

SULIT



First Semester Examination
2017/2018 Academic Session

January 2018

MGM561 - STATISTICAL METHODS FOR RESEARCH
[Kaedah Statistik Untuk Penyelidikan]

Duration : 3 hours
(Masa : 3 jam)

Please check that this examination paper consists of **FIFTEEN (15)** pages of printed material before you begin the examination.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **LIMA BELAS (15)** muka surat yang bercetak sebelum anda memulakan peperiksaan ini].*

Instructions : Answer **FIVE (5)** questions.

Arahan : Jawab **LIMA (5)** soalan.]

In the event of any discrepancies, the English version shall be used.

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah digunapakai].

...2/-
SULIT

Question 1

- (a) Improperly filled orders are costly problem for mail-order houses. To estimate the mean loss per incorrectly filled order, a large firms plan to sample n incorrectly filled orders and to determine the added cost associated with each one. The firm estimates that the added cost is between \$53 and \$430. How many incorrectly filled orders must be sampled to estimate the mean additional cost using a 95% confidence interval of width \$25?

[7 marks]

- (b) A standardized math test is given to 30 students in the form 4. Their scores with descriptive statistics and a stem-and-leaf plot are as follows:

| | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|
| 44 | 49 | 62 | 45 | 51 | 59 | 57 | 55 | 70 | 64 | 54 |
| 58 | 65 | 75 | 43 | 42 | 67 | 63 | 71 | 54 | 60 | 53 |
| 40 | 49 | 52 | 54 | 61 | 42 | 38 | 50 | | | |

Descriptive Statistics: test score

| | | | | | | | | |
|------------|-------|----|------|--------|-------|---------|-------|--------|
| Variable | Mean | SE | Mean | TrMean | StDev | Minimum | Q1 | Median |
| test score | 54.90 | | 1.78 | 54.73 | 9.75 | 38.00 | 48.00 | 54.00 |

| | | |
|------------|---------|-------|
| Variable | Maximum | Q3 |
| test score | 75.00 | 62.25 |

Stem-and-leaf of test score N = 30

Leaf Unit = 1.0

```

1  3  8
6  4  02234
9  4  599
(7) 5  0123444
14 5  5789
10 6  01234
5  6  57
3  7  01
1  7  5

```

- (i) Interpret the amount of variability in the data using the Empirical Rule.
- (ii) Charlie scored 62 on the test. What is his relative standing in the class?
- (iii) Compute the Interquartile Range (IQR), hence validates is there any outliers?

[13 marks]

Soalan 1

- (a) Borang pesanan yang salah diisi menyebabkan masalah kos pada syarikat pesanan pos. Untuk menganggar purata kerugian bagi setiap borang pesanan yang salah diisi ini, sebuah syarikat besar merancang untuk membuat pensampelan borang pesanan yang salah diisi untuk menganggarkan kos tambahan bagi setiap borang tersebut. Syarikat menganggarkan bahawa kos tambahan ialah antara \$53 dan \$430. Berapakah bilangan borang pesanan yang salah diisi diperlukan untuk persampelan bagi menganggar selang keyakinan 95% kos tambahan dengan panjang selang sebanyak \$25.

[7 markah]

- (b) Satu ujian matematik yang piawai diberikan kepada 30 pelajar dalam Tingkatan 4. Markah mereka dengan statistik deskriptif dan rajah tangkai dan daun seperti berikut diperolehi:

| | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|
| 44 | 49 | 62 | 45 | 51 | 59 | 57 | 55 | 70 | 64 | 54 |
| 58 | 65 | 75 | 43 | 42 | 67 | 63 | 71 | 54 | 60 | 53 |
| 40 | 49 | 52 | 54 | 61 | 42 | 38 | 50 | | | |

Descriptive Statistics: test score

| | | | | | | | |
|------------|-------|---------|--------|-------|---------|-------|--------|
| Variable | Mean | SE Mean | TrMean | StDev | Minimum | Q1 | Median |
| test score | 54.90 | 1.78 | 54.73 | 9.75 | 38.00 | 48.00 | 54.00 |

| | | |
|------------|---------|-------|
| Variable | Maximum | Q3 |
| test score | 75.00 | 62.25 |

Stem-and-leaf of test score N = 30

Leaf Unit = 1.0

| | | |
|-----|---|---------|
| 1 | 3 | 8 |
| 6 | 4 | 02234 |
| 9 | 4 | 599 |
| (7) | 5 | 0123444 |
| 14 | 5 | 5789 |
| 10 | 6 | 01234 |
| 5 | 6 | 57 |
| 3 | 7 | 01 |
| 1 | 7 | 5 |

- (i) Tafsir jumlah variasi di dalam data menggunakan Petua Empirik.
- (ii) Charlie mendapat markah 62 dalam ujian tersebut. Apakah kedudukan perbandingannya di dalam kelas itu?
- (iii) Kira julat antara kuartil (IQR), seterusnya tentusahkan adakah terdapat titik terpencil?

[13 markah]

Question 2

- (a) The following table represents the data on the number of small factories in which industrial injuries resulted claims for compensation between April 2003 and March 2004.

| Number of Claims | Frequency |
|------------------|------------|
| 0 | 144 |
| 1 | 91 |
| 2 | 32 |
| 3 | 11 |
| 4 | 2 |
| 5+ | 0 |
| Total | 280 |

Propose an appropriate distribution for these data and test to see if it is consistent with the data (use $\alpha = 0.05$).

[10 marks]

- (b) It is reported that offspring of users of a certain drug may have a higher incidence of birth defects than the general population. To obtain information about a possible relationship between this drug and birth defects, 100 offspring of female rats are fed the drug and 100 offspring from untreated female rats are examined. The results are given below:

| Rats | Progeny | |
|-----------|---------------|--------|
| | Birth Defects | Normal |
| Treated | 30 | 70 |
| Untreated | 20 | 80 |

Analyze these data at $\alpha = 0.05$. Is this a test of homogeneity or independence? State your conclusion from this study.

[10 marks]

Soalan 2

- (a) Jadual berikut mengandungi data mewakili bilangan perusahaan kecil yang mana kemalangan industri menyebabkan tuntutan pampasan antara bulan April 2003 sehingga bulan Mac 2004.

| <i>Bilangan Tuntutan</i> | <i>Kekerapan</i> |
|--------------------------|------------------|
| 0 | 144 |
| 1 | 91 |
| 2 | 32 |
| 3 | 11 |
| 4 | 2 |
| 5+ | 0 |
| Jumlah | 280 |

Cadangkan taburan yang bersesuaian bagi data tersebut dan uji untuk melihat adakah ianya konsisten dengan data (guna $\alpha = 0.05$).

[10 markah]

- (b) Dilaporkan bahawa pengguna sejenis ubat berkemungkinan tinggi mendapat anak dengan kecacatan semasa lahir berbanding populasi yang umum. Untuk mendapat maklumat bagi kemungkinan hubungan antara penggunaan ubat dengan kecacatan semasa lahir, 100 anak daripada tikus betina yang diberikan ubat dan 100 anak daripada tikus betina yang tidak diberikan ubat, diperiksa. Hasilnya seperti ditunjukkan di bawah:

| <i>Tikus</i> | <i>Anak</i> | |
|--------------------------|--------------|---------------|
| | <i>Cacat</i> | <i>Normal</i> |
| <i>Diberi ubat</i> | 30 | 70 |
| <i>Tidak diberi ubat</i> | 20 | 80 |

Analisis data tersebut dengan $\alpha = 0.05$. Adakah ini merupakan Ujian Homogenus atau Ujian Ketaksandaran? Nyatakan kesimpulan dari kajian ini.

[10 markah]

Question 3

- (a) The following values were computed from the length of life of two brands of light bulbs (in hour):

| | <i>Brand A</i> | <i>Brand B</i> |
|-------------------------|----------------|----------------|
| n | 9 | 25 |
| \bar{y} | 1560 | 1573 |
| $\Sigma(y - \bar{y})^2$ | 440 | 1860 |

- (i) Is there a difference in the variability of lifetimes for the two brands of bulbs? (Use $\alpha = 0.02$).
- (ii) Find the 98% confidence interval on the ratio of the two variabilities.

[10 marks]

- (b) An oncologist wants to evaluate the usefulness of the CAT scan for uterine tumor diagnosis. For 12 women with fibroid tumors, certain measurements are taken by CAT scan techniques prior to surgery and then compared with other measurements taken on the tumors in the pathology laboratory after they had been surgically removed. Suppose the paired measurements on tumor mass are:

| Patient | A | B | C | D | E | F |
|----------------|----------|----------|----------|----------|----------|----------|
| CAT scan, x | 18 | 17 | 28 | 20 | 11 | 24 |
| Pathology, y | 20 | 4 | 25 | 16 | 19 | 21 |
| Patient | G | H | I | J | K | L |
| CAT scan, x | 16 | 15 | 19 | 24 | 23 | 13 |
| Pathology, y | 22 | 10 | 23 | 27 | 18 | 11 |

and the statistics computed are

$$\Sigma(x - \bar{x})^2 = 278; \Sigma(y - \bar{y})^2 = 498;$$

$$\frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = 0.723; \quad \frac{[\Sigma(x - \bar{x})(y - \bar{y})]^2}{\Sigma(x - \bar{x})^2} = 108.58$$

- (i) Find the sample correlation coefficient.
- (ii) State the most logical hypotheses about the correlation between the CAT scan measurement of tumor mass and that obtained at pathology.
- (iii) Test your hypothesis in (ii) for an $\alpha = 0.05$
- (iv) Do you think the relationship would be useful in being able to use the CAT scan information to predict fibroid tumor mass prior to surgery?

[15 marks]

Soalan 3

- (a) Nilai berikut, diperolehi daripada jangka hayat (dalam jam) mentol lampu bagi dua jenama lampu :

| | <i>Jenama A</i> | <i>Jenama B</i> |
|-------------------------|-----------------|-----------------|
| n | 9 | 25 |
| \bar{y} | 1560 | 1573 |
| $\Sigma(y - \bar{y})^2$ | 440 | 1860 |

- (i) Adakah terdapat perbezaan dalam variabiliti terhadap jangka hayat lampu bagi kedua-dua jenama tersebut?
- (ii) Dapatkan selang keyakinan 98% terhadap nisbah kedua variabiliti itu.

[10 markah]

- (b) Seorang ahli onkologi mahu mentaksir kepentingan imejan CAT bagi diagnosis ketumbuhan uterin. Untuk 12 wanita yang mempunyai ketumbuhan fibroid, pengukuran diambil menggunakan imejan CAT sebelum pembedahan dan dibandingkan dengan ukuran ketumbuhan dalam makmal patologi selepas pembedahan. Andaikan ukuran bagi ketumbuhan adalah seperti berikut:

| <i>Patient</i> | A | B | C | D | E | F |
|---------------------|----------|----------|----------|----------|----------|----------|
| <i>CAT scan, x</i> | 18 | 17 | 28 | 20 | 11 | 24 |
| <i>Pathology, y</i> | 20 | 4 | 25 | 16 | 19 | 21 |
| <i>Patient</i> | G | H | I | J | K | L |
| <i>CAT scan, x</i> | 16 | 15 | 19 | 24 | 23 | 13 |
| <i>Pathology, y</i> | 22 | 10 | 23 | 27 | 18 | 11 |

dan statistik berikut diperolehi;

$$\Sigma(x - \bar{x})^2 = 278; \Sigma(y - \bar{y})^2 = 498;$$

$$\frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = 0.723; \quad \frac{[\Sigma(x - \bar{x})(y - \bar{y})]^2}{\Sigma(x - \bar{x})^2} = 108.58$$

- (i) Cari pekali korelasi sampel tersebut.
- (ii) Nyatakan hipotesis yang paling logik tentang korelasi antara ukuran oleh imejan CAT dengan ukuran secara patologi.
- (iii) Uji hipotesis anda dalam (ii) pada $\alpha = 0.05$.
- (iv) Pada pendapat anda adakah hubungan ini berguna untuk membolehkan penggunaan imejan CAT untuk meramal saiz ketumbuhan fibroid sebelum pembedahan?

[15 markah]

Question 4

- (a) Data from a sample of 10 pharmacies are used to examine the relation between prescription sales volume and the percentage of prescription ingredients purchased directly from the supplier. The sample data are shown here:

| Pharmacy | Sales Volume, y (in RM 1000) | % of Ingredients Purchased Directly, x |
|----------|--------------------------------|--|
| 1 | 25 | 10 |
| 2 | 55 | 18 |
| 3 | 50 | 25 |
| 4 | 75 | 40 |
| 5 | 110 | 50 |
| 6 | 138 | 63 |
| 7 | 90 | 42 |
| 8 | 60 | 30 |
| 9 | 10 | 5 |
| 10 | 100 | 55 |

- (i) Find the least-squares estimates for the regression line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ (use informations given in **Figure 1**).
- (ii) Predict sales volume for a pharmacy that purchases 15% of its prescription ingredients directly from the supplier.
- (iii) Interpret the value of $\hat{\beta}_1$ in the context of the problem.

[10 marks]

- (b) To evaluate cigarettes, the Federal Trade Commission (FTC) uses “smoking machines” that measure the tar, nicotine, and carbon monoxide in each cigarette. Carbon monoxide has been linked to heart disease, tar has been linked to cancer, and nicotine is addictive. Suppose that the amount of tar (measured in milligrams) is recorded for 25 cigarettes randomly selected from each of four brands (brand A, brand B, brand C and brand D). Hence, by referring to **Figure 2**,

- (i) Discuss on distribution and normality of the data.
- (ii) By using the given outputs of the ANOVA analysis, determine whether one brand is “lowest” in tar.

[10 marks]

Soalan 4

- (a) Data dari sampel 10 buah farmasi digunakan untuk mengkaji hubungan antara jumlah jualan preskripsi dengan peratusan pembelian langsung bahan-bahan diskripsi daripada pembekal. Data sampel ditunjukkan seperti berikut:

| <i>Farmasi</i> | <i>Jumlah Jualan, y (dalam RM 1000)</i> | <i>% Pembelian Langsung Bahan-Bahan, x</i> |
|----------------|---|--|
| 1 | 25 | 10 |
| 2 | 55 | 18 |
| 3 | 50 | 25 |
| 4 | 75 | 40 |
| 5 | 110 | 50 |
| 6 | 138 | 63 |
| 7 | 90 | 42 |
| 8 | 60 | 30 |
| 9 | 10 | 5 |
| 10 | 100 | 55 |

- (i) Cari penganggar kuasa dua terkecil bagi garis regresi $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ (gunakan maklumat diberi di **Rajah 1**).
- (ii) Ramal jumlah jualan bagi suatu farmasi yang membeli 15% bahan-bahan secara langsung daripada pembekal.
- (iii) Taksir nilai $\hat{\beta}_1$ dalam konteks permasalahan yang diberikan.

[10 markah]

- (b) Untuk mengkaji rokok, Suruhanjaya Perdagangan Persekutuan (FTC) menggunakan 'mesin rokok' yang mengukur kandungan tar, nikotin and karbon monoksida bagi setiap rokok. Karbon monoksida dikaitkan dengan penyakit hati, tar dikaitkan dengan kanser dan nikotin menyebabkan ketagihan. Katakan jumlah tar (diukur dalam milligram) direkodkan bagi 25 rokok yang dipilih secara rawak masing-masing daripada empat jenama (jenama A, jenama B, jenama C dan jenama D). Maka itu, dengan merujuk **Rajah 2**,

- (i) Bincangkan taburan dan kenormalan data tersebut.
- (ii) Berdasarkan analisis varians, tentukan sama ada suatu jenama itu mengandungi tar yang terendah.

[10 markah]

Question 5

The municipal solid waste of a large city wants to investigate ways to reduce the amount of recyclable materials that are placed in the city's landfill. From an analysis of recycling records from other cities, it is determined that if the average weekly amount of recyclable material is more than five kilograms per household a commercial recycling firm could make a profit collecting the material. To determine the feasibility of the recycling plan, a random sample of 25 households is selected. The weekly weight of recyclable material (in kg/week) for each household is given here:

1.1 1.2 2.1 2.6 2.7 2.9 3.6 3.9 4.2 4.3 4.5 4.7
5.3 5.6 5.8 6.5 6.7 6.7 7.8 7.8 14.2 25.9 29.5
34.8 43.8

- (a) Determine an appropriate measure of the amount of recyclable waste from a typical house hold in the city and hence, construct its 95% confidence interval (the sample mean, \bar{y} is 9.53, use $\alpha = 0.05$).

[6 marks]

- (b) Find a 95% confidence interval for the median using the normal approximation.

[6 marks]

- (c) Hence, gives comment on the approximation value versus the exact value.

[3 marks]

Soalan 5

Majlis Pembuangan Sisa Pepejal sebuah bandar besar mahu mengkaji cara mengurangkan bahan kitar semula yang diletakkan di lokasi pembuangan. Daripada analisis rekod kitar semula sebuah bandar yang lain, adalah didapati bahawa jika purata mingguan jumlah bahan kitar semula melebihi 5 kg bagi setiap isirumah, syarikat kitar semula boleh mendapat untung daripada kutipan barangan tersebut. Untuk melihat ketersauran cadangan kitar semula itu, 25 isirumah dipilih secara rawak. Berat bahan kitar semula (dalam kg/minggu) bagi setiap isirumah diberikan seperti berikut:

| | | | | | | | | | | | |
|------|------|-----|-----|-----|-----|-----|-----|------|------|------|-----|
| 1.1 | 1.2 | 2.1 | 2.6 | 2.7 | 2.9 | 3.6 | 3.9 | 4.2 | 4.3 | 4.5 | 4.7 |
| 5.3 | 5.6 | 5.8 | 6.5 | 6.7 | 6.7 | 7.8 | 7.8 | 14.2 | 25.9 | 29.5 | |
| 34.8 | 43.8 | | | | | | | | | | |

- (a) Tentukan ukuran yang bersesuaian bagi jumlah sisa buangan kitar semula isirumah bandar itu, kemudian, bina selang keyakinan 95% nya (min sampel, \bar{y} ialah 9.53, guna $\alpha = 0.05$).

[6 markah]

- (b) Cari selang keyakinan 95% bagi median menggunakan penghampiran normal..

[6 markah]

- (c) Maka, berikan komen terhadap nilai penghampiran dengan nilai sebenar.

[3 markah]

Figure 1: Informations related to Question 4(a)

Rajah 1: Maklumat berkaitan Soalan 4(a)

| | y | x | $y - \bar{y}$ | $x - \bar{x}$ | $(x - \bar{x})(y - \bar{y})$ | $(x - \bar{x})^2$ |
|---------------|-------------|-------------|---------------|---------------|------------------------------|-------------------|
| | 25 | 10 | -46.3 | -22.8 | 1101.94 | 566.44 |
| | 55 | 18 | -16.3 | -15.8 | 257.54 | 249.64 |
| | 50 | 25 | -21.3 | -8.8 | 187.44 | 77.44 |
| | 75 | 40 | 3.7 | 6.2 | 22.94 | 38.44 |
| | 110 | 50 | 38.7 | 16.2 | 626.94 | 262.44 |
| | 138 | 63 | 66.7 | 29.2 | 1947.64 | 852.64 |
| | 90 | 42 | 18.7 | 8.2 | 153.34 | 67.24 |
| | 60 | 30 | -11.3 | -3.8 | 42.94 | 14.44 |
| | 10 | 5 | -61.3 | -28.8 | 1765.44 | 829.44 |
| | 100 | 55 | 28.7 | 21.2 | 608.44 | 449.44 |
| Totals | 713 | 338 | 0 | 0 | 6714.60 | 3407.60 |
| Means | 71.3 | 33.8 | | | | |

Figure 2 : Minitab output for Question 4

Rajah 2: Output Minitab untuk Soalan 4

One-way ANOVA: brandA, brandB, brandC, brandD

| Source | DF | SS | MS | F | P |
|--------|----|---------|---------|------|-------|
| Factor | 3 | 0.09260 | 0.03087 | 3.72 | 0.014 |
| Error | 96 | 0.79670 | 0.00830 | | |
| Total | 99 | 0.88930 | | | |

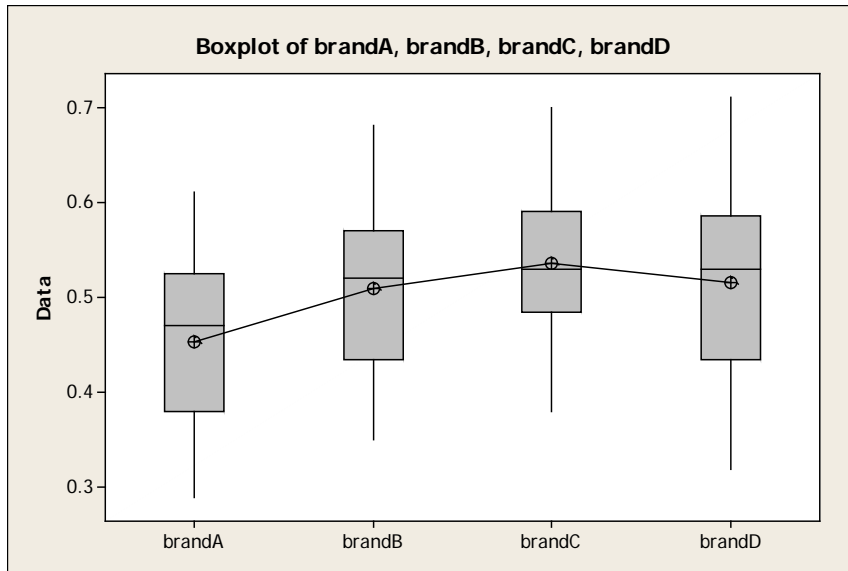
S = 0.09110 R-Sq = 10.41% R-Sq(adj) = 7.61%

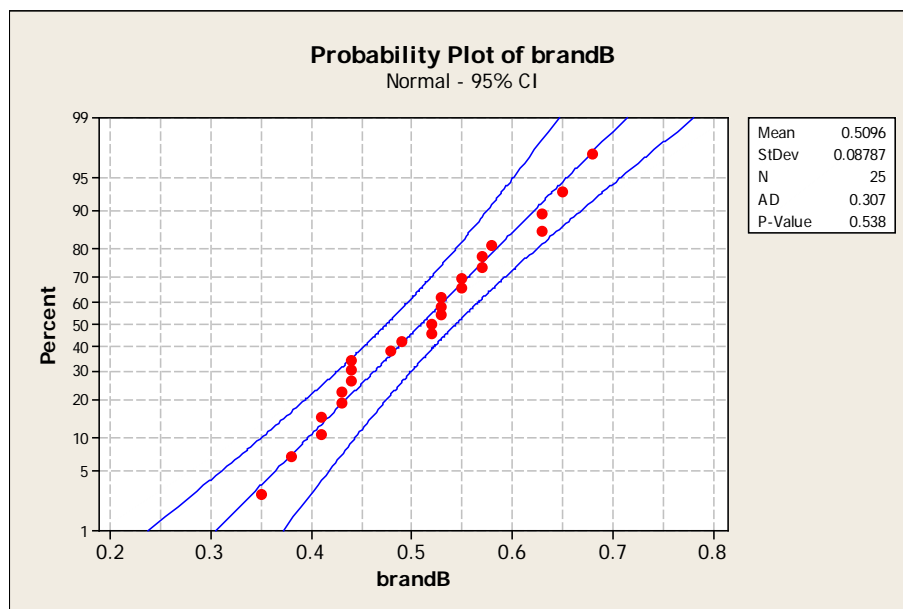
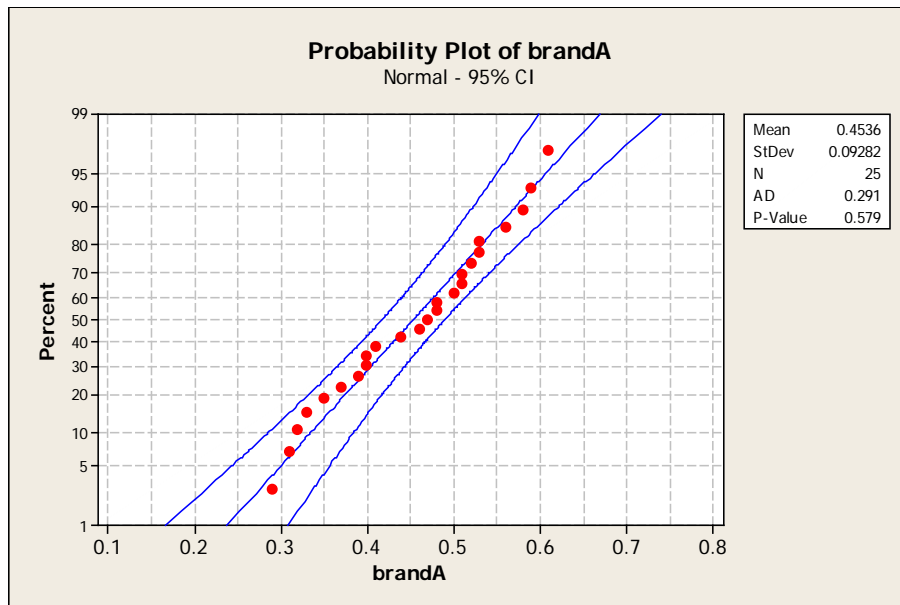
Individual 95% CIs For Mean Based on Pooled StDev

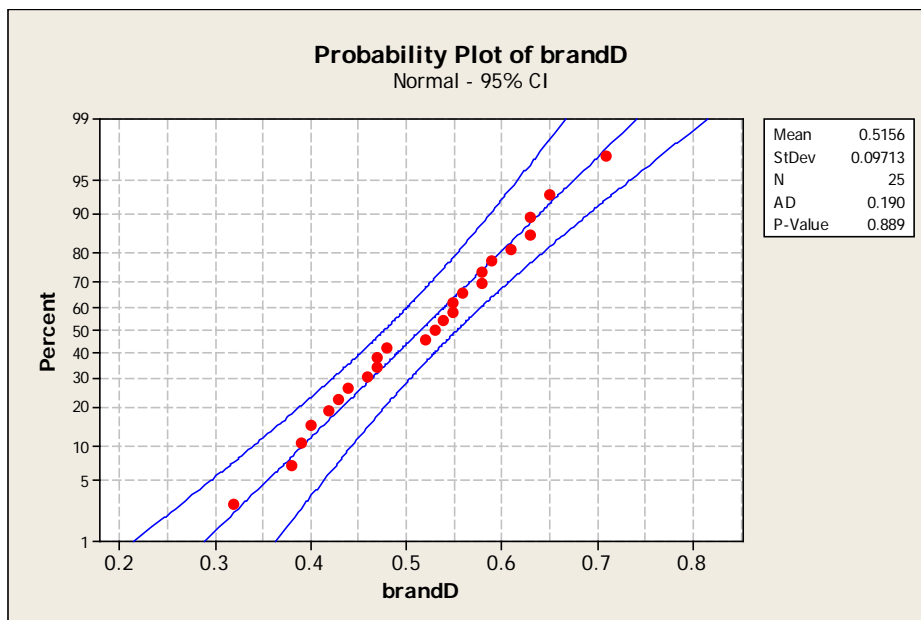
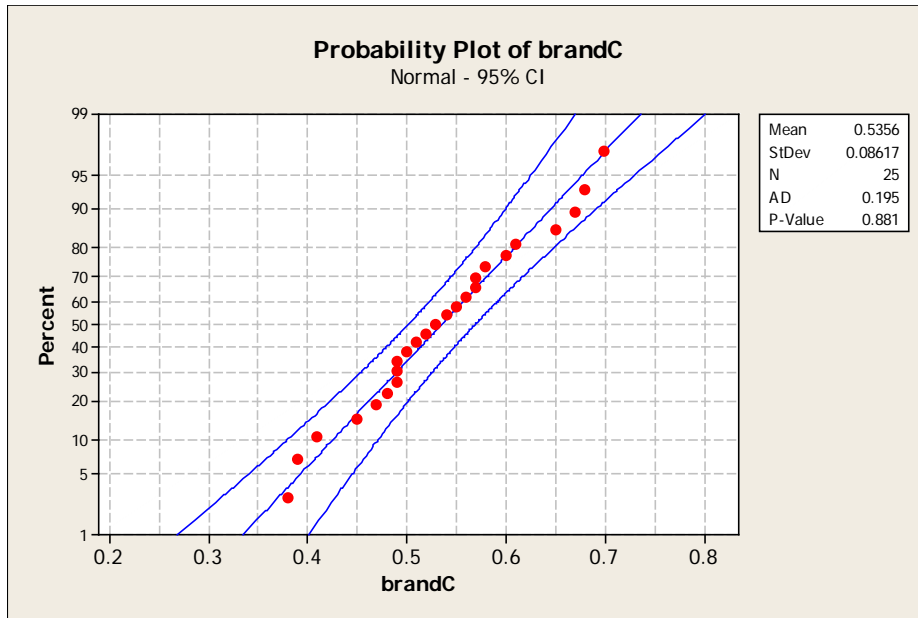
| Level | N | Mean | StDev |
|--------|----|---------|---------|
| brandA | 25 | 0.45360 | 0.09282 |
| brandB | 25 | 0.50960 | 0.08787 |
| brandC | 25 | 0.53560 | 0.08617 |
| brandD | 25 | 0.51560 | 0.09713 |

0.440 0.480 0.520 0.560

Pooled StDev = 0.09110







-oo000oo-