University of Waterloo STAT 221 – W. H. Cherry

Assignment 8

A8 – 1. A new procedure is being investigated for measuring calcium content. The following table shows the *actual* calcium content (x) for each of ten samples, and the calcium content (y) given by the new procedure:

x	4.0	8.0	12.5	16.0	20.0	25.0	31.0	36.0	40.0	40.0
y	3.7	7.8	12.1	15.6	19.8	24.5	31.1	35.5	39.4	39.5.

(a) Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of $\mathbf{Y}$ on $\mathbf{X}$.

(b) Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of σ for these data.

(c) Assess how well the straight-line regression model fits the data; give your reasons completely but concisely.

(d) Using an appropriate 95% confidence interval, check whether the respondent (or study) population regression of $\mathbf{Y}$ on $\mathbf{X}$ can be considered to pass through the origin.

(e) Find a 90% *confidence* interval for the *mean* $\mu_Y(x_j = 23.25)$ representing ${}_{reg}\overline{\mathbf{Y}}_i(\mathbf{X}_i)$, the respondent ((or study) population regression *average* calcium content given by the new procedure when the actual calcium content is specified by explanatory variate $\mathbf{X}_i = 23.25$ units.

A8 – 2. Observations on the volume of algae cells were made to determine their growth rate in a particular medium. Because many natural growth processes are approximately exponential with time, it seemed appropriate to relate the *logarithms* of the volumes (rather than the volumes themselves) to time. The log(volume f cells) (y) in the medium was recorded on each of eight consecutive days (x), as follows:

x	1	2	3	4	5	6	7	8
y	3.538	3.828	4.349	4.833	4.911	5.297	5.566	6.036.

(a) Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of $\mathbf{Y}$ on $\mathbf{X}$.

(b) Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of σ for these data.

(c) Assess how well the straight-line regression model fits the data; give your reasons completely but concisely.

(d) Find a 99% confidence interval for the growth rate of the logarithm of the algae volume in the respondent (or study) population.

(e) How would the regression analysis of these data be affected if the days were numbered 0, 1, ..., 7 instead of 1, 2, ..., 8? Explain briefly.

A8 – 3. Theory suggests that a linear relationship should exist between the shearing strength of steel bolts and their diameters. The following table gives the diameter (x inches) and shearing strength (y units) for nine bolts made of a particular type of steel:

x	⅛	¼	⅜	½	⅝	¾	⅞	1	1½
y	47	72	97	126	165	186	233	257	311.

(a) Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of $\mathbf{Y}$ on $\mathbf{X}$.

(b) Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of σ for these data.

(c) Assess how well the straight-line regression model fits the data; give your reasons completely but concisely.

(d) Test the hypothesis the intercept of the respondent (or study) population regression of $\mathbf{Y}$ on $\mathbf{X}$ is zero.

(e) Find a 95% *confidence* interval for the *mean* $\mu_Y(x_j = 1¼)$ representing ${}_{reg}\overline{\mathbf{Y}}_i(\mathbf{X}_i)$, the responent (or study) population regression *average* shearing strength of bolts with diameter specified by explanatory variate $\mathbf{X}_i = 1¼$ inches.

(f) Find a 95% *prediction* interval for the *random variable* $Y(x_j = 1¼)$ representing ${}_{reg}\mathbf{Y}_i(\mathbf{X}_i)$, the respondent (or study) population regression shearing strength of an equiprobably-selected *individual* bolt with diameter specified by explanatory variate $\mathbf{X}_i = 1¼$ inches.

(g) **BONUS:** Repeat the question omitting the *last* observation (for bolt diameter 1½ inches).

A8 – 4. Samples selected equiprobably (of the same size) of families from each of eight cities of various population sizes (x people) were questioned about how much they spent per year on food, clothing and housing (y dollars). The city population sizes and the *total* family responses were as follows [both in *thousands* (of people or dollars)]:

x	30	50	75	100	120	150	175	200
y	65	77	79	80	81	82	84	90.

(a) Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of $\mathbf{Y}$ on $\mathbf{X}$.

(b) Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of σ for these data.

(c) Assess how well the straight-line regression model fits the data; give your reasons completely but concisely.

(*continued overleaf*)

A8 – 4. (cont.) (d) Give a *point* estimate for the average total expenditure, for a sample of families of the same size as that used in the survey, for a city of population size **(i)** 125,000; **(ii)** 10,000;
discuss briefly how close to the true value you expect these two point estimates to be.

A8 – 5. Ethylene oxide has been found to be an effective fumigant in granaries for the control of pests such as mice, beetles, ants, crickets, etc. To investigate whether the use of ethylene oxide affected the nutritional quality of the grain, ten genetically-identical rats, of the same age and of approximately equal weight, were selected equiprobably and two were fed a fixed weight of grain that had *not* been treated with ethylene oxide. Pairs of the remaining eight rats were fed the same fixed weight of grain that *had* been fumigated with ethylene oxide for 1, 2, 3 and 4 units of time. The controlled feedings were made for 1 week, and the differences in the initial and final weights of the rats were recorded; the results were:

Fumigation time (t)	0	0	1	1	2	2	3	3	4	4
Weight loss (w)	1	2	9	14	49	63	88	137	485	401.

(a) Prepare a properly-labelled scatter diagram of these data and show on it the estimated regression of **W** on **T**.

(b) Give the ANOVA table, coefficient of determination, correlation coefficient and estimate of σ for these data.

(c) Assess how well the straight-line regression model fits the data; give your reasons completely but concisely.

(d) Give a *point* estimate for the average weight loss of rats, similar to those used in the experiment, fed for one week on the same fixed weight of grain that had been fumigated with ethylene oxide for (i) 2½ or (ii) 10 units of time; discuss briefly how close to the true value you expect the two point estimates to be.

(e) Find the estimated regression of ln**W** on **T**; set out the corresponding ANOVA table, find the coefficient of determination, correlation coefficient and estimate of σ, and assess how well the *transformed* regression model fits the data.

(f) Repeat (d) for the transformed model in (e).

Assignment 7 (continued)

A7 – 12.

Year	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937
x	1.350	1.960	2.270	2.483	2.730	3.091	3.647	4.620	5.497	6.260	7.012	7.618	8.131	8.593
y	8	8	9	10	11	11	12	16	18	19	20	21	22	23 ;

for these data: $\sum_{j=1}^{n} x_j = 65.262$, $\sum_{j=1}^{n} y_j = 208$, $\sum_{j=1}^{n} x_j y_j = 1{,}148.08$;
$\sum_{j=1}^{n} x_j^2 = 385.193\,966$; $\sum_{j=1}^{n} y_j^2 = 3{,}490$; $n = 14$.

(a) Prepare a properly-labelled scatter diagram of these data.

(b) Calculate the value of the correlation coefficient of x and y.

(c) Comment briefly on a claim based on the data that listening to the radio could result in confinement to a psychiatric institution.