Assignment 4

A4 – 1. A government inspector wishes to estimate the average weight of fill for cereal boxes packaged by a certain factory. The cereal is available to her in cartons containing 12 boxes each. The inspector obtains five cartons by equiprobable selecting (EPS) and measures the weight of fill for every box in these cartons, with the results at the right. From these data, find an approximate 90% confidence interval for the average weight of fill per box:

CARTON	OUNCES OF FILL											
1	16.1	15.9	16.1	16.2	15.9	15.8	16.1	16.2	16.0	15.9	15.8	16.0
2	15.9	16.2	15.8	16.0	16.3	16.1	15.8	15.9	16.0	16.1	16.1	15.9
3	16.2	16.0	15.7	16.3	15.8	16.0	15.9	16.0	16.1	16.0	15.9	16.1
4	15.9	16.1	16.2	16.1	16.1	16.3	15.9	16.1	15.9	15.9	16.0	16.0
5	16.0	15.8	16.3	15.7	16.1	15.9	16.0	16.1	15.8	16.0	16.1	15.9

(a) on the basis of the clustered design actually used to collect the data;

(b) ignoring the clustering and assuming that the data were obtained by EPS of $n = 60$ *individual* boxes. Comment briefly on how the intervals in (a) and (b) compare.

(c) What assumption is necessary about the total number of cartons packaged by the factory? Is it a reasonable assumption? Explain briefly.

A4 – 2. As part of a safety campaign, a taxi company wishes to estimate the number of unsafe tires on its fleet of 175 taxis. Twenty-five taxis, obtained by equiprobable selecting (EPS), show the numbers of unsafe tires (out of the four tires in use) per taxi as shown at the right.

0 4 1 3 1   2 4 0 1 2   2 0 1 1 2   2 4 1 0 0   3 1 2 2 1.

(a) Estimate the *proportion* of unsafe tires in use on the company's taxis, and find an estimate of the standard deviation of this estimate.

(b) Estimate the *number* of unsafe tires in use on the company's taxis, and find an estimate of the standard deviation of this estimate.

(c) Explain briefly whether you expect the estimates in (a) and (b) to be more or less precise than an estimate based on choosing, by EPS, 100 of the 700 tires in use on the company's taxis. Check your answer by an appropriate calculation based on the data given above.

A4 – 3. Holes of a uniform size are dug at haphazardly placed locations in a section of a dry stream bed. In the soil from each hole, the number of insect larvae, and the number of these which are of species *A*, are counted. The results are:

Hole	1	2	3	4	5	6	7	8
Number of larvae	150	42	93	104	37	93	120	85
Number of species *A* larvae	61	21	37	56	22	44	72	34.

(a) Find an approximate 95% confidence interval for the proportion of species *A* among the insect larvae in the soil of the stream bed.

(b) Explain briefly whether you would expect the estimate in (a) to be more or less precise than an estimate based on obtaining, by equiprobable selecting, the same number of insect larvae from *one* larger soil sample. Check your answer by an appropriate calculation based on the data given above.

A4 – 4. A study population consists of $\mathbf{N} = 5$ elements which have the values of two responses ($\mathbf{Y}$ and $\mathbf{X}$) as in the table below.

(a) Find the population *ratio*: $\mathbf{R} = \overline{\mathbf{Y}}/\overline{\mathbf{X}} \equiv {}_{T}\mathbf{Y}/{}_{T}\mathbf{X}$.

(b) List all the possible samples of size 2, under equiprobable selecting (EPS), from this population.

(c) For each sample, tabulate the value of the *sample* ratio: $r = \overline{y}/\overline{x} \equiv {}_{T}y/{}_{T}x$; also tabulate the probability of selecting each sample.

(d) Use the probability function in (c) to find the mean and then the *estimating bias* of the random variable R representing the sample ratio under EPS.

i	$\mathbf{Y}$	$\mathbf{X}$
1	1	3
2	3	5
3	3	4
4	2	4
5	5	6

(e) Use equation (2.15.7) on the second side (page 2.114) of Figure 2.15 of the Course Materials to calculate an approximate value of the estimating bias of R; discuss briefly how this value compares to that found in (d).

A4 – 5. The members of a teachers' federation are concerned about the salary increases given to high school teachers in a particular school board. Fifteen teachers are obtained by equiprobable selecting from an alphabetical list of all high school teachers in the board; all 15 teachers are then interviewed to determine their salaries for this year and the previous one; the complete data are given overleaf at the top of page #0.20-2.

(a) Use these data to estimate the ratio of average present-to-past salaries for the 750 high school teachers in the board, and give an appropriate measure of the precision of your estimate.

(b) Use these data to estimate the average of the ratio of present-to-past salaries for the 750 high school teachers in the board, and give an appropriate measure of the precision of your estimate.

(c) Comment briefly on the precision of your estimate in (a), the reason(s) for this precision, and whether you believe it to be an accurate assessment of the precision of your estimate.

(*continued overleaf*)

A4 - 5. (cont.)

TEA-CHER (j)	YEAR'S SALARY ($) Past ($x_j$)	Present (y_j)	TEA-CHER (j)	YEAR'S SALARY ($) Past ($x_j$)	Present (y_j)
1	27,000	28,000	9	27,080	28,110
2	33,500	34,700	10	26,985	27,985
3	38,960	40,420	11	40,760	42,185
4	49,780	51,375	12	32,180	33,500
5	31,775	32,980	13	45,960	47,615
6	25,540	26,610	14	35,030	36,395
7	39,455	40,835	15	36,555	37,910 ;
8	26,080	27,125			

for these data:

$\sum_{j=1}^{n} x_j = 516{,}640$, $\sum_{j=1}^{n} x_j^2 = 18{,}580{,}122{,}800$;

$\sum_{j=1}^{n} y_j = 535{,}745$, $\sum_{j=1}^{n} y_j^2 = 19{,}964{,}401{,}275$;

$\sum_{j=1}^{n} x_j y_j = 19{,}259{,}771{,}450$.

A4 – 6. Traders on the futures market are interested in relative prices of certain commodities rather than in specific price levels; these relative prices can be presented in terms of a ratio. One important ratio in agriculture is the cattle/hog ratio. From 64 trading days in 1977, the cattle and hog prices were recorded on 18 of the days obtained by equiprobable selecting, with the following results:

DAY (j)	CATTLE (x_j)	HOGS (y_j)	DAY (j)	CATTLE (x_j)	HOGS (y_j)
1	42.40	47.80	11	38.45	44.30
2	41.40	48.60	12	37.80	43.90
3	39.60	48.20	13	37.20	42.70
4	39.45	46.75	14	37.60	43.25
5	37.00	46.50	15	37.50	44.55
6	37.80	45.40	16	36.90	45.10
7	38.55	47.30	17	37.30	45.00
8	38.60	48.20	18	38.60	45.25 ;
9	38.80	49.40			
10	39.65	49.40			

for these data:

$\sum_{j=1}^{n} x_j = 831.6$, $\sum_{j=1}^{n} x_j^2 = 38{,}495.09$;

$\sum_{j=1}^{n} y_j = 694.6$, $\sum_{j=1}^{n} y_j^2 = 26{,}841.60$;

$\sum_{j=1}^{n} x_j y_j = 32{,}125.1125$.

Use these data to estimate the ratio of cattle/hog average prices for the 64-day period, and give an appropriate measure of the precision of your estimate. Discuss briefly whether you believe the latter is accurate.

A4 – 7. (a) Rework question **A4 – 2**(a) overleaf by treating it as one of estimating a *ratio* under EPS of individual *elements* (tires), rather than as estimating a *proportion* from EPS of *clusters* of equal size (taxis).

(b) On the basis of your findings in (a), or otherwise, briefly compare and contrast the mathematical structure of questions **A4 – 2** and **A4 – 3** overleaf from a data analytic perspective.

A4 – 8. In a sample survey, the study population attribute of interest for a response variate $\mathbf{Y}$ is the *reciprocal* of the average, represented by response variate $\mathbf{W}$ as shown at the *upper* right. One hundred units (or elements) obtained by equiprobable selecting from the study (or respondent) population yield an average of $\overline{y} = 4.15$ units and a standard deviation of 0.27 units, based on complete response. The estimator of $\mathbf{W}$ is the random variable W shown the *lower* right.

$$\mathbf{W} = \frac{1}{\overline{\mathbf{Y}}}$$

$$W = \frac{1}{\overline{Y}}$$

(a) Use a *second*-order univariate Taylor series expansion to find an approximate expression for the *estimating bias* of W.

(b) Use a *first*-order univariate Taylor series expansion to find an approximate expression for the *standard deviation* of W.

(c) Ignoring the estimating bias and assuming that the sampling fraction is negligible, use the result obtained in (b) to find an approximate 99% confidence interval for $\mathbf{W}$.

(d) Again assuming the sampling fraction to be negligible, find an approximate 99% confidence interval for $\overline{\mathbf{Y}}$ and then use it to obtain such an interval for $\mathbf{W}$.

(e) Compare the intervals obtained in (c) and (d) and comment briefly.