Environmental Statistics

Sheet No.1

1- **Earthquake records.** Measurements of engineering interest have been recorded during earthquakes in Japan and in other parts of the world since 1800. One of the critical recordings is of apparent relative density, RDEN. *After the commencement of a strong earthquake, a saturated fine, loose sand undergoes vibratory motion and consequently the sand may liquefy without retaining any shear strength, thus behaving like a dense liquid. This will lead to failures in structures supported by the liquefied sand. These are often catastrophic. The standard penetration test is used to measure RDEN.* Another measurement taken to estimate the prospect of liquefaction is that of the intensity at which the ground shakes. This is the peak surface acceleration of the soil during the earthquake, ACCEL.

RDEN (%)	ACCEL (units of g)	RDEN (%)	ACCEL (units of g)	RDEN (%)	ACCEL (units of g)
53	0.219	30	0.138	50	0.313
64	0.219	72	0.422	44	0.224
53	0.146	90	0.556	100	0.231
64	0.146	40	0.447	65	0.334
65	0.684	50	0.547	68	0.419
55	0.611	55	0.204	78	0.352
75	0.591	50	0.170	58	0.363
72	0.522	55	0.170	80	0.291
40	0.258	75	0.192	55	0.314
58	0.250	53	0.292	100	0.377
43	0.283	70	0.299	100	0.434
32	0.419	64	0.292	52	0.350
40	0.123	53	0.225	58	0.334

Note: g denotes acceleration due to gravity (9.81 m/s^2).

Compute the sample mean $\bar{x}$, standard deviation $\hat{s}$, and the coefficient of skewness, $g1$, for RDEN and ACCEL. Plot the scatter diagram and calculate the correlation coefficient r . What conclusions can be reached?

2- **Flood discharge.** Annual maximum flood flows in the certain River at Italy, over a 61-year period from 1918 to 1978 are given in the second column of Table below. Compute the sample mean $\bar{x}$ and standard deviation $\hat{s}$. Sketch a histogram and the cumulative

relative frequency diagram. Flood embankments along the banks of the river can withstand a flow of 5000 m^3/s. What is the probability that this will be exceeded during a 12-month period?

Table E.7.2 Hydrologic data of the Po basin at Pontelagoscuro, northern Italy, from 1918 to 1978

Year	Annual maximum flow (m^3/s)	Annual minimum flow (m^3/s)	Annual minimum 7-day flow (m^3/s)	Mean annual flow (m^3/s)	Rainfall (mm)	Runoff (mm)
1918	5390	910	944	2010	1133	904
1919	4240	826	856	1440	999	648
1920	7220	969	1005	2400	1501	1080
1921	3000	569	590	1220	807	549
1922	2590	380	394	1070	1051	481
1923	2980	563	580	1280	969	576
1924	3920	749	813	1400	997	630
1925	3460	696	713	1530	1090	688
1926	8850	740	757	2040	1356	918
1927	3760	825	861	1630	1133	733
1928	8600	538	563	1800	1171	810
1929	2220	428	463	1090	876	490
1930	5400	607	618	1660	1159	747
1931	3700	508	527	1180	993	531
1932	4150	555	571	1420	1112	639
1933	4690	437	463	1310	1128	589
1934	6810	855	875	2050	1345	922
1935	6620	529	565	1750	1290	787
1936	6620	787	797	2310	1259	1039
1937	7700	668	675	2130	1529	958
1938	4380	287	305	1150	940	517
1939	3900	745	800	1780	1196	801
1940	5420	424	447	1350	1046	607
1941	6870	720	749	1860	1218	837
1942	4600	366	383	1160	948	522
1943	3270	310	311	987	896	444
1944	3660	306	307	905	950	407
1945	6830	304	306	916	846	412
1946	5130	655	669	1510	1011	679
1947	5460	588	594	1300	1096	585
1948	6630	711	735	1610	1100	724
1949	7220	275	278	967	922	435
1950	3260	400	411	1020	978	459
1951	8940	830	837	2200	1496	990
1952	4200	450	510	1110	913	499
1953	7400	520	550	1370	1046	616
1954	4450	440	456	1500	1100	675
1955	2400	423	441	1060	886	477
1956	5090	426	445	1220	1028	549
1957	6990	540	544	1390	1215	625
1958	5680	425	466	1320	1142	594
1959	7730	470	524	1900	1422	855
1960	6510	939	981	2620	1654	1179
1961	4880	424	450	1330	987	598
1962	4540	354	361	1070	909	481
1963	6430	732	737	1980	1362	891
1964	5630	444	461	1370	1026	616
1965	6110	321	363	1300	1015	585

3- **Flood discharge.** The following are the annual maximum flows in m^3/s in the Colorado River for the 52-year period from 1878 to 1929:

1980	1130	3120	2120	1700	2550	8500	3260	3960	2270
1700	1570	2830	2120	2410	2550	1980	2120	2410	2410
1420	1980	2690	3260	1840	2410	1840	3120	3290	3170
1980	4960	2120	2550	4250	1980	4670	1700	2410	4550
2690	2270	5660	5950	3400	3120	2070	1470	2410	3310
3230	3090								

Compute the mean $\bar{x}$ and standard deviation $\hat{s}$. Sketch a histogram and the relative frequency diagram.

4- **Population growth.** From past records, the population of an urban area has doubled every 10 years. Currently, it has a population of 200,000. An engineer needs to make an estimate of the requirements for water supply during the next 23 years. What maximum population does one assume for the period?

5- **Soil erosion.** Measurements taken on farmlands of the amounts of soil washed away by erosion suggest a relationship with flow rates.

Flow (L/s)	0.31	0.85	1.26	2.47	3.75
Soil eroded (kg)	0.82	1.95	2.18	3.01	6.07

Draw a plot of the data. Comment on the results.

6- **Water quality.** Water quality measurements are taken daily on the certain River in England. The concentrations of chlorides and phosphates in solution, given below in milligrams per liter, are determined over a 30-day period.

Chloride: 64.0, 66.0, 64.0, 62.0, 65.0, 64.0, 64.0, 65.0, 65.0, 67.0, 67.0, 74.0, 69.0, 68.0, 68.0, 69.0, 63.0, 68.0, 66.0, 66.0, 65.0, 64.0, 63.0, 66.0, 55.0, 69.0, 65.0, 61.0, 62.0, 62.0
Phosphate: 1.31, 1.39, 1.59, 1.68, 1.89, 1.98, 1.97, 1.99, 1.98, 2.15, 2.12, 1.90 1.92, 2.00, 1.90, 1.74, 1.81, 1.86, 1.86, 1.65, 1.58, 1.74, 1.89, 1.94, 2.07, 1.58, 1.93, 1.72, 1.73, 1.82

Compare the coefficients of variation *v*. Draw a scatter diagram and compute the correlation coefficient *r*. Comment on the results. Do you see any role in this association for predictive purposes?

7- **Air pollution.** On 13 April 1994, the following concentration of pollutants were recorded at eight stations of the monitoring system for pollution control located in the downtown area of Milan, Italy:

	Station							
	Aquileia	Cenisio	Juvara	Liguria	Marche	Senato	Verziere	Zavattari
NO_2 ($\mu g/m^3$)	130	130	115	120	135	142	90	116
CO (mg/m^3)	2.9	4.4	3.6	4.1	3.3	5.7	4.8	7.3

Compare the coefficients of variation *v* of the pollutants and determine their correlation *r*.

8- **Storm rainfall.** The analysis of storm data is essential for predicting flood hazards in urban areas. Annual maximum rainfall depths (in millimeters) recorded at Genoa University in Italy, for durations varying from 5 minutes to 3 hours, are presented here for the years 1974–1987.

	Duration (min)								
Year	5	10	20	30	40	50	60	120	180
1974	12.1	19.5	28.8	30.5	32.4	35.5	38.7	48.0	51.6
1975	10.1	14.9	26.7	31.2	34.7	38.2	40.2	55.0	56.0
1976	17.9	20.0	31.1	37.2	41.1	51.0	55.7	67.1	80.6
1977	20.0	32.6	52.6	72.4	90.1	108.8	118.9	146.5	157.3
1978	5.1	13.6	16.0	21.3	24.1	24.6	25.0	40.7	49.9
1979	20.5	26.1	36.3	46.1	49.3	50.3	55.6	65.2	90.1
1980	10.0	15.7	20.9	25.0	30.5	38.0	40.1	58.0	63.8
1981	12.0	27.9	47.9	56.0	70.0	80.0	89.4	106.9	114.2
1982	10.0	14.4	20.0	23.3	25.1	26.4	27.2	34.3	41.2
1983	10.0	12.1	17.3	19.2	22.1	27.3	32.7	54.4	66.5
1984	20.1	32.8	60.0	65.7	76.1	92.8	105.7	122.3	122.3
1985	7.6	8.1	13.0	16.5	21.6	25.3	25.3	27.0	32.3
1986	8.7	11.7	20.0	22.9	26.1	26.3	27.6	41.1	56.7
1987	24.6	36.7	56.7	73.9	93.9	110.1	128.5	180.8	188.7

Compute the mean $\bar{x}$ and standard deviation $\hat{s}$ and coefficient of skewness $g1$ for each duration. Are there some regularities in the growth of these statistics with increasing duration? Comment on the results and the physical relevance to storm characteristics.

9- **Carbon dioxide.** The records of atmospheric trace gases are used in the study of global climatic changes. Monthly carbon dioxide concentrations (in parts per million in volumes) recorded at Mount Cimone, Italy, from 1980 to 1988 is given here.

	Month											
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1980	340.87	339.83	342.27	342.51	338.27	335.52	330.14	328.81	331.17	335.03	339.05	340.43
1981	341.47	343.11	342.39	342.51	339.49	335.28	330.77	330.30	333.55	336.80	339.41	343.18
1982	341.70	344.38	345.68	345.70	340.80	336.66	334.65	332.40	335.15	339.26	341.19	345.18
1983	342.38	346.18	345.00	344.24	342.32	338.34	336.03	335.00	336.57	339.86	343.97	345.61
1984	346.32	349.44	351.33	350.50	346.43	344.35	346.29	335.19	337.59	342.26	344.88	346.91
1985	349.92	348.17	350.62	350.61	345.93	341.43	337.67	337.16	339.40	344.07	349.49	347.40
1986	349.41	351.41	352.29	350.75	348.37	342.96	337.22	338.53	340.90	346.28	348.95	350.52
1987	351.94	353.75	354.79	352.61	350.39	347.38	341.64	341.64	342.19	345.60	350.39	352.36
1988	353.13	355.02	354.96	354.51	352.20	346.71	342.60	344.60	343.66	348.99	352.42	353.27

Compute the mean $\bar{x}$ and standard deviation $\hat{s}$ for each year (by rows) and for each month (by columns). Because the temporal evolution of the annual mean indicates that carbon dioxide increases (probably resulting in global warming), compute the annual rate of increase. Comment on the results.

10- Sea waves. Because of scarcity of records, the characteristics of sea waves are often derived from other climatological data. They investigated two different strategies for model calibration, called "no. 1" and "no. 2" in the table presented here. The table also includes the observed and the simulated values of the height of the highest sea wave and of its period for measurements taken from August 1977 to September 1978.

		Simulated values			
Measured values		Calibration strategy no. 1		Calibration strategy no. 2	
Height (m)	Period (s)	Height (m)	Period (s)	Height (m)	Period (s)
2.26	6.1	1.81	5.4	1.54	5.8
3.10	4.3	2.93	6.8	2.54	6.4
3.22	5.7	3.24	7.2	2.80	6.7
3.84	7.7	3.18	7.1	2.69	6.6
2.56	5.3	2.74	6.6	2.32	6.1
2.74	5.7	3.49	7.4	3.00	6.9
2.28	4.9	2.12	5.8	1.80	5.4
3.88	6.7	5.10	9.0	4.43	8.4
2.49	5.0	2.14	5.8	1.81	5.4
4.22	6.9	4.45	8.8	3.77	7.7
2.01	5.0	2.57	6.4	2.19	5.9
2.77	5.9	2.68	6.5	2.27	6.0
3.61	6.5	3.86	7.8	3.36	7.3
3.51	7.4	4.02	8.0	3.51	7.5
2.52	5.0	3.39	7.3	2.95	6.9
2.12	5.1	2.61	6.5	2.21	6.0
2.73	6.5	2.22	6.0	1.88	5.5
3.30	5.4	4.05	8.0	3.49	7.5

Draw a scatter diagram to compare the observed and simulated values of wave heights and periods. Compute the correlation coefficients r. Compute the deviations of the simulated data from the observed data, and find the mean $\bar{x}_1$, standard deviation $\hat{s}_1$, and coefficient of variation v of these deviations. Do these results indicate which of the two investigated strategies provides the better representation of sea waves from climatological data?

Solution of CHAPTER 1

1- RDEN: 61.0, 17.39, 0.63; ACCEL: 0.327, 0.142, 0.70; r = 0.98.

2- 5408, 1735; *p* = 0.54.

3- 2837, 1301.

4- 985,000.

5-

6- 4.9, 10.7% *r* = 0.027.

7- 1.2, 3.0%; *r* = −0.15.

8- Increase with duration.

9- 1.65 per year.

10-