AN INFECTIOUS PROBLEM SET

EPIDEMIOLOGY PART 1

Imagine that you were an epidemiologist working in a government laboratory during the summer in 1980; you notice a curious report about Kaposi's sarcoma, a rare skin cancer, occurring in young men. You become alerted because Kaposi's sarcoma normally is a slowly developing, non-lethal cancer in elderly men, but it was ferociously virulent in the cases that are described in the report (Kaposi's sarcoma often becomes combined with *Pneumocystis carinii* pneumonia, a rare protozoan-based disease); you also recall that one or two other cases had been reported recently. Being a Computational Biologist, you decide to collect all available data and monitor the problem. By December 1981, you have amassed the following data:

Table 1. Newly and cumulatively reported cases for Kaposi's sarcoma (or related unusual opportunistic infections in men).

Year	Quarter	New Cases	Cumulative Cases	
1979	-	-	-	
	1	2	2	
	2	5	7	
	3	3	10	10
1980	4	8	18	
	5	12	30	
	6	9	39	
	7	16	55	55
1981	8	29	84	
	9	40	124	
	10	75	199	
	11	88	287	287

(these data actually were derived from Morbidity and Mortality Weekly Reports (MMWR), Center for Disease Control (CDC), Atlanta, Georgia; the data have been pooled for quarterly periods, which are numbered for convenience).

You become convinced that something serious is transpiring.

1. Please plot the ‘raw’ data and examine the resulting pattern. Please use regression analysis to determine an equation relating cumulatively reported cases N and time t (measured in three-month periods; use time as the independent variable).

2. Please determine an equation relating newly reported cases M and time t.

3. Using the equations that you obtained in responding to points 1 and 2, please predict N and M for the last quarter in 1982 (*i.e.*, for t = 15).

EPIDEMIOLOGY PART 2

Imagine that 1982 has just ended (the Falkland Islands conflict between Great Britain and Argentina has been resolved; Micheal Jackson's album *Thriller* topped the charts, Madonna's debut album soon will be released, and Britney Spears just celebrated her first birthday; and the first permanent artificial heart has been surgically implanted in Dr. Barney Clark). The nasty new disease was given a name in July: Acquired Immune Deficiency Syndrome – AIDS. You have updated the data:

Table 2. Newly and cumulatively reported cases for Acquired Immune Deficiency Syndrome - AIDS, from 1979 to mid 1983 (the data were derived from MMWRs).

Year	Quarter	New Cases	Cumulative Cases	
1979	-	-	-	
	1	2	2	
	2	5	7	
	3	3	10	10
1980	4	8	18	
	5	12	30	
	6	9	39	
	7	16	55	55
1981	8	29	84	
	9	40	124	
	10	75	199	
	11	88	287	287
1982	12	163	450	
	13	171	621	
	14	278	899	
	15	339	1238	1238
1983	16	485	1723	
	17	612	2335	

4. Please assess the accuracy for your projections in responding to point 3 from Part 1.

5. Please comment on whether you used an appropriate model in Part 1. A viral population rampaging without control is expected to spread in an exponential manner.

6. Please fit to the data exponential equations for N(t) and M(t) for the first (11 if you used a power function in Part 1 and then) 17 quarters.

7. Please predict new and cumulative cases for the last quarter in 1984, (*i.e.*, t = 23).

8. Please plot the 'raw' data. Then, please calculate and plot N(t) and M(t) for each quarter up to t = 23, using the equations you have determined in responding to point 6. Please assess the accuracy for the model that you are using.

EPIDEMIOLOGY PART 3

Imagine that 1993 has just begun (the Toronto Maple Leafs will be eliminated in the semi-finals by the Los Angeles Kings; Michael Jackson's album *Dangerous* will reveal that he is coasting on his past successes, Madonna's album *Erotica* and book *SEX* will reveal that she is in her prime, and Britney Spears is about to join the Mickey Mouse Club along with Christina Aguilera and Justin Timberlake; and the second-generation genetic map for the human genome project has been established). The most-recent data compilation that is available to you is the summary for 1992 (the CDC has been falling behind due to budget cuts – what else is new?), which appears on the following page.

9. Please assess the accuracy for your projections in responding to point 7 from Part 2.

10. Please comment on whether you used an appropriate model in Part 2.

11. Just as population growth ultimately is limited, so, too, is disease spread. Perhaps the disease infiltrated throughout the entire population, all susceptible individuals became infected, or some advance in medical technology checked it. Whatever the reason for the constraint, you expect that a logistic model would describe accurately the epidemic. To fit a logistic equation to the data, you require some information about the carrying capacity K. Lacking specialised information, you suppose that K = 350000 for N(t) and 20000 for M(t). Please fit to the data a logistic model.

12. Please plot the 'raw' data. Then, please calculate and plot N(t) and M(t) from 1979 to 1992 (annually, to minimise keytapping and mouseclicking).

13. Considering all the data, please compare your predictions using as models any among the power function, exponential equation, and logistic equation (whatever you used). Remember that a model at best approximates reality. Please state which among these models best describes the real situation.

Table 3. . Newly and cumulatively reported cases for Acquired Immune Deficiency Syndrome - AIDS, from 1979 to 1992.

Year	Quarter	New Cases	Cumulative Cases	
1979	-	-	-	
	1	2	2	
	2	5	7	
	3	3	10	10
1980	4	8	18	
	5	12	30	
	6	9	39	
	7	16	55	55
1981	8	29	84	
	9	40	124	
	10	75	199	
	11	88	287	287
1982	12	163	450	
	13	171	621	
	14	278	899	
	15	339	1238	1238
1983	16	485	1723	
	17	612	2335	
	18	663	2998	
	19	685	3683	3683
1984	20	815	4498	
	21	1047	5545	
	22	1165	6710	
	23	1418	8128	8128
1985	24	1599	9727	
	25	2042	11769	
	26	2217	13986	
	27	2391	16377	16377
1986	28	2814	19191	
	29	3233	22424	
	30	3459	25883	
	31	3426	29309	29309
1987	32	4572	33881	
	33	4505	38386	
	34	4654	43040	
	35	7339	50379	50379
1988	36	7407	57786	
	37	7880	65666	
	38	8201	73867	
	39	7513	81380	81380
1989	40	7892	89272	

	41	8556	97828	
	42	8969	106797	
	43	8305	115102	115102
1990	44	10191	125293	
	45	11142	136435	
	46	11790	148225	
	47	8472	156697	156697
1991	48	10357	167054	
	49	10666	177720	
	50	12507	190227	
	51	10142	200369	200369
1992	52	11890	212259	
	53	11116	223375	
	54	11531	234906	
	55	10935	245841	245841

EPIDEMIOLOGY PART 4

The discrete version for the logistic equation

$$N_{t+1} = r_{difference} N_t (K - N_t)$$

may be used to model systems that persist discontinuously, such as annual plant or insect species or viruses (*e.g.*, 'cold and flu seasons'). The equation involves iteration: the value for the initial population size N_0 is used to calculate N_1; N_1 is used to calculate N_2; …; N_t is used to calculate N_{t+1}.

A convenient means to model growing populations for which a carrying capacity K is known is to rescale the N relative to K; this is achieved by assigning the carrying capacity as the value when N reaches 100% or, numerically, K = 1.00. The discrete version for the logistic equation then becomes

$$N_{t+1} = r_{difference} N_t (1 - N_t)$$

Please consider a growing population like the Human immunodeficiency virus HIV that you considered in questions 1 to 8. One could consider how it might grow under different conditions, by tinkering with assigning a variety of values for N_0 and $r_{difference}$ in the difference equation.

14. Please start with $N_0 = 0.5$ and $r \in (1, 3)$ (*i.e.*, a number between 1 and 3 … 'pick a number, any number') and calculate $N_1, N_2, \ldots, N_t$ until the population size 'reaches equilibrium' (*i.e.*, stabilises). Please describe as concisely as you can how the population grows under this condition; plotting $\{t, N_t\}$ pairs might help you visualise and interpret the result.

Please describe as concisely as you can the result that would be obtained if the same r but different N_0 (*i.e.*,. greater and less than 0.5) were used.

15. Please choose $r \in (3, 3.6)$ and plot values for N_t as t increases; please describe as concisely as you can how the population grows under this condition. Please choose $r > 3.6$ and plot values for N_t as t increases; please describe as concisely as you can how the population grows under this condition (choose your words carefully).