Assignment 4 STATS 707 Computational introduction to statistics Due 4:00pm Thur 2 June

1. An environmental scientist identified a point source for E. Coli at the edge of a stream. She then measured y =E. Coli, in colony forming units per 100 ml water, at different distances, in feet, downstream from the point source. Suppose she obtains the following pairs of (x, y).

x	100	150	250	250	400	650	1000	1600
y	21	20	24	17	18	10	11	9

   (a) Transform the x values to $x' = \log_{10} x$ and plot the scatter diagram of y versus x'.

   (b) Fit a straight line regression to the transformed data.

   (c) Obtain a 90% confidence interval for the slope of the regression line.

   (d) Estimate the expected y value corresponding to $x = 300$ and give a 95% confidence interval.

2. As part of the multilab study, four fabrics are tested for flammability at the National Bureau of Standards. The following burn times in seconds are recorded after a paper tab is ignited on the hem of a dress made of each fabric.

Fabric 1	Fabric 2	Fabric 3	Fabric 4
17.8	11.2	11.8	14.9
16.2	11.4	11.0	10.8
17.5	15.8	10.0	12.8
17.4	10.0	9.2	10.7
15.0	10.4	9.2	10.7

   (a) State the statistical model and present the ANOVA table. With $\alpha = 0.05$, test the null hypothesis of no difference in the degree of flammability for the four fabrics.

   (b) If the null hypothesis is rejected, construct simultaneous confidence intervals to determine the fabric(s) with the lowest mean burn time.

   (c) Plot the residuals and comment on the plausibility of the assumptions.

   (d) If the tests had been conducted one at a time on a single mannequin, how would you have randomized the fabrics tested in this experiment?

3. Many industrial air pollutants adversely affect plants. Sulphur dioxide causes leaf damage in the form of intraveinal bleaching in many sensitive plants. In a study of the effect of a given concentration of sulphur dioxide in the air on three types of garden vegetables, 40 plants of each type are exposed to the pollutant under controlled greenhouse conditions. The frequencies of severe leaf damage are recorded in the following table:

	Leaf damage		
	Severe	Moderate or none	Total
Lettuce	32	8	40
Spinach	28	12	40
Tomato	19	21	40
Total	79	41	120

   Analyze these data to determine if the incidence of severe leaf damage is a like for the three types of plants. In particular:

   (a) Formulate the null hypothesis.

   (b) Test the null hypothesis with $\alpha = 0.05$.

   (c) Construct three individual 98% confidence intervals.

4. The following record shows a classification of 41,208 births in Wisconsin (courtesy of Professor Jerome Klotz). Test the goodness of fit of the model that births are uniformly distributed over all 12 months of the year. Use $\alpha = 0.01$.

Jan.	3478	July	3476
Feb.	3333	Aug.	3495
Mar.	3771	Sept.	3490
Apr.	3542	Oct.	3331
May	3479	Nov.	3188
June	3304	Dec.	3321
		Total	41208

5. An investigation is conducted to determine if the exposure of women to atomic fallout influences the rate of birth defects. A sample of 500 children born to mothers who were exposed to the atomic explosion in Hiroshima is to be studied. A sample of 400 children from a Japanese island far removed from the site of the atomic explosion forms the control group. Suppose that the following data for the incidence of birth defects are obtained:

	Birth defect		
	Present	Absent	Total
Mother exposed	84	416	500
Mother not exposed	43	357	400
Total	127	773	900

   (a) Do these data indicate that there is a different rate of incidence of birth defects for the two groups of mothers?

   (b) Use the normal test for testing the equality of two population proportions with a two-sided alternative. Verify the relation $\chi^2 = z^2$ by comparing their numerical values.