FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF DESIGN AND MANUFACTURING TECHNOLOGY

REPEAT EXAMINATION PAPER 2016

MODULE CODE: MS5431

SEMESTER: Autumn 2017

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 a random variable? Explain the difference between a continuous random variable and a discrete random variable.Support your answer with examples of each.

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

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

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

(v) (4 marks) What is a randomised block design?

(vi) (4 marks) Distinguish between correlation and regression when analysing the relationshop between two variables.

(vii) (4 marks) What are the key components that need to be identified when designing an 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) Compare and contrast parametric and non-parametric tests. Under what circumstances would you consider using a non-parametric 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)

Specimens of milk from dairies in four different districts are assayed for their concentrations of the radioactive isotope Strontium-90. The results, in picocuries per litre, are as shown in the table below.

District 1	District 2	District 3	District 4
17.02	17.72	18.87	18.27
17.17	17.41	17.74	17.75
17.65	17.81	17.70	18.05
18.56	17.55	17.70	17.91
17.35	17.94	17.20	19.42
18.34	17.35	18.40	18.42
17.25	17.33	18.18	19.22

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 concentration levels of the compound and each district. Explain your conclusions. Make reference to the Post-hoc procedures in the Minitab output.

Remark: *The "Group" variable corresponds to the various districts.*

One-way ANOVA: y versus Group

Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Group	4	1, 2, 3, 4

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Group	3	27.23	9.077	4.90	0.007
Error	28	51.85	1.852		
Total	31	79.08			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.36079	34.43%	27.41%	14.36%

Means

Group	N	Mean	StDev	95% CI
1	8	16.715	0.602	(15.729, 17.701)
2	8	16.219	1.799	(15.233, 17.204)
3	8	18.489	1.758	(17.503, 19.474)
4	8	18.001	0.846	(17.016, 18.987)

Pooled StDev = 1.36079

Figure 1: Minitab Out for Q2 Part A

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Group	N	Mean	Grouping	
3	8	18.489	A	
4	8	18.001	A	B
1	8	16.715	A	B
2	8	16.219		B

Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
2 - 1	-0.496	0.680	(-2.353, 1.361)	-0.73	0.884
3 - 1	1.774	0.680	(-0.083, 3.631)	2.61	0.065
4 - 1	1.286	0.680	(-0.571, 3.143)	1.89	0.255
3 - 2	2.270	0.680	(0.413, 4.127)	3.34	0.012
4 - 2	1.782	0.680	(-0.075, 3.640)	2.62	0.064
4 - 3	-0.488	0.680	(-2.345, 1.370)	-0.72	0.890

Individual confidence level = 98.92%

Figure 2: Minitab Out for Q2 Part A - Continued

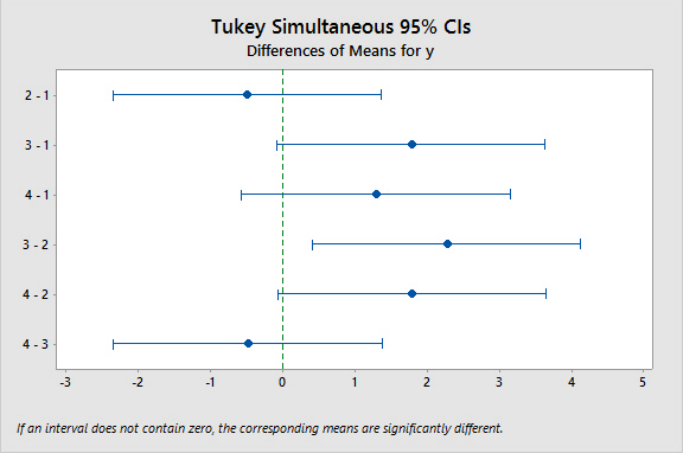


Figure 3: Post Hoc Tests - Plot 1

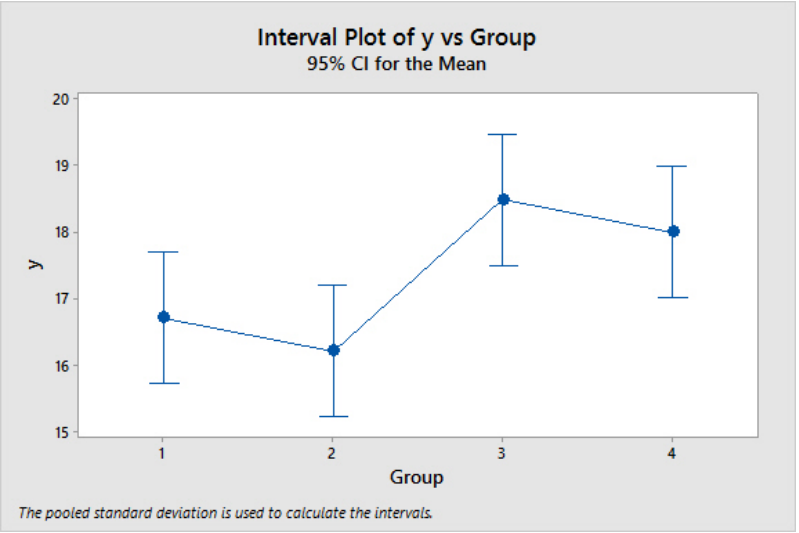


Figure 4: Post Hoc Tests - Plot 2

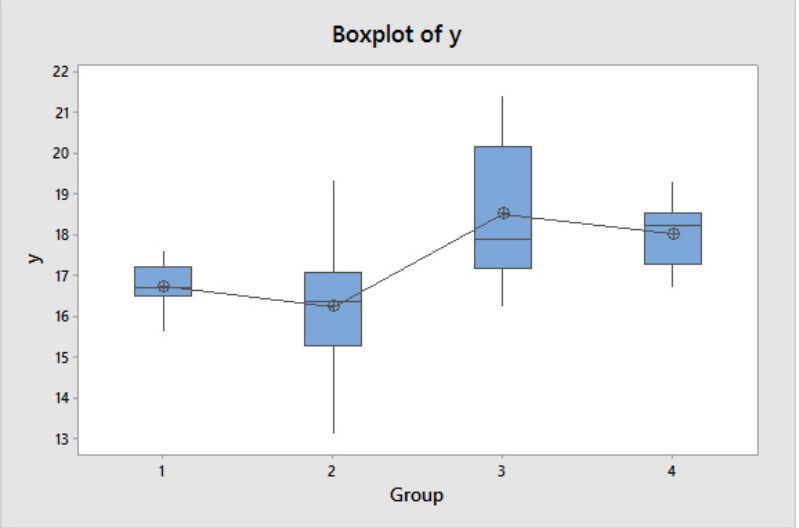


Figure 5: Post Hoc Tests - Plot 3

Part B - Two Way ANOVA (10 Marks)

An experiment is conducted to study how long different digital camera batteries last. The aim is to find out whether there is a difference in terms of battery life between four brands of batteries using 5 different cameras. Each battery was tried 4 times with each camera. You are also interested in determining if there is an interaction between the two variables.
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 battery lifetimes and both of the factors. Explain your conclusions.

General Linear Model: Lifetime versus Battery, Camera

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Battery	Fixed	4	1, 2, 3, 4
Camera	Fixed	5	1, 2, 3, 4, 5

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Battery	3	39643	13214	5.85	0.001
Camera	4	12857	3214	1.42	0.237
Battery*Camera	12	60413	5034	2.23	0.021
Error	60	135462	2258		
Total	79	248375			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
47.5152	45.46%	28.19%	3.04%

Figure 6: Minitab Output for Question 2 Part B

Question 3 - Data Analysis

Part A - Normal Distribution (6 Marks)

Suppose we have a manufacturing process that is designed to produce a product with a mean weight of 1003 grams. The weight of the products are normally distributed, with a standard deviation of 1.2 grams.

(a) (2 Marks) What percentage of products will have a weight exceeding 1005 grams?

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

(c) (2 Marks) The production manager reports that at least 99% of the output is between 1000g and 1006g. Do you agree with this statement? Justify your answer with the appropriate calculations.

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?

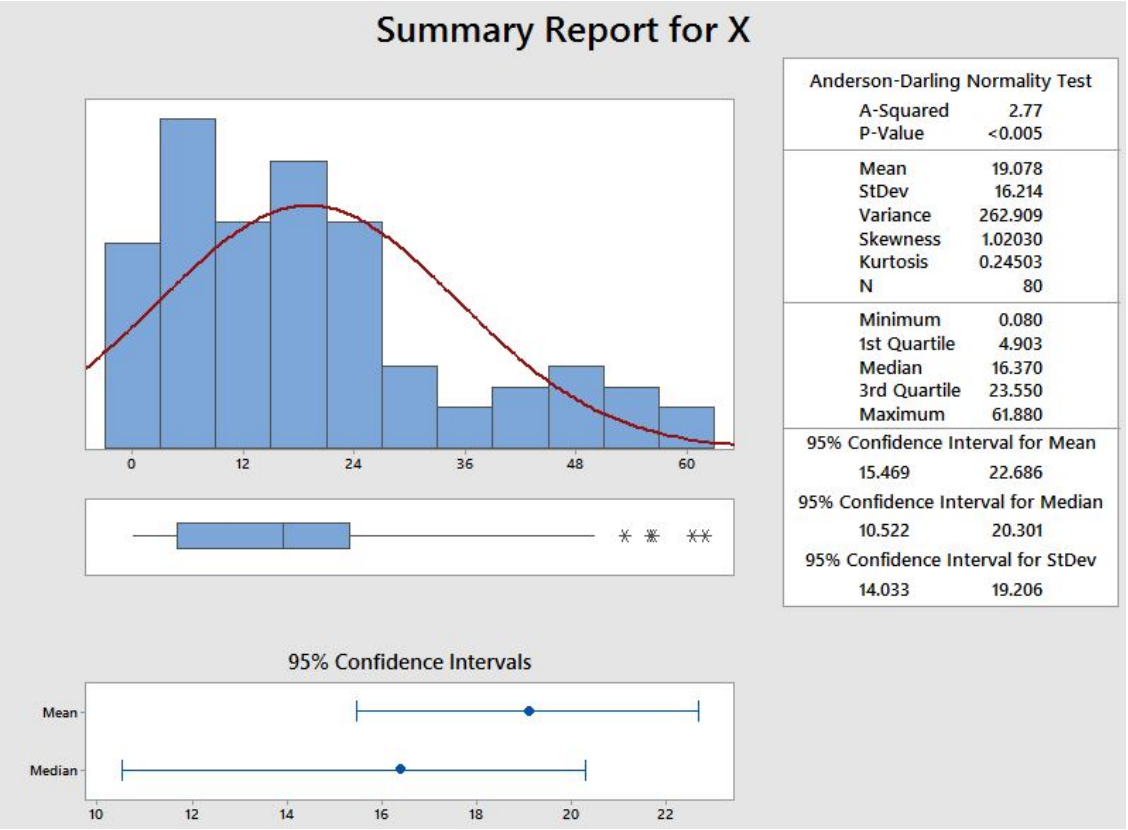
(b) (1 Mark) What is the relevance of this analysis as part of an overall statistical study.

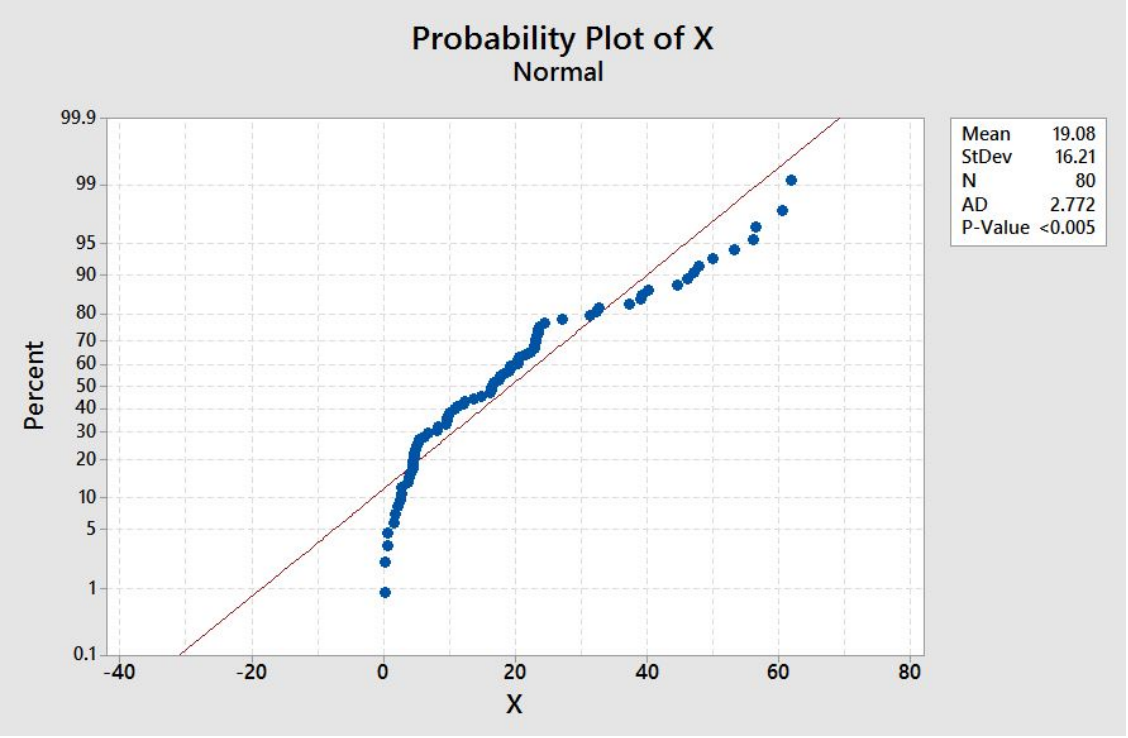
(c) (3 Marks) What is the conclusion of this analysis for the Variable X? Justify your answer with reference to 3 separate indications.

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

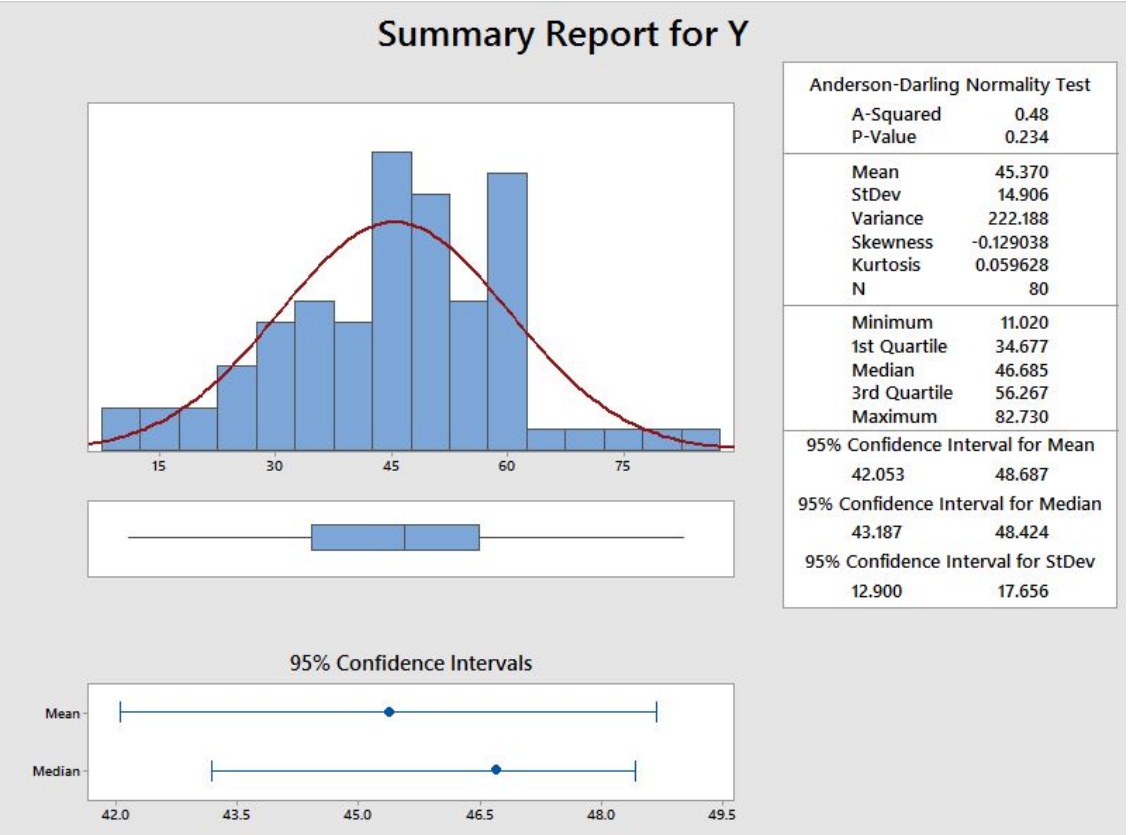
Important: *Question 3 comprises a third part: Part C. This part is presented in subsequent pages.*

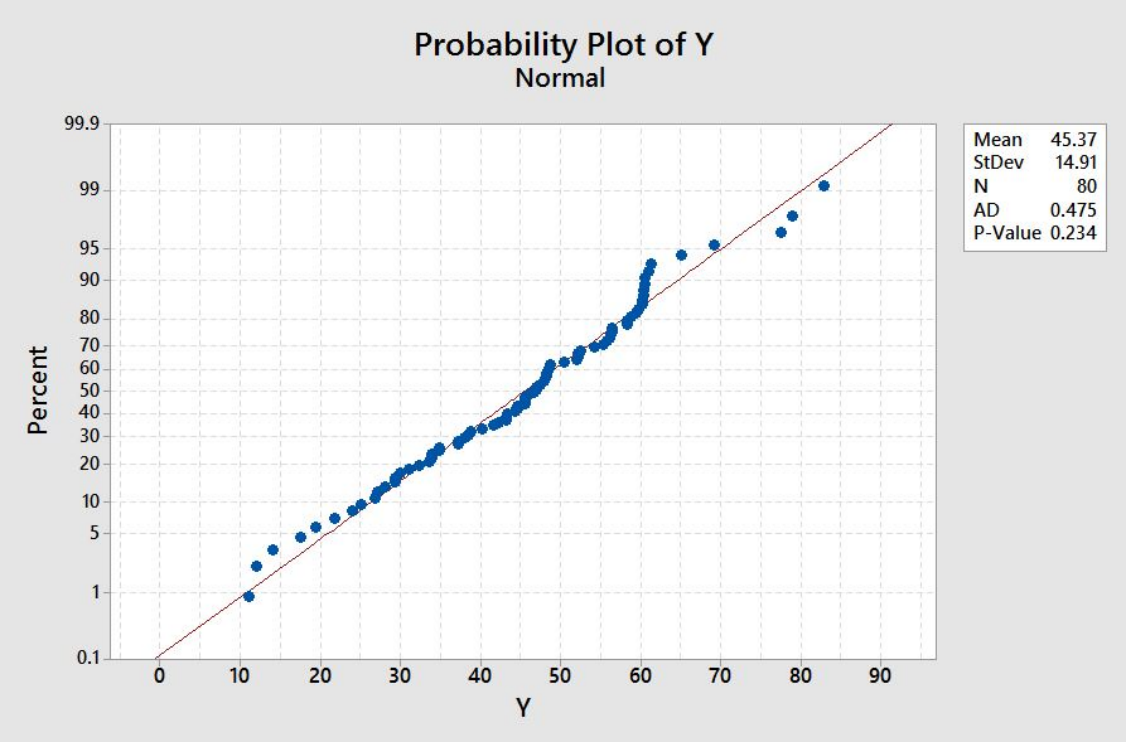
Question 3 - Part B - Minitab Output for Variable X





Question 3 - Part B - Minitab Output for Variable Y





Part C - 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 (75% pure)	14	1.8 2.0 2.1 2.1 2.1 2.2 2.3 2.3 2.3 2.4 2.4 2.5 2.6 2.7
B (60% pure)	10	1.9 1.9 2.0 2.1 2.1 2.2 2.2 2.5 2.7 2.8

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.

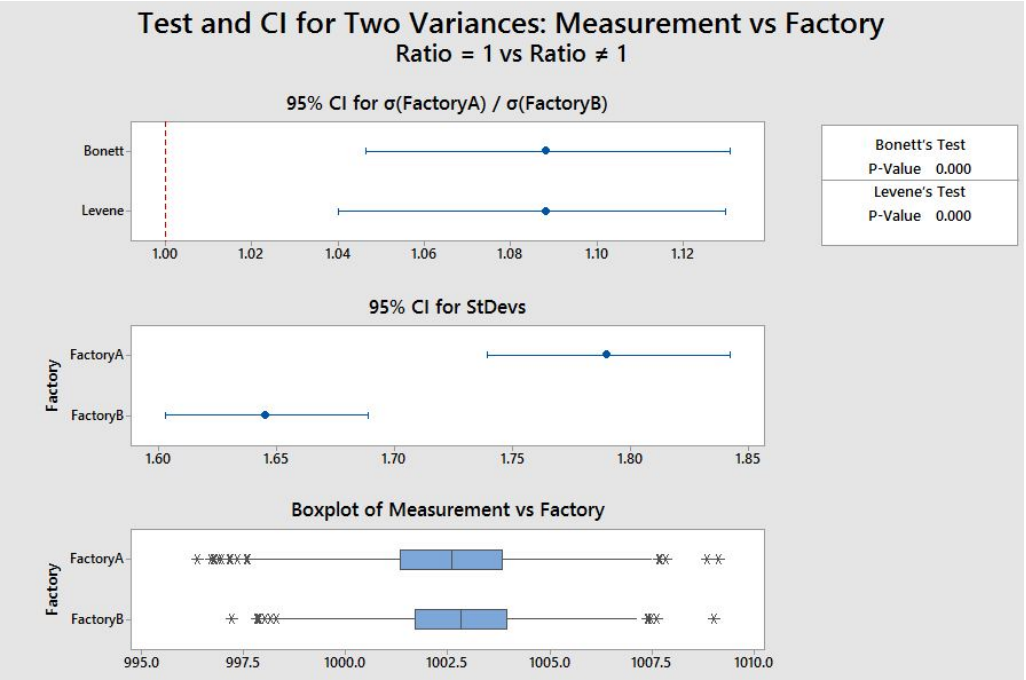
Measurement data from both factories has been compiled and analysed, with the following Minitab output has been created.

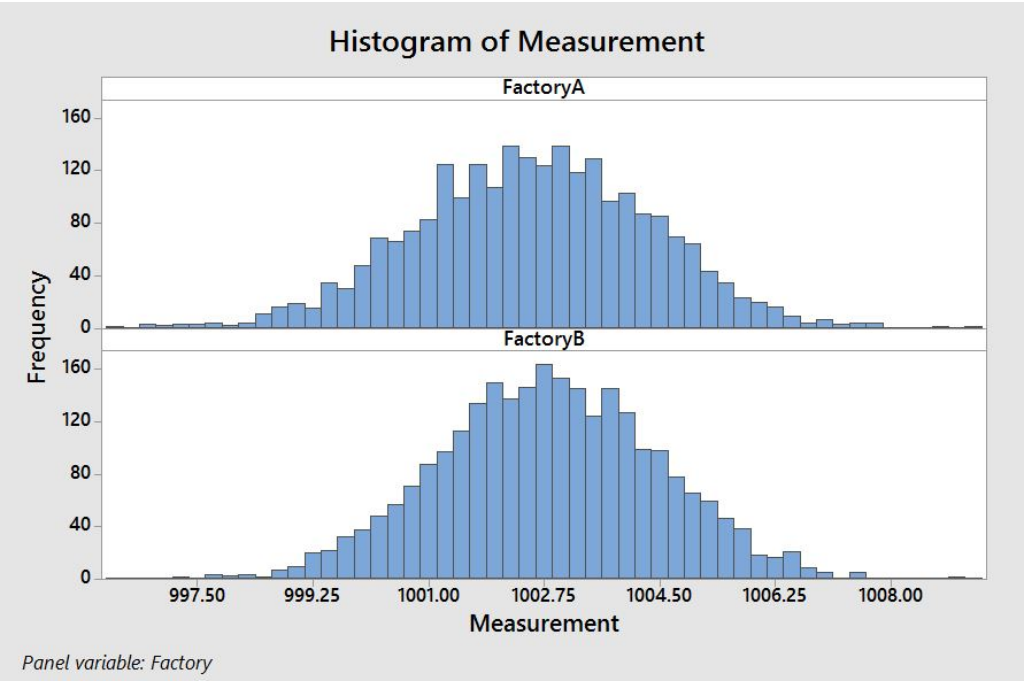
Using the printout, write a brief report on what the analysis tells you about the comparison of manufacturing processes in both factories. Explain your reasoning clearly. Formally state any null and alternative hypotheses where relevant.

```
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 4 - 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
              ....  .....   .....

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

Process C       20     35      45  100
             .....  .....   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 16 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
15	14	54
16	9	42

(Please turn to the next page for questions.)

(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     1.81     ....    0.606
Knots      .....   0.0333     6.02    0.000  1.00


Regression Equation

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

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?

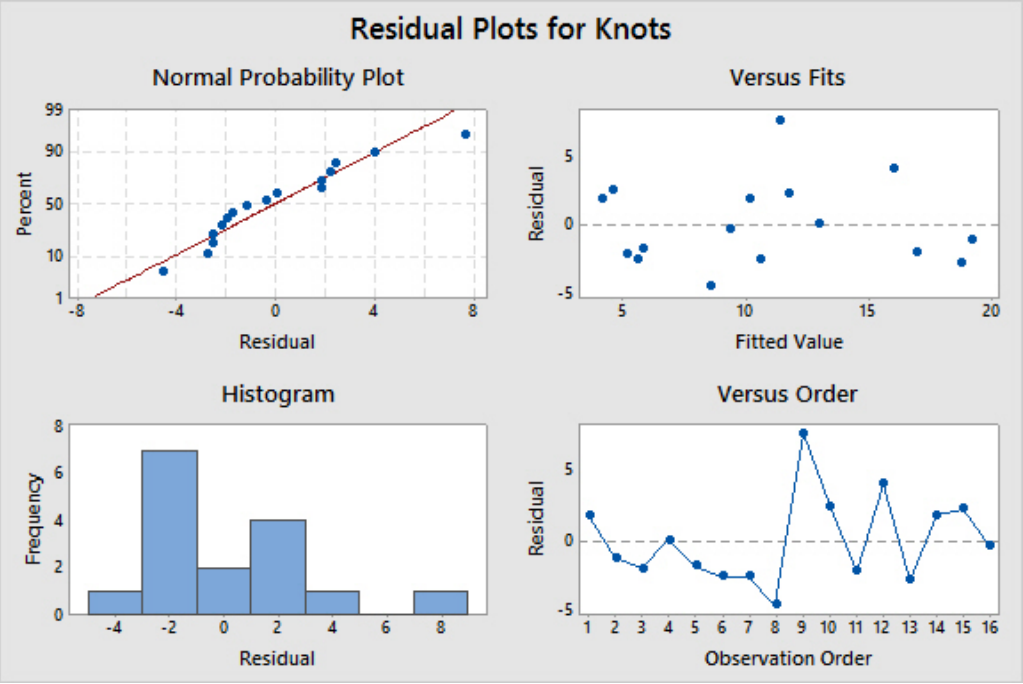


Figure 7: Diagnostic Plots for Regression Model in Question 5 Part A

Part C - Multiple Linear Regression (8 Marks)

A logistic companu is researching full efficiency in its fleet of vehicles. The company measured the fuel consumption and 10 aspects of automobile design and performance for 500 vehicles.

The variables are:

mpg Miles/(US) gallon

cyl Number of cylinders

disp Displacement (cu.in.)

hp Gross horsepower

drat Rear axle ratio

wt Weight (1000 lbs)

qsec 1/4 mile time

vs V engine or a straight engine design

am Transmission (0 = automatic, 1 = manual)

gear Number of forward gears

carb Number of carburetors

The data 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 fuel consumption. Which variable or variables appear to be good predictors of fuel consumption? Would you refine the model in the light of these results? If so, what changes would you make?

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.20345	86.90%	86.63%	86.28%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0000	0.0985	0.00	1.000	
cyl	-0.111	0.217	-0.51	0.607	15.37
disp	0.01334	0.00370	3.60	0.000	21.62
hp	-0.02148	0.00451	-4.76	0.000	9.83
wt	-3.715	0.393	-9.46	0.000	15.16
carb	-0.199	0.172	-1.16	0.246	7.91
gear	0.655	0.309	2.12	0.035	5.36
drat	0.787	0.339	2.32	0.021	3.37
qsec	0.821	0.151	5.42	0.000	7.53
vs	0.318	0.436	0.73	0.467	4.97
am	2.520	0.426	5.91	0.000	4.65

Regression Equation

mpg = 0.0000 - 0.111 cyl + 0.01334 disp - 0.02148 hp - 3.715 wt - 0.199 carb + 0.655 gear + 0.787 drat + 0.821 qsec + 0.318 vs + 2.520 am

Figure 8: Minitab Output for Regression Model

Fits and Diagnostics for Unusual Observations

Obs	mpg	Fit	Resid	Std Resid	
53	4.244	-0.434	4.678	2.14	R
86	10.352	4.345	6.008	2.75	R
101	1.081	6.057	-4.976	-2.28	R
111	7.264	13.352	-6.088	-2.84	R
114	5.907	10.907	-5.000	-2.30	R
115	1.288	-3.449	4.737	2.16	R
145	-4.706	-0.229	-4.477	-2.06	R
183	-7.223	-1.402	-5.820	-2.67	R
188	-12.663	-7.717	-4.945	-2.28	R
268	-5.795	-1.436	-4.359	-2.00	R
290	6.117	1.717	4.400	2.01	R
304	12.313	7.526	4.787	2.21	R
313	13.649	9.021	4.628	2.16	R
314	0.752	-3.705	4.457	2.04	R
316	-11.547	-7.213	-4.334	-2.00	R
320	9.090	2.447	6.643	3.07	R
333	-11.404	-6.209	-5.194	-2.40	R
348	5.567	-2.183	7.750	3.56	R
373	-9.860	-3.242	-6.618	-3.03	R
377	6.117	0.555	5.563	2.57	R
422	2.935	7.504	-4.568	-2.09	R
451	0.081	-4.469	4.549	2.08	R
457	1.271	5.650	-4.379	-2.01	R
474	-3.096	1.471	-4.567	-2.09	R
485	3.239	-1.167	4.405	2.01	R
499	7.300	1.228	6.072	2.77	R
500	0.113	-4.662	4.775	2.20	R

R Large residual

Figure 9: Minitab Output for Regression Model - Continued