National Exams December 2003

98-Env-A5: Environmental Data Collection and Analysis

3 hours duration

NOTES:

1. If doubt exists as to the interpretation of any question, the candidate is urged to submit with the answer paper, a clear statement of any assumptions made.

2. This is an open book exam. Any non-communicating calculator is permitted. However, you must indicate the type of calculator being used, i.e., write the name and model designation of the calculator, on the first inside left hand sheet, of the exam work book.

3. Answer Question 1 and any 4 other questions.

4. All questions are of equal value. Part marks are as shown.

5. Use the statistical tables and graph papers provided.

Tables Provided:

Standard normal table
t-distribution table
F-distribution table

This question must be answered.

1. a) List the two properties which make Simple Random Sampling (SRS) the standard by which other sampling methods are judged. Explain why SRS is not always used in practice. Give an example where other sampling schemes may be better. [5 marks]

   b) List 5 typical characteristics of environmental data [5 marks].

   c) Indicate whether each of the statements below is true or false: [1 mark each]

   (i) For statistical significance, the α-value must be greater than the p-value.

   (ii) The Pearson r is a measure of linear correlation only.

   (iii) We can increase the power of a statistical test by decreasing α or increasing the sample size n.

   (iv) As the sample size of a set of data increases, the data tend to be normally distributed.

   (v) Two events that are mutually exclusive can also be independent.

   (vi) The mean square error is a measure of precision of a statistical estimate.

   d) What is meant by the central limit theorem and how can it be demonstrated on a computer? [4 marks]

2. Air pollution concentrations at two locations were compared. Eleven measurements were taken at each location and the following values in parts per million of CO_2 were obtained:

Location A:	66.3	63.5	64.9	61.8	64.3	64.7	65.1	64.5	68.4	63.2	67.4
Location B:	71.3	60.4	64.6	63.9	68.8	70.1	64.8	68.9	65.8	66.2	69.2

It is desired to show that the pollution values obtained at location A are significantly less than that at location B at the 5% significance level. The following three tests were carried out in using a computer.

Test 1: Paired T for Location A - Location B (Paired test)

	N	Mean	StDev	SE Mean
Location A	11	64.918	1.885	0.568
Location B	11	66.727	3.238	0.976
Difference	11	-1.809	3.016	0.909

T-Test of mean difference = 0 (vs < 0): P-Value = 0.037

Test 2: Two-sample T for Location A vs Location B (One-tailed test)

	N	Mean	StDev	SE Mean
Location A	11	64.92	1.88	0.57
Location B	11	66.73	3.24	0.98

```
Difference = mu Location A - mu Location B
Estimate for difference:  -1.81
T-Test of difference = 0 (vs <):   P-Value = 0.064  DF = 16
```

Test 3: Two-sample T for Location A vs Location B (Two-tailed test)

	N	Mean	StDev	SE Mean
Location A	11	64.92	1.88	0.57
Location B	11	66.73	3.24	0.98

```
Difference = mu Location A - mu Location B
Estimate for difference:  -1.81
T-Test of difference = 0 (vs not =): P-Value = 0.129  DF = 16
```

a) Which of the above results is the appropriate one? Explain why this is so. What is the conclusion from the test? [10 marks]

b) What is the main assumption of the t-test used? Use a simple graphical method to verify the assumption of the t-test used? [5 marks]

c) Calculate the value of the test statistic for each of the above tests. [5 marks]

3. Explain the following statistical terms. Illustrate your explanation with an example.

a) Random and Fixed effect models as used in ANOVA. [4 marks]

b) Fractional factorial design as used in the design of experiments. [4 marks]

c) Cluster sampling as used in sampling design. [4 marks]

d) Method of Moments as used in parameter estimation. [4 marks]

e) Power of a test as used in hypothesis testing. [4 marks]

4. Ten values of a random variable X are:

2 5 8 120 9 13 45 24 57 16

The summary statistics from obtained from a computer package are given below.

N	Mean	Median	StDev	Minimum	Maximum	Q1	Q3
10	29.9	14.5	36.4	2.0	120.0	7.3	48.0

a) Draw accurately a dotplot and a standard boxplot for the data. Indicate clearly where the fences are for the boxplot. Indicate whether any outliers exist. [10 marks]

b) From the numerical summary and the boxplot, what can you conclude about the characteristics of the data? Compute also the quartile skew and coefficient of skewness. [10 marks]

5. A 2^3 factorial design was used to study the impact of Factors A, B, and C on the response Y, of an environmental system. Two replicates were used. The responses in Yate's standard order are as follows:

Treatment Combination	Response, Y
(1)	0.64, 0.63
a	0.45, 0.47
b	0.72, 0.74
ab	0.65, 0.66
c	0.59, 0.57
ac	0.55, 0.56
bc	0.73, 0.75
abc	0.78, 0.80

a) Estimate all main and all interaction effects. [5 marks]

b) Determine the effects that are considered important using a graphical method and ANOVA. [10 marks]

c) Develop a regression equation to predict Y using the important effects identified in Part (b). Explain how you would check whether all the assumptions of regression have been fulfilled. [5 marks]

6. If the annual precipitation amount (X) in a certain city is normally distributed with a mean of 50 cm and a coefficient of variation of 20%, determine the following:

a) $P(X \leq 30 \text{ cm})$ [4 marks]

b) $P(40 \text{ cm} \leq X \leq 55 \text{ cm})$ [4 marks]

c) Probability that X is within 5 cm of the mean annual precipitation. [4 marks]

d) $P(X > 60 | 30 \leq X \leq 70)$ [4 marks]

e) The value of X_o such that the probability of the annual precipitation exceeding X_o is only ¼ that of not exceeding X_o. [4 marks]

7. a) Consider 2 events A and B. If P(A) = 0.55 and P(A'.B)=0.30, and P(B'|A) = 0.80, what are :P(A'.B'), P(A|B), and P(A' U B)? [5 marks]

b) If $\mu_X = 5$ and $\sigma_X = 2$, what is the mean of $4X^2 - 3X + 2$? [5 marks]

c) The life of an instrument is a random variable T which has a cumulative distribution function (CDF) given by the equation:

$F(t) = 1 - \exp(-t/2) \qquad t > 0$

Where t is measured in years. Determine the probability that a newly installed instrument will last more than 4 years. [5 marks]

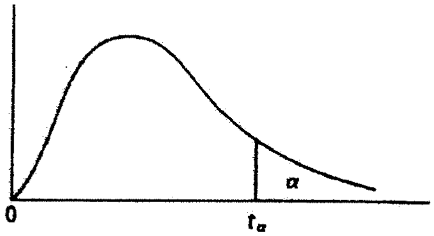
d) Determine the median and interquartile range (IQR) of the life of the instrument. [5 marks]

Standard Normal Distribution Table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Auxiliary table of quantiles of the normal d.f.

$1-\alpha$	$z(1-\alpha)$	$1-\alpha$	$z(1-\alpha)$	$1-\alpha$	$z(1-\alpha)$
.50	0	.91	1.341	.995	2.576
.55	.126	.92	1.405	.999	3.090
.60	.253	.93	1.476	.9995	3.291
.65	.385	.94	1.555	.9999	3.719
.70	.524	.95	1.645	.99995	3.891
.75	.674	.96	1.751	.99999	4.265
.80	.842	.97	1.881	.999995	4.417
.85	1.036	.98	2.054	.999999	4.753
.90	1.282	.99	2.326	.9999999	5.199



Critical Values of the F-Distribution

$f_{0.05}(v_1, v_2)$

v_2	v_1 = 1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

*Reproduced from Table 18 of *Biometrika Tables for Statisticians*, Vol. I, by permission of E. S. Pearson and the Biometrika Trustees.

t-distribution table

Critical Values of the t-Distribution

v	α 0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

10 X 10 TO THE CENTIMETER 18 X 25 CM.