UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2005/2006

April/May 2006

MAA 161E – Statistics for Science Students
[Statistik Untuk Pelajar Sains]

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of ELEVEN pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi SEBELAS muka surat yang bercetak sebelum anda memulakan peperiksaan ini].

Instructions: Answer **all four** [4] questions.

*[**Arahan:** Jawab **semua empat** [4] soalan].*

1. (a) The following table shows the telephone bills of 100 households in a residential area in City A for a duration of one month.

Phone Bill (RM)	0-49	50-99	100-149	150-199	200-299	300-499
No. of households	30	20	10	10	15	15

$\sum x_i f_i = 14950$, $\sum x_i^2 f_i = 3916275$

(i) Determine the mean, median and standard deviation for the phone bills of the above households

(ii) Determine the value of X, if the telephone bill of 20% of households is at least RM X.

(iii) Use the Chebyshev Theorem to determine the interval which contains at least 20% of the household.

(b) Events A and B are such that $P(A) = \frac{2}{5}$ and $P(B) = \frac{1}{4}$ and $P\left[(A \cap \overline{B}) \cup (\overline{A} \cap B)\right] = \frac{1}{6}$. Find $P(A \cap B)$ and determine whether A and B are independent.

(c) A factory produces electronic equipments and receives 40% of the supplies from supplier A, 25% from supplier B, and 35% from supplier C. The percentage of faulty electronic components supplied by supplier A, B and C are respectively 5%, 2% and 1%.

(i) Find the probability that an electronic component chosen at random from all the suppliers is faulty.

(ii) Find the probability that from two electronic components chosen at random from all the suppliers; at least one component is faulty.

(iii) If an electronic component chosen is faulty, what is the probability that it is from supplier B?

(d) A continuous random variable X has the probability density function given by

$$f(x) = \begin{cases} k, & 2 < x < 3 \\ k(x-2), & 3 \le x < 4 \\ 0, & otherwise \end{cases}$$ where k is a constant.

(i) Determine the value of k.

(ii) Calculate $P(2.5<X<3.5)$

(iii) Determine the median of X.

[100 marks]

1. (a) *Jadual yang berikut menunjukkan bil telefon daripada 100 keluarga di sebuah kawasan kediaman yang terletak di Bandar A dalam jangkamasa sebulan.*

Bil Telefon (RM)	*0-49*	*50-99*	*100-149*	*150-199*	*200-299*	*300-499*
Bilangan keluarga	*30*	*20*	*10*	*10*	*15*	*15*

$$\sum x_i f_i = 14950 \quad , \quad \sum x_i^2 f_i = 3916275$$

(i) *Tentukan min, median dan sisihan piawai bagi bil-bil telefon daripada keluarga di atas.*

(ii) *Tentukan nilai X, jika bil telefon daripada 20% keluarga adalah sekurang-kurangnya RM X.*

(iii) *Gunakan Teorem Chebyshev untuk menentukan selang yang mengandungi sekurang-kurangnya 20% daripada keluarga.*

(b) *Peristiwa A dan B adalah sedemikian rupa sehingga* $P(A) = \frac{2}{5}$ *and* $P(B) = \frac{1}{4}$ *dan* $P[(A \cap \overline{B}) \cup (\overline{A} \cap B)] = \frac{1}{6}$. *Cari* $P(A \cap B)$ *dan tentukan samada A dan B adalah tak bersandar.*

(c) *Sebuah kilang pengeluar alat elektronik menerima 40% bekalan komponen elektronik dari pembekal A, 25% daripada pembekal B, dan 35% daripada pembekal C. Peratusan komponen elektronik yang cacat yang dibekalkan oleh pembekal A, B and C masing-masing ialah 5%, 2% dan 1%.*

(i) *Cari kebarangkalian satu komponen elektronik yang dipilih secara rawak daripada semua pembekal yang diterima oleh kilang itu didapati cacat.*

(ii) *Cari kebarangkalian bahawa daripada dua komponen elektronik yang dipilih secara rawak daripada kesemua pembekal; sekurang-kurangnya satu komponen elektronik cacat,*

(iii) *Sekiranya satu komponen elektronik cacat dipilih, apakah kebarangkalian bahawa komponen itu daripada pembekal B?*

(d) *Suatu pembolehubah rawak selanjar X mempunyai fungsi ketumpatan kebarangkalian:*

$$f(x) = \begin{cases} k, & 2 < x < 3 \\ k(x-2), & 3 \le x < 4 \\ 0, & \text{selainya} \end{cases}$$

iaitu k ialah pemalar.

(i) *Tentukan nilai k*
(ii) *Hitung* $P(2.5<X<3.5)$
(iii) *Tentukan nilai median bagi X.*

[100 markah]

2. (a) An accounts clerk is given the job of counting the overtime claims of workers in a company. The probability that the clerk makes a mistake in calculating a claim payment is 0.2.
   (i) Find the probability that exactly two out of five claims are faulty.
   (ii) Find the probability that at least two out of five claims are faulty.
   (iii) After a week on the job, the clerk performs better and the probability of making a faulty claim becomes 0.15. Using a suitable approximation, find the probability, that in a sample of 120 claims, there are less than 15 faults.
   (iv) After another week on the job, the clerk makes mistake in the calculation with probability of 0.05. Using a suitable approximation, find the probability that in a sample of 50 claims, there are at most 2 faults.

   (b) The time it takes an ordinary mouse to run through a particular maze is assumed to be normally distributed with a mean of 15 seconds and a standard deviation of 3 seconds.
   (i) What percent of the mice have times between 10 and 20 seconds?
   (ii) The top 10% in speed will be selected for another experiment. What time do they have to beat to be selected for the next experiment?
   (iii) Suppose 4 mice are selected at random, what is the probability that the total length of time for the 4 mice will be less than 55 seconds?

   (c) Two independent samples with means $\overline{X}_1$ and $\overline{X}_2$ with sizes 30 and 35 respectively are taken from two infinite populations which have Poisson distributions with means $\mu_1 = 9$ and $\mu_2 = 10$
   (i) Find the probability that the sample mean $\overline{X}_1$ will exceed $\overline{X}_2$.
   (ii) Find the probability that the difference between the two samples means will be less than 0.5

[100 marks]

2. (a) Seorang kerani akaun diberi tugas mengira bayaran tuntutan lebih masa pekerja sebuah syarikat. Kebarangkalian kerani itu membuat kesilapan penghitungan suatu bayaran tuntutan ialah 0.2.
   (i) Cari kebarangkalian tepat dua daripada lima bayaran tuntutan silap dikira.
   (ii) Cari kebarangkalian sekurang-kurangnya dua daripada lima bayaran tuntutan silap dikira.

.../5-

(iii) *Selepas seminggu kerani itu menjalankan tugas tersebut, prestasinya bertambah baik dan kebarangkalian dia membuat kesilapan penghitungan suatu bayaran tuntutan menjadi 0.15. Dengan menggunakan penghampiran yang sesuai, cari kebarangkalian, dalam satu sampel 120 bayaran tuntutan, terdapat kurang daripada 15 bayaran tuntutan yang silap dikira.*

(iv) *Selepas seminggu lagi, kebarangkalian kerani itu membuat kesilapan mengira suatu bayaran tuntutan berkurang menjadi 0.05. Dengan menggunakan penghampiran yang sesuai, cari kebarangkalian, dalam satu sampel 50 bayaran tuntutan, terdapat selebih-lebihnya 2 bayaran tuntutan yang silap dikira.*

(b) *Masa yang lazimnya diambil oleh tikus untuk berlari di sepanjang rangkaian lorong dianggap bertaburan normal dengan min 15 saat dan sisihan piawai 3 saat.*

(i) *Berapa peratus tikus yang mengambil masa antara 10 dan 20 saat?*

(ii) *Kelajuan mereka yang 10% teratas akan dipilih untuk ujikaji lain. Apakah masa yang mereka perlu atasi agar dipilih untuk ujikaji selanjutnya?*

(iii) *Andaikan 4 ekor tikus dipilih secara rawak, apakah kebarangkalian bahawa jumlah masa bagi 4 ekor tikus tersebut akan kurang dari 55 saat?*

(c) *Dua sampel rawak tak bersandar yang masing-masing mempunyai min* $\overline{X}_1$ *dan* $\overline{X}_2$ *serta bersaiz 30 dan 35 diambil daripada dua populasi tak terhingga yang bertabur secara Poisson dengan min* $\mu_1 = 9$ *dan* $\mu_2 = 10$.

(i) *Cari kebarangkalian bahawa min sampel* $\overline{X}_1$ *melebihi* $\overline{X}_2$.

(ii) *Cari kebarangkalian bahawa beza antara dua min sampel adalah kurang daripada 0.5*

[100 markah]

3. (a) The mass of a certain chocolate bar produced in a factory has a normal distribution. A random sample of 160 chocolate bars is weighed. The mass of each chocolate bar, x grams, is recorded. The results are summarized by $\sum x = 800$ and $\sum x^2 = 4640$.

(i) Calculate the unbiased estimate for the population mean μ and the population variance σ^2. Obtain a 95% confidence interval for the population mean.

(ii) When the chocolate bars in the sample are examined, it is found that 20 of them are contaminated. Calculate a 90% confidence interval for the population proportion, p of chocolate bars that are not contaminated.

(iii) Suppose k such samples are taken. The 90% confidence interval for p for each sample is constructed. What is the expected least value of k if the factory wishes to have 180 of these intervals to contain p?

(b) A type of steel rod produced by Company A is known to have an average length of 36 cm and standard deviation 0.05 cm if the machine used to produce the rod functions properly. The quality control section of the company takes a sample of 40 rods every week to calculate the mean length and test the null hypothesis H_0: $\mu = 36$ cm as compared to the alternative hypothesis H_A: $\mu \neq 36$ cm at 1% level of significance. If H_0 is rejected, the machine will be stopped for inspection. A sample of 40 rods is inspected and produced a mean length of 36.015 cm. Based on the sample, will the machine be stopped for inspection?

(c) To compare two programs for training industrial workers to perform a skilled job, 20 workers are included in an experiment. Of these, 10 are selected at random to be trained by method 1; the remaining 10 workers are to be trained by method 2. After completion of training, all the workers are subjected to a time-and-motion test that records the speed of performance of a skilled job. The following data are obtained:

	Time (in minutes)									
Method 1	15	20	11	23	16	21	18	16	27	24
Method 2	23	31	13	19	23	17	28	26	25	28

(i) Can you conclude from the data that the mean time for the job is significantly less after training with method 1 than after training with method 2? Test at $\alpha = 0.05$.

(ii) State the assumptions for the population distributions.

(iii) Construct a 95% confidence interval for the population mean difference in job times between the two methods.

[100 marks]

3. (a) *Jisim seketul coklat tertentu yang dihasilkan dari kilang bertaburan normal. Satu sampel rawak yang terdiri daripada 160 ketul coklat telah ditimbang. Jisim bagi setiap ketul coklat, x gram, telah direkodkan. Keputusannya telah diringkaskan dengan* $\Sigma x = 800$ *dan* $\Sigma x^2 = 4640$.

(i) *Hitung anggaran yang saksama bagi min populasi* μ *dan varians* σ^2. *Bina selang keyakinan 95% bagi min populasi.*

(ii) *Apabila ketulan coklat dari sampel telah diteliti, didapati bahawa 20 daripadanya telah dicemari. Hitung selang keyakinan 90% bagi kadaran populasi, p ketulan coklat yang tidak dicemari.*

(iii) *Katakan k sampel telah diambil. Selang keyakinan 90% bagi p untuk setiap sampel telah dibina. Sekurang-kurangnya apakah nilai k yang dijangka jika kilang tersebut berhasrat untuk mempunyai 180 selang yang mengandungi p?*

(b) *Sejenis rod besi yang dihasilkan oleh Syarikat A diketahui mempunyai panjang purata 36 cm dengan sisihan piawainya 0.05 cm jika mesin digunakan untuk membuat rod besi itu berfungsi dengan baik. Bahagian kawalan kualiti dalam syarikat itu mengambil suatu sampel rawak 40 batang rod besi pada setiap minggu, menghitung min panjang rod-rod besi itu, dan menguji hipotesis nol* H_0: $\mu = 36$ *cm berlawanan dengan hipotesis alternative* H_A: $\mu \neq 36$ *cm pada aras keertian 1%. Jika* H_0 *ditolak, mesin itu akan dihentikan untuk penyelarasan. Suatu sampel 40 barang rod besi yang diperiksa menghasilkan min panjang 36.015 cm. Berdasarkan sampel itu, adakah mesin tersebut memerlukan penyelarasan?*

(c) *Untuk membandingkan dua program bagi melatih pekerja latihan industri untuk melaksanakan suatu pekerjaan dengan mahir, 20 orang pekerja telah dipilih dalam suatu ujikaji. Daripada jumlah ini , 10 pekerja telah dipilih secara rawak untuk dilatih dengan kaedah 1, manakala selebihnya dilatih dengan kaedah 2. Selepas tamat latihan, kesemua pekerja menjalani ujian masa dan gerak-geri yang mencatatkan kelajuan pelaksanaan pekerjaan tersebut. Data yang berikut diperoleh:*

	Masa (minit)									
Kaedah 1	15	20	11	23	16	21	18	16	27	24
Kaedah 2	23	31	13	19	23	17	28	26	25	28

(i) *Bolehkah anda simpulkan daripada data bahawa min masa untuk melaksanakan pekerjaan tersebut setelah dilatih dengan kaedah 1 adalah kurang daripada setelah dilatih dengan kaedah 2. Uji pada* $\alpha = 0.05$.

(ii) *Nyatakan anggapan untuk taburan populasi.*

(iii) *Bina selang keyakinan 95% bagi perbezaan min populasi masa melaksanakan pekerjaan antara dua kaedah.*

[100 markah]

4. (a) Two different chemicals are being tested for use on the nylon mantles used in camping lanterns. Evidence shows that each chemical produces about the same average burn time. The one with the smaller variance is the more desirable since it produces a more dependable mantle. The following observations are obtained on the length of life in hours of mantles treated with the respective chemicals.

Chemical X: $\sum x = 198.88$, $\sum x^2 = 4173.952$, $n_X = 11$

Chemical Y: $\sum y = 228.8$, $\sum y^2 = 4485.9966$, $n_Y = 13$

(i) Is there sufficient evidence to conclude that the true standard deviation of the burn time for mantles produced using chemical Y exceed 5 hours (use $\alpha = 0.05$) ?

(ii) Find a 90% confidence interval on σ_X^2 / σ_Y^2 .

(iii) Based on part (ii) which chemical would you suggest to be used? Explain briefly.

(b) A company produces steel sheets. The manufacturing process is said to be in control if 92% of the steel sheets are without fault, 5% has a fault and 3% has 2 or more faults. The quality control inspector regularly checks the steel sheets for any fault. If the distribution of the fault in a sample differs significantly from the distribution of the percentage mentioned, the process is stopped and an inspection is made. From a random sample of 300 steel sheets taken, the following frequency distribution for faults is obtained:

Number of faults	None	One	Two or more
Number of steel sheets	262	24	14

At α = 1%, does the sample above gives enough evidence to suggest the process needs to be inspected?

(c) A study is conducted to examine the relationship between relative humidity (X), where $28\% \le X \le 75\%$ and the percentage of evaporation of the solvent in paint while spray painting (Y). These data are obtained:

$$\sum_{i=1}^{25} x^2 = 76308.53\ ,\ \sum_{i=1}^{25} x = 1314.90\ ,\ \sum_{i=1}^{25} xy = 11824.44\ ,\ \sum_{i=1}^{25} y = 235.7\ ,$$
$$\sum_{i=1}^{25} y^2 = 2322.96$$

(i) Find the estimated line of regression.

(ii) Test: $H_0 : \beta_1 = 0$
$H_A : \beta_1 \neq 0$ at $\alpha = 0.01$.

(iii) Predict the percentage evaporation of the paint solvent on a day in which the relative humidity is 50%.

(iv) Would the model be useful in predicting the percentage of evaporation if the relative humidity is 85%? Explain.

(v) Estimate the value of the coefficient of determination. Explain your result.

[100 marks]

4. (a) *Dua bahan kimia yang berbeza telah diuji untuk kegunaan bagi menutupi nilon yang digunakan bagi lampu perkhemahan. Terdapat bukti menunjukkan bahawa setiap bahan kimia menghasilkan hampir sama purata masa membakar. Varians yang lebih kecil adalah dikehendaki kerana ia boleh menghasilkan lebihan tutupan pada nilon. Hasil dari pengawasan telah diperoleh panjang masahayat dalam jam bagi proses menutupi nilon dengan bahan kimia masing-masing.*

Bahan Kimia X:	$\Sigma x = 198.88$	$\Sigma x^2 = 4173.952$	$n_X = 11$
Bahan Kimia Y:	$\Sigma y = 228.8$	$\Sigma y^2 = 4485.9966$	$n_Y = 13$

...**/9-**

(i) *Adakah terdapat bukti yang mencukupi untuk menyimpulkan bahawa sisihan piawai yang sebenar bagi masa membakar menutupi nilon yang dihasilkan dengan menggunakan bahan kimia Y melebihi 5 jam (guna* $\alpha = 0.05$*)?*

(ii) *Cari selang keyakinan 90% bagi* σ_X^2 / σ_Y^2

(iii) *Bahan kimia yang manakah yang anda cadangkan? Jawapan anda mesti berdasarkan jawapan pada bahagian (ii). Terangkan secara ringkas.*

(b) *Suatu syarikat membuat kepingan-kepingan besi. Proses pembuatan kepingan-kepingan besi dikatakan dalam keadaan terkawal jika 92% daripada kepingan besi tiada kecacatan, 5% daripadanya mempunyai satu kecacatan dan 3% daripadanya mempunyai 2 atau lebih kecacatan. Pemeriksa kawalan mutu dalam syarikat itu kerap mengambil sampel-sampel rawak kepingan besi untuk diperiksa bagi sebarang kecacatan. Jika taburan bagi kecacatan untuk suatu sampel adalah berbeza secara bererti daripada taburan peratusan yang dinyatakan itu, proses dihentikan dan penyelenggaraan dibuat. Daripada suatu sampel rawak 300 kepingan besi yang diambil, berikut adalah taburan kekerapan bagi kecacatan:*

Bilangan kecacatan	*Tiada*	*Satu*	*Dua atau lebih*
Bilangan kepingan besi	262	24	14

Pada $\alpha = 1\%$*, adakah sampel tersebut memberikan bukti yang cukup untuk menyatakan bahawa proses itu memerlukan penyelenggaraan?*

(c) *Satu kajian dijalankan untuk mengkaji perhubungan antara kelembaban relatif (X), dengan* $28\% \leq X \leq 75\%$ *dan peratusan penyejatan dari larutan mengecat semasa semburan pada lukisan (Y). Data yang berikut diperoleh:*

$$\sum_{i-1}^{25} x^2 = 76308.53 \quad \sum_{i-1}^{25} x = 1314.90 \quad \sum_{i-1}^{25} xy = 11824.44 \quad \sum_{i-1}^{25} y = 235.7 \quad \sum_{i-1}^{25} y^2 = 2322.96$$

(i) *Cari anggaran garis regresi.*

(ii) *Uji:* $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$ *pada aras* $\alpha = 0.01$.

(iii) *Ramal peratusan penyejatan dari larutan mengecat pada hari kelembaban relatifnya ialah 50%.*

(iv) *Bolehkah model digunakan dalam menganggar peratusan penyejatan jika kelembaban relatif ialah 85%? Jelaskan.*

(v) *Anggarkan nilai pekali penentuan. Jelaskan jawapan anda.*

[100 markah]

APPENDIX

FORMULA
MAA 161E – Statistics For Science Students

Confidence Interval:

$$\overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\overline{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$\overline{d} \pm t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n_d}}$$

$$b \pm t_{\frac{\alpha}{2}} s_b$$

$$(\overline{X}-\overline{Y}) \pm t_{\alpha/2}\, S_p \sqrt{\frac{1}{n_x}+\frac{1}{n_y}}$$

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$(\hat{p}_x-\hat{p}_y) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x}+\frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

$$(\overline{X}-\overline{Y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_x^{\,2}}{n_x}+\frac{\sigma_y^{\,2}}{n_y}}$$

$$(\overline{X}-\overline{Y}) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_x^{\,2}}{n_x}+\frac{s_y^{\,2}}{n_y}}$$

$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$$

$$\left(\frac{s}{1+\frac{Z_{\frac{\alpha}{2}}}{\sqrt{2n}}}, \frac{s}{1-\frac{Z_{\frac{\alpha}{2}}}{\sqrt{2n}}}\right)$$

$$\left(\frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2},(v_2,v_1)}, \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2},(v_2,v_1)}\right)$$

Test Statistic:

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

$$T = \frac{\overline{X}-\mu}{s/\sqrt{n}}$$

$$T = \frac{\overline{d}-\mu_d}{s_d/\sqrt{n_d}}$$

$$T = \frac{b-\beta_1}{s_b}$$

$$T = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$Z = \frac{s-\sigma}{\sigma/\sqrt{2n}}$$

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^{\,2}}{n_x}+\frac{\sigma_y^{\,2}}{n_y}}}$$

$$T = \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{S_p^{\,2}\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}$$

$$S_p^{\,2} = \frac{(n_x-1)s_x^{\,2}+(n_y-1)s_y^{\,2}}{n_x+n_y-2}$$

$$F = \frac{s_x^2}{s_y^2}$$

$$Z = \frac{(\hat{p}_x-\hat{p}_y)-(p_x-p_y)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}$$

$$T = \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$$

$$dk = \frac{\left(\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}\right)^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x-1}+\frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y-1}}$$

Regression / Correlation Analysis

$$s_e = \sqrt{\frac{S_{YY} - bS_{XY}}{n-2}} \quad ; \quad s_b = \frac{s_e}{\sqrt{S_{XX}}} \quad ; \quad r = \frac{S_{XY}}{\sqrt{S_{XX}.S_{YY}}}$$

- ooo O ooo -