L1022 Statistics for Economics and Finance

Lecture 1

Course Organization

- 2nd and 3rd year course
- Introduction to statistical principles and techniques
- Core text: Barrow
- Emphasis on practical applications
- Assessment: project

Introduction

- Key definitions
  - **Population**: the entire set of observations/census
  - **Sample**: a sub-group of the population
  - **Parameter**: the true value of a characteristic of the population
    - denoted by Greek characters e.g. μ and σ^2
  - **Statistic**: an estimate of the parameter calculated using the sample
    - denoted by normal characters e.g. $\overline{x}$ and s^2

Descriptive Statistics

- Descriptive statistics summarize a mass of information
- We may use graphical and/or numerical methods
- Examples of graphical methods are the bar chart, XY chart, pie chart and histogram (see seminar 1 and workshop 1 for practice)
- Examples of numerical methods are averages and standard deviations

Numerical Techniques

- We examine measures of
  - Location
  - Dispersion

Measures of Location

- Mean – strictly the arithmetic mean, the well known ‘average’
- Median – e.g. the income of the person in the middle of the distribution
- Mode – e.g. the level of income that occurs most often
- These different measures can give different answers…

The Mean of the Income Distribution

Person	1	2	3	4	5	6	7	8	9	10
income	15	15	20	25	45	55	70	85	125	250

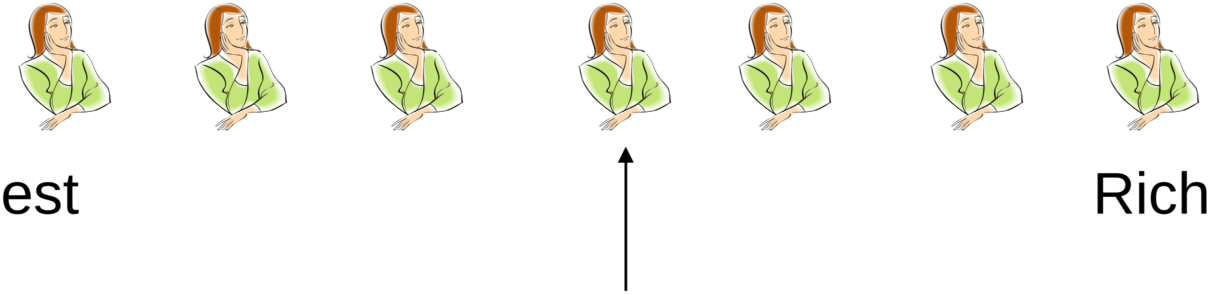
$$\boxed{\mu = \frac{\sum x}{n} = \frac{705}{10} = 70.5}$$

Mean income is therefore £70,500 per year

NB: use μ if data is the whole population , or $\bar{x}$ if the data is a sample (same formula).

The Median

- The income of the 'middle person' – i.e. the one located halfway through the distribution



Poorest

Richest

This person's income

- The median is little affected by outliers, unlike the mean

Calculating the Median

- We have 10 observations in the sample, so the person 5.5 in rank order has the median wealth. This person is somewhere between £45,000 and £55,000

Person	1	2	3	4	5	6	7	8	9	10
income	15	15	20	25	45	55	70	85	125	250

- Hence the median income is £50,000 per year
- Q: what happens to the median if the richest person's income is doubled to £500?
- Q: what happens to the mean?

The Mode

- The mode is the observation with the highest frequency
- For our data we have a single mode at £15,000
- It is possible to have a sample or population with no mode, or more than 1 mode
  - E.g. two modes: bimodal

Measures of Dispersion

- Range – the difference between smallest and largest observation. Not very informative for most purposes
- Variance – based on all observations in the population or sample

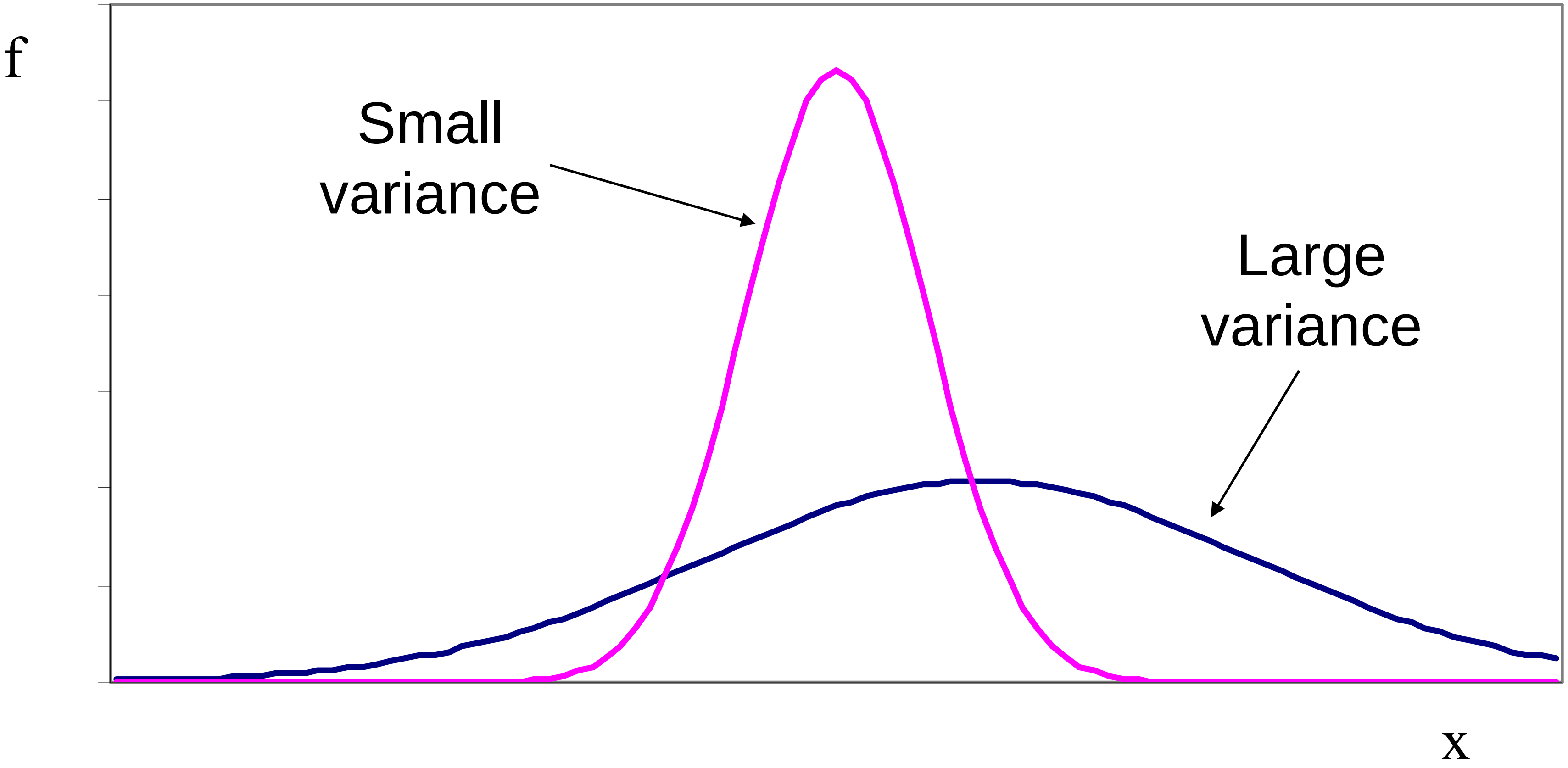
The Variance

- The variance is the average of all squared deviations from the mean:

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

- The larger this value, the greater the dispersion of the observations
- NB: use σ^2 for population variance; for sample variance use s^2 and divide by n-1 rather than by n

The Variance (cont.)



Calculating the Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

i	1	2	3	4	5	6	7	8	9	10
x	15	15	20	25	45	55	70	85	125	250
x^2	225	225	400	625	2025	3025	4900	7225	15625	62500

$$\sum_{i=1}^{10} x_i^2 = 225 + 225 + ...62500 = 96775; n\bar{x}^2 = 10 \times 70.5^2 = 49705$$

$$s^2 = \frac{96775 - 49705}{9} = 5230$$

NB: Variance is in £^{2}, so we use the square root, known as the standard deviation, *s*. *s*=72.318, i.e. £72,318.

Standard Deviation

- Useful to help us estimate

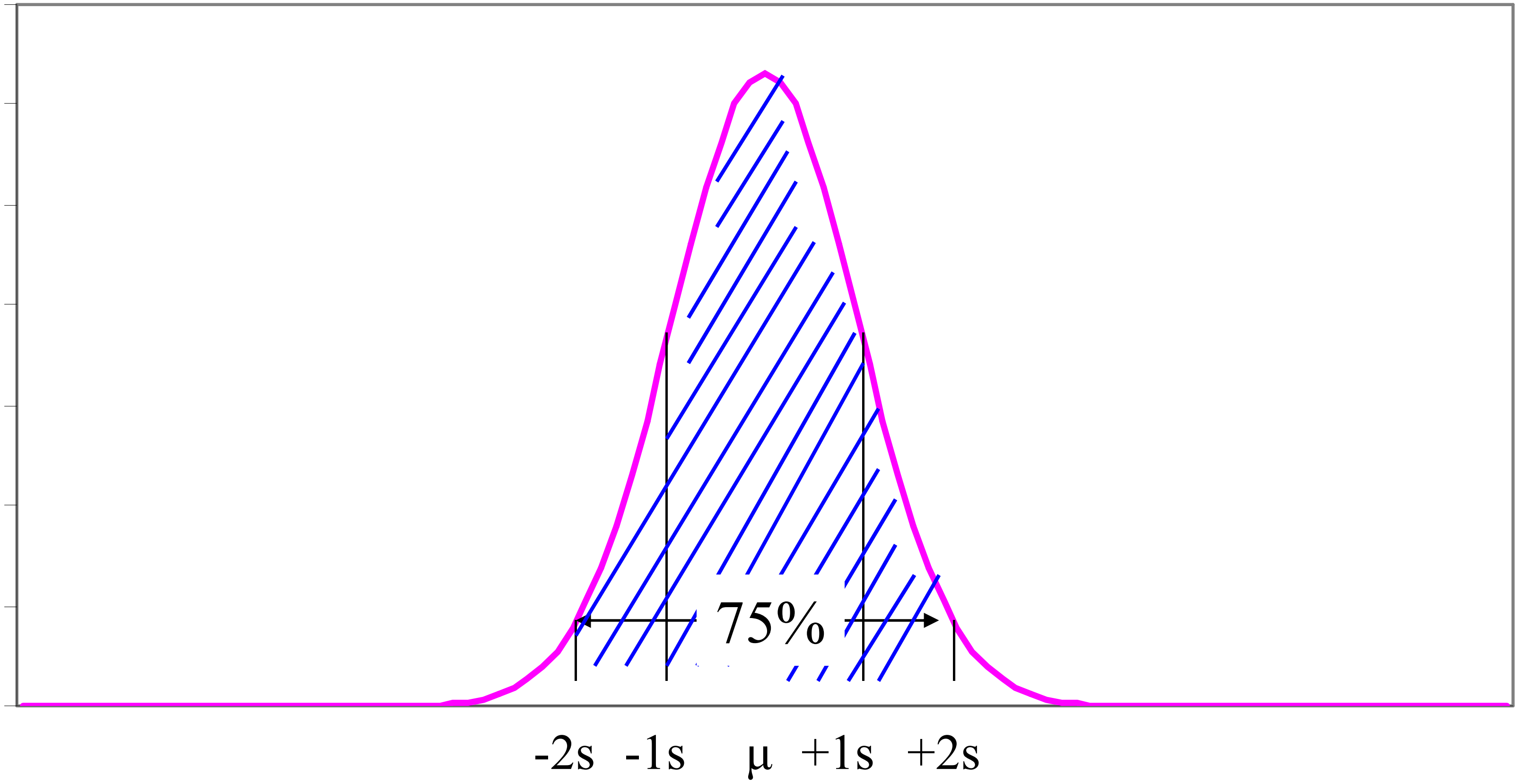
  a) The % of obs. that lie within a given number of standard deviations above or below the mean (2 rules)

  b) Where a particular observation lies relative to the mean

Chebyshev's Rule

- 100(1-1/**k**^{2})% of observations lie within **k** standard deviations above and below the mean.

e.g. 100*[1-1/(2^2)]%=75% of obs. lie within 2 s.d.s either side of the mean



Empirical Rule

- *If* the underlying distribution is Normal (more next week), then
  - 68% of observations lie within ± 1 st. devs
  - 95% of observations lie within ± 2 st. devs
  - 99% of observations lie within ± 3 st. devs

z-scores

- z-scores tell us how many standard deviations an observation lies above or below the mean

$$z = \frac{x - \mu}{\sigma}$$

  - z>0 means that the observation lies above the mean
  - z<0 means that the observation lies below the mean

- e.g. $\mu = 55$ and $\sigma = 10$. What is z-score of 65?

$$z = \frac{65 - 55}{10} = 1$$

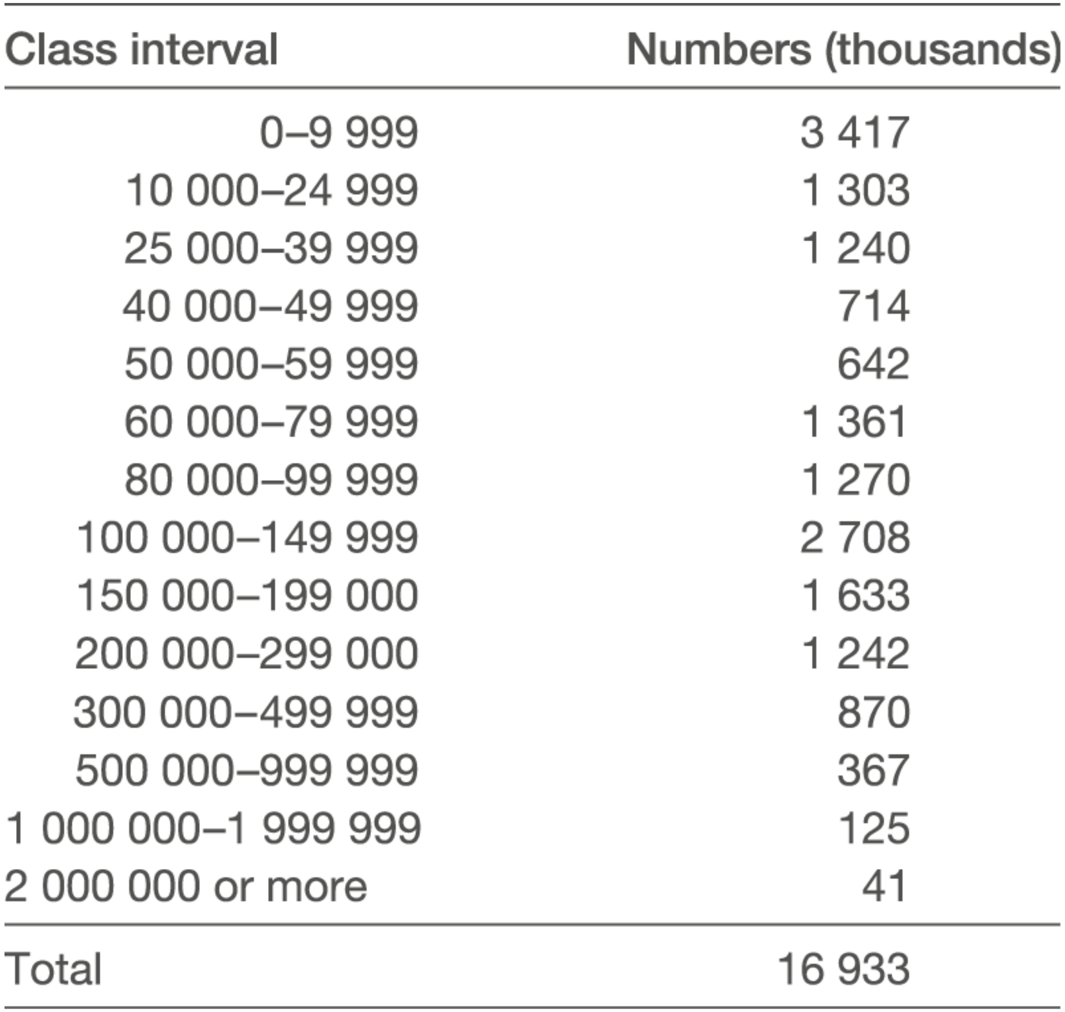
- Thus, 65 is exactly 1 st.dev. above the mean

Summary

- We can use graphical and numerical measures to summarise data
- The aim is to simplify without distorting the message
- Summary measures of location [mean, median, mode] and dispersion [variance, standard deviation, z-scores] provide a good description of the data

Appendix: calculating summary statistics when the data is grouped

Data on Wealth in the UK



Class interval	Numbers (thousands)
0–9 999	3 417
10 000–24 999	1 303
25 000–39 999	1 240
40 000–49 999	714
50 000–59 999	642
60 000–79 999	1 361
80 000–99 999	1 270
100 000–149 999	2 708
150 000–199 000	1 633
200 000–299 000	1 242
300 000–499 999	870
500 000–999 999	367
1 000 000–1 999 999	125
2 000 000 or more	41
Total	16 933

Table 1.3 The distribution of wealth, UK, 2001

The mean of the Wealth Distribution

Range	mid-point, x	f	fx
0–	5.0	3,417	17,085.0
10,000–	17.5	1,303	22,802.5
25,000–	32.5	1,240	40,300.0
40,000–	45.0	714	32,130.0
50,000–	55.0	642	35,310.0
60,000–	70.0	1,361	95,270.0
80,000–	90.0	1,270	114,300.0
100,000–	125.0	2,708	338,500.0
150,000–	175.0	1,633	285,775.0
200,000–	250.0	1,242	310,500.0
300,000–	400.0	870	348,000.0
500,000–	750.0	367	275,250.0
1,000,000–	1500.0	125	187,500.0
2,000,000–	3000.0	41	123,000.0
Total		16,933	2,225,722.5

$$\mu = \frac{\sum fx}{\sum f} = \frac{2{,}225{,}722.5}{16{,}933} = 133.443$$

Calculating the Median

- 16,933 observations, hence person 8,466.5 in rank order has the median wealth
- This person is somewhere in the £60-80k interval

Range	Frequency	Cumulative frequency
0–	3,417	3,417
10,000–	1,303	4,720
25,000–	1,240	5,960
40,000–	714	6,674
50,000–	642	7,316
60,000–	1,361	8,677
80,000–	1,270	9,947
:	:	:

Number with wealth less than £60k

Number with wealth less than £80k

Calculating the Median (cont.)

- To find the precise median, use

$$x_L + (x_U - x_L)\left\{\frac{\frac{N}{2} - F}{f}\right\}$$

$$= 60 + (80 - 60)\left\{\frac{\frac{16{,}933}{2} - 7{,}316}{1{,}361}\right\} = 76.907$$

- Median wealth is £76,907

The Mode (cont.)

- For grouped data, the mode corresponds to the interval with greatest frequency density

Range	Frequency	Class width	Frequency density
0–	3,417	10,000	0.3417
10,000–	1,303	15,000	0.0869
25,000–	1,240	15,000	0.0827
40,000–	714	10,000	0.0714
50,000–	642	10,000	0.0642

Modal class

Mode = £0–10,000

The Variance

- The variance is the average of all squared deviations from the mean:

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{\sum f}$$

- The larger this value, the greater the dispersion of the observations

Calculation of the Variance

Range	Mid-point x (£000)	Frequency, f	Deviation (x - μ)	(x - μ)^2	f(x - μ)^2
0–	5.0	3,417	- 126.4	15,987.81	54,630,329.97
10,000–	17.5	1,303	- 113.9	12,982.98	16,916,826.55
25,000–	32.5	1,240	- 98.9	9,789.70	12,139,223.03
40,000–	45.0	714	- 86.4	7,472.37	5,335,274.81
50,000–	55.0	642	- 76.4	5,843.52	3,751,537.16
60,000–	70.0	1,361	- 61.4	3,775.23	5,138,086.73
80,000–	90.0	1,270	- 41.4	1,717.51	2,181,241.95
100,000–	125.0	2,708	- 6.4	41.51	112,411.42
150,000–	175.0	1,633	43.6	1,897.22	3,098,162.88
200,000–	250.0	1,242	118.6	14,055.79	17,457,288.35
300,000–	400.0	870	268.6	72,122.92	62,746,940.35
500,000–	750.0	367	618.6	382,612.90	140,418,932.52
1,000,000–	1500.0	125	1,368.6	1,872,948.56	234,118,569.53
2,000,000–	3000.0	41	2,868.6	8,228,619.88	337,373,415.02
Total		16,933			895,418,240.28

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{\sum f} = \frac{895{,}418{,}240.28}{16{,}933} = 52{,}880.07$$

The Standard Deviation

- The variance is measured in ‘squared £s’ (because we used squared deviations)
- Hence take the square root to get back to £s This gives the standard deviation

$$\sigma = \sqrt{52{,}880.07} = 229.957$$

or £229,957

Sample Measures

- For sample data, use

$$s^2 = \frac{\sum f\left(x - \bar{x}\right)^2}{n-1}$$

to calculate the sample variance

- This gives an unbiased estimate of the population variance
- Take the square root of this for the sample standard deviation