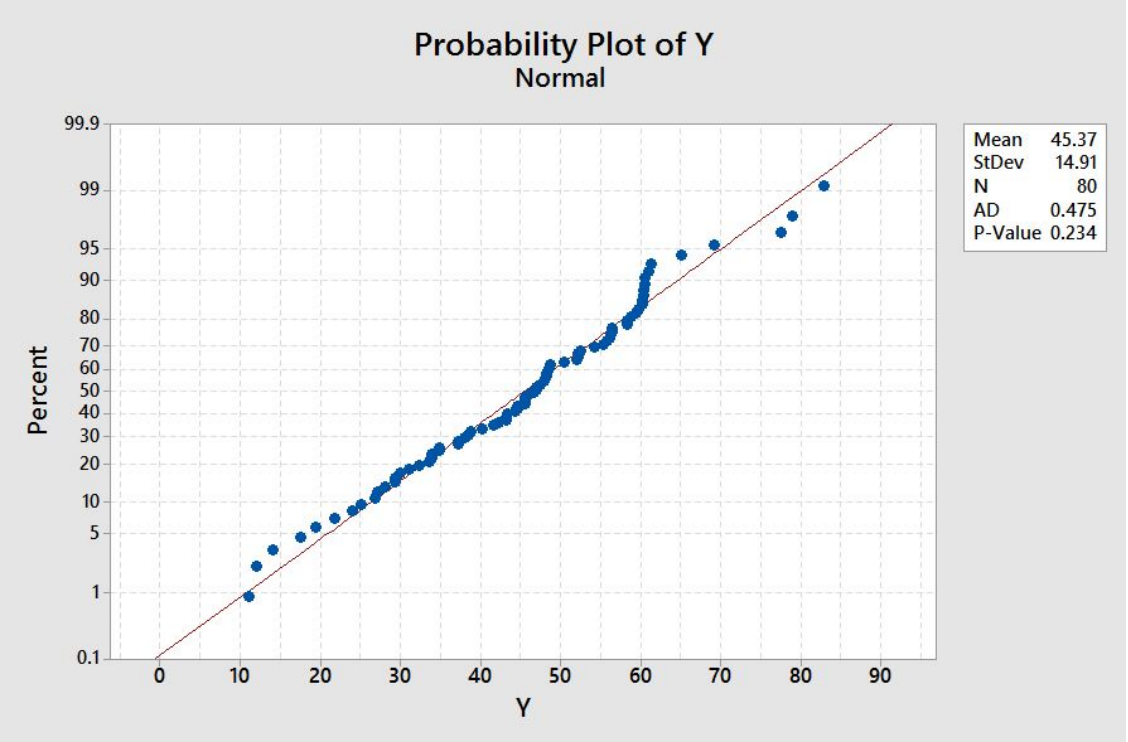
Question 3 - Part B - Minitab Output for Variable X





Question 3 - Part B - Minitab Output for Variable Y





Part C - Normal Distribution (6 Marks)

Suppose we have a manufacturing process that is designed to produce a product with a mean weight of 1002.5 grams.
The weight of the products are normally distributed, with a standard deviation of 1.25 grams.

(a) (2 Marks) What percentage of products will have a weight exceeding 1004.1 grams?

(b) (2 Marks) What percentage of products will be less than 1000g?

(c) (2 Marks) The production manager reports that less than 99% of the output is between 1000g and 1005g. Do you agree with this statement? Justify your answer with the appropriate calculations.

Question 4 - Introduction to Inference

Part A - Two Sample Mean Test (12 Marks)

Suppose a company manufactures a particular product in two facilities: Factory A and Factory B. The management wants to make sure that the Factory B, which has been recently constructed, is manufacturing components to the same specifications as Factory A.

Measurement data from both factories has been compiled and analysed, with the following Minitab output has been created.

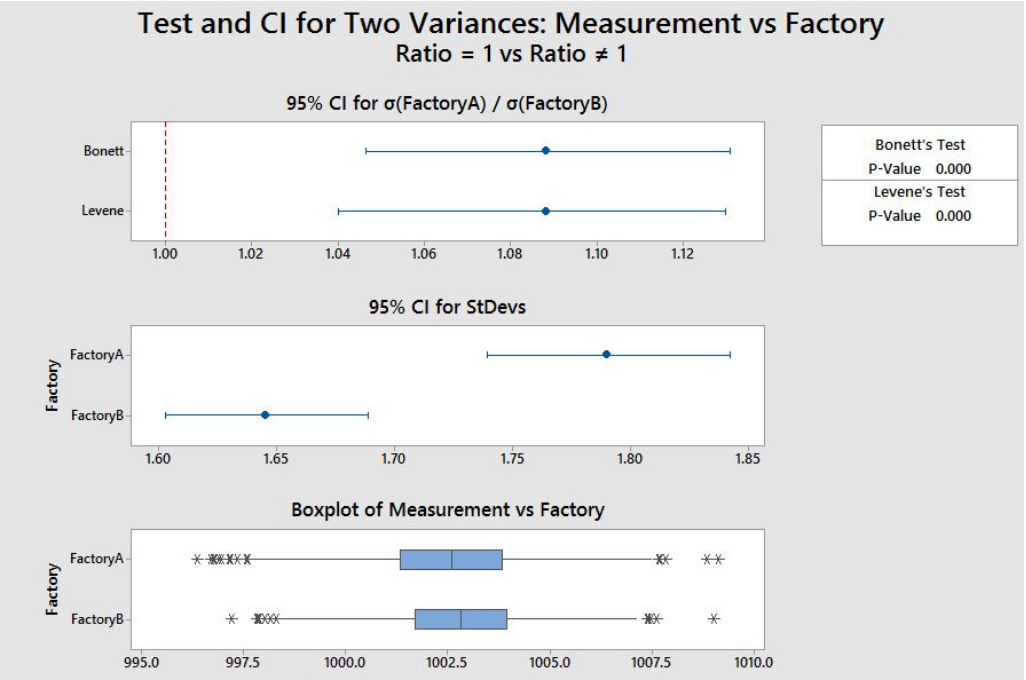
Using the printout, write a brief report on what the analysis tells you about the comparison of manufacturing processes in both factories. Explain your reasoning clearly. Formally state any null and alternative hypotheses where relevant.

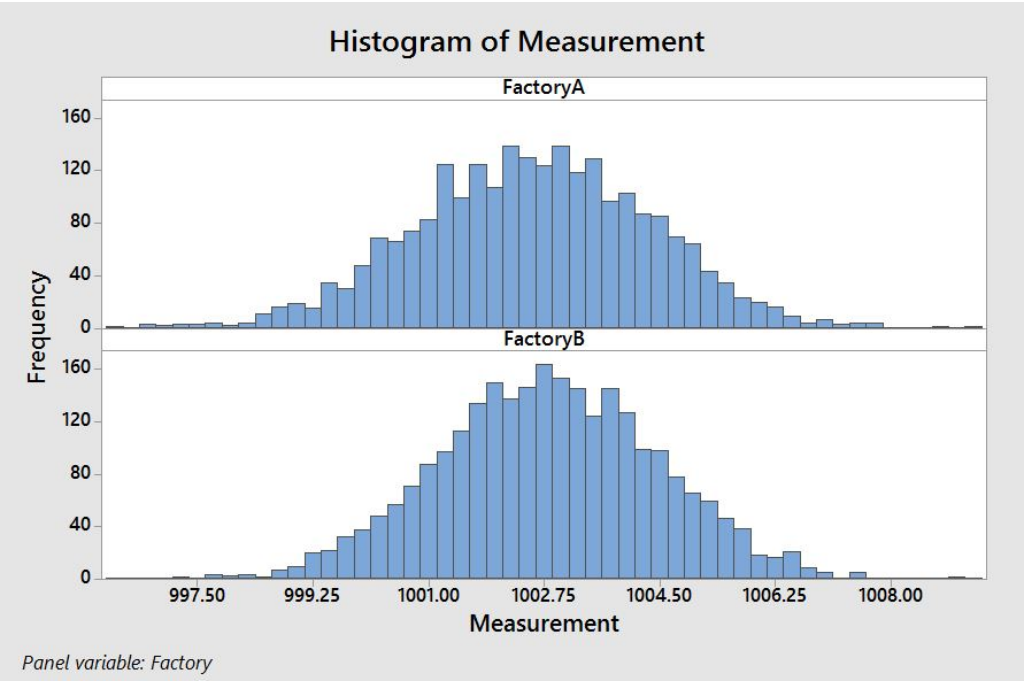
```
Two-Sample T-Test and CI: Measurement, Factory

Two-sample T for Measurement

Factory      N     Mean  StDev  SE Mean
FactoryA   240  1002.60   1.76    0.036
FactoryB   260  1002.72   1.64    0.032


Difference = mu (FactoryA) - mu (FactoryB)
Estimate for difference:  -0.1248
95% upper bound for difference:  -0.0456

T-Test of difference = 0 (vs "<0"):
T-Value = -2.59  P-Value = 0.005  DF = 498
Both use Pooled StDev = 1.7011
```

```
Test and CI for Two Variances: Measurement vs Factory

Method

Null hypothesis         s(FactoryA) / s(FactoryB) = 1
Alternative hypothesis  s(FactoryA) / s(FactoryB) not equal 1
Significance level      alpha = 0.05
```

```
Statistics

95% CI for
Factory      N  StDev  Variance      StDevs
FactoryA   240  1.789     3.202  (1.739, 1.842)
FactoryB   260  1.645     2.705  (1.603, 1.689)

Ratio of standard deviations = 1.088
Ratio of variances = 1.183


95% Confidence Intervals

CI for
CI for StDev      Variance
Method       Ratio           Ratio
Bonett  (1.046, 1.131)  (1.095, 1.279)
Levene  (1.040, 1.130)  (1.082, 1.276)


Tests

Test
Method  DF1   DF2  Statistic  P-Value
Bonett    1   498      14.58    0.000
Levene    1   498      14.58    0.000
```

Question 2 - Part B - Minitab Output for Part A





Part B - Chi Square Tests (8 Marks)

The research department of an agriculture company are compare the outcomes of pollutant side effects of a fertlizer treatment process on the local water supply.

- There are three research sites. Each of the fertilizer processes is carried out at one of three sites.
- There are 100 water quality monitors at each site. The results from each quality monitor is classified as "None", "Mild" and "Severe".

Process	None	Mild	Severe	Total
A	30	45	25	100
B	35	45	20	100
C	20	35	45	100
Total	85	125	90	300

The minitab out is tabulated below.

```
Chi-Square Test for Association: C1, Worksheet columns

Rows: C1   Columns: Worksheet columns

              None   Mild  Severe  All

Process A       30     45      25  100
              ....  41.67   30.00

Process B       35     45      20  100
             28.33  41.67   .....

Process C       20     35      45  100
             28.33  41.67   30.00

All             85    125      90  300

Cell Contents:      Count
                    Expected count


Pearson Chi-Square = 17.384, DF = ......., P-Value = 0.002
Likelihood Ratio Chi-Square = 17.094, DF = ......, P-Value = 0.002
```

(i) (2 Marks) Provide a brief description of the statistical analysis being carried out here.

(ii) (2 Marks) Two expected count values have been removed from the output. State what these values should be. Show your workings.

(iii) (1 Mark) The "degrees of freedom" value has been removed from the output. State what the degrees of freedom should be.

(iv) (3 Marks) State your conclusions about this procedure. State the null and alternative hypothesis.

Question 5 - Regression Models

Part A - Regression Analysis (8 Marks)

A wood scientist wishes to determine if there is a relationship between the number of knots in a piece of wood and its tensile strength. A random selection of 14 timber beams were analysed and the results are given in the table below. Following this is a scatter plot of the data.

Sample number	Number of knots	Tensile Strength N.MS
1	6	1.6
2	18	9.1
3	15	8.0
4	13	6.0
5	4	2.4
6	8	4.8
7	3	2.3
8	4	3.8
9	19	5.2
10	7	1.8
11	3	2.1
12	20	7.5
13	16	8.9
14	12	4.6

(i) (2 Marks) Fill in the missing values in the coefficients table of the Minitab printout.

(ii) (1 Mark) State the regression equation, as estimated by Minitab.

(iii) (3 Marks) Briefly explain what the coefficients table of the Minitab printout tells you about this relationship.

(iv) (1 Mark) Estimate the tensils strength for the case where there is 10 knots on the piece of wood.

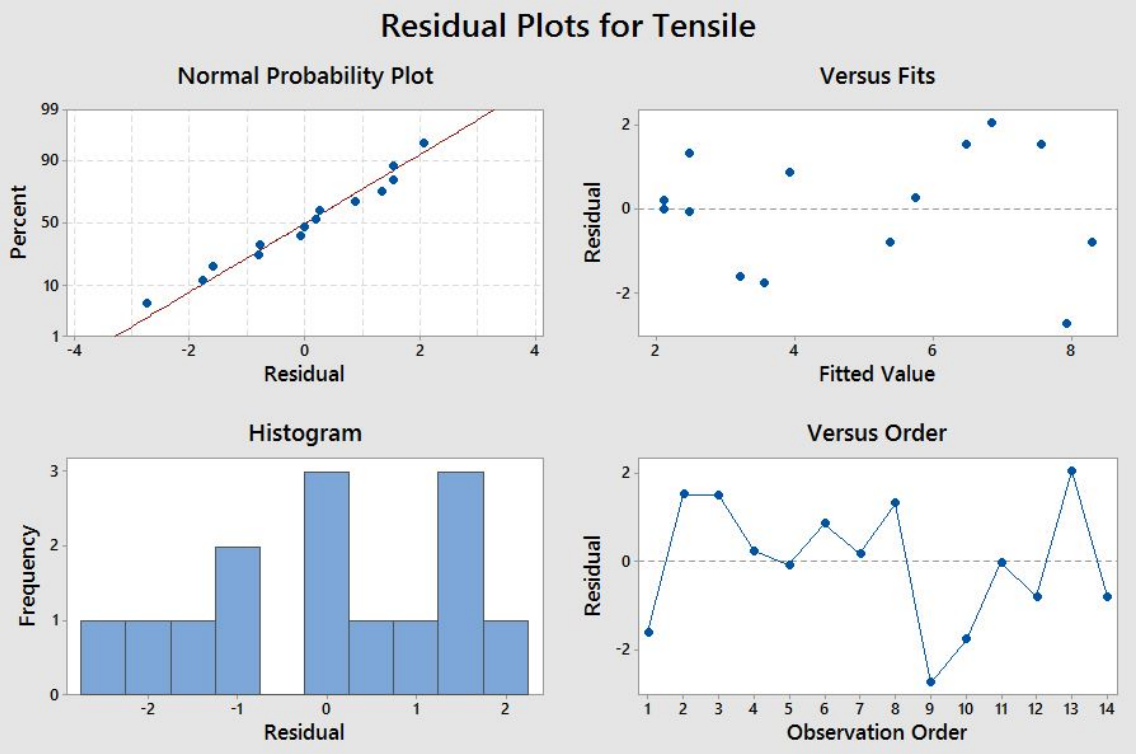
(v) (1 Mark) How good is this model, in your estimation. Refer to any relevant output in the Minitab printout.

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.46988  72.40%     70.10%      61.30%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant   1.016    0.790     ....    0.223
Knots      .....   0.0649     5.61    0.000  1.00


Regression Equation

.............................


Fits and Diagnostics for Unusual Observations

Obs  Tensile    Fit   Resid  Std Resid
  9    5.200  7.932  -2.732      -2.09  R

R  Large residual
```

Part B - Analysis of Residual Plots (4 Marks)

Minitab can provide four diagnostic plots to help you appraise the quality of a linear model. The diagnostic plots for the regression analysis described in Part A are presented below.

What is the purpose of the diagnostic plots and what does each one tell you?



Part C - Multiple Linear Regression (8 Marks)

A multinational software company that sells products in several sales regions, each assigned to a single sales representative, is interested in the variables that influence sales.

The variables are:

Sales: Total sales credited to the sales representative

Time: Length of time employed in months

Poten: Market potential; total industry sales in units for the sales region

AdvExp: Advertising expenditure in the sales region

Share: Market share; weighted average for the past 4 years

Change: Change in market share over the previous 4 years

Accounts: Number of accounts assigned to the sales representative

Work: Workload; a weighted index based on annual purchases and concentrations of accounts.

Rating: Sales representatives overall rating on eight performance dimensions; an aggregate rating on a 1-7 scale.

A random sample of 50 sales regions resulted in a data set that was then entered into Minitab and the following printout was generated.

Write a brief report analysing the printout. In your report comment on how well the model explains the variability in sales. Which variable or variables appear to be good predictors of sales? Would you refine the model in the light of these results? If so, what changes would you make?

```
Regression Analysis: Sales versus Time, Poten, AdvExp, Share,
Change, Accounts, Work, Rating

.........

Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
167.199  60.42%     52.70%      35.20%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant    2221      574     3.87    0.000
Time       18.07     8.30     2.18    0.035  1.10
Poten       2.43     2.17     1.12    0.270  1.22
AdvExp     0.266    0.199     1.34    0.189  1.16
Share     -13.35     2.51    -5.32    0.000  1.53
Change     -1.37     1.32    -1.04    0.307  1.09
Accounts   10.16     2.37     4.29    0.000  1.50
Work        2.33     2.23     1.04    0.302  1.16
Rating      60.2     12.8     4.69    0.000  1.10


Regression Equation

Sales = 2221 +18.07Time +2.43Poten +0.266AdvExp -
13.35Share +1.37Change +10.16Accounts +2.33Work +60.2Rating


Fits and Diagnostics for Unusual Observations

Obs  Sales     Fit   Resid  Std Resid
  4  680.9  1011.2  -330.3      -2.28  R
 32  825.2   534.9   290.4       2.32  R

R  Large residual
```