Probability and statistics

 info@amttraining.com

DATA

How would you organise the data?

99	88
100	88
82	89
95	90
78	92
86	92
88	94
90	82
82	86
84	89

SORTING THE LIST

- We can describe the data set better if we know the difference between the highest and lowest point
- We call this the range

RANGE

The range is between 78 and 100

78	89
82	89
82	90
82	90
84	92
86	92
86	94
88	95
88	99
88	100

FREQUENCY

- We can summarise the data by calculating the number records within a number range:

Range	Frequency
80	1
85	4
90	9
95	4
100	2

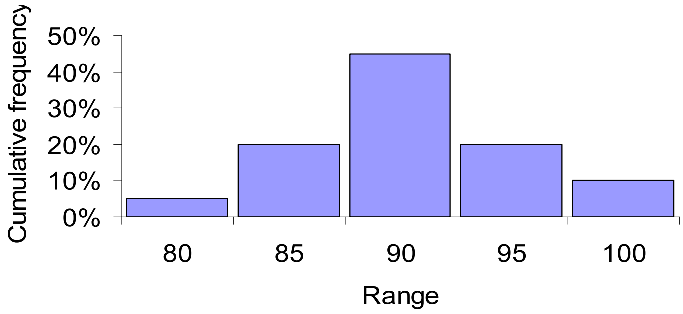
RELATIVE FREQUENCY

- Relative frequency is the frequency as a percentage of the total observations

Range	**Frequency**	**Relative frequency**
80	1	5%
85	4	20%
90	9	45%
95	4	20%
100	2	10%
Total	20	

HISTOGRAM

A histogram is a graph of the relative frequency



 7 AdkinsMatchett&Toy A Wilmington Company

SUMMATION NOTATION

- The sum of the measurements whose typical member is χ_1, beginning with the number χ_1 and ending with the number χ_n

$$\sum_{i=1}^{6} \chi_i$$

1 + 2 + 3 + 4 + 5 + 6 = 21

$\chi_1 + \chi_2 + \chi_3 + \chi_4 + \chi_5 + \chi_6 = 21$

 8 AdkinsMatchett&Toy A Wilmington Company

CENTRAL TENDENCY

- Central tendency measures the tendency of the data set to cluster or centre around a numerical value
- Popular examples include
  - Mean
  - Median

MEAN

The sum of the measurements divided by the number of the measurements

Find the mean of the following numbers:

3, 5, 2, 5, 6, 3

$$\frac{\sum_{i=1}^{6} \chi_i}{6} = mean$$

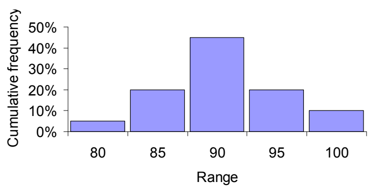
MEDIAN

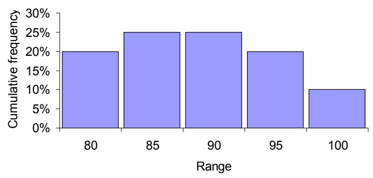
- The median of a data set is the middle number when the measurements are arranged in an ascending or descending order
- Use when looking at large data sets
- When you have an even number average to two central numbers

Find the median of the following :
3, 5, 2, 5, 6, 3

STANDARD DEVIATION

- Measuring central tendency is incomplete without looking at the *spread* of the data set
- Look at the following two histograms





- Which gives you the most reliable prediction?

STANDARD DEVIATION

- The following two data sets have the same mean

  2, 3, 4, 5, 6 mean = 4

  3, 4, 4, 4, 5 mean = 4

- Calculate each number's deviation from the mean

STANDARD DEVIATION

Deviation from the mean

Data set 1	2	3	4	5	6
Mean	4	4	4	4	4
Deviation	-2	-1	0	1	2

Data set 2	3	4	4	4	5
Mean	4	4	4	4	4
Deviation	-1	0	0	0	1

The sum of the deviations is equal

A SINGLE MEASURE OF DEVIATION

- We can't take the mean of the differences – that would add up to zero
- Instead we can take differences and square them

Data set 1	2	3	4	5	6
Mean	4	4	4	4	4
Deviation	-2	-1	0	1	2
Squared deviation	4	1	0	1	4
Data set 2	3	4	4	4	5
Mean	4	4	4	4	4
Deviation	-1	0	0	0	1
Squared deviation	1	0	0	0	1

SAMPLE VARIANCE

- Now calculate the sample variance by dividing by the number of data points minus 1 (a statistical adjustment for samples)

Data set 1

(4 + 1 + 0 + 1 + 4) ÷ (5-1) = 2.5

Data set 2

(1 + 0 + 0 + 0 + 1) ÷ (5-1) = 0.5

AT LAST… THE STANDARD DEVIATION

- Take the positive square root of the sample deviance:

  Data set 1

  $\sqrt{2.5} = 1.58$

  Data set 2

  $\sqrt{0.5} = 0.71$

STANDARD DEVIATION SHORTCUT

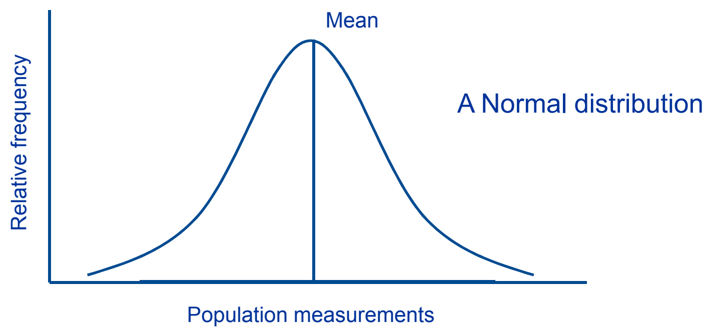
Here is a full formula for the standard deviation

Mean $$\bar{x} = \frac{\sum x_i}{n}$$

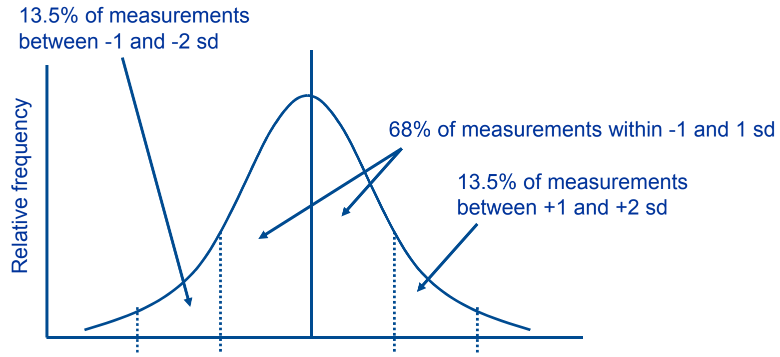
Standard deviation $$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

NORMAL DISTRIBUTION

- In many situations in life (the stock market, population height etc) the relative frequency distribution is mound shaped and symmetric:



NORMAL DISTRIBUTION



CALCULATING

1. Now we can calculate the range of measurements within certain standard deviations of the mean (assuming a normal distribution)
2. If the mean height of the male population is 5.9 ft and the standard deviation is 0.5 feet
3. Then 68% of the population would be between 5.9 – 0.5 = 5.4 and 5.9 + 0.5 = 6.4 in height

Z SCORES

- Another measure of relative standing are z scores which use the mean and standard deviation

$$\text{Z score} = \frac{\text{Measurement - mean}}{\text{Standard deviation}}$$

- The z score measures the number of standard deviations from the mean

Z SCORE EXAMPLE

- 1,000 financial analysts have a mean income of 60,000.
- The standard deviation is 10,000
- If Jane Flight's income is 65,000 how does she compare with the mean?

  (65,000 – 60,000) ÷ 10,000 = 0.5

  Jane is 0.5 standard deviations above the mean

IF THE DISTRIBUTION IS NORMAL

- Approximately 68% of financial analysts will have a z score between -1 and 1
- Approximately 95% of financial analysts will have a z score between -2 and 2
- Approximately 99.7% of financial analysts will have a z score between -3 and 3

APPLICATION

- Assume the mean stock market return is 12% and its relative frequency distribution is normal
- What is the probability of a gain of over 32% in one year? Assume the market's standard deviation is 20%

APPLICATION

1. Calculate the z score:

$$\frac{32\% - 12\%}{20\%} = 1$$

2. Therefore a 32% gain is 1 standard deviation from the mean
3. Assuming a normal distribution there is a 16% chance of the stock market giving you a gain of 32% or higher in one year


AdkinsMatchett&Toy
A Wilmington Company
Please visit us on:
www.amttraining.com
www.crunchthenumbers.com
Expert financial training
Please note:
All materials are the intellectual property of AMT Training.
Offices: London, New York, Hong Kong