Problem1: statistical computing

Part1:

44 A random variable X is said to follow a lognormal distribution if $Y = \log(X)$ follows a normal distribution. The lognormal is sometimes used as a model for heavy-tailed skewed distributions.

a Calculate the density function of the lognormal distribution.

b Examine whether the lognormal roughly fits the following data (Robson 1929), which are the dorsal lengths in millimeters of taxonomically distinct octopods.

110	15	60	54	19	115	73
190	57	43	44	18	37	43
55	19	23	82	175	50	80
65	63	36	16	10	17	52
43	70	22	95	20	41	17
15	12	11	29	29	61	22
40	17	26	30	16	116	28
32	33	29	27	16	55	8
11	49	82	85	20	67	27
44	16	6	35	17	26	32
76	150	21	5	6	51	75
23	29	64	22	47	9	10
28	18	84	52	130	50	45
12	21	73				

45 **a** Generate samples of size 25, 50, and 100 from a normal distribution. Construct probability plots and hanging rootograms. Do this several times to get an idea of how probability plots behave when the underlying distribution is really normal.

b Repeat part (a) for a chi-square distribution with 10 df.

c Repeat part (a) for $Y = Z/U$, where $Z \sim N(0, 1)$ and $U \sim U[0, 1]$ and Z and U are independent.

d Repeat part (a) for a uniform distribution.

e Repeat part (a) for an exponential distribution.

f Can you distinguish between the normal distribution of part (a) and the subsequent nonnormal distributions?

46 Suppose that a sample is taken from a symmetric distribution whose tails decrease more slowly than those of the normal distribution. What would be the qualitative shape of a normal probability plot of this sample?

47 The Cauchy distribution has the probability density function

$$f(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right), \qquad -\infty < x < \infty$$

What would be the qualitative shape of a normal probability plot of a sample from this distribution?

Part 2:

40 Olson, Simpson, and Eden (1975) discuss the analysis of data obtained from a cloud seeding experiment. A cloud was deemed "seedable" if it satisfied certain criteria; for each seedable cloud a decision was made at random whether or not to actually seed. The nonseeded clouds are referred to as control clouds. The following table presents the rainfall from 26 seeded and 26 control clouds. Make Q-Q plots for rainfall versus rainfall and log rainfall versus log rainfall. What do these plots suggest about the effect, if any, of seeding?

Seeded Clouds							
129.6	31.4	2745.6	489.1	430.0	302.8	119.0	4.1
92.4	17.5	200.7	274.7	274.7	7.7	1656.0	978.0
198.6	703.4	1697.8	334.1	118.3	255.0	115.3	242.5
32.7	40.6						

Control Clouds							
26.1	26.3	87.0	95.0	372.4	0.01	17.3	24.4
11.5	321.2	68.5	81.2	47.3	28.6	830.1	345.5
1202.6	36.6	4.9	4.9	41.1	29.0	163.0	244.3
147.8	21.7						

41 Construct a nonparametric confidence interval for a quantile x_p by using the same reasoning as in the derivation of a confidence interval for a median.

42 In a study of the natural variability of rainfall, the rainfall of summer storms was measured by a network of rain gauges in southern Illinois for the years 1960–1964 (Changnon and Huff, in LeCam and Neyman 1967). The following tables give the average amount of rainfall (in inches) from each storm, by year.

1960							
.02	.001	.001	.12	.08	.42	1.72	.05
.01	.01	.003	.001	.003	.27	.001	.06
.05	2.13	.04	1.10	.02	.001	.14	.08
.21	:07	.32	.24	.29	.001	.29	1.13
.003	.01	.19	.002	.01	.04	.002	.07
.45	.01	.18	.67	.003	.01	.04	.002

1961							
.49	.02	.02	.34	.14	.37	.33	.33
.35	.01	.50	.76	1.06	.002	.06	.16
.27	.25	.29	.02	.05	.46	.07	.41
.02	.08	.21	.01	.44	.02	.05	.11
1.50	.003	.18	.01	.002	.24	.01	.75
.01	.14	.13	.01	.01	.27	.45	1.78

Part 3:

43 This and the next two problems are based on discussions and data in Le Cam and Neyman (1967), which is devoted to the analysis of weather modification experiments. The examples illustrate some ways in which principles of experimental design have been used in this field. During the summers of 1957 through 1960, a series of randomized cloud-seeding experiments were carried out in the mountains of Arizona. Of each pair of successive days, one day was randomly selected for seeding to be done. The seeding was done during a two-hour to four-hour period starting at midday, and rainfall during the afternoon was measured by a network of 29 gauges. The data for the four years are given in the table below (in inches). Observations in this table are listed in chronological order.

a Analyze the data for each year and for the years pooled together to see if there appears to be any effect due to seeding. You should use graphical descriptive methods to get a qualitative impression of the results and hypothesis tests to assess the significance of the results.

b Why should the day on which seeding is to be done be chosen at random rather than just alternating seeded and unseeded days? Why should the days be paired at all, rather than just deciding randomly which days to seed?

1957		1958		1959		1960	
Seeded	Unseeded	Seeded	Unseeded	Seeded	Unseeded	Seeded	Unseeded
0	.154	.152	.013	.015	0	0	.010
.154	0	0	0	0	0	0	0
.003	.008	0	.445	0	.086	.042	.057
.084	.033	.002	0	.021	.006	0	0
.002	.035	.007	.079	0	.115	0	.093
.157	.007	.013	.006	.004	.090	0	.183
.010	.140	.161	.008	.010	0	.152	0
0	.022	0	.001	0	0	0	0
.002	0	.274	.001	.055	0	0	0
.078	.074	.001	.025	.004	.076	0	0
.101	.002	.122	.046	.053	.090	0	0
.169	.318	.101	.007	0	0	0	0
.139	.096	.012	.019	0	.078	.008	0
.172	0	.002	0	.090	.121	.040	.060
0	0	.066	0	.028	1.027	.003	.102
0	.050	.040	.012	0	.104	.011	.041
				.032	.023		
				.133	.172		
				.083	.002		
				0	0		

Part 4:

27 The following table gives the survival times (in hours) for animals in an experiment whose design consisted of three poisons, four treatments, and four observations per cell.

a Conduct a two-way analysis of variance to test the effects of the two main factors and their interaction.

b Box and Cox (1964) analyzed the reciprocals of the data, pointing out that the reciprocal of a survival time can be interpreted as the rate of death. Conduct a two-way analysis of variance, and compare to the results of part (a). Comment on how well the standard two-way analysis of variance model fits and on the interaction in both analyses.

	Treatment							
Poison	A		B		C		D	
I	3.1	4.5	8.2	11.0	4.3	4.5	4.5	7.1
	4.6	4.3	8.8	7.2	6.3	7.6	6.6	6.2
II	3.6	2.9	9.2	6.1	4.4	3.5	5.6	10.0
	4.0	2.3	4.9	12.4	3.1	4.0	7.1	3.8
III	2.2	2.1	3.0	3.7	2.3	2.5	3.0	3.6
	1.8	2.3	3.8	2.9	2.4	2.2	3.1	3.3