UNIVERSITY *of* LIMERICK

OLLSCOIL LUIMNIGH

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF MATHEMATICS & STATISTICS

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: **MA4104** **SEMESTER: Spring 2013/14**

MODULE TITLE: **Business Statistics** **DURATION: 2.5 hours**

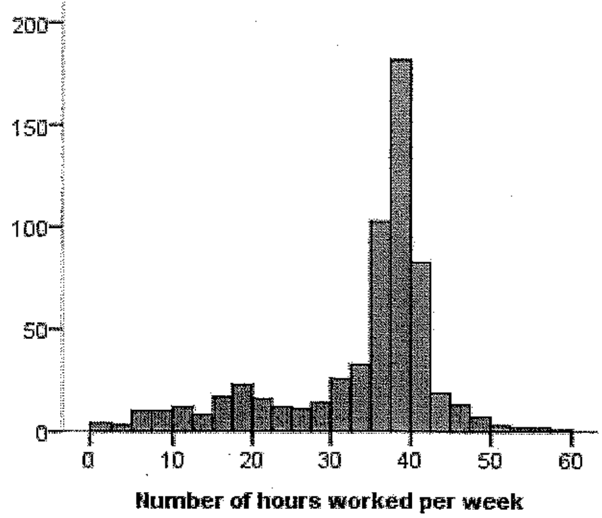
LECTURER: **Dr. Helen Purtill**

EXTERNAL EXAMINERS: **Prof. Brendan Murphy**

INSTRUCTIONS TO CANDIDATES:

- **Answer all 4 questions.**
- **All questions carry equal marks.**
- **Calculators may be used.**
- **Relevant statistical tables are attached to this paper.**
- **A set of formulae is attached to this paper.**
- **The exam is worth 80% of the final grade for this module.**

Q1. (a) The following histogram and descriptive statistics are produced for the number of hours worked per week by a sample of 614 employees from a particular industry.



Number of hours worked per week	
Mean	33.51
Median	37.51
Standard Deviation	10.08
Interquartile range	7.27

(i) Classify the number of hours worked per week by data type and scale of measurement.

(ii) How would you describe the shape of the distribution?

(iii) What are the appropriate measures of centrality and dispersion for these data? Justify your answer.

(7 marks)

(b) Monthly office rents in a particular city are found to be normally distributed with a mean of €350 and standard deviation of €45.

(i) What percentage of monthly rents are greater than €380?

(ii) What percentage of monthly rents are between €320 and €390?

(iii) 15% of rents are higher than what value?

(iv) Use a control chart to identify which of the following rents are unusually high or low; €230, €200, €480, €370.

(12 marks)

(Question 1 is continued on overleaf)

(c) A large hotel wanted to determine how its customers rated the hotel. A postal questionnaire was sent to a sample of 300 hotel customers, selected at random from the customer address list. The hotel received 187 completed questionnaires.

(i) Identify the population, sampling frame and sample.
(ii) What method of sampling was used?
(iii) What was the response rate for the questionnaire?

(6 marks)

Q2. (a) A new weight-watching company, Weight Reducers International, advertises that those who join will lose an average of 4kg within a month. A random sample of 50 people who joined the new weight reduction program had a mean weight loss of 3.5kg with a standard deviation of 1.4kg.

(i) At the 5% level of significance, test whether average weight loss within a month is less than 4kg.

(ii) Construct and interpret a 95% confidence interval for the average weight loss within a month of all customers of Weight Reducers International.

(10 marks)

(b) The marketing department in a large online business would like to estimate the proportion of customers who use promotional coupons when making a purchase. In a sample of 300 customers, 86 have used a coupon. Construct and interpret a 95% confidence interval for the proportion of customers who use a coupon when making a purchase.

(5 marks)

(c) A new company has developed a petrol additive that is supposed to improve the fuel efficiency of a car's engine. To test the effectiveness of the product, 5 cars are randomly selected. Each car sampled is driven both with and without the additive. The resulting petrol mileages, in miles per gallon are recorded:

Car	With Additive	Without Additive
1	26.7	24.9
2	22.0	18.8
3	32.4	27.7
4	13.7	13.0
5	19.8	17.8

Has the petrol additive increased mileage? Test this hypothesis using a 5% significance level. ($\bar{d} = 2.48$, $s_d = 1.53$)

(8 marks)

Q3. (a) A business would like to compare heating costs associated with their offices in two locations (Location A and Location B). A random sample of weekly heating costs (€) was taken from each location and the sample data is summarised as:

Location A	Location B
Sample size = 11	Sample size = 19
Mean = 63.7	Mean = 52.4
Standard Deviation = 7.4	Standard Deviation = 6.8

At the 5% significance level, test whether weekly costs differ between the two groups (you may assume equal population variances in both groups). Interpret the results.

(13 marks)

(b) In a sample survey, 122 people were asked if they supported the government decision to increase the price of wine. The following data was obtained:

	Age		
Opinion	**<25 years**	**≥25 years**	**Total**
Agree	15	12	27
Neutral	10	15	25
Disagree	25	45	70
Total	50	72	122

(i) What percentage of respondents agree with the increase?

(1 mark)

(ii) What percentage of people aged under 25 years agree with the increase?

(1 mark)

(iii) Test whether there is an association between age and opinion, stating clearly the null and alternative hypotheses, test statistic and decision. What conclusions would you make following the analysis? (Note: $\chi^2_{2,0.05} = 5.991$)

(8 mark)

(iv) What condition must be satisfied for the Chi-square test to be valid? Does the data in this question satisfy this condition? Justify your answer.

(2 mark)

Q4. The amount of electricity used in a home is thought to depend on the living area of the home (measured in square metres). In order to investigate this relationship a researcher selected a random sample of 80 homes and for each home recorded the living area (Area) and the number of kilowatt hours of electricity used in a year (Consumption).

Let X denote the area and Y denote the consumption of electricity. The data for the 80 homes was used to calculate the following quantities:

$$\bar{x} = 150, \quad \bar{y} = 3200, \quad S_{XX} = 200, \quad S_{YY} = 430000, \quad S_{XY} = 3600$$

(i) Compute the Pearson Correlation Coefficient. Interpret this value.

(3 marks)

(ii) Determine the least squares regression equation. Interpret the meaning of the slope of the regression line.

(4 marks)

(iii) Use the regression line to estimate the electricity consumption for a house with a living area of 100 square metres.

(2 marks)

(iv) Using the 5% significance level, test the significance of the slope. Interpret the result.

(6 marks)

(v) Calculate and interpret a 95% confidence for the slope of the regression line.

(4 marks)

(vi) Calculate and interpret the coefficient of determination.

(3 marks)

(vii) What assumptions should a regression model satisfy? How can these assumptions be checked?

(3 marks)

FORMULAE

Notation:

Means	Proportions
μ = population mean σ = population standard deviation $\bar{x}$ = sample mean s = sample standard deviation n = sample size	p = population proportion $\hat{p}$ = sample proportion

1. Standard error of sample mean ($\bar{x}$):

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

2. Standard error of sample proportion ($\hat{p}$) for a confidence interval :

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

3. Standard error of $\bar{x}_1 - \bar{x}_2$ (large samples, population variances unknown):

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

4. Standard error of $\bar{x}_1 - \bar{x}_2$ (small samples):

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

5. Chi-Square Test Statistic :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

O_i = observed frequency in cell i,
E_i = expected frequency in cell i under the Null Hypothesis

Regression and Correlation

6. Sums of Squares : $S_{XY} = \sum(x_i - \bar{x})(y_i - \bar{y})$

$$S_{XX} = \sum (x_i - \bar{x})^2$$

$$S_{YY} = \sum (y_i - \bar{y})^2$$

7. Pearson's correlation coefficient :

$$r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

8. Regression Line:

$$\hat{y} = b_0 + b_1 x$$

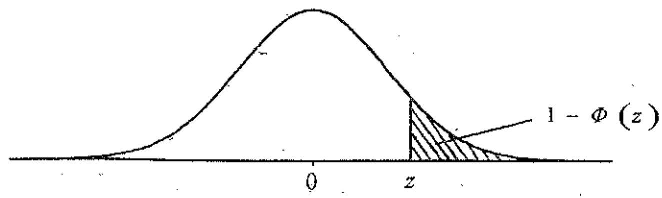
where $b_1 = \frac{S_{XY}}{S_{XX}}$ and $b_0 = \bar{y} - b_1\bar{x}$

Standard error of b_1: $SE(b_1) = \sqrt{\frac{s^2}{S_{XX}}}$ where $s^2 = \frac{SSE}{n-2}$ and $SSE = S_{YY} - b_1 S_{XY}$

Table 3 Areas in Upper Tail of the Normal Distribution

The function tabulated is $1 - \Phi(z)$ where $\Phi(z)$ is the cumulative distribution function of a standardised Normal variable, z.

Thus $1-\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_z^{\infty} e^{-z^2/2}$ is the probability that a standardised Normal variate selected at random will be greater than a value of $z \left(= \frac{x-\mu}{\sigma}\right)$



$\frac{x-\mu}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.000108	.000104	.000100	.000096	.000092	.000088	.000085	.000082	.000078	.000075
3.8	.000072	.000069	.000067	.000064	.000062	.000059	.000057	.000054	.000052	.000050
3.9	.000048	.000046	.000044	.000042	.000041	.000039	.000037	.000036	.000034	.000033
4.0	.000032									

5.0 → 0.000 000 286 7   5.5 → 0.000 000 019 0   6.0 → 0.000 000 001 0

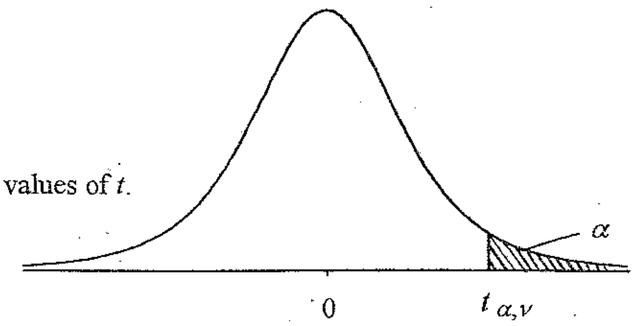
Table 7 Percentage Points of the t Distribution

The table gives the value of $t_{\alpha,\nu}$ – the 100α percentage point of the t distribution for ν degrees of freedom.

The values of t are obtained by solution of the equation:

$$\alpha = \Gamma[\tfrac{1}{2}(\nu+1)][\Gamma(\tfrac{1}{2}\nu)]^{-1}(\nu\pi)^{-1/2}\int_t^{\infty}(1+x^2/\nu)^{-(\nu+1)/2}\,dx$$

Note: The tabulation is for one tail only, that is, for positive values of t.
For $|t|$ the column headings for α should be doubled.



$\alpha =$	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
$\nu = 1$	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

This table is taken from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research*, reprinted by permission of Addison Wesley Longman Ltd. Also from Table 12 of *Biometrika Tables for Statisticians*, Volume 1, by permission of Oxford University Press and the Biometrika Trustees.