Cork Institute of Technology

Bachelor of Science (Honours) in Herbal Science – Stage 2

(NFQ Level 8)

Summer 2007

Biostatistics & Data Analysis

(Time: 3 Hours)

Instructions ;Answer five questions.

Examiners: Mr. J. Mulhare
Dr. D. Corrigan
Mr. E. Walsh

Q1. In a trial on fifty patients it was found that the times, to the nearest minute, taken for the effect of a drug to wear off were as given in the following table;

128	119	95	107	124	128	147	98	108	117
113	109	124	111	107	139	136	123	112	132
146	128	103	125	114	109	100	131	113	111
124	131	133	131	88	118	116	98	113	127
104	111	112	104	152	117	122	117	95	117

(i) Construct a *Grouped Frequency Distribution* of the above data…groups (80-89),(89-90) etc.

(ii) Hence calculate *the mean and, standard deviation* of the data.

(iii) Construct a cumulative frequency curve (an ogive) of the data..

(iv) Estimate the median and first and third quartiles using the ogive.

(v) Construct a boxplot of the above data .Do you detect any outliers?

(5 x 8 marks)

Q2. (a) In a certain population the percentages of children who contract measles, mumps and rubella before adulthood are 20%, 30% and 15% respectively. Given that these events occur independently in the population, draw a Venn diagram depicting these probabilities.

What is the probability that a person chosen at random will contract;

(i) all three infections;

(ii) measles or rubella

(iii) only mumps or only rubella

(iv) at least one of these infections. (5 x 4 marks)

(b) A micro-organism reproduces by dividing into two after intervals. After one division it waits in 80% of cases one hour before dividing again, independently of its previous history, and in the remaining 20 % of cases it waits two hours. If a biologist isolates one organism immediately after a division, what are the possible number of organisms that could be present just before the end of (i) two hours; (ii) three hours, and what are the probabilities of these numbers of organisms occurring?

(14 marks)

If an organism is introduced into the environment in which it has a 10% chance of it being eaten during any one hour, what is the probability of there being two organisms present just before the end of two hours?

(6 marks)

Q3. (a) A particular drug is known to effect a cure in 40% of patients treated for a particular ailment. If ten such patients are treated with this drug, calculate, using the binomial distribution, the probability that;

(i) six patients;

(ii) fewer than three patients;

(iii) at least two patients …

…..are cured. (15 marks)

(b) Hits occur at a "web site" at random according to the Poisson distribution with an average of five hits per hour.

Determine the probability that;

(i) four hits;

(ii) more than three hits……..…occur in an hour.

What is the probability that three hits occur in half an hour?

(15 marks)

(c) A box contains twelve balls, four of which are red while the remaining eight balls are blue. Five balls are selected at random. Use the hypergeometric distribution to determine the probability that of the five selected;

(i) three are red;

(ii) at least one is red.

(10 marks)

Q4. A health drink is sold in bottles with a nominal marked volume of 500 ml.

The volumes contained in the bottles are in fact normally distributed with a mean of 510 ml. and a standard deviation of 6 ml.. A bottle of the drink is selected at random. What is the probability that it contains;

(i) less than 518 ml.;

(ii) less than 500 ml.;

(iii) between 516 and 520 ml.; (15 marks)

Between what two values do the middle 90% of volumes lie? (5 marks)

If the standard deviation remains at 6 ml. what must the mean be set to if 10% of bottles contain more than 520 ml? (10 marks)

Where the mean is maintained at 510 what must the standard deviation be set to if less than 5% of bottles contain less than the marked quantity? (10 marks)

Q5. (a) Out of the 100 workers surveyed, 23 travel to work by means of public transport. Calculate the (i) 90% ; (ii) 99.8% confidence limits for the population **proportion** (**P**) of workers who use public transport to travel to work. (12 marks)

In estimating the proportion of workers who use public transport, how large should a survey of workers be if you want the resulting sample proportion to differ from the true population proportion by less than 1% (i.e 0.01), with a 98% degree of certainty?
(8 marks)

(b) Distinguish between one-tailed and two-tailed hypothesis tests.
A manufacturer claims that not more than 20% of his products are defective.
In a sample of 40 of this product 11 are found to be defective.
Test if we can reject the manufacture's claim based on this sample.
Conduct the test at a 1% level of significance. (20 marks)

Q6. (a) To test the effect of a particular fertilizer on the yield of corn, eight sites were chosen in different locations and the sites divided in two halves such that the soil conditions in both halves were as close to each other as possible. One half of each site was treated with the fertilizer while the other half was not. The same type of corn seed was sown equally throughout the eight sites. At harvest, the yield, in tons per acre, from each half site was measured and is shown in the following table.

Site Number	1	2	3	4	5	6	7	8
Yield without fertilizer	3.6	2.4	2.7	2.3	3.2	3.7	2.8	2.5
Yield with fertilizer	4.3	3.0	2.8	2.7	3.5	3.5	2.5	2.9

Test the hypothesis that the fertilizer contributes to a higher yield, at a 5% level of significance. (20 marks)

(b) In a sample survey of workers in an industry, the workers are divided into clerical and industrial categories. The following table gives the "sick-days" taken by the workers in each category in the sample.

Category of worker	**Sick-Days**		
	None	One	More than one
Industrial	139	212	149
Clerical	66	94	40

Do these figures suggest a relationship between the category of worker and the number of sick-days taken? Test at a 5% level of significance using the chi-squared distribution. (20 marks)

Q7. The following table gives the values of the packed cell volume, *x mm*, and the related red blood cell count *y millions*, of eight dogs tested in a trial;

Packed Cell Volume, *x mm*	46	43	56	48	42	35	58	40
Red Blood Cell count, *y millions*	6.6	6.4	8.7	7.5	6.9	5.8	9.6	6.2

Given $\sum x = 368$, $\sum y = 57.7$, $\sum xy = 2723.6$, $\sum x^2 = 17358$, $\sum y^2 = 428.31$

You are required to

(i) Plot the data on a scatter diagram and estimate the regression line

(ii) Evaluate the correlation coefficient [*r*] and test its significance.

(iii) Evaluate the coefficient of determination. Interpret this value.

(iv) Find the equation of the least squares regression line through the data and estimate the expected value of Y when X=50. (4 by 10 marks)

STATISTICAL FORMULAE

Central Tendency and Dispersion

Mean of an array; $\bar{x} = \frac{\sum x}{n}$

Mean of a frequency distribution $\bar{x} = \frac{\sum fx}{\sum f}$. $\bar{x} = a + \frac{\sum fd}{\sum f}$

Standard Deviation of an array: $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$ or $\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

Standard Deviation of a frequency distribution $\sigma = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}}$ or

$= \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

Median $= L_M + C_M\left(\frac{\frac{1}{2}N - F_{M-1}}{f_M}\right)$

Mode $= L_M + C_M\left[\frac{f_M - f_{M-1}}{2f_M - (f_{M-1} + f_{M+1})}\right]$

Skewness $= \frac{3(\text{Mean - Median})}{\text{S.D.}}$

Coefficient of Variation $= \frac{\sigma}{\bar{x}} \times 100$

Probability distributions

Binomial Distribution; $P(r,n) = {}^{n}C_r\, p^r q^{n-r}$ ...mean = np ; variance = npq

Poisson Distribution; $P(x) = \frac{\lambda^x . e^{-\lambda}}{x!}$; $\lambda = mean = np$

Hypergeometric Distribution ; $P(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ or $= \frac{{}^{k}C_x \times {}^{N-k}C_{(n-x)}}{{}^{N}C_n}$

Standard Normal Units; $Z = \frac{X - \mu}{\sigma}$

Baye's Rule (conditional probability) $P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(B)}$

<u>**Central Limit Formulae**</u>

$x \sim N\{\mu, \sigma\} \Rightarrow z = \dfrac{x - \mu}{\sigma}$; $\bar{x} \sim N\left\{\mu, \dfrac{\sigma}{\sqrt{n}}\right\} \Rightarrow z_{\bar{x}} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$;

$t_{\bar{x}} = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

Confidence Limits for the population mean: $\bar{x} \pm z_c \dfrac{\sigma}{\sqrt{n}}$ ...σ known

$(\bar{x}_1 - \bar{x}_2) \sim N\left\{(\mu_1 - \mu_2), \sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}\right\}$ ....σ s known

$(\bar{x}_1 - \bar{x}_2) \sim t\left\{(\mu_1 - \mu_2), s_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right\}$

...........σ s estimated.....$s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 2)s_2^2}{n_1 + n_2 - 2}$

$p \sim N\left\{P, \sqrt{\frac{PQ}{n}}\right\}$; Confidence Limits for the population proportion: $p \pm z_c \sqrt{\dfrac{pq}{n}}$

$(p_1 - p_2) \sim N\left\{(P_1 - P_2), \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}\right\}$

A $(100 \cdot \alpha)$% **Confidence Interval for a population variance** σ^2 ;

$\dfrac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \dfrac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$

Chi-Squared Distribution .... $\chi^2 = \sum \frac{(O-E)^2}{E}$

Regression and Correlation

Least Squares Regression Line : $y = a + bx$;

$$S_{xx} = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}; \qquad S_{yy} = \Sigma(y-\bar{y})^2 = \sum y^2 - \frac{\left(\sum y\right)^2}{n};$$

$$S_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$b = \frac{S_{xy}}{S_{xx}} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}; \qquad a = \bar{y} - b\bar{x}; \qquad \bar{y} = a + b\bar{x}$$

$$\text{or} \qquad a = \frac{\sum y - b\sum x}{n}$$

Coefficient of Correlation:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} . S_{yy}}} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}};$$

[illegible]

Area under the Normal Curve

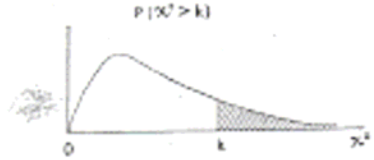
$$P(z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}z^2}\, dz .$$

z	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
0·0	0·5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0·1	0·5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0·2	0·5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0·3	0·6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0·4	0·6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0·5	0·6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0·6	0·7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0·7	0·7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0·8	0·7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0·9	0·8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1·0	0·8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1·1	0·8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1·2	0·8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1·3	0·9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1·4	0·9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1·5	0·9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1·6	0·9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1·7	0·9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1·8	0·9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1·9	0·9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2·0	0·9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2·1	0·9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2·2	0·9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2·3	0·9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2·4	0·9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2·5	0·99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2·6	0·99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2·7	0·99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2·8	0·99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2·9	0·99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3·0	0·99865	99869	99874	99878	99882	99886	99889	99893	99897	99900
3·1	0·99903	99906	99910	99913	99916	99918	99921	99924	99926	99929
3·2	0·99931	99934	99936	99938	99940	99942	99944	99946	99948	99950
3·3	0·99952	99953	99955	99957	99958	99960	99961	99962	99964	99965
3·4	0·99966	99968	99969	99970	99971	99972	99973	99974	99975	99976
3·5	0·99977	99978	99978	99979	99980	99981	99981	99982	99983	99983
3·6	0·99984	99985	99985	99986	99986	99987	99987	99988	99988	99989
3·7	0·99989	99990	99990	99990	99991	99991	99992	99992	99992	99992
3·8	0·99993	99993	99993	99994	99994	99994	99994	99995	99995	99995
3·9	0·99995	99995	99996	99996	99996	99996	99996	99996	99997	99997

[illegible] χ^2-DISTRIBUTION

	·99	·95	·50	·20	·10	·05	·025	·01
1	·0002	·0039	·45	1·64	2·71	3·84	5·02	6·63
2	·020	·103	1·39	3·22	4·61	5·99	7·38	9·21
3	·115	·352	2·37	4·64	6·25	7·81	9·35	11·34
4	·30	·71	3·36	5·99	7·78	9·49	11·14	13·28
5	·55	1·15	4·35	7·29	9·24	11·07	12·83	15·09
6	·87	1·64	5·35	8·56	10·64	12·59	14·45	16·81
7	1·24	2·17	6·35	9·80	12·02	14·07	16·01	18·48
8	1·65	2·73	7·34	11·03	13·36	15·51	17·53	20·09
9	2·09	3·33	8·34	12·24	14·68	16·92	19·02	21·67
10	2·56	3·94	9·34	13·44	15·99	18·31	20·48	23·21
11	3·05	4·57	10·34	14·63	17·28	19·68	21·92	24·72
12	3·57	5·23	11·34	15·81	18·55	21·03	23·34	26·22
13	4·11	5·89	12·34	16·98	19·81	22·36	24·74	27·69
14	4·66	6·57	13·34	18·15	21·06	23·68	26·12	29·14
15	5·23	7·26	14·34	19·31	22·31	25·00	27·49	30·58
16	5·81	7·96	15·34	20·47	23·54	26·30	28·85	32·00
17	6·41	8·67	16·34	21·61	24·77	27·59	30·19	33·41
18	7·01	9·39	17·34	22·76	25·99	28·87	31·53	34·81
19	7·63	10·12	18·34	23·90	27·20	30·14	32·85	36·19
20	8·26	10·85	19·34	25·04	28·41	31·41	34·17	37·57
21	8·90	11·59	20·34	26·17	29·62	32·67	35·48	38·93
22	9·54	12·34	21·34	27·30	30·81	33·92	36·78	40·29
23	10·20	13·09	22·34	28·43	32·01	35·17	38·08	41·64
24	10·86	13·85	23·34	29·55	33·20	36·42	39·36	42·98
25	11·52	14·61	24·34	30·68	34·38	37·65	40·65	44·31
26	12·20	15·38	25·34	31·79	35·56	38·89	41·92	45·64
27	12·88	16·15	26·34	32·91	36·74	40·11	43·19	46·96
28	13·56	16·93	27·34	34·03	37·92	41·34	44·46	48·28
29	14·26	17·71	28·34	35·14	39·09	42·56	45·72	49·59
30	14·95	18·49	29·34	36·25	40·26	43·77	46·98	50·89
40	22·16	26·51	39·34	47·27	51·81	55·76	59·34	63·69
50	29·71	34·76	49·33	58·16	63·17	67·50	71·42	76·15
60	37·48	43·19	59·33	68·97	74·40	79·08	83·30	88·38
70	45·44	51·74	69·33	79·71	85·53	90·53	95·02	100·43
80	53·54	60·39	79·33	90·41	96·58	101·88	106·63	112·33
90	61·75	69·13	89·33	101·05	107·57	113·15	118·14	124·12
100	70·06	77·93	99·33	111·67	118·50	124·34	129·56	135·81

$P\,[\chi^2 > k]$



t-DÁILEADH						t-DISTRIBUTION
ν \ %	20	10	5	2	1	0·2
1	3·078	6·314	12·706	31·821	63·657	318·310
2	1·886	2·920	4·303	6·965	9·925	22·327
3	1·638	2·353	3·182	4·541	5·841	10·213
4	1·533	2·132	2·776	3·747	4·604	7·173
5	1·476	2·015	2·571	3·365	4·032	5·893
6	1·440	1·943	2·447	3·143	3·707	5·208
7	1·415	1·895	2·365	2·998	3·499	4·785
8	1·397	1·860	2·306	2·896	3·355	4·501
9	1·383	1·833	2·262	2·821	3·250	4·297
10	1·372	1·812	2·228	2·764	3·169	4·144
11	1·363	1·796	2·201	2·718	3·106	4·025
12	1·356	1·782	2·179	2·681	3·055	3·930
13	1·350	1·771	2·160	2·650	3·012	3·852
14	1·345	1·761	2·145	2·624	2·977	3·787
15	1·341	1·753	2·131	2·602	2·947	3·733
16	1·337	1·746	2·120	2·583	2·921	3·686
17	1·333	1·740	2·110	2·567	2·898	3·646
18	1·330	1·734	2·101	2·552	2·878	3·610
19	1·328	1·729	2·093	2·539	2·861	3·579
20	1·325	1·725	2·086	2·528	2·845	3·552
21	1·323	1·721	2·080	2·518	2·831	3·527
22	1·321	1·717	2·074	2·508	2·819	3·505
23	1·319	1·714	2·069	2·500	2·807	3·485
24	1·318	1·711	2·064	2·492	2·797	3·467
25	1·316	1·708	2·060	2·485	2·787	3·450
26	1·315	1·706	2·056	2·479	2·779	3·435
27	1·314	1·703	2·052	2·473	2·771	3·421
28	1·313	1·701	2·048	2·467	2·763	3·408
29	1·311	1·699	2·045	2·462	2·756	3·396
30	1·310	1·697	2·042	2·457	2·750	3·385
40	1·303	1·684	2·021	2·423	2·704	3·307
60	1·296	1·671	2·000	2·390	2·660	3·232
120	1·289	1·658	1·980	2·358	2·617	3·160
∞	1·282	1·645	1·960	2·326	2·576	3·090

MATAMAITIC FHEIDHMEACH

Aonad faid: méadar (m)
Aonad mais: cileagram (kg)
Aonad fórsa: Niútan (N) = kgm/s^2
Aonad oibre: giúl (J) = Nm
Aonad cumhachta: vata (W) = J/s

Gluaiseacht i líne faoi luasghéarú tairiseach: $v = u + ft$; $s = ut + \frac{1}{2}ft^2$; $v^2 = u^2 + 2fs$

Fuinneamh Poitéinsiúil: mgh. Fuinneamh cinéiteach (obair déanta): $\frac{1}{2}mv^2$.
I gcóras inchoimeádach: Fuinneamh poitéinsiúil + fuinneamh cinéiteach = méid tairiseach.

Móimintam chathéala: $m\vec{v}$
Spreagadh fórsa = athrú san móimintam.
Comhéifeacht chúitimh (comhéifeacht leaisteachais):

$$e = -\left(\frac{\text{luas gaolmhar indhiaidh comhbhualadh}}{\text{luas gaolmhar roimh comhbhualadh}}\right)$$

Luasghéarú lárnaimsitheach: $\frac{v^2}{r} = \omega^2 r$

Dlí Hooke le haghaidh téada leaistigh: $p = kx$ (p an fórsa, x an searradh, tairiseach don téad é k).

Meánlár:
Stua, gath r; uillinn 2θ ag an lár: $\frac{r \sin\theta}{\theta}$ ón lárphointe.

Teascóg ciorcail; gath r, uillinn 2θ: $\frac{2}{3}r\frac{\sin\theta}{\theta}$ ón lárphointe.

Lann triantánach: $\frac{1}{3}$ ón mbonn feadh an mheánlíne.

Meáchanlár:
Leathsféar, gath r: $\frac{3}{8}r$ ón lárphointe.
Sceall leathsféarach, gath r: $\frac{1}{2}r$ ón lárphointe.
Dronn-chón ciorclach, airde h: $\frac{1}{4}h$ ón bhonn.

Brú ag pointe i leacht: ρgh.
Sá ar dhromchla atá báite: achar × brú ag an meánlár.

Móiminti na táimhe:
Bata aonfhoirmeach, fad $2l$: lár: $\frac{1}{3}ml^2$; foirceann: $\frac{4}{3}ml^2$
Fonsa aonfhoirmeach, gath r: lár: mr^2; lárlíne: $\frac{1}{2}mr^2$
Diosca aonfhoirmeach, gath r: lár: $\frac{1}{2}mr^2$; lárlíne: $\frac{1}{4}mr^2$
Dlúth-sféar aonfhoirmeach, gath r: lár líne: $\frac{2}{5}mr^2$