FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF MATHEMATICS AND STATISTICS

END OF SEMESTER EXAMINATION PAPER 2015

MODULE CODE: MA4505

SEMESTER: Autumn 2015

MODULE TITLE: Applied Statistics For Administration 1

DURATION OF EXAM: 2.5 hours

LECTURER: Mr. Kevin O'Brien

GRADING SCHEME: 100 marks
70% of module grade

EXTERNAL EXAMINER: Prof. J. King

INSTRUCTIONS TO CANDIDATES

Scientific calculators approved by the University of Limerick can be used.
Formula sheet and statistical tables provided at the end of the exam paper.
There are 5 questions in this exam. Students must attempt any 4 questions.

Question 1 Inference Procedures and Distributions

Question 1 Part A (5 Marks)

(i.) (3 Marks) Provide a brief description for three tests from the family of Grubb's Outliers Tests. Include in your description a statement of the null and alternative hypothesis for each test

(ii.) (2 Marks) Describe any required assumptions for these tests, and the limitations of these tests.

Question 1 Part B (5 Marks)

The typing speeds for one group of 12 Engineering students were recorded both at the beginning of year 1 of their studies. The results (in words per minute) are given below:

149	146	112	142	168	153
137	161	156	165	170	159

Use the Dixon Q-test to determine if the lowest value (118) is an outlier. You may assume a significance level of 5%.

(i.) (1 Mark) State the null and alternative Hypothesis for this test.

(ii.) (2 Marks) Compute the test statistic

(iii.) (1 Mark) State the appropriate critical value.

(iv.) (1 Mark) What is your conclusion to this procedure.

Question 1 Part C (15 Marks)

The following data give the recovery of bromide from spiked samples of vegetable matter, measured using a gas-liquid chromatographic method. The same amount of bromide was added to each specimen. The units for all measurements are mg/g

	Data	Sample Mean	Sample Variance
Tomato	$\{777, 790, 759, 790, 770, 758, 764\}$	774	142.666
Cucumber	$\{782, 773, 778, 765, 789, 797, 782\}$	781	106
Asparagus	$\{786, 783, 781, 785, 789, 797, 782\}$	785	64.666
Overall		780	115.7

The following questions will result in the completion of the ANOVA Table on the bottom of this page. The $p-$value is already provided.

(i.) (3 Marks) Compute the Between Groups Sum of Squares. (Show your workings.)

(ii.) (3 Marks) Compute the Within Groups Sum of Squares. (Show your workings.)

(iii.) (2 Marks) Compute the Total Sum of Squares. (Show your workings.)

(iv.) (2 Marks) State the degrees of freedom for the ANOVA Table.

(v.) (1 Mark) Compute both of the Mean Square values.

(vi.) (1 Mark) Compute the test statistic for this procedure (i.e. the F-value.)

(vii.) (3 Marks) This analysis is used to assess if there is any difference between the mean levels of bromide for each of the three types of vegetable. What is your conclusion? Clearly state the null and alternative hypothesis.

Source	DF	SS	MS	F	p-value
Between	?	?	?	?	0.154
Within	?	?	?		
Total	?	?			

Question 2 Chi-Squared and Normal Distribution

Question 2 Part A (15 Marks)

A market research survey was carried out to assess preferences for three brands of chocolate bar: A, B, and C. The study group was categorised by age group to determine any difference in preferences.

	X	Y	Z	Total
Children	50	70	80	200
Teenagers	90	50	20	160
Adults	140	120	100	360

(i.) (2 Marks) Formally state the null and alternative hypotheses.

(ii.) (4 Marks) Compute the cell values expected under the null hypothesis.

(iii.) (4 Marks) Compute the test statistic.

(iv.) (2 Marks) State the appropriate critical value for this hypothesis test.

(v.) (3 Marks) Discuss your conclusion to this test, supporting your statement with reference to appropriate values.

Question 2 Part B (10 Marks)

Assume that the diameter of a critical component is normally distributed with a mean of 250mm and a standard deviation of 15mm. You are required to estimate the approximate probabilities of the following measurements occurring on an individual component:

(i.) (3 Mark) Greater than 245mm.

(ii.) (3 Marks) Less than 265mm.

(iii.) (4 Marks) Between 245mm and 265mm.

You are required to show all of your workings. Use the normal tables to determine the probabilities for the above exercises. Name any rule or techniques you are using.

Question 3. Two Way ANOVA Procedures (25 marks)

Question 3 Part A (10 Marks)

A supermarket buys a particular product from five suppliers (V,W,X,Y and Z), and conducts regular tasting tests by three expert panels. Various characteristics are scored and an analysis of the totals of these scores is made.

	V	W	X	Y	Z
Panel 1	22	25	23	23	24
Panel 2	21	16	16	19	16
Panek 3	23	22	19	23	22

You are given the following information:

- $S_r^2 = 8.9733$
- $S_c^2 = 1.0777$
- The variance of scores in the table above is $\text{VAR}(y) = 9.0666$

The following questions will result in the completion of the ANOVA Table on the bottom of this page. The $p-$values for both tests are already provided.

(i.) (4 Marks) Complete the Sum of Squares column. (Show your workings.)

(iv.) (1 Marks) State the degrees of freedom for the ANOVA Table.

(v.) (1 Marks) Compute the Mean Square values.

(vi.) (1 Marks) Compute the test statistics for this procedure (i.e. the F-values).

(vii.) (3 Marks) What is your conclusion? Clearly state the null and alternative hypotheses for both tests.

Source	DF	SS	MS	F	p-value
Factor A	?	?	?	?	0.00205 **
Factor B	?	?	?	?	0.43290
Error	?	?	?		
Total	?	?			

Question 3 Part B (9 Marks)

An engineer is designing a battery for use in a device that will be subjected to some extreme variations in temperature. The engineer can select the type of plate material for the battery. There are three possible choices for material.

Temperature can be controlled in the product development laboratory for the purposes of a test. The engineer decides to test all three place materials at three temperature levels : 15, 70, and 125 degrees.

Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order.

The following partial ANOVA table resulted from this analysis.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	$F-$Value	$p-$value
Material	?	?	?	?	4.16e-05 ***
Temperature	?	?	2250	?	0.00256 **
Interaction	?	6000	?	?	0.00379 **
Error	?	?	300		
Total	?	27600			

(i.) (6 Marks) Complete the ANOVA Table. Show your workings.

(ii.) (3 Marks) For the two main effects and the interaction effect, state your conclusion to this procedure.

Marking Scheme : There are 2 Marks awarded for the Degrees of Freedom, 2 Marks awarded for Sums of Squares, 1 Mark awarded for the Mean Square Values and 1 Mark awarded for the F-value.

Question 3 Part C (6 Marks)

Six analysts each made seven determinations of the paracetamol content of the same batch of tablets. The results are shown below. There are 42 determinations in total.

Analyst	Content						
A	84.32	84.61	84.64	84.62	84.51	84.63	84.51
B	84.24	84.13	84.00	84.02	84.25	84.41	84.30
C	84.29	84.28	84.40	84.63	84.40	84.68	84.36
D	84.14	84.48	84.27	84.22	84.22	84.02	84.33
E	84.50	83.91	84.11	83.99	83.88	84.49	84.06
F	84.70	84.36	84.61	84.15	84.17	84.11	83.81

The following R output has been produced as a result of analysis of these data:

```
Analysis of Variance Table

Df   Sum Sq   Mean Sq   F value    Pr(>F)
Analyst        5   0.8611   0.17222     4.236   0.00394 **
Residuals     36   1.4635   0.04065
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

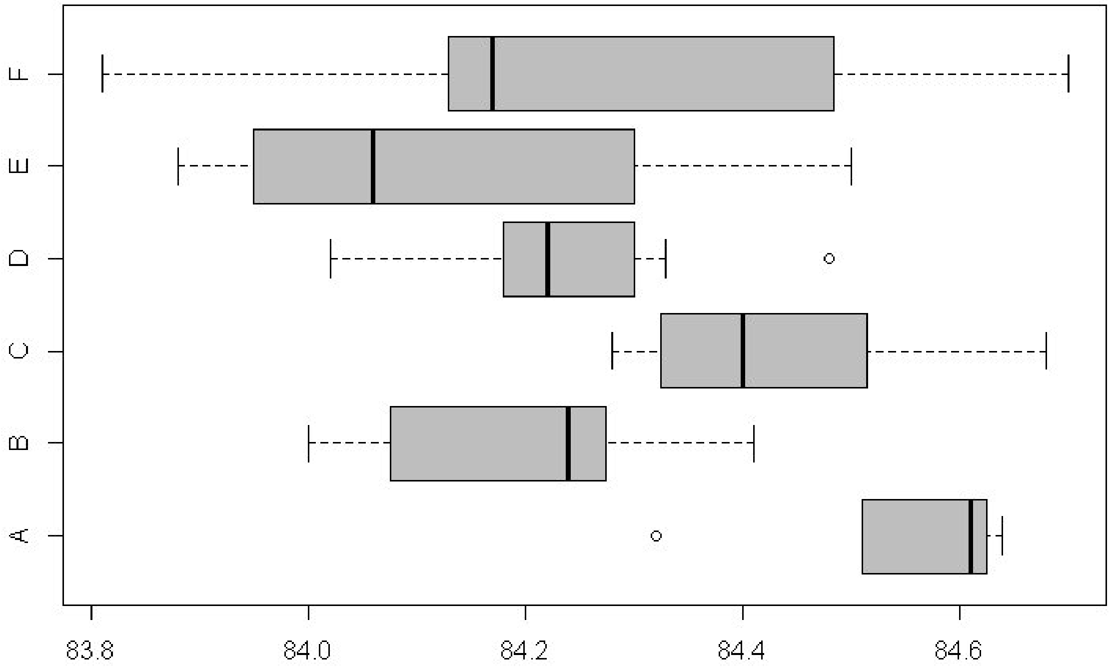
The R code and graphical procedures, below and on the following pages, are relevant to checking whether the underlying assumptions are met for this ANOVA model.

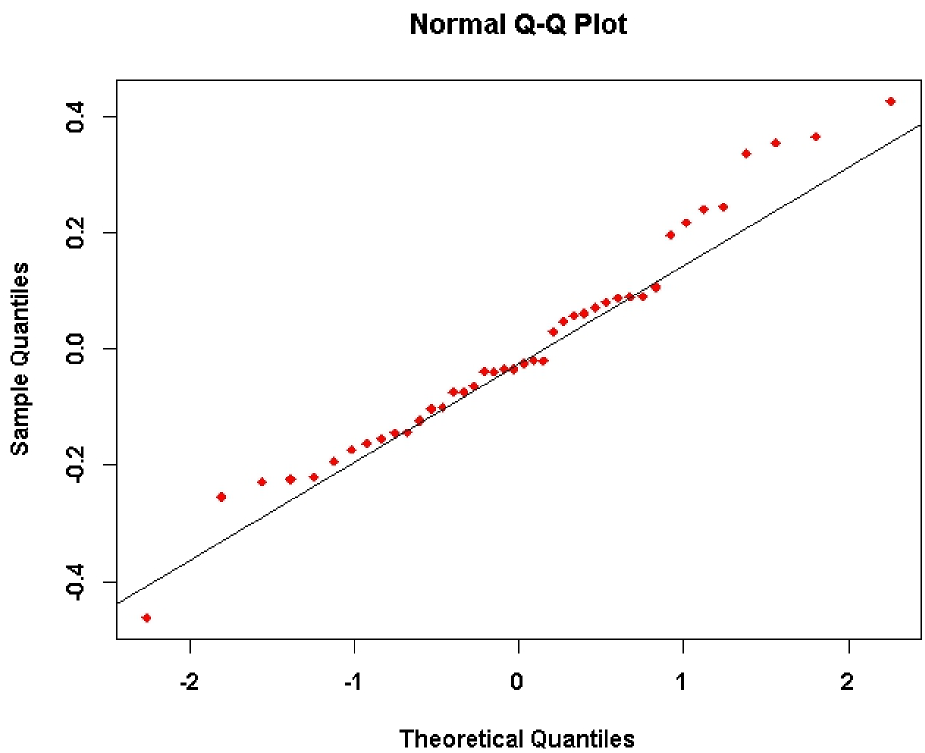
(i.) (2 Marks) What are the assumptions underlying ANOVA? (You may limit your answer to testable assumptions).

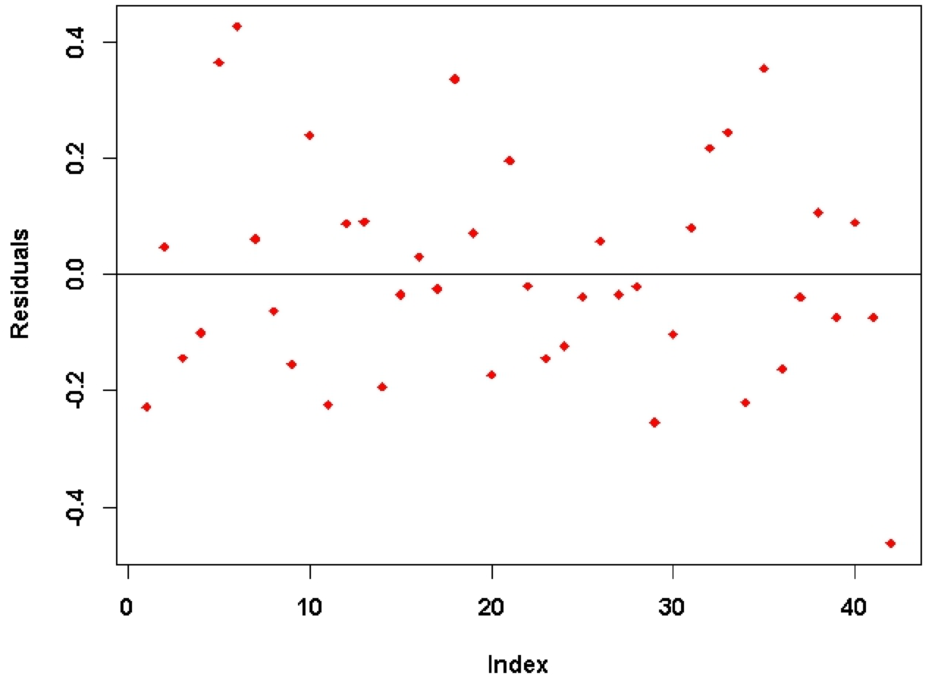
(ii.) (4 Marks) Assess the validity of these assumptions for the this ANOVA model.

```
Shapiro-Wilk normality test

data:  Residuals
W = 0.9719, p-value = 0.3819
```

```
Bartlett test of homogeneity of variances

data:  Experiment
Bartlett's K-squared = 105.9585, df = 1, p-value < 2.2e-16
```


F
E
D
C
B
A
83.8
84.0
84.2
84.4
84.6


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
0.4
0.2
0.0
-0.2
-0.4
-2
-1
0
1
2



Residuals
0.4
0.2
0.0
-0.2
-0.4
0
10
20
30
40
Index

Question 4. Linear Models (25 Marks)

Question 4 Part A (7 Marks)

The mercury level of several tests of sea-water from costal areas was determined by atomic-absorption spectrometry. The results obtained are as follows

Concentration in μg l^{-1}	0	10	20	30	40	50	60	70	80	90	100
Absorbance	0.321	0.834	1.254	1.773	2.237	2.741	3.196	3.678	4.217	4.774	5.261

The analysis of the relationship between concentration and absorbance is obtained in R and presented below.

```
summary(model)

Call:
lm(formula = y ~ x)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.2933636  0.0234754   12.50 5.45e-07
x           0.0491982  0.0003968  123.98 7.34e-16
---

Residual standard error: 0.04162 on 9 degrees of freedom
Multiple R-squared: 0.9994,     Adjusted R-squared: 0.9993
F-statistic: 1.537e+04 on 1 and 9 DF,  p-value: 7.337e-16

confint(model)
                 2.5 %     97.5 %
(Intercept) 0.24025851 0.34646876
x           0.04830054 0.05009582
```

(i.) (2 Marks) State the regression equation for this model. (Use 4 decimal places.)

(ii.) (2 Marks) State the 95% confidence interval for the slope and the intercept coefficients. Interpret this intervals with respect to any relevant hypothesis tests.

(iv.) (1 Mark) Using the regression equation, predict a value for absorbance when the concentration level is 25 mg/l.

(iv.) (2 Marks) The following piece of `R` code gives us a statistical metric. What is this metric? What is it used for? How should it be interpreted.

```
> AIC(model)
[1] -34.93389
```

Question 4 Part B (12 Marks)

Given the AIC for every possible candidate model in the table below, use both **Forward Selection** and **Backward Selection** to determine the best fitting linear regression model for predicting values of y with predictor variables x_1, x_2, x_3, x_4 and x_5.

(Marking Scheme : 6 Marks for both procedures. You may support your answer with sketchs.)

Variables	AIC	Variables	AIC
$\emptyset$	200	x1, x2, x3	74
		x1, x2, x4	75
x1	150	x1, x2, x5	79
x2	145	x1, x3, x4	72
x3	135	x1, x3, x5	85
x4	136	x1, x4, x5	95
x5	139	x2, x3, x4	83
		x2, x3, x5	82
x1, x2	97	x2, x4, x5	78
x1, x3	81	x3, x4, x5	85
x1, x4	94		
x1, x5	88	x1, x2, x3, x4	93
x2, x3	87	x1, x2, x3, x5	120
x2, x4	108	x1, x2, x4, x5	104
x2, x5	87	x1, x3, x4, x5	101
x3, x4	105	x2, x3, x4, x5	89
x3, x5	82		
x4, x5	86	x1, x2, x3, x4, x5	100

Question 4 Part C (6 Marks)

Suppose we have a regression model, described by the following equation

$$\hat{y} = 28.81 + 6.45x_1 + 7.82x_2$$

We are given the following pieces of information.

- The standard deviation of the response variance y is 10 units.
- There are 53 observations.
- The *Coefficient of Determination* (also known as the *Multiple R-Squared*) is 0.75.

Complete the *Analysis of Variance* Table for a linear regression model. The required values are indicated by question marks.

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Regression	?	?	?	?	$< 2.2e^{-16}$
Error	?	?	?		
Total	?	?			

Marking Scheme for Table Completion : There are 2 Marks awarded for the Degrees of Freedom, 2 Marks awarded for Sums of Squares, 1 Mark awarded for the Mean Square Values and 1 Mark awarded for the F-values.

Question 5. Statistical Process Control (25 marks)

Question 5 Part A (12 Marks)

The **Nelson Rules** are a set of eight decision rules for detecting "out-of-control" or non-random conditions on control charts. These rules are applied to a control chart on which the magnitude of some variable is plotted against time. The rules are based on the mean value and the standard deviation of the samples.

(i) (4×3 Marks) Discuss any four of these rules, and how they would be used to detect "out of control" processes. Support your answer with sketch.

In your answer, you may make reference to the following properties of the Normal Distribution. Consider the random variable X distributed as

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

where μ is the mean and σ^2 is the variance of an random variable X.

- $\Pr(\mu - 1\sigma \leq X \leq \mu + 1\sigma) = 0.6827$
- $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545$
- $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973$

Question 5 Part B (6 Marks)

A normally distributed quality characteristic is monitored through the use of control charts. These charts have the following parameters. All charts are in control.

	LCL	Centre Line	UCL
$\bar{X}$-Chart	1995	2000	2005
R-Chart	0	21	44.394

(i.) (2 Marks) What sample size is being used for this analysis?

(ii.) (2 Marks) Estimate the mean of the process standard deviations $\bar{s}$.

(iii.) (2 Marks) Compute the control limits for the process standard deviation chart (i.e. the s-chart).

Question 5 Part C (7 Marks)

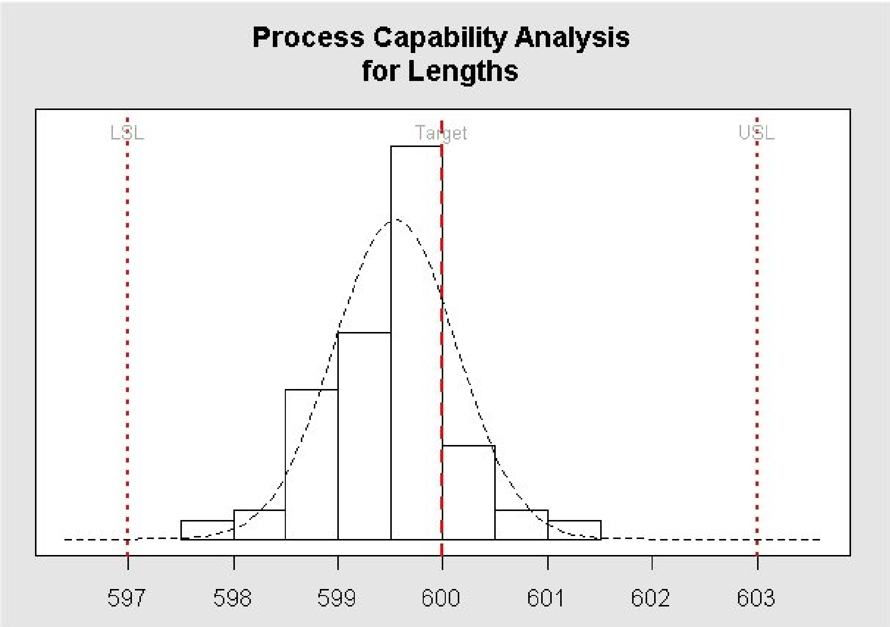
An automobile assembly plant concerned about quality improvement measured sets of five camshafts on twenty occasions throughout the day. The specifications for the process state that the design specification limits at 600±3mm.

(i.) (4 Marks) Determine the *Process Capability Indices* C_p and C_{pk}, commenting on the respective values. Use the `R` code output on the following page.

(ii.) (2 Mark) Explain why there would be a discrepancy between C_p and C_{pk}. Illustrate your answer with sketches.

(iii.) (1 Mark) Comment on the graphical output of the *Process Capability Analysis*, also presented on the next page.

```
Process Capability Analysis

Call:
process.capability(object = obj,
spec.limits = c(597, 603))
Number of obs = 100          Target = 600
Center = 599.548         LSL = 597
StdDev = 0.5846948       USL = 603

Capability indices:
    Value   2.5%  97.5%
Cp    ...
Cp_l  ...
Cp_u  ...
Cp_k  ...
Cpm   1.353  1.134  1.572
Exp<LSL 0%   Obs<LSL 0%
```



Formulas and Tables

Critical Values for Dixon Q Test

N	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
3	0.941	0.97	0.994
4	0.765	0.829	0.926
5	0.642	0.71	0.821
6	0.56	0.625	0.74
7	0.507	0.568	0.68
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568
11	0.392	0.444	0.542
12	0.376	0.426	0.522
13	0.361	0.41	0.503
14	0.349	0.396	0.488
15	0.338	0.384	0.475
16	0.329	0.374	0.463

Critical Values for Chi Square Test

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
1	2.705	3.841	6.634	10.827
2	4.605	5.991	7.378	9.21
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.07	12.833	15.086
6	10.645	12.592	14.449	16.812
7	12.017	14.067	16.013	18.475
8	13.362	15.507	17.535	20.09
9	14.684	16.919	19.023	21.666
10	15.987	18.307	20.483	23.209

Two Way ANOVA

$$MS_A = c \times S_r^2$$
$$MS_B = r \times S_c^2$$

Control Limits for Control Charts

$$\bar{\bar{x}} \pm 3\frac{\bar{s}}{c_4\sqrt{n}}$$

$$\bar{s} \pm 3\frac{c_5\bar{s}}{c_4}$$

$$\left[\bar{R}D_3, \bar{R}D_4\right]$$

Process Capability Indices

$$\hat{C}_p = \frac{\text{USL} - \text{LSL}}{6s}$$

$$\hat{C}_{pk} = \min\left[\frac{\text{USL} - \bar{x}}{3s}, \frac{\bar{x} - \text{LSL}}{3s}\right]$$

$$\hat{C}_{pm} = \frac{\text{USL} - \text{LSL}}{6\sqrt{s^2 + (\bar{x} - T)^2}}$$

Factors for Control Charts

Sample Size (n)	c4	c5	d2	d3	D3	D4
2	0.7979	0.6028	1.128	0.853	0	3.267
3	0.8862	0.4633	1.693	0.888	0	2.574
4	0.9213	0.3889	2.059	0.88	0	2.282
5	0.9400	0.3412	2.326	0.864	0	2.114
6	0.9515	0.3076	2.534	0.848	0	2.004
7	0.9594	0.282	2.704	0.833	0.076	1.924
8	0.9650	0.2622	2.847	0.82	0.136	1.864
9	0.9693	0.2459	2.970	0.808	0.184	1.816
10	0.9727	0.2321	3.078	0.797	0.223	1.777
11	0.9754	0.2204	3.173	0.787	0.256	1.744
12	0.9776	0.2105	3.258	0.778	0.283	1.717
13	0.9794	0.2019	3.336	0.770	0.307	1.693
14	0.9810	0.1940	3.407	0.763	0.328	1.672
15	0.9823	0.1873	3.472	0.756	0.347	1.653
16	0.9835	0.1809	3.532	0.750	0.363	1.637
17	0.9845	0.1754	3.588	0.744	0.378	1.622
18	0.9854	0.1703	3.64	0.739	0.391	1.608
19	0.9862	0.1656	3.689	0.734	0.403	1.597
20	0.9869	0.1613	3.735	0.729	0.415	1.585
21	0.9876	0.1570	3.778	0.724	0.425	1.575
22	0.9882	0.1532	3.819	0.720	0.434	1.566
23	0.9887	0.1499	3.858	0.716	0.443	1.557
24	0.9892	0.1466	3.895	0.712	0.451	1.548
25	0.9896	0.1438	3.931	0.708	0.459	1.541