MA208 Quantitative Techniques for Business

Lecture 4: Statistics ctd.

Dr Kirsten Pfeiffer

School of Mathematics, Applied Mathematics and Statistics
NUI Galway

Using Standard Deviation

Example

The Wechsler Adult Intelligence Scale involves an IQ test designed so that the mean score is 100 and the standard deviation is 15. Use the range rule of thumb to find the minimum and maximum "usual" IQ scores. Then determine whether an IQ score of 135 would be considered "unusual".

Minimum "usual" value = mean − 2 × standard deviation

$= 100 - 2(15) = 70$

Maximum "usual" value = $\mu - 2 \times \sigma$

$= 100 + 2(15) = 130$

So we expect that typical IQ scores fall between 70 and 130.
As 135 does not fall within those limits, it would be considered as unusual IQ.

Using Standard Deviation

Note:

- The standard deviation measures the *variation* among data values.
- Values close together have a small standard deviation, but values with much more variation have a larger standard deviation.
- For many data sets, a value is *unusual* if it differs from the mean by more than two standard deviations.

Chebyshev's Inequality

Another concept that is helpful in understanding or interpreting a value of a standard deviation is Chebyshev's Inequality, which applies to *any* data set.

Chebyshev's Inequality

At least $1 - \frac{1}{k^2}$ of the distribution's values are within k standard deviations of the mean, where k is any positive number greater than 1.

For $k = 2$ and $k = 3$ we get:

- At least $\frac{3}{4}$ (or 75 %) of all values lie within 2 standard deviations of the mean.
- At least $\frac{8}{9}$ (or 89 %) of all values lie within 3 standard deviations of the mean.

Chebyshev's Inequality

Example

IQ scores have a mean of 100 and a standard deviation of 15. What can we conclude from Chebyshev's Inequality?

Solution

- At least $\frac{3}{4}$ (or 75%) of IQ scores are within 2 standard deviations of the mean, that is at least 75% are between 70 and 130.
- At least $\frac{8}{9}$ (or 89%) of IQ scores are within 3 standard deviations of the mean, i.e. lie between 55 and 145.

Coefficient of Variation

When comparing variation in two different sets of data, the standard deviations should be compared only if the two sets of data use the same scale and units and they have approximately the same mean. If the means are substantially different, or if the sample uses different scales or measurement units, we can use the coefficient of variation (cv). This measure describes the standard deviation relative to the mean and is usually expressed as percentage:

Coefficient of Variation

$$cv = \frac{\sigma}{\overline{x}} \times 100\%$$

When two distributions are compared, the distribution with the largest coefficient of variation has the greatest dispersion.
Note: The coefficient of variation is often used as a measure for economic inequality.

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.

$$cv = \frac{\sigma}{\mu} \cdot 100\,\% = \frac{0}{100} \cdot 100\,\% = 0\,\%$$

2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.

$$cv = \frac{8.16}{100} \cdot 100\,\% = 8.16\,\%$$

3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.

$$cv = \frac{30.78}{27.875} \cdot 100\,\% = 110.4\,\%$$

Coefficient of Variation

Exercise

Given the following data set:

$$\{20, 20, 21, 24, 25, 28\}$$

Find the standard deviation and the coefficient of variation.

Solution

$$\mu = \frac{20+20+21+24+25+28}{6} = 23$$

σ:

x	$x-\mu$	$(x-\mu)^2$
20	-3	9
20	-3	9
21	-2	4
24	1	1
25	2	4
28	5	25
		52

$$\sigma = \sqrt{\frac{52}{6-1}} = 3.22$$

$$cv = \frac{3.22}{23} \cdot 100\% = 14\%$$

Exercise

Question from 2008 exam paper

The scores of twenty-one exam students are

72	93	97	52	57	69	69
84	49	72	68	68	98	51
65	84	79	55	64	71	78

(i) Present the data using a stem-and-leaf display.

(ii) Present the data using a histogram with classes $0-9$, $10-19$, ..., $90-99$.

(iii) What is the median exam score? What is the upper quartile score and the lower quartile score?

(iv) Consider the sample consisting of the first five exam scores. What is the sample mean, what is the standard deviation and what is the coefficient of variation?

Exercise

Question 1(b) from 2018 exam paper

Computers from a particular company are found to last on average for three years without any hardware malfunction, with a standard deviation of two months. At least what percent of the computers last between 31 months and 41 months?

See tutorial!