OLLSCOIL NA hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II EXAMINATIONS 1998-99

MA 113, MA 228 – STATISTICS
PROBABILITY AND STATISTICS II

Dr. M.G. Kenward
Dr. J.N. Sheahan

Time Allowed: **Two** hours
Answer question 1 and any *two* of questions 2–4.
(Each of questions 2 – 4, but not necessarily parts therein, carries equal marks.)
This examination carries a weight of 46% of the course.

1.

Assume that the population of heights of people has (an approximately) normal distribution with unknown mean μ and standard deviation $\sigma = 3.0$.This question mainly concerns the z-test for testing the null hypothesis $H_0 : \mu = 68.0$ against the alternative hypothesis $H_1 : \mu \neq 68.0$ using $\alpha = 0.05$, and a confidence interval for μ. Note that three values from the standard normal tables that you should need somewhere below are: $z_{0.025} = 1.96$, $z_{0.0228} = 2.0$ and $z_{0.05} = 1.645$.

Consider using the z-test along with $\alpha = 0.05$ to test the alternatives $H_0 : \mu = 68.0$, $H_1 : \mu \neq 68.0$. Suppose that a random sample of $n = 9$ people was chosen, and their heights analysed by MINITAB. The output is as follows:

TEST OF $H_0 : \mu = 68.0$ versus $H_1 : \mu \neq 68.0$

n	mean	st dev	se mean	z-value	p-value
9	70.0	3.30	1.0	2.0	0.0456

a. Based on the printout above, **should** H_0 be rejected? Give briefly a reason for your answer.

b. **Show** how the z-value was calculated (from some other entries that precede it in the above table).

c. **Define** the p-value of any statistical test, and for the output above, **show** how the p-value can be calculated from the z-value.

d. **Define** the power of the z-test of the above alternatives at a value of μ different from 68.0. **Give**, *without* justification, answers to the following questions:

Is it necessary to be given the actual 9 data points (or at least their mean 70.0) to calculate the power of the test?

Is the power of the test at $\mu = 70.0$ larger than, smaller than, or equal to, the power at $\mu = 71.0$? Is the power of the test at $\mu = 70.0$ larger than, smaller than, or equal to, the power at $\mu = 66.0$?

e. **Derive** the formula $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ for a 95% confidence interval for μ, **calculate** this interval for the data above, and **interpret** it carefully.

2.

a. Recall that $X_1, X_2, ..., X_n$ is a random sample if these variables are independent and have the same distribution. For $n = 2$, explain briefly but clearly what X_1 and X_2 being *independent* and having the *same distribution* mean.

b. State the distribution of the mean $\bar{X}$ of a random sample if (i) the underlying random variable X has a $N(\mu, \sigma^2)$ distribution, and (ii) any distribution with mean μ and positive variance σ^2.

c. Define the mean $\bar{X}$ and variance S^2 of a random sample and show how the mean μ and variance σ^2 can be roughly defined in terms of these quantities.

d. The mean of a random sample of size 81 is used to estimate the mean of a normal population that has $\sigma = 27$. With what probability can we assert that the error of this estimate will be less than 9 (i.e. what approximately is $P(|\bar{X} - \mu| < 9)$?) if we use (i) Chebychev's inequality, (ii) the Central Limit Theorem.

Note : $z_{0.1587} = 1.0, z_{0.025} = 1.96, z_{0.0228} = 2.0, z_{0.0013} = 3.0$.

Turn over ⇒

3.

Imagine that you are a shipping magnate who wishes to purchase one of two brands of paint to apply to all your ships. Naturally, you will choose the paint that leads to less rust on average. Let μ_1 and μ_2 respectively denote the population mean amount of rust per square metre of ships' surface with brand 1 and brand 2. Suppose the alternatives to be tested using $\alpha = 0.05$ are $H_0 : \mu_1 = \mu_2$, and $H_1 : \mu_1 > \mu_2$.

The design used on your behalf involves applying brand 1 paint to 3 randomly chosen ships in Galway Harbour and brand 2 paint to 3 ships in Dublin Harbour. Data is then collected after a year by taking rust measurements (from square metres of the ships' surfaces.) The data are as follows:

	Amount of rust		
Brand 1 paint	45	46	47
Brand 2 paint	42	43	44

a. **State** the two assumptions concerning the populations of amounts of rust that should hold to justify the independent samples *t*-test used to test the above alternatives.

b. **Perform** in detail the independent samples *t*-test of the alternatives above, using $\alpha = 0.025$. To save you time, note that the two samples have the same variance! *Note*: One of the following critical points is relevant: $t_{4,0.025} = 2.776$, $t_{6,0.025} = 2.447$.

c. A colleague claimed that you could get more information about the comparison of the two types of paint if instead of the design used above you applied paint 1 to the port side of three ships in Galway Harbour and paint 2 to the starboard side of the *same* three ships. Briefly but clearly **explain** why your colleague is correct. You must answer this question in a mature way, both in the context of describing the factor that the paired design suggested by your colleague would block out, and the statistical sense in which the latter design is 'better' for a given level of significance.

4.

a A random sample of 200 Canadian voters was taken and each of them was classified according to political affiliation and also according to opinion concerning a particular current Canadian foreign-policy issue. The observed frequencies are as follows:

	OPINION ON POLICY			
POLITICAL AFFILIATION	**Approve**	**Do not approve**	**No opinion**	TOTAL
Conservative	**60**	**15**	**25**	100
Other	**40**	**35**	**25**	100
TOTAL	100	50	50	200

Use $\alpha = 0.05$ to test to see if, in the population of all Canadians, the variables "political affiliation" and "opinion on policy" are independent.

Note: One of the following critical points is relevant:

$$\chi^2_{1,\,0.05} = 3.84, \chi^2_{2,\,0.05} = 5.99, \chi^2_{3,\,0.05} = 7.81, \chi^2_{4,\,0.05} = 9.49.$$

b Concerning the linear regression model $\mu_{Y|x} = \alpha + \beta x$ relating a response random variable Y to a non-stochastic input variable x:

i Explain each of the three terms $\mu_{Y|x}$, α and β in the above population model.

ii Derive the formulae for the least squares estimates a and b of α and β based on n data points (x_i, y_i), $i = 1, 2, \ldots, n$.

J.N.S.