AGRICULTURAL STATISTICS

ACS220S

PRACTICAL P01

This practical is concerned with the statistical and graphical description of data.

Aims

1. Understand how to summarize data graphically (e.g. by drawing appropriate plots) and numerically (e.g. by selecting an appropriate measure of central tendency).
2. Understand the concept of an outlier.
3. Be able to describe a sample in terms of form and location, and compare it to other samples.

Objectives

1. Use a calculator to find the sample mean, median, uncorrected sum of squares, correction term, corrected sum of squares, variance, standard deviation, standard error of the mean
2. Construct and interpret a histogram and a stem and leaf plot.
3. Identify possible outliers.
4. Compare two or more samples using graphical techniques.
5. Summarize samples verbally.

NUMERICAL SUMMARIES

Exercise 1

Familiarize yourself with the following symbols

$\bar{x}$ $\sum x$ $\sum x^2$ s σ SS_x

$$Sum\,of\ observations = \sum x$$

$$Mean,\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$Uncorrected\ SS_x = \sum x^2\,(sum\,of\ the\,squared\ observations)$$

$$\left(\sum x\right)^2 = (sum\ of\ observations)\,squared$$

$$Correction\ Term,\ CT_x = \frac{\left(\sum x\right)^2}{n}$$

$$Corrected\ Sum\ of\ Squares,\ SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$Variance,\ s^2 = \frac{SS_x}{n-1}$$

$$Std\ Deviation,\ s = \sqrt{s^2}$$

$$Std\ Error\ of\ the\ mean,\ sem = \frac{s}{\sqrt{n}}$$

Take the following numbers:

8.3 7.9 5.5 5.8 6.1 7.2

Calculate the

a. Mean
b. Median
c. Variance
d. Standard deviation
e. Uncorrected sum of squares
f. Corrected sum of squares

Exercise 2

If $X_1 = 3.1\text{kg}$, $X_2 = 3.4\text{kg}$, $X_3 = 3.6\text{kg}$, $X_4 = 3.7\text{kg}$, $X_5 = 4.0\text{kg}$, what is the value of

$(a)\sum_{i=1}^{4} x_i$

$(b)\sum_{i=2}^{4} x_i$

$(c)\sum_{i=1}^{5} x_i$

(d) *Uncorrected Sum of Squares* $\sum_{i=1}^{5} x_i^2$

$(e)\left(\sum_{i=1}^{5} x_i^2\right)^2$

(f) *Corrected Sum of Squares,* SS_x

(g) *mean*

(h) *variance*

(*i*) *standard deviation*

(*j*) *median*

Exercise 3

The following are leaf weights in grams of Tobacco seedlings two weeks after germination.

1.2 1.6 1.8 1.4 2. 4 2.0 2.2

Calculate the mean, the median and the standard deviation of the weights.

Exercise 4

Now for this sample.....

15.6 24.2 18.8 19.3 22.3 18.5

Calculate the

- **a.** Median
- **b.** Mean
- **c.** Uncorrected sum of squares
- **d.** Correction term
- **e.** Sum of squares
- **f.** Variance
- **g.** Standard deviation
- **h.** Standard error of the mean

GRAPHICAL SUMMARIES

Exercise 6

The survival times (in months) for piglets from the onset of symptoms of severe anaemia are given below:

6	20	29	15	7	32	36
17	15	19	35	10	16	39
27	14	10	16	12	13	16
9	18	33	30	29	31	27

- **a.** What is the range of the data?
- **b.** Use Sturge's rule (1+3.322Logn) to find the number of classes.
- **c.** Find the class width.
- **d.** Using the classes $5 < x \leq 10$, $10 < x \leq 15$, e.t.c., make a frequency distribution of the data.
- **e.** Draw a histogram.

Exercise 7

The following data give the results of a sample survey. The researcher wanted to know if commercial farmers in the Kunene region were prepared to share their expertise with newly resettled farmers. The responses were N for ”No”, Y for ”Yes” and D for ”not sure”.

D	N	N	Y	Y	Y	N	Y	D	Y
Y	Y	Y	Y	N	Y	Y	N	N	Y
N	Y	Y	N	D	N	Y	Y	Y	Y
Y	Y	N	N	Y	Y	N	N	D	Y

a. Prepare a frequency distribution table.
b. Draw a pie chart showing the percentage distributions.
c. Draw a bar graph.

Exercise 7

The gross annual income (in *N*$000000*s*) from various divisions of a large farm in Tsumeb is given below:

Division	Amount
Dairy	5.04
Crops	3.20
Poultry	5.5
Piggery	6.25
Other	7.08

Draw a pie chart to illustrate this information. Show the percentage income from each division. Comment on the pie chart.

Exercise 8

The thicknesses (in mm) for tomato plant stems, three weeks after germination are given below:

3.4	3.21	3.26	3.37	3.4	3.35	3.4	3.48	3.3	3.38	3.27
3.35	3.28	3.39	3.44	3.29	3.38	3.38	3.4	3.38	3.44	3.29
3.37	3.41	3.45	3.44	3.35	3.35	3.46	3.31	3.33	3.47	3.33
3.37	3.31	3.51	3.36	3.32	3.33	3.43	3.39	3.39	3.28	3.33
3.25	3.28	3.3	3.41	3.39	3.33	3.27	3.34	3.33	3.42	3.35

a. Determine an appropriate frequency distribution for the above data.
b. Draw a histogram of the data. Interpret the histogram.

Exercise 9

Consider now following weights arranged in the form of a frequency distribution.

weight	frequency
1.85 - < 1.95	2
1.95 - < 2.05	1
2.05 - < 2.15	2
2.15 - < 2.25	3
2.25 - < 2.35	5
2.35 - < 2.45	6
2.45 - < 2.55	4
2.55 - < 2.65	3
2.65 - < 2.75	1

a. Calculate the mean weight.
b. Calculate the median weight
c. Calculate the standard deviation of the weights.

Exercise 10

Are the following nominal, ordinal, interval, or ratio data? Explain your answers.

a. Social Security numbers.
b. The number of students at the Polytechnic.
c. Temperatures measured in Celsius.
d. Temperatures measured on the Kelvin scale.
e. Military ranks.

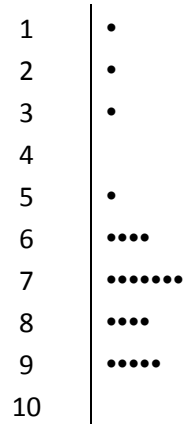
DOTPLOTS

This is a quick technique, used extensively in this course. Mark out a horizontal axis – the scale is a matter of convenience and judgement. Then simply mark with a dot (cross) the position of each data value above the axis. If two or more dots coincide, or are very close, place one above the other.

Exercise 11

Consider this sample of 24 observations...

7.93	8.37	6.92	7.13	6.60	3.28
7.51	9.71	6.78	7.30	9.81	7.88
2.97	8.23	5.47	6.98	9.43	8.29
1.36	9.30	8.20	9.69	7.16	7.32



A dotplot shows where the actual data points are, relative to each other. It is a very good way of summarizing data graphically.

An ***outlier*** is an observation which appears to be inconsistent with the rest of the data, generally because it takes an extreme value. Exercise 11 has three observations which might be outliers. Outliers often originate from an error in observation or recording of data. However, it is important to consider the possibility that an extreme observation might in fact be genuine, e.g.: because it was drawn from a skewed distribution. You have to use your own judgement.

a. Construct a stem-and-leaf plot for the above data and use it to identify the outliers.
b. Where there are apparent outliers, the sample median may be better than the sample mean as a measure of central tendency. Explain why.

Exercise 12

A botanist counted the average number of hairs per square mm on the surface of a sample of 25 plants (all of the same species). The results are shown below.

15.15	11.36	11.04	11.22	11.14
14.33	11.25	12.15	11.57	13.70
10.90	13.37	11.65	8.92	13.20
9.88	9.84	11.70	14.47	11.62
15.56	13.36	13.18	10.34	9.70

For this sample, calculate the

a. Median
b. Mean
c. Uncorrected Sum of Squares
d. Correction term
e. Sum of Squares
f. Variance
g. Standard deviation

h. Quote the mean with its standard error
i. Construct a stem and leaf plot
j. Construct a dot plot
k. Construct a histogram for the data
l. Comment on the apparent nature of the data

Exercise 13

An entomologist measured the length (mm) of the wing cases of 30 beetles. The results were....

24.9	23.2	24.9	22.4	25.7	24.9
27.9	23.5	26.8	27.2	20.5	22.8
22.5	22.8	20.2	23.4	23.6	31.6
24.3	26.8	23.2	25.6	23.3	21.8
25.1	24.4	23.7	21.2	26.2	25.1

For this sample, calculate the

a. Median
b. Mean
c. Uncorrected Sum of squares
d. Correction term
e. Sum of Squares
f. Variance
g. Standard deviation
h. Quote the mean with its standard error of the mean
i. Construct a stem and leaf plot
j. Construct a histogram for the data
k. Comment on the apparent nature of the data

GRAPHING MORE THAN ONE SAMPLE

If we want to compare two or more samples (e.g. experimental results) it is sensible to start off by graphing them on a common scale.

Exercise 14

Twenty chicks were assigned to each of three diet treatments: two experimental dietary supplements (coded A and B) and a control (no supplement).

The purpose of the experiment was to establish if there was any evidence that the supplements affected the rate of growth of the chicks. The response variable measured was a dimensionless growth index. The results are shown in the table below.

A	B	Control
17.50	18.30	20.90
15.20	13.70	22.40
14.80	14.00	20.70
23.70	11.80	21.40
21.90	18.10	19.90
15.40	20.60	21.40
22.80	16.00	20.50
15.60	17.20	20.20
18.10	16.60	22.70
17.50	19.30	21.10
21.10	18.80	18.00
14.00	16.30	20.00
14.60	17.70	21.70
15.60	17.10	20.60
24.90	17.70	20.60
26.50	18.30	20.00
21.60	19.80	22.30
18.30	14.30	22.10
12.00	18.90	22.70
17.30	16.80	21.00

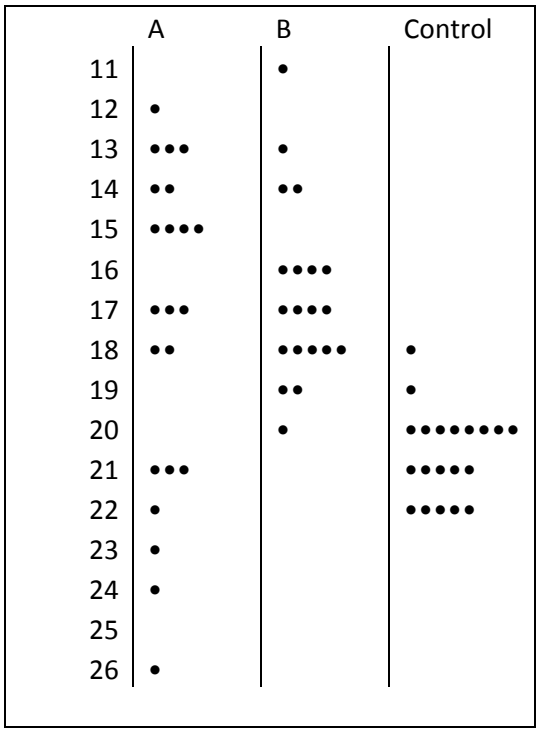
There are several ways in which these results can be analysed, but first we must obtain a summary description and plot the data.

	mean	**median**	**min**	**max**
A	18.42	17.50	12.00	26.50
B	17.06	17.45	11.80	20.60
Control	21.01	20.10	18.00	22.70

We can see that all the sample means are different, but that the median values of A and B are rather similar. The ranges of the three samples are also different. Plot will help to elucidate this....

The simplest graphical comparison is to dotplot all three samples on the same scale.

Clearly these samples differ. The Control group is clustered more or less, symmetrically around a central value of about 21.0 (both mean and median are similar).



Although the average growth index of chicks fed on the experimental diets is less than those in the Control group, this is not the whole story.

If the median is used to compare samples, then A and B are almost the same. If the mean is used, then group B is less than the group B.

We can also see that one of the effects of the diet - especially A - may have been to increase the variation found among the chicks.

The position of a sample on the x-axis is known as its location, and its shape is known as its form. For this example, we can say that the location of the A and B groups differs from the Control group (average growth index is reduced), and that the form of all three groups differs. In other words, we have some evidence that the diet supplements may have affected chick growth, both with respect to the Control and to each other. The evidence is not conclusive, of course. The

next step would be to perform a statistical test which would help us to assess the probability that these results could have occurred by chance.

Exercise 15

Examine each of the three pairs of data below. State if they differ in either form or location and explain your answer. Use dotplots, but use other techniques only if yu have time.

a	
(i)	(ii)
49.3	71.7
59.8	72.9
62.4	71.3
59.1	68.6
53.3	67.1
56.0	72.3
63.3	75.3
68.0	68.0
63.1	70.5
62.2	68.7
64.7	63.5
63.5	67.6
63.0	58.0
59.5	69.6
58.6	71.6
	73.1
	79.4
	63.2

b	
(i)	(ii)
9.6	9.8
10.9	10.5
9.9	10.0
11.4	10.7
11.9	10.9
9.1	9.5
9.3	9.6
10.2	10.1
8.1	9.1
10.1	10.0
8.9	9.4
9.8	9.9
11.1	10.6
10.3	10.2
8.6	9.3
10.4	10.2
9.7	9.8
10.7	10.4
10.6	10.3
9.4	9.7

c	
(i)	(ii)
4.8	4.8
4.0	5.1
5.3	5.8
4.7	4.4
4.7	5.4
4.1	5.2
4.0	5.6
5.6	6.0
6.0	4.2
8.2	6.8
3.7	4.6
4.9	3.2
4.2	3.7
6.6	4.0
4.5	4.9
	6.3