
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2011/2012 Academic Session

June 2012

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

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **all eight** [8] questions.

[Arahan: Jawab **semua lapan** [8] soalan.]

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

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

1. (a) Briefly explain the difference between a census and a sample survey. Why is conducting a sample survey preferable to conducting a census?
- (b) Here are the data prices of 12 used upright pianos in the classified section of newspaper.

2100 500 100 1100 2200 650 1200 450 1200 1600 695 2300

- (i) Find the mean, variance and standard deviation. Are the values of these summary measure population parameters or sample statistics?
- (ii) Prepare a box-and-whisker plot.
- (iii) Would the distribution best be described as left-skewed, having low outliers, fairly symmetric, right-skewed, or having high outliers?

[15 marks]

1. (a) *Terangkan secara ringkas perbezaan antara banci dan tinjauan sampel. Mengapa menjalankan tinjauan sampel lebih diutamakan daripada banci?*

- (b) *Berikut adalah harga data untuk 12 buah piano terpakai yang dikelaskan daripada satu surat khabar.*

2100 500 100 1100 2200 650 1200 450 1200 1600 695 2300

- (i) *Cari min, varians dan sisihan piawai. Adakah nilai-nilai ini ukuran ringkasan untuk parameter populasi atau statistik sampel?*
- (ii) *Sediakan satu plot “kotak dan misai”.*
- (iii) *Adakah taburan tersebut boleh digambarkan sebagai pencong ke kiri, mempunyai titik terpencil rendah, agak simetri, pencong ke kanan atau mempunyai titik terpencil tinggi?*

[15 markah]

2. A random sample of 400 college students was asked if college athletes should be paid. The following table gives a two-way classification of the responses:

	Should Be paid	Should Not Be Paid
Student athlete	90	20
Student nonathlete	210	90

- (a) If one student is randomly selected from these 400 students, find the probability that this student
- (i) is in favour of paying college athletes,
 - (ii) favours paying college athletes given that the student selected is a nonathlete,
 - (iii) is an athlete and favour paying student athletes,
 - (iv) is a nonathlete or is against paying student athletes.
- (b) Are the events “student athlete” and “should be paid” independent? Are they mutually exclusive? Explain why or why not.

[10 marks]

2. Satu sampel rawak 400 orang pelajar kolej telah ditanyai sama ada atlet kolej perlu dibayar. Jadual berikut memberikan pengelasan dua hala untuk maklum balas tersebut:

	Perlu Dibayar	Tidak Perlu Dibayar
Pelajar Atlet	90	20
Pelajar Bukan Atlet	210	90

- (a) Jika seorang pelajar dipilih secara rawak daripada 400 orang pelajar, cari kebarangkalian bahawa pelajar ini
- (i) memihak kepada membayar atlet kolej,
 - (ii) memihak kepada membayar atlet kolej jika diketahui bahawa pelajar yang ditayai adalah pelajar bukan atlet,
 - (iii) adalah seorang atlet dan memihak kepada membayar pelajar atlet,
 - (iv) bukan seorang atlet atau tidak memihak kepada membayar pelajar atlet.

- (b) Sama ada peristiwa "pelajar atlet" dan "perlu dibayar" tak bersandar? Adakah dua peristiwa itu saling eksklusif? Jelaskan mengapa atau mengapa tidak.

[10 markah]

3. The number of theft cases reported every day at a small police station has a Poisson distribution with mean 2.

- (a) Show that the probability that, on a particular day, at least one theft is reported is 0.865.
- (b) Find the probability that the total number of theft cases reported on 3 particular days is at least 3.
- (c) Find the probability that for 5 particular days, no theft cases were reported on 2 days.
- (d) Find the probability that, for 100 particular days, at least one theft is reported on less than 80 days.

[10 marks]

3. Bilangan kes kecurian yang dilaporkan setiap hari di sebuah balai polis yang kecil mempunyai taburan Poisson dengan min 2.

- (a) Tunjukkan bahawa kebarangkalian, pada suatu hari tertentu, sekurang-kurangnya satu kecurian dilaporkan adalah 0.865.
- (b) Cari kebarangkalian bahawa jumlah bilangan semua kecurian yang dilaporkan pada 3 hari tertentu adalah sekurang-kurangnya 3.
- (c) Cari kebarangkalian bahawa, untuk 5 hari tertentu, tiada kes kecurian dilaporkan pada tepat 2 hari.
- (d) Cari kebarangkalian bahawa, untuk 100 hari tertentu, sekurang-kurangnya satu kes kecurian dilaporkan pada kurang daripada 80 hari.

[10 markah]

4. A random variable X is normally distributed with mean 10 and standard deviation 2.

- (a) Find $P(X > 13)$ and $P(9 < X < 13)$.
- (b) Find the possible values of a if the probability that X lies between a and 12 is 0.15.
- (c) If three values of X are chosen at random, calculate the probability that two of the values are less than 12 and the other is more than 12.

[15 marks]

4. Suatu pembolehubah rawak X bertaburan normal dengan min 10 dan sisihan piawai 2.

- (a) Cari $P(X > 13)$ dan $P(9 < X < 13)$.
- (b) Cari nilai-nilai yang mungkin bagi a jika kebarangkalian bahawa X terletak di antara a dan 12 ialah 0.15.
- (c) Jika tiga nilai X dipilih secara rawak, hitungkan kebarangkalian bahawa dua nilai tersebut adalah kurang daripada 12 dan yang lain adalah lebih daripada 12.

[15 markah]

5. A group of veterinarians wants to test a new canine vaccine for Lyme disease. (Lyme disease is transmitted by the bite of an infected deer tick.) In an area that has a high incidence of Lyme disease, 100 dogs are randomly selected (with their owner's permission) to receive the vaccine. Over a 12 months period, these dogs are periodically examined by veterinarians for symptoms of Lyme disease. At the end of 12 months, 10 of these 100 dogs are diagnosed with the disease. During the same 12-months period, 18% of the unvaccinated dogs in the area have been found to have Lyme disease. Let p be the proportion of all potential vaccinated dogs who would contract Lyme disease in this area.

- (a) Find 95% confidence interval for p .
- (b) Does 18% lie within your confidence interval of part a? Does this suggest the vaccine might or might not be effective to some degree?
- (c) Write a brief critique of this experiment, pointing out anything that may have distorted the results or conclusions.

[8 marks]

5. *Sekumpulan doktor veterinar mahu menguji vaksin baru untuk penyakit Lyme. (Penyakit Lyme anjing yang disebarkan melalui gigitan pijat rusa yang dijangkiti.) Di kawasan yang mempunyai insiden penyakit Lyme yang tinggi, 100 anjing yang dipilih secara rawak (dengan kebenaran pemilik mereka) untuk menerima vaksin. Sepanjang tempoh 12 bulan, anjing-anjing ini diperiksa secara berkala oleh doktor veterinar untuk gejala penyakit Lyme. Pada akhir 12 bulan, 10 daripada 100 anjing itu didiagnosis untuk penyakit itu. Sepanjang tempoh 12 bulan yang sama, 18% daripada anjing-anjing yang tidak divaksinasi di kawasan tersebut telah didapati mempunyai penyakit Lyme. Katakan p ialah kadaran semua anjing divaksin yang berpotensi yang akan menghidap penyakit Lyme di kawasan ini.*

- (a) Cari selang keyakinan 95% bagi p .*
- (b) Adakah 18% terletak dalam selang keyakinan anda dalam bahagian a? Adakah ini mencadangkan vaksin ini mungkin atau mungkin tidak berkesan ke tahap yang tertentu?*
- (c) Rumuskan kritikan ringkas tentang eksperimen ini, nyatakan sebarang sebab yang mungkin memutarbelitkan keputusan atau kesimpulan.*

[8 markah]

6. Employees of a large corporation are concerned about the declining quality of medical service provided by their group health insurance. A random sample of 100 office visits by employees of this corporation to primary care physicians during 2004 found that the doctors spent an average of 19 minutes with each patient. This year a random sample of 108 such visits showed that doctors spent an average of 15.5 minutes with each patient. Assume that the standard deviations for the two populations are 2.7 and 2.1 minutes, respectively.

- (a) Construct a 95% confidence interval for the difference between the two population means for these two years.
- (b) Using the 2.5% level of significance, can you conclude that the mean spent by doctors with each patient is lower for this year than for 2004?
- (c) What would your decision be in part b if the probability of making a Type I error were zero? Explain.

[15 marks]

6. *Pekerja-pekerja sebuah syarikat besar mengambil berat tentang kemerosotan kualiti perkhidmatan perubatan yang disediakan oleh insurans kesihatan kumpulan mereka. Satu sampel rawak 100 lawatan ke pejabat penjagaan pakar perubatan utama oleh pekerja perbadanan ini sepanjang tahun 2004 mendapati bahawa doktor memeriksa setiap pesakit dengan purata 19 minit. Pada tahun ini, satu sampel rawak 108 lawatan itu menunjukkan bahawa doktor memeriksa setiap pesakit dengan purata 15.5 minit. Andaikan bahawa sisihan piawai bagi dua populasi adalah masing-masing 2.7 dan 2.1 minit.*
- (a) *Bina satu selang keyakinan 95% bagi perbezaan antara dua min populasi bagi kedua-dua tahun.*
- (b) *Menggunakan aras keertian 2.5%, bolehkah anda membuat kesimpulan bahawa min yang digunakan oleh doktor-doktor untuk setiap pesakit adalah lebih rendah bagi tahun ini berbanding dengan tahun 2004?*
- (c) *Apakah keputusan anda dalam bahagian b jika kebarangkalian untuk ralat Jenis I adalah sifar? Jelaskan.*

[15 markah]

7. A survey of 1000 drivers this year showed that 29% of the people send text messages while driving. Last year a survey of 1000 drivers showed that 17% of those sent text messages while driving. At $\alpha = 0.01$, can it be concluded that there has been an increase in the number of drivers who text while driving?

[12 marks]

7. *Satu kajian 1000 pemandu pada tahun ini menunjukkan bahawa 29% rakyat menghantar mesej teks semasa memandu. Pada tahun yang lepas, satu kajian 1000 pemandu menunjukkan bahawa 17% daripada mereka menghantar mesej teks ketika memandu. Pada $\alpha = 0.01$, bolehkah disimpulkan bahawa terdapat peningkatan dalam bilangan pemandu yang teks semasa memandu?*

[12 markah]

8. Each person in a random sample of 500 registered voters (200 urban, 200 suburban, and 100 rural residents) was asked his or her opinion about the governor's proposed legislation. Does the sample evidence shown in the table below support the hypothesis "Voters within the different residence groups have different opinion about the governor's proposal"? Use $\alpha = 0.05$.

Residence	Governor's Proposal	
	Favour	Oppose
Urban	143	57
Suburban	98	102
Rural	13	87

[15 marks]

8. Setiap orang dalam satu sampel rawak 500 pengundi berdaftar (200 di bandar, 200 di pinggir bandar, dan 100 di luar bandar) telah diminta pendapat mereka mengenai perundangan yang dicadangkan oleh gabenor. Adakah sampel ini yang ditunjukkan di dalam jadual di bawah mempunyai bukti yang cukup untuk menyokong hipotesis "Pengundi di dalam kumpulan kediaman yang berlainan mempunyai pendapat yang berbeza mengenai cadangan Gabenor"? Gunakan $\alpha = 0.05$.

Tempat kediaman	Cadangan Gabenor	
	Menyokong	Menentang
Bandar	143	57
Pinggir Bandar	98	102
Luar Bandar	13	87

[15 markah]

APPENDIX

$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x}{n}$ $s^2 = \frac{\sum (x^2 f) - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}$ $= \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$	$S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $\bar{p} = \frac{X + Y}{n_x + n_y}$
Confidence Intervals: 1. $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ 2. $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ 3. $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$ 4. $(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$	5. $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}$ 6. $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1 - \hat{p}_x)}{n_x} + \frac{\hat{p}_y(1 - \hat{p}_y)}{n_y}}$ 7. $\frac{(n-1)s^2}{\chi_{\alpha/2}^2}$ to $\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}$

Test Statistics:

$$1. Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ or } Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$2. T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$3. T = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$$

$$4. Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$5. Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

$$6. T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$$

$$7. T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

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

$$8. Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}}$$

$$9. Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\bar{p}(1-\bar{p}) \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$$

$$10. \chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$11. F = \frac{s_1^2}{s_2^2}$$

$$12. \chi^2 = \sum \frac{(O - E)^2}{E}, \quad E = np$$