THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT1302/STAT2602 PROBABILITY AND STATISTICS II

December 19, 2015 **Time: 9:30 a.m. - 11.30 a.m.**

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Answer ALL NINE questions. Marks are shown in square brackets.

You may assume that a population is normally distributed when necessary.
Final numerical answers should be either exact or correct to 4 decimal places unless otherwise specified.

1. The table below records the sales of a convenience food, in thousands of dollars, for each of five regions during a month before and a month after an advertising campaign respectively.

   (a) Determine, at the 5% level of significance, whether an increase in the mean amount of one-month sales has occurred after the advertising campaign. **[10 marks]**

   (b) Construct a 95% confidence interval (correct to 3 decimal places) for the difference between the mean amounts of one-month sales before and after the advertising campaign respectively. **[4 marks]**

Region	A	B	C	D	E	Sample mean	Sample variance
Sales before campaign	2.4	2.6	3.9	2.0	3.2	2.82	0.552
Sales after campaign	3.0	2.5	4.0	4.1	4.8	3.68	0.847

[Total: 14 marks]

2. Suppose a population follows a distribution which has a density function

$$f(x;\theta) = (\ln\theta)\theta^{-x}1_{(0,+\infty)}(x), \qquad \text{where } \theta \text{ is a constant greater than 1.}$$

   (a) Find an estimator of θ by the method of moments. **[4 marks]**

   (b) Find an estimator of θ by the method of maximum likelihood. **[7 marks]**

   (c) Find a sufficient statistic for θ. **[4 marks]**

   (d) Find the Cramér-Rao lower bound for estimating θ. **[6 marks]**

[Total: 21 marks]

3. A study is to be made to estimate the proportion of residents in a certain city who favour the construction of a nuclear power plant.

   (a) How large a sample is needed if one wants the sample proportion to be, with probability 0.99, within 0.03 of the true proportion of residents in the city who favour the construction of the nuclear power plant? **[4 marks]**

   (b) A sample of 1000 residents is drawn and 234 of them favour the construction of the nuclear power plant. Someone claims that less than 25% of the residents in the city favour the construction of the nuclear power plant. Using p-value, conclude at the 10% level of significance whether the claim is acceptable. **[9 marks]**

[Total: 13 marks]

4. A random sample of fifteen workers from a vacuum flask assembly line was selected and each of these workers was asked to assemble a vacuum flask. The times (in seconds) taken to complete these tasks are given below.

109.2	146.2	127.9	92.0	108.5	91.1	109.8	114.9
115.3	99.0	112.8	130.7	141.7	122.6	119.9	

(a) Construct a 90% confidence interval (correct to integers) for the variance of the time taken to assemble a randomly selected vacuum flask. **[6 marks]**

(b) At the 10% level of significance, will you conclude that the standard deviation of the time taken to assemble a randomly selected vacuum flask is 20? **[4 marks]**

[Total: 10 marks]

5. Suppose there is a population with a density function $\dfrac{x^{\delta-1}e^{-x/\theta}}{\theta^{\delta}\,\Gamma(\delta)}1_{(0,+\infty)}(x)$, where δ and θ are positive constants, and δ is known. Suppose we want to test

$$H_0\colon \theta = \theta_0 \quad \text{against} \quad H_1\colon \theta = \theta_1, \quad \text{where } 0 < \theta_1 < \theta_0.$$

(a) Derive a likelihood ratio test based on a sample of size n. **[4 marks]**

(b) Suppose $\delta = 2$, $\theta_0 = 5$ and $\theta_1 = 4$. A sample $\{1, 5, 6\}$ is drawn from the population. Based on this sample and the result of (a), conclude whether H_0 should be rejected at the 5% level of significance. [Hint: use chi-square distributions.] **[6 marks]**

[Total: 10 marks]

6. Suppose X_1, X_2 and X_3 constitute a random sample from a population with mean μ and variance σ^2. Suppose we want to estimate μ.

(a) Determine the value of k such that $\dfrac{2X_1 - X_2 + X_3}{k}$ is an unbiased estimator of μ. **[2 marks]**

(b) Find the efficiency of $\dfrac{2X_1 - X_2 + X_3}{k}$ relative to $\dfrac{X_2 + X_3}{2}$. **[4 marks]**

(c) Which of these two estimators is relatively more efficient? **[1 mark]**

[Total: 7 marks]

7. Five brands of orange juice are displayed side by side in several supermarkets in a large city. It was noted that, in one day, 32 customers picked Brand A, 40 picked Brand B, 25 picked Brand C, 35 picked Brand D and 48 picked brand E. In this city, can you conclude at the 10% significance level that there is a preferred brand of orange juice?

[Total: 9 marks]

8. A statistics professor wants to select a statistical software package for her course. One of the most important features, according to the professor, is the ease with which students learn to use the software. She has narrowed the selection to two possibilities: software A, a menu-driven statistical package with some high-powered techniques, and software B, a spread sheet that has the capability of performing most techniques. She asks statistics students selected at random to choose one of the two packages. She gives each student a statistics problem to solve by computer and the appropriate manual. The amount of time (in minutes) each student needs to complete the assignment was recorded and the summaries are as follows.

Software	Sample size	Sample mean	Sample standard deviation
A	23	74.91	24.53
B	15	53.40	8.58

Can the professor conclude from these data that the two software packages differ in the amount of time needed to learn how to use them? (Use the 1% significance level.)

[Total: 9 marks]

9. Suppose 25% of bottled-water brands fill their bottles with just tap water. Using normal approximation, find the probability that, in 65 randomly selected bottled-water brands, 20 or more contain just tap water.

[Total: 7 marks]

Formulae

$\mathrm{M}_X(t) = \mathrm{E}(e^{tX})$. $\quad \left.\left(\dfrac{\mathrm{d}^r}{\mathrm{d}t^r}\mathrm{M}_X(t)\right)\right|_{t=0} = \mu'_r$.

Normal population $\Rightarrow \quad \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, $\quad \dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim \mathrm{t}(n-1)$.

$\mathrm{b}(\hat{\Theta}) = \mathrm{E}(\hat{\Theta}) - \theta$. $\qquad \mathrm{E}([\hat{\Theta} - \theta]^2) = \mathrm{Var}(\hat{\Theta}) + [\mathrm{b}(\hat{\Theta})]^2$.

$\mathrm{I}(\theta) = \mathrm{E}\left(\left[\dfrac{\partial \ln \mathrm{f}(X;\theta)}{\partial \theta}\right]^2\right) = \mathrm{E}\left(-\dfrac{\partial^2 \ln \mathrm{f}(X;\theta)}{\partial \theta^2}\right)$. $\qquad \mathrm{Var}(\hat{\Theta}) \geq \dfrac{1}{n\,\mathrm{I}(\theta)}$.

$\mathrm{f}(x_1, x_2, \ldots, x_n; \theta) = \mathrm{g}(\mathrm{u}(x_1, x_2, \ldots, x_n), \theta) \cdot \mathrm{h}(x_1, x_2, \ldots, x_n)$.

$\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$. $\quad \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$. $\quad \dfrac{(n-1)S^2}{\sigma_0^2}$. $\quad \dfrac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$.

$\dfrac{\overline{X}-\overline{Y}-\delta}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$. $\quad \dfrac{\overline{X}-\overline{Y}-\delta}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$, $S_p^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

$\dfrac{\overline{X}-\overline{Y}-\delta}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$, $df \approx \left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2 \bigg/ \left[\dfrac{(s_1^2/n_1)^2}{n_1-1}+\dfrac{(s_2^2/n_2)^2}{n_2-1}\right]$. $\quad \dfrac{S_1^2}{S_2^2}$, $f_{1-\alpha,m,n} = \dfrac{1}{f_{\alpha,n,m}}$.

$\dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$, $\hat{p} = \dfrac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$. $-2\ln\Lambda \approx \sum_i \dfrac{(O_i-E_i)^2}{E_i} = \sum_i \dfrac{O_i^2}{E_i} - n$.

$\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$. $\quad \bar{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$. $\quad \left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$. $\quad \hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$. $\quad \bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2,n_1+n_2-2}\, s_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$. $\quad \bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2,df}\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$.

$\left(\dfrac{s_1^2}{s_2^2}\cdot\dfrac{1}{f_{\alpha/2,n_1-1,n_2-1}}, \dfrac{s_1^2}{s_2^2}\cdot f_{\alpha/2,n_2-1,n_1-1}\right)$. $\quad \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$.

Chi-Square Distribution Table ($\chi^2_{\alpha,df}$)

	α									
df	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.60
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.73	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58	32.80

************ END OF PAPER ************