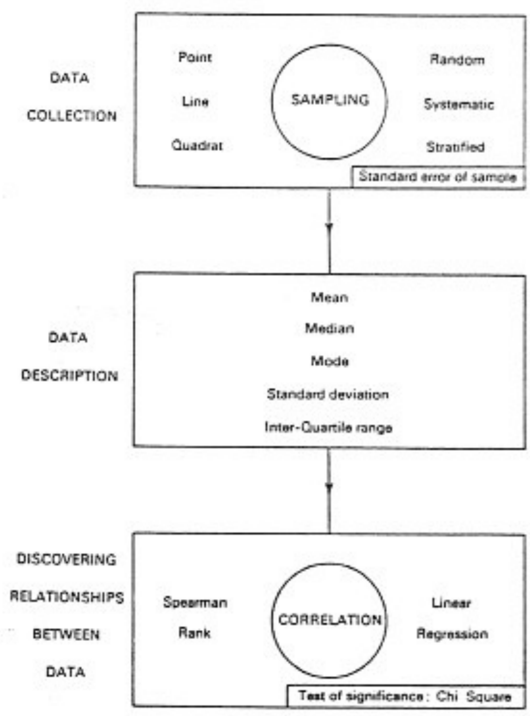
# STATISTICAL TECHNIQUES FOR GEOGRAPHICAL INVESTIGATIONS



# STATISTICAL ANALYSIS OF GEOGRAPHICAL DATA

The first stage in the analysis of data you have collected in the field is STATISTICAL ANALYSIS. Next comes the detailed written analysis and interpretation

## Background information on statistical Techniques

[redacted]

## Measures of Central Tendency.

When you have collected a set of data or a sample the first task is to find a middle value.

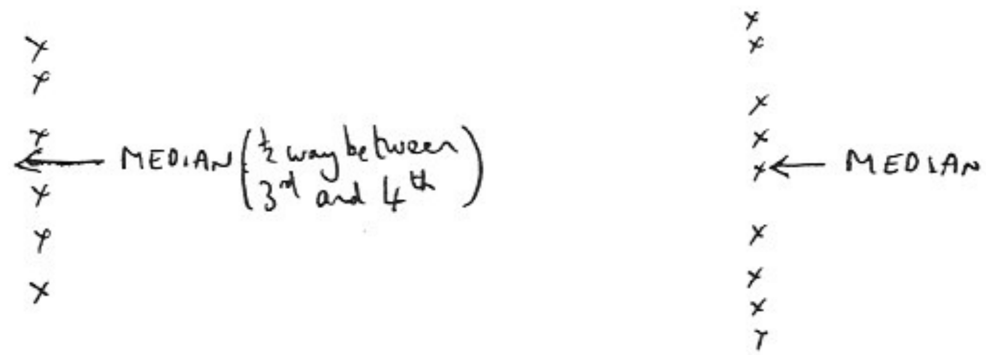
a. Mean/average $\bar{x} = \frac{\Sigma x}{n}$

$x$ = data
$\Sigma$ = sum of
$n$ = sample size.

b. Median. The mid point of a set of data.
The data is ranked in descending order and the middle value is the median.

AN EVEN NUMBER OF VALUES

AN ODD NUMBER OF VALUES




The median is not influenced by extreme values.

e. Mode. The most frequently occuring value.

The data may need to be grouped into classes to give a Modal class.

The number of classes is imp.

use $5 \times \log n$ $n$ = no of values.

this gives $\simeq$ no of classes.

class size $\simeq \dfrac{\text{Range of data}}{\text{no of classes.}}$

---

Rarely do the Mean, Mode and Median coincide exactly (only with a normal distribution)

## Measures of Spread / dispersion

After finding a 'middle' value we need to find out by how much the data spreads around this 'middle' value

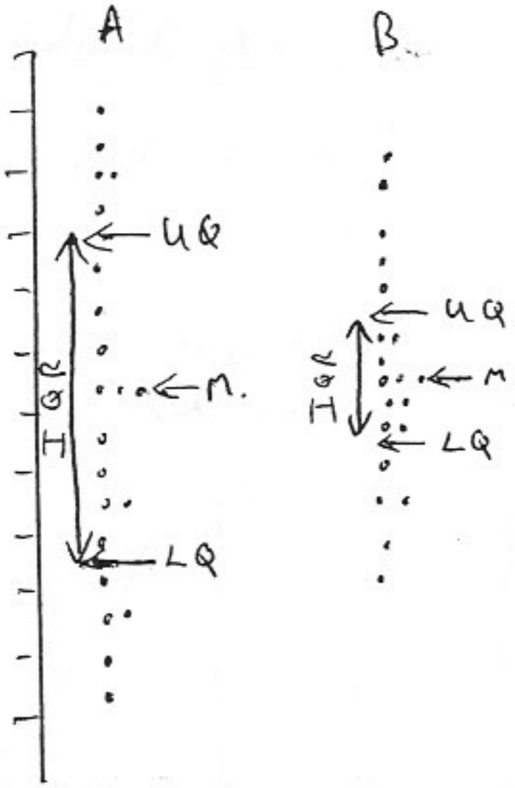
**Range** Max value − Min value.

Not a very good value since it emphasises extremes.

## Inter-Quartile Range

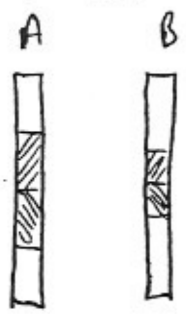
A measure of spread around the median.

- A dispersion diagram is drawn.
- Median is marked
- Upper Quartile marked
- Lower Quartile marked
- Inter Quartile range calculated.




Data Sets A and B have the same median, but A has a greater spread as shown by the larger IQR.

A visual representation of the IQR can be use



## Standard Deviation

The most powerful measure of spread. Measures spread around the mean

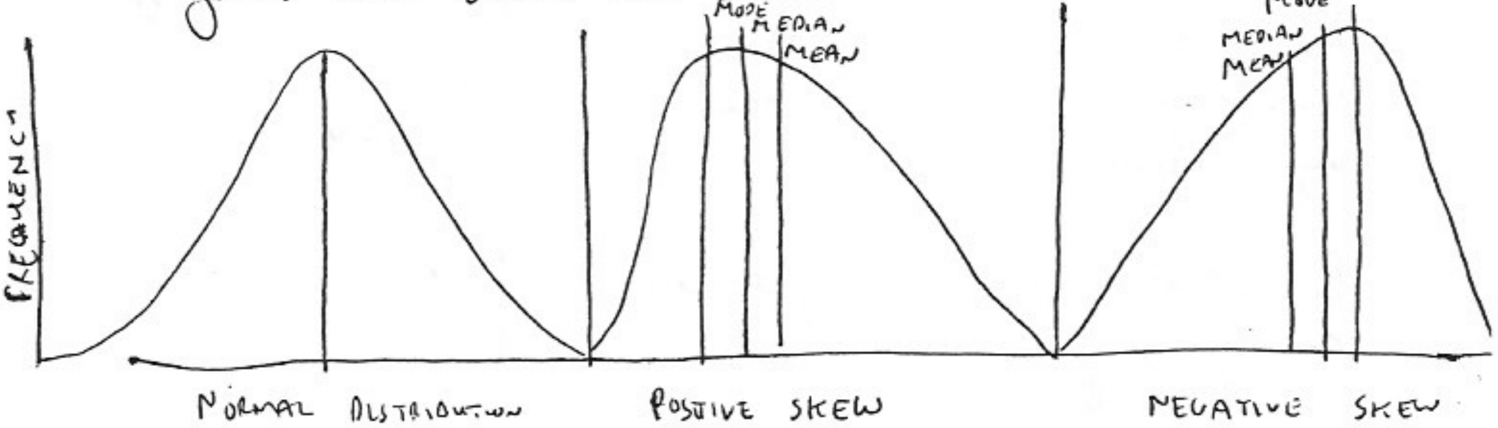
$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$$

Good measures for comparing data since they produce %'s are the Coefficient of variation and the index of variability.
In all cases a larger number or % means a greater degree of spread of data.

## Frequency Distributions

Frequency distributions or Cumulative frequency curves can also be used

The shape of the frequency distribution says a great deal about the data




NORMAL DISTRIBUTION — MEAN = MODE = MEDIAN

POSITIVE SKEW — MEAN > MEDIAN > MODE

NEGATIVE SKEW — MEAN < MEDIAN < MODE

# STATISTICAL TECHNIQUES

- Random number Table
- Mean, mode, median.
- Frequency distribution, Cumulative frequency curve
- Measures of dispersion. Inter quartile range
  Quartile Deviation,
  Coefficient of variation
  Index of variability
  Standard Deviation

- Sampling error
- size of the sample
- Testing for statistical difference between samples
- 't' test significance table
- critical values for Spearmans Rank correlation Coefficient

## D1 Random Numbers

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 80 30 | 23 64 | 67 96 | 21 33 | 36 90 | 03 91 | 69 33 | 90 13 | 34 43 | 02 19 |
| 61 29 | 89 61 | 32 08 | 12 62 | 26 08 | 42 00 | 31 73 | 31 30 | 30 61 | 34 11 |
| 23 33 | 61 01 | 02 21 | 11 81 | 51 32 | 36 10 | 23 74 | 50 31 | 90 11 | 73 52 |
| 94 21 | 32 92 | 93 50 | 72 67 | 23 20 | 74 59 | 30 30 | 43 66 | 75 32 | 27 97 |
| 87 61 | 92 69 | 01 60 | 28 79 | 74 76 | 86 06 | 39 29 | 73 85 | 03 27 | 50 57 |
| 37 56 | 19 18 | 03 42 | 86 03 | 85 74 | 44 81 | 86 45 | 71 16 | 13 52 | 35 56 |
| 64 86 | 66 31 | 55 04 | 88 40 | 10 30 | 84 38 | 06 13 | 58 83 | 62 04 | 63 52 |
| 22 69 | 22 69 | 58 45 | 49 23 | 09 81 | 98 84 | 05 04 | 75 99 | 27 70 | 72 79 |
| 23 22 | 14 22 | 64 90 | 10 26 | 74 23 | 53 91 | 27 73 | 78 19 | 92 43 | 68 10 |
| 42 38 | 59 64 | 72 96 | 46 57 | 89 67 | 22 81 | 94 56 | 69 84 | 18 31 | 06 39 |
| 17 18 | 01 34 | 10 98 | 37 48 | 93 86 | 88 59 | 69 53 | 78 86 | 37 26 | 85 48 |
| 39 45 | 69 53 | 94 89 | 58 97 | 29 33 | 29 19 | 50 94 | 80 57 | 31 99 | 38 91 |
| 43 18 | 11 42 | 56 19 | 48 44 | 45 02 | 84 29 | 01 78 | 65 77 | 76 84 | 88 85 |
| 59 44 | 06 45 | 68 55 | 16 65 | 66 13 | 38 00 | 95 76 | 50 67 | 67 65 | 18 83 |
| 01 50 | 34 32 | 38 00 | 37 57 | 47 82 | 66 59 | 19 50 | 87 14 | 35 59 | 79 47 |
| 79 14 | 60 35 | 47 95 | 90 71 | 31 03 | 85 37 | 38 70 | 34 16 | 64 55 | 66 49 |
| 01 56 | 63 68 | 80 26 | 14 97 | 23 88 | 59 22 | 82 39 | 70 83 | 48 34 | 46 48 |
| 25 76 | 18 71 | 29 25 | 15 51 | 92 96 | 01 01 | 28 18 | 03 35 | 11 10 | 27 84 |
| 23 52 | 10 83 | 45 06 | 49 85 | 35 45 | 84 08 | 81 13 | 52 57 | 21 23 | 67 02 |
| 91 64 | 08 64 | 25 74 | 16 10 | 97 31 | 10 27 | 24 48 | 89 06 | 42 81 | 29 10 |
| 80 86 | 07 27 | 26 70 | 08 65 | 85 20 | 31 23 | 28 99 | 39 63 | 32 03 | 71 91 |
| 31 71 | 37 60 | 95 60 | 94 95 | 54 45 | 27 97 | 03 67 | 30 54 | 86 04 | 12 41 |
| 05 83 | 50 36 | 09 04 | 39 15 | 66 55 | 80 36 | 39 71 | 24 10 | 62 22 | 21 53 |
| 98 70 | 02 90 | 30 63 | 62 59 | 26 04 | 97 20 | 00 91 | 28 80 | 40 23 | 09 91 |
| 82 79 | 35 45 | 64 53 | 93 24 | 86 55 | 48 72 | 18 57 | 05 79 | 20 09 | 31 46 |
| 37 52 | 49 55 | 40 65 | 27 61 | 08 59 | 91 23 | 26 18 | 95 04 | 98 20 | 99 52 |
| 48 16 | 69 65 | 69 02 | 08 83 | 08 83 | 68 37 | 00 96 | 13 59 | 12 16 | 17 93 |
| 50 43 | 06 59 | 56 53 | 30 61 | 50 21 | 29 06 | 49 60 | 90 38 | 31 43 | 19 25 |
| 89 31 | 62 79 | 45 73 | 71 72 | 77 11 | 28 80 | 72 35 | 75 77 | 24 72 | 98 43 |
| 63 29 | 90 61 | 86 39 | 07 38 | 38 85 | 77 06 | 10 23 | 30 84 | 07 95 | 30 76 |

# ANALYSIS OF SHINGLE SAMPLE - CHESIL BEACH (SITE 11)

Mean long axis = 67.3 mm.

Sample

Mode = 66mm

Median = 66mm.

87

84

82

80 80

76

72 72

70

69

68 68

67 67

66 66 66 66 --- Median

65

64 64

62

60 60

58 58

56 56

55

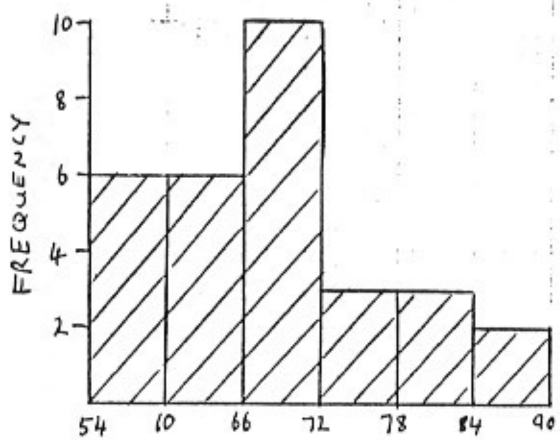
54

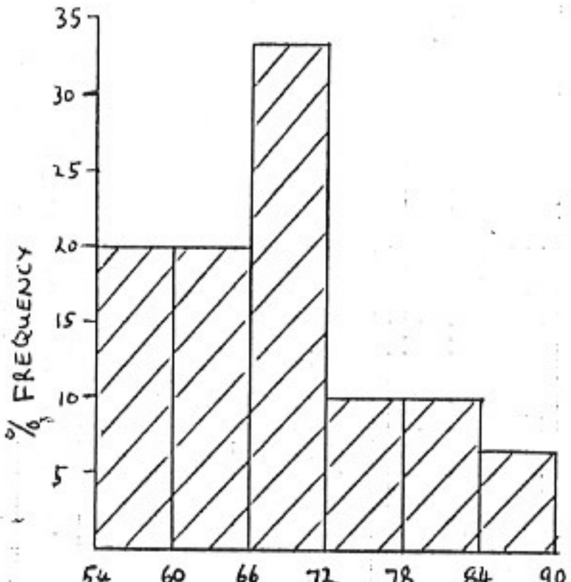
Grouped data:-

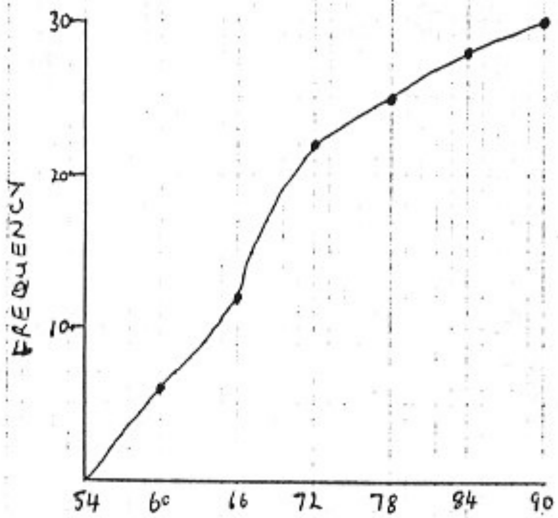
| CLASS LIMITS | FREQUENCY F | % F | CUMULATIVE F | CUMULATIVE % F |
|---|---|---|---|---|
| 54-59.9 | 6 | 20 | 6 | 20 |
| 60-65.9 | 6 | 20 | 12 | 40 |
| 66-71.9 | 10 | 33.33 | 22 | 73.33 |
| 72-77.9 | 3 | 10 | 25 | 83.33 |
| 78-83.9 | 3 | 10 | 28 | 93.33 |
| 84-89.9 | 2 | 666 | 30 | 100 |

SITE 11

FREQUENCY DISTRIBUTION




CUMULATIVE FREQUENCY CURVE

CUMULATIVE % FREQUENCY CURVE

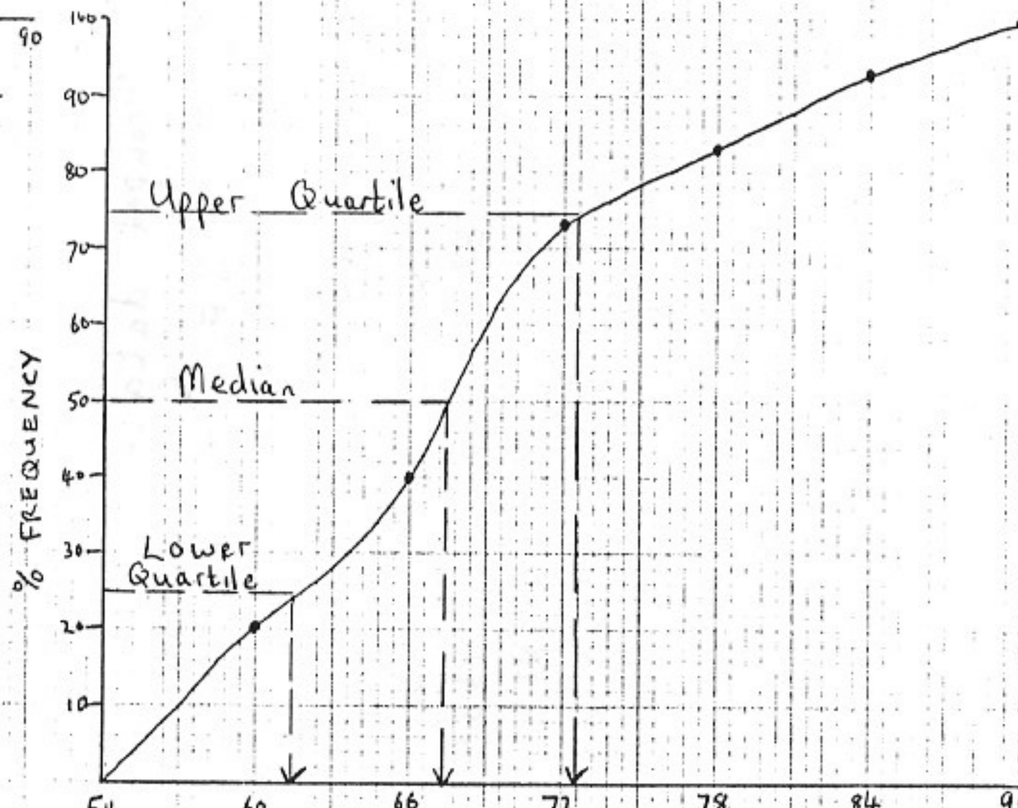
Approximate values:

Median = 67·5 mm

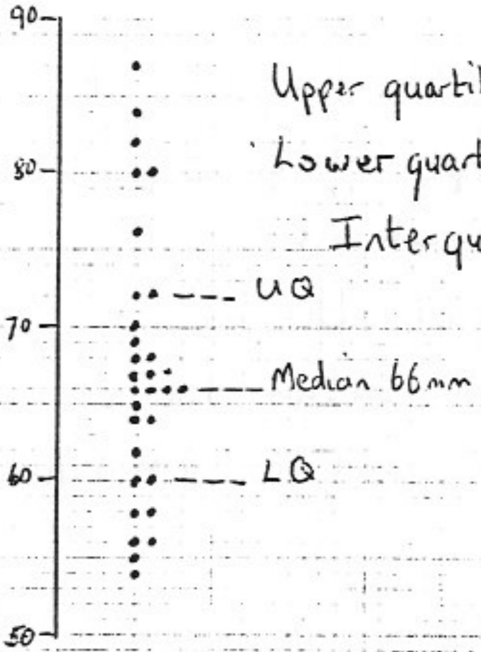
Upper quartile = 72·6 mm

Lower quartile = 61·5 mm

$\simeq$ I Q R = 11·1 mm



## SITE 11 Measures of Dispersion




Upper quartile = 72

Lower quartile = 60

Interquartile range = 12

Quartile deviation

$= \frac{12}{2} = \underline{\underline{6}}$

Coefficient of variation :-

$= \frac{6}{\text{mean}} \times 100$

$= \frac{8.68}{67.3} \times 100$

$= \underline{\underline{12.9\%}}$

Index of variability :-

$= \frac{\text{Quartile deviation}}{\text{median}} \times 100$

$= \frac{6}{66} \times 100$

$= \underline{\underline{9.1\%}}$

STANDARD DEVIATION $\sigma$

$\bar{x} = 67.3\text{mm}$

Long axis

| $x$ | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 70 | 2.7 | 7.29 |
| 84 | 16.7 | 278.89 |
| 60 | -7.3 | 53.29 |
| 67 | -0.3 | 0.09 |
| 87 | 19.7 | 388.09 |
| 67 | -0.3 | 0.09 |
| 58 | -9.3 | 86.49 |
| 66 | -1.3 | 1.69 |
| 66 | -1.3 | 1.69 |
| 72 | 4.7 | 22.09 |
| 68 | 0.7 | 0.49 |
| 56 | -11.3 | 127.69 |
| 80 | 12.7 | 161.29 |
| 82 | 14.7 | 216.09 |
| 58 | -9.3 | 86.49 |
| 66 | -1.3 | 1.69 |
| 72 | 4.7 | 22.09 |
| 66 | -1.3 | 1.69 |
| 68 | 0.7 | 0.49 |
| 62 | -5.3 | 28.09 |
| 56 | -11.3 | 127.69 |
| 55 | -12.3 | 151.29 |
| 65 | -2.3 | 5.29 |
| 64 | -3.3 | 10.89 |
| 54 | -13.3 | 176.89 |
| 64 | -3.3 | 10.89 |
| 60 | -7.3 | 53.29 |
| 69 | 1.7 | 2.89 |
| 80 | 12.7 | 161.29 |
| 76 | 8.7 | 75.69 |

$\Sigma(x-\bar{x})^2 = 2261.9$

$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{2261.9}{30}}$

$= \sqrt{75.396}$

$\underline{\underline{\sigma = 8.68}}$

## SAMPLING ERROR

Although we have a mean size for the long axis at each site, we have only taken a sample of the very large total population of pebbles at each site.
We must work out the probability that the actual mean of pebble sizes will lie between certain limits around the sample mean size. This is called the STANDARD ERROR

$$\text{Standard error of the mean (SE)} = \frac{\sigma}{\sqrt{n}}$$

we only have the sample standard deviation (s).

$$\therefore SE = \frac{s}{\sqrt{n}}$$

**SITE 1**

$SE = \frac{2.58}{5.48} = 0.47$

sample mean = 11.5mm

∴ At the 68% level of confidence the actual mean lies between 11.03 mm and 11.97 mm.

- At the 95% level of confidence the actual mean lies between. 10.56mm and 12.44mm
- At the 99% level of confidence the actual mean lies between 10.09 mm and 12.91 mm.

**SITE 11**

$SE = \frac{8.68}{5.48} = 1.58$

Sample mean = 67.3 mm

∴ 68% 65.72mm and 68.88mm

∴ 95% 64.14 and 70.46

∴ 99% 62.56 and 72.04

# SIZE OF THE SAMPLE

Can be used in 2 ways

1. After a preliminary pilot survey has been taken to determine what size the main sample needs to be for a given (chosen) sample error.
2. As a critical comment about your work in calculating what size of sample you <u>should</u> have taken to achieve an acceptable sample error.

$$SE \text{ of mean} = \frac{s}{\sqrt{n}}$$

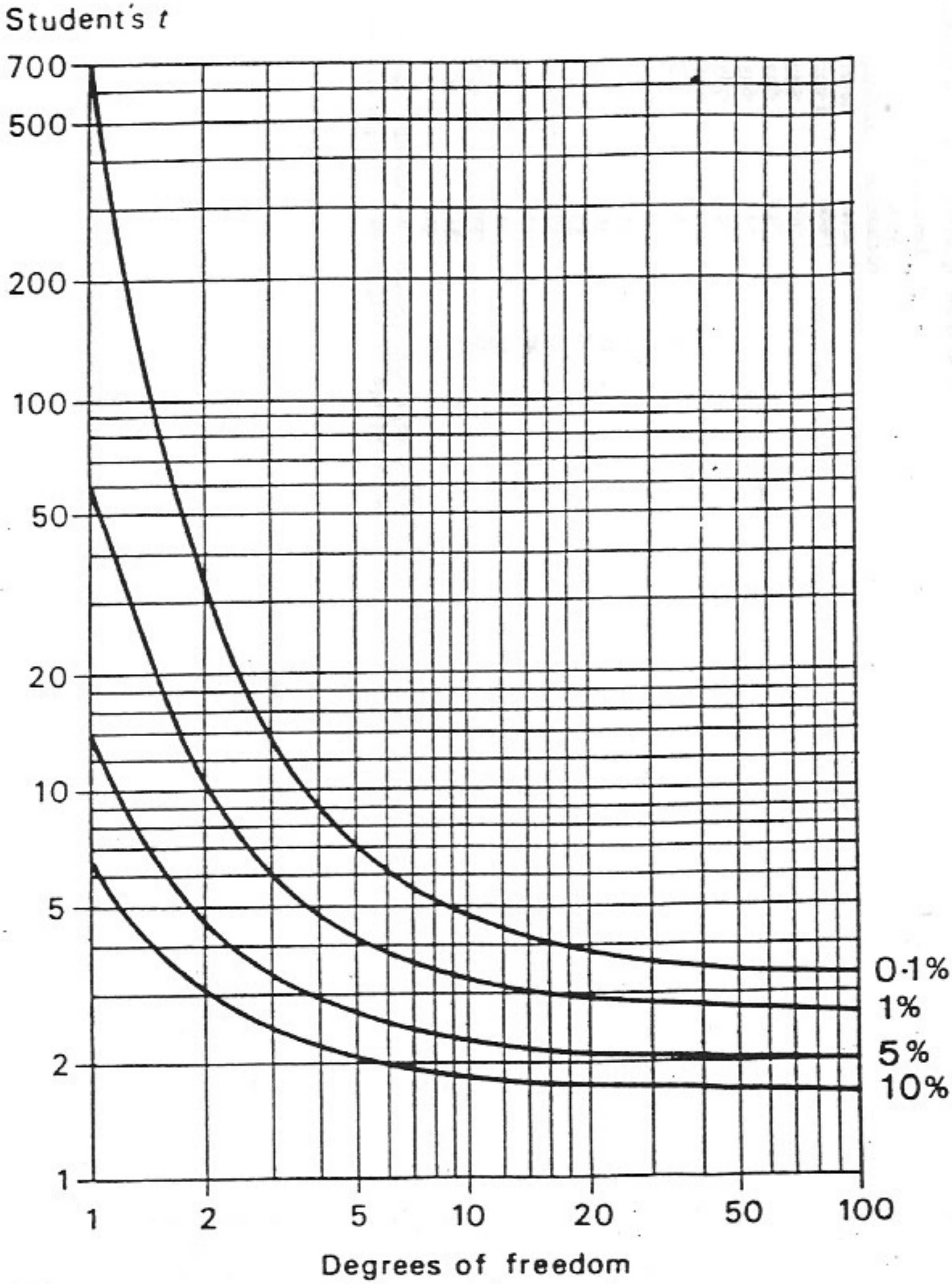
s = standard deviation

$$\therefore \quad n = \left(\frac{s}{SE}\right)^2$$

∴ using s, the standard deviation of the pilot sample and SE the chosen allowable standard error n the sample size needed can be found.

eg $s = 12$
SE allowable $= 2$

$$n = \left(\frac{12}{2}\right)^2 = 6^2 = \underline{36}$$


Student's t
700
500
200
100
50
20
10
5
2
1
1
2
5
10
20
50
100
Degrees of freedom
0·1%
1%
5%
10%

| N | 95% confidence level | 99% confidence level |
|---|---|---|
| 4 | 1·000 | 1·000 |
| 5 | ·900 | ·943 |
| 6 | ·829 | ·893 |
| 7 | ·714 | ·833 |
| 8 | ·643 | ·783 |
| 9 | ·600 | ·746 |
| 10 | ·564 | ·712 |
| 12 | ·506 | ·645 |
| 14 | ·456 | ·601 |
| 16 | ·425 | ·564 |
| 18 | ·399 | ·534 |
| 20 | ·377 | ·508 |
| 22 | ·359 | ·485 |
| 24 | ·343 | ·465 |
| 26 | ·329 | ·448 |
| 28 | ·317 | ·432 |
| 30 | ·306 | |

*Source:* Olds, E. G., in *Annals of Mathematical Statistics* (1938), vol. 9, pages 133–148, and (1949), vol. 20, pages 117–118.