

OLLSCOIL NA hEIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER 1, 2001/2002

Data Management (EC515)

M.Sc. in Econ., Degree Examinations

Professor H. Dixon
Professor M.P. Cuddy
Mr. Eoghan Garvey

Time allowed: Two and a half hours.

Instructions: Answer 8 questions only - 4 from Section 1 and 4 from Section 2. Each question is worth equal marks. You are permitted to consult your textbook (or any other book) during the exam.

SECTION 1

1. A machine part has an average diameter of .25 of an inch, with a standard deviation of .06. A sample of 36 parts is taken every so often and measured. If the sample mean is greater than .265 of an inch or less than .235 of an inch the machine is be re-adjusted.

- i) what is the probability that the machine will continue operating without adjustment if the population mean remains as is?
- ii) what is the probability that the machine will be adjusted unnecessarily (i.e. when the true mean is still the same)?
- iii) what is the probability that the machine will not be adjusted if the population mean actually shifts to .26 of an inch?
- iv) what is the probability that that the machine will be re-adjusted if the true mean shifts to .235 of an inch?

Indicate whether the above situations describe no errors, a Type 1 error or a Type 2 error. Identify the probabilities in each question as α , $1 - \alpha$, β or $1 - \beta$.

2. Given the following data on two variables: X = 10, 13, 14, 18, 12, 16, 17, 20 and Y = 12, 16, 12, 17, 12, 19, 10, 18, conduct t tests as if

- i) the two samples were taken independently, and as if
- ii) this was just one sample of matching pairs.

Explain the difference in results.

3. From a total of 1000 single people sampled, 225 said their favourite kind of evening was a night in watching tv. From a sample of 1500 married people, the number was 460.

At the .01 level, is there a significant difference in the proportions of single and married people who like to sit in of an evening?

4. Can you justify using $\sum (x_i - \bar{x})^2 / (n - 1)$ rather than $\sum (x_i - \bar{x})^2 / n$ as an estimator of population variance.

5. A sample of employees in a company were asked to signal their approval for one or other of three radio stations to play over the sound system. The results are given for three staff grades. Does it seem, at the usual level of significance, as if there is a relationship between the station preferred and the grade?

	Lyric FM	Radio 2	Galway Bay FM
Manager	10	13	29
Admin.	39	40	19
Production Staff	81	57	22

6. This is a bivariate distribution of Income (X) and Vacation Expenditure (Y) (both given in thousands of pounds):

		X		
		20	30	40
Y	1	.28	.03	0
	2	.08	.15	.03
	3	.04	.06	.06
	4	0	.06	.15
	5	0	0	.03
	6	0	0	.03

Fill in the following:

- the three values for the conditional mean of Y, given X,
- the three values for the conditional variance of Y, given X
- the 6 conditional probabilities of Y, given that X is 40.

SECTION 2

7. Two variables y and x are believed to be related by the following stochastic equation:

$y = \alpha + \beta x + u$, where u is the usual random disturbance, with zero mean and constant variance. Two different researchers estimate the slope parameter using OLS. Their results were as follows:

RESEARCHER 1

y x

4	3
4.5	3
4.5	3
3.5	3
4.5	4
4.5	4
5.5	4
5	4

$$\hat{y} = 1.875 + .75x$$

(1.2) (.339)

$$r^2 = .45$$

$$\hat{\sigma} = .48$$

RESEARCHER 2

y x

2	1
2.5	1
2.5	1
1.5	1
11.5	10
10.5	10
10.5	10
11	10

$$\hat{y} = 1.5 + .97x$$

(0.27) (.038)

$$r^2 = .99$$

$$\hat{\sigma} = .48$$

Can you explain why the standard error of the slope estimate for the first researcher is larger than the standard error of the slope coefficient for the second researcher (the standard errors are in brackets, underneath the coefficients)?

8. Indexes of mobility within 3 industries are constructed (a value of zero indicates a person didn't change company within a ten year period while a value of 100 indicates a notional maximum 'degree of movement').

Stock Exchange

4
17
8
20
16

Service

3
12
40
17
31
19

Air Transport

30
38
46
40
21

It is unlikely that the indexes are normally distributed. Using the .05 level, determine whether the mobility indexes are equal across the three industries.

9. Discuss and compare some of the properties of the sample mean and the sample median, with particular reference to least squares, least absolute deviations and maximum likelihood procedures.

10. A city sends surveyors out to value business premises for rates payments. The city manager send out 4 new surveyors to 5 premises each to compare and check their results.

	SURVEYOR			
	Lynch	Dolan	Joyce	Quinn
Premises1	£53	£55	£49	£45
Premises2	£50	£51	£52	£53
Premises3	£48	£52	£47	£53
Premises4	£70	£68	£65	£64
Premises5	£84	£89	£92	£86

i) Is there a difference in the treatment means?

ii) Is there a difference in the block means?

Can we conclude that there is a difference in the surveyors (at 5% significance)?

11. A simple regression model finds that: $\hat{y} = 901.247 + .1686x$. From the regression estimation, the following is known:

$$\bar{x} = 2921.2857; S_{yx} = 936.85; \sum_{i=1}^n (x_i - \bar{x})^2 = 37,357,090.86 \text{ and } n = 14.$$

Set up a 95% prediction interval estimate i) for an individual x, valued at 4000 and ii) for an average x, also valued at 4000.

12. i) Why is an F test of $\mu_1 = \mu_2 = \mu_3$ generally preferable to three t tests of the equality of each of the three pairs of means?

ii) A researcher testing the correlation between asbestos exposure and lung cancer chooses a level of significance of 10%. Why so high?

iii) If, in the simple linear regression, the standard error of the OLS slope estimator rises with the variance of the x variable, should we just get rid of the middle values of x so as to increase the variance of x and if so, why so or if not, why not?

iv) Show that t-squared = F, when there is just one degree of freedom in the numerator.