



Univariate descriptives

Guest lecturer: Miquel Pellicer

Johan A. Elkink
School of Politics & International Relations
University College Dublin

17 September 2018

Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References



1 Graphical descriptives

2 Frequency tables

3 Central tendency

4 Variation



**Graphical
descriptives**

Frequency
tables

Central
tendency

Variation

References

1 Graphical descriptives

2 Frequency tables

3 Central tendency

4 Variation



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

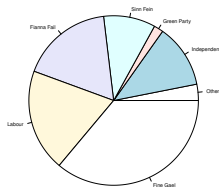
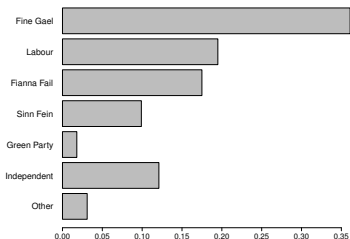
Graphical presentations of single variables.

Categorical	pie-charts barplots
Scale	histogram density plot boxplot

Categorical variables

For categorical variables, it is often useful to look at the number of cases or the proportion of cases in a particular category.

Barplots and **pie charts** are useful for this.



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

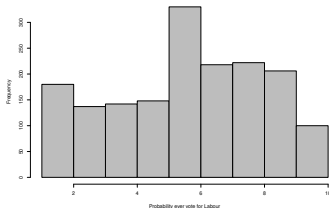


Histogram

For continuous (or scale) variables, we often want to get an idea of the **distribution** of values. How many low, medium, high values?

Histograms are useful to get an impression.

- bin the data using equal-distance cut-off points
- then produce a barplot of the number in each bin.





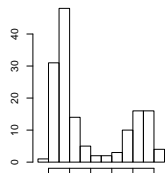
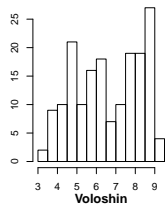
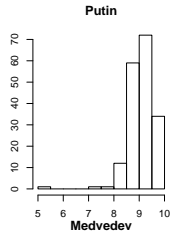
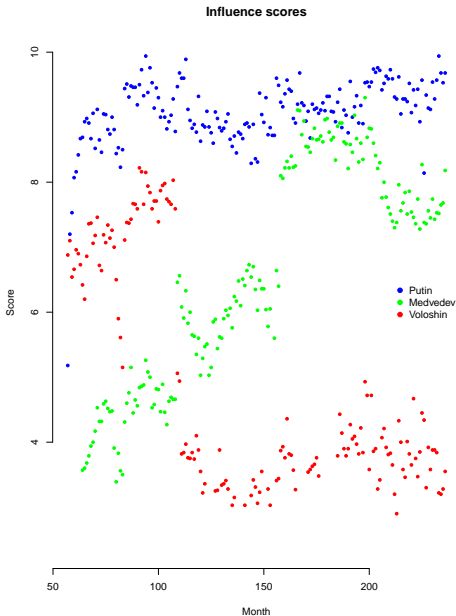
Graphical
descriptives

Frequency
tables

Central
tendency

Variation

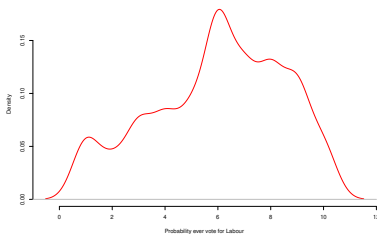
References



Density plot



A density plot is a smoothed version of a histogram, based on a non-parametric estimation of the shape of the distribution.



(Irish National Election Study 2011)

Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

For graphs of distributions (histogram, density plot, boxplot, etc.) you want to get an impression of:

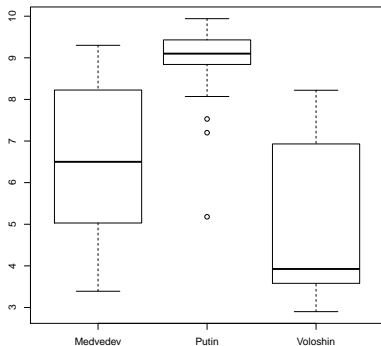
- the **shape** of the distribution;
- the **center** and **spread** of the distribution;
- the presence of **outliers**.

(Moore, 2003, 12)

Boxplot

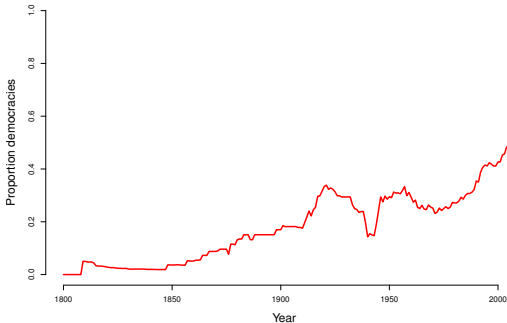
Another way of looking at the distribution of a continuous variable is to find out where the lowest 25% are located, where the lowest 50% are located, and where the top 25% are located.

A plot that shows this is the **boxplot**.



Time plot

When data is measured over time, another useful plot is a time plot, to see trends over time.



(Polity IV)

Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

1 Graphical descriptives

2 Frequency tables

3 Central tendency

4 Variation

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Democracy is preferable to any other kind of government	1924	80.2%
In some circumstances, a non-democratic government can be preferable	182	7.6%
For someone like me, it doesn't matter what kind of government we have	124	5.2%
Missing / don't know	168	7.0%
Total	2398	100.0%

Example



Graphical
descriptives

**Frequency
tables**

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Example frequency table



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

value	count	percentage
0	2	17
2	1	8
3	2	17
5	1	8
6	2	17
7	3	25
9	1	8
total	12	100%

Example summary



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

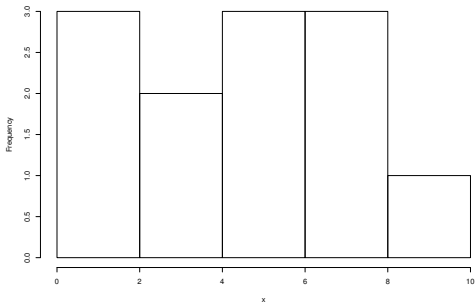
Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References



Outline



Graphical
descriptives

Frequency
tables

**Central
tendency**

Variation

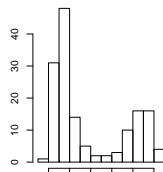
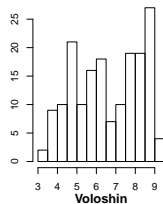
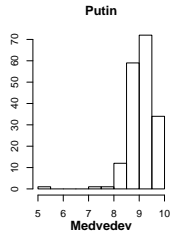
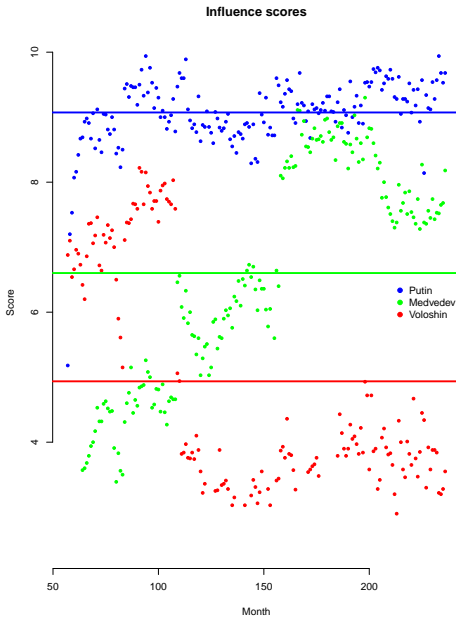
References

1 Graphical descriptives

2 Frequency tables

3 Central tendency

4 Variation





Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

The **mode** is the category with the highest frequency.

Example



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Example mode



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

The mode is the value that most often occurs, i.e. 7.

Median



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

The **median** is the value where 50% of the cases has a lower value on this variable and 50% a higher value.



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

The **median** is the value where 50% of the cases has a lower value on this variable and 50% a higher value.

If N is uneven: the middle value after sorting in ascending order.

If N is even: the average of the two middle values after sorting in ascending order.

Example



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Example median



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

First, sort the data:

0	0	2	3	3	5	6	6	7	7	7	9
---	---	---	---	---	---	---	---	---	---	---	---

Example median



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

First, sort the data:

0	0	2	3	3	5	6	6	7	7	7	9
---	---	---	---	---	---	---	---	---	---	---	---

Then, find the centre:

0	0	2	3	3	5	6	6	7	7	7	9
---	---	---	---	---	---	---	---	---	---	---	---

Example median



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

First, sort the data:

0	0	2	3	3	5	6	6	7	7	7	9
---	---	---	---	---	---	---	---	---	---	---	---

Then, find the centre:

0	0	2	3	3	5	6	6	7	7	7	9
---	---	---	---	---	---	---	---	---	---	---	---

The median is $\frac{5+6}{2} = 5.5$.

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

The **mean** is the sum of all values, divided by the number of values.

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

Example



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Example mean

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

$$\bar{x} = \frac{6+2+3+0+7+9+6+7+5+3+7+0}{12} = \frac{55}{12} = 4.6$$

Measures of central tendency



Graphical
descriptives

Frequency
tables

**Central
tendency**

Variation

References

Measures of central tendency provide information about the centre of a distribution, roughly put: “what is a typical value for this variable?”

Measures of central tendency

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Measures of central tendency provide information about the centre of a distribution, roughly put: “what is a typical value for this variable?”

Different measures are available for different levels of measurement:

	mode	median	mean
nominal	x		
ordinal	x	x	
scale	(x)	x	x

Mean or median?

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Outliers is a general term of values that are very far from the main values of the distribution.

- For a symmetric distribution, median and mean are the same.
- The more **skewed** the distribution, the more mean and median differ.
- Mean is sensitive to outliers, while the median is not.
- Mean has better understood mathematical properties.

Exercise



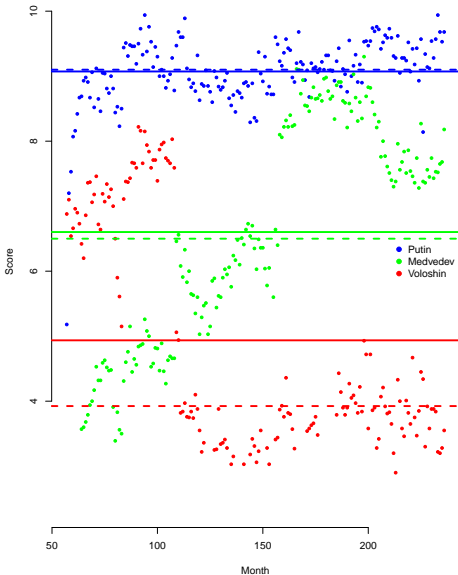
Calculate appropriate measures of central tendency for each of the following variables related to party membership:

gender	class	years in party	gender	class	years in party
M	High	32	M	Medium	17
M	Low	32	M	Low	50
M	Low	25	M	Medium	25
F	High	12	F	High	10
F	Medium	21	F	Medium	33
M	Low	37	F	Low	15
F	Low	31			

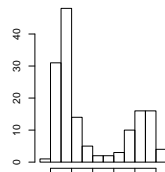
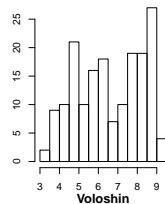
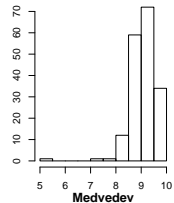
(Healey, 1996, 81)



Influence scores



Putin



Outline



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

1 Graphical descriptives

2 Frequency tables

3 Central tendency

4 **Variation**

Measures of dispersion

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Measures of dispersion provide an indication of the amount of variation or heterogeneity in a variable.

The **range** is the highest value minus the lowest value.

The **interquartile range** (IQR) is the range between the lowest 25% and the top 25%. A boxplot typically provides the median and the IQR, with some indication of outliers.

Example



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Example range



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

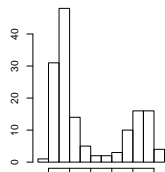
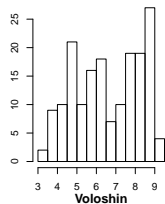
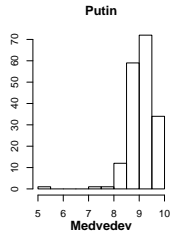
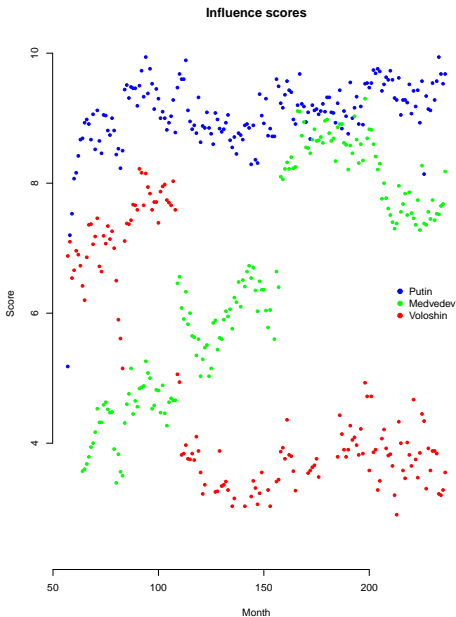
$$\text{Range: } 9 - 0 = 9.$$



$$\text{Var}(x) = s_x^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

Note that many software packages do not calculate this sample variance, but the unbiased estimator of the population variance:

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}$$



Example mean

Graphical
descriptivesFrequency
tablesCentral
tendency**Variation**

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

$$\bar{x} = \frac{6+2+3+0+7+9+6+7+5+3+7+0}{12} = \frac{55}{12} = 4.6$$

Example variance



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

$$\bar{x} = \frac{6+2+3+0+7+9+6+7+5+3+7+0}{12} = \frac{55}{12} = 4.6$$

$$s_x^2 = \frac{1}{N} \sum_{i=1}^N (x - \bar{x})^2 = \frac{1}{12} \sum_{i=1}^N (x - 4.6)^2$$

Graphical
descriptivesFrequency
tablesCentral
tendency**Variation**

References

Example variance



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

$$\begin{aligned}s_x^2 &= \frac{1}{N} \sum_{i=1}^N (x - \bar{x})^2 = \frac{1}{12} \sum_{i=1}^N (x - 4.6)^2 \\ &= \frac{1}{12} ((1.4)^2 + (-2.6)^2 + (-1.6)^2 + (-4.6)^2 + (2.4)^2 + (4.4)^2 \\ &\quad + (1.4)^2 + (2.4)^2 + (0.4)^2 + (-1.6)^2 + (2.4)^2 + (-4.6)^2)\end{aligned}$$

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Example variance



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

$$\begin{aligned}
 s_x^2 &= \frac{1}{N} \sum_{i=1}^N (x - \bar{x})^2 = \frac{1}{12} \sum_{i=1}^N (x - 4.6)^2 \\
 &= \frac{1}{12} ((1.4)^2 + (-2.6)^2 + (-1.6)^2 + (-4.6)^2 + (2.4)^2 + (4.4)^2 \\
 &\quad + (1.4)^2 + (2.4)^2 + (0.4)^2 + (-1.6)^2 + (2.4)^2 + (-4.6)^2) \\
 &= \frac{1}{12} (2.0 + 6.7 + 2.5 + 21.0 + 5.8 + 19.5 + 2.0 + 5.8 + 0.2 + 2.5 \\
 &\quad + 5.8 + 21.0) = \frac{94.9}{12} = 7.9
 \end{aligned}$$

Variance: sample data

Graphical
descriptivesFrequency
tablesCentral
tendency**Variation**

References

Country	Asylum seekers	$(x - \bar{x})$	$(x - \bar{x})^2$
Denmark	2.3		
Finland	3.6		
Ireland	4.3		
Norway	5.4		
Netherlands	12.4		
Belgium	16.0		
Sweden	17.5		
Germany	28.9		
United Kingdom	30.5		
France	50.1		
Total	170.8		

$$s^2 = \frac{\sum_{i=1}^N (x - \bar{x})^2}{N}$$

Variance: sample data

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Country	Asylum seekers	$(x - \bar{x})$	$(x - \bar{x})^2$
Denmark	2.3	$2.3 - 17.1 = -14.8$	
Finland	3.6	$3.6 - 17.1 = -13.5$	
Ireland	4.3	$4.3 - 17.1 = -12.8$	
Norway	5.4	$5.4 - 17.1 = -11.7$	
Netherlands	12.4	$12.4 - 17.1 = -4.7$	
Belgium	16.0	$16.0 - 17.1 = -1.1$	
Sweden	17.5	$17.5 - 17.1 = 0.4$	
Germany	28.9	$28.9 - 17.1 = 11.8$	
United Kingdom	30.5	$30.5 - 17.1 = 13.4$	
France	50.1	$50.1 - 17.1 = 33$	
Total	170.8		

$$s^2 = \frac{\sum_{i=1}^N (x - \bar{x})^2}{N}$$

Variance: sample data

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Country	Asylum seekers	$(x - \bar{x})$	$(x - \bar{x})^2$
Denmark	2.3	$2.3 - 17.1 = -14.8$	
Finland	3.6	$3.6 - 17.1 = -13.5$	
Ireland	4.3	$4.3 - 17.1 = -12.8$	
Norway	5.4	$5.4 - 17.1 = -11.7$	
Netherlands	12.4	$12.4 - 17.1 = -4.7$	
Belgium	16.0	$16.0 - 17.1 = -1.1$	
Sweden	17.5	$17.5 - 17.1 = 0.4$	
Germany	28.9	$28.9 - 17.1 = 11.8$	
United Kingdom	30.5	$30.5 - 17.1 = 13.4$	
France	50.1	$50.1 - 17.1 = 33$	
Total	170.8	0	

$$s^2 = \frac{\sum_{i=1}^N (x - \bar{x})^2}{N}$$

Variance: sample data

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Country	Asylum seekers	$(x - \bar{x})$	$(x - \bar{x})^2$
Denmark	2.3	$2.3 - 17.1 = -14.8$	$(-14.8)^2 = 219.0$
Finland	3.6	$3.6 - 17.1 = -13.5$	$(-13.5)^2 = 182.3$
Ireland	4.3	$4.3 - 17.1 = -12.8$	$(-12.8)^2 = 163.8$
Norway	5.4	$5.4 - 17.1 = -11.7$	$(-11.7)^2 = 136.9$
Netherlands	12.4	$12.4 - 17.1 = -4.7$	$(-4.7)^2 = 22.1$
Belgium	16.0	$16.0 - 17.1 = -1.1$	$(-1.1)^2 = 1.2$
Sweden	17.5	$17.5 - 17.1 = 0.4$	$(0.4)^2 = 0.2$
Germany	28.9	$28.9 - 17.1 = 11.8$	$(11.8)^2 = 139.2$
United Kingdom	30.5	$30.5 - 17.1 = 13.4$	$(13.4)^2 = 179.6$
France	50.1	$50.1 - 17.1 = 33$	$(33)^2 = 1089$
Total	170.8	0	2133.3

$$s^2 = \frac{\sum_{i=1}^N (x - \bar{x})^2}{N}$$

Variance: sample data

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Country	Asylum seekers	$(x - \bar{x})$	$(x - \bar{x})^2$
Denmark	2.3	$2.3 - 17.1 = -14.8$	$(-14.8)^2 = 219.0$
Finland	3.6	$3.6 - 17.1 = -13.5$	$(-13.5)^2 = 182.3$
Ireland	4.3	$4.3 - 17.1 = -12.8$	$(-12.8)^2 = 163.8$
Norway	5.4	$5.4 - 17.1 = -11.7$	$(-11.7)^2 = 136.9$
Netherlands	12.4	$12.4 - 17.1 = -4.7$	$(-4.7)^2 = 22.1$
Belgium	16.0	$16.0 - 17.1 = -1.1$	$(-1.1)^2 = 1.2$
Sweden	17.5	$17.5 - 17.1 = 0.4$	$(0.4)^2 = 0.2$
Germany	28.9	$28.9 - 17.1 = 11.8$	$(11.8)^2 = 139.2$
United Kingdom	30.5	$30.5 - 17.1 = 13.4$	$(13.4)^2 = 179.6$
France	50.1	$50.1 - 17.1 = 33$	$(33)^2 = 1089$
Total	170.8	0	2133.3

$$s^2 = \frac{\sum_{i=1}^N (x - \bar{x})^2}{N} = \frac{2133.3}{10} = 213.3.$$

Standard deviation



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

Standard deviation:

$$s_x = \sqrt{\text{Var}(x)} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

Example variance



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

$$\begin{aligned}
 s_x^2 &= \frac{1}{N} \sum_{i=1}^N (x - \bar{x})^2 = \frac{1}{12} \sum_{i=1}^N (x - 4.6)^2 \\
 &= \frac{1}{12} ((1.4)^2 + (-2.6)^2 + (-1.6)^2 + (-4.6)^2 + (2.4)^2 + (4.4)^2 \\
 &\quad + (1.4)^2 + (2.4)^2 + (0.4)^2 + (-1.6)^2 + (2.4)^2 + (-4.6)^2) \\
 &= \frac{1}{12} (2.0 + 6.7 + 2.5 + 21.0 + 5.8 + 19.5 + 2.0 + 5.8 + 0.2 + 2.5 \\
 &\quad + 5.8 + 21.0) = \frac{94.9}{12} = 7.9
 \end{aligned}$$

Example standard deviation

Graphical
descriptivesFrequency
tablesCentral
tendency**Variation**

References

6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Standard deviation: $s = \sqrt{s^2} = \sqrt{7.9} = 2.8$.

Variance: sample data

Graphical
descriptivesFrequency
tablesCentral
tendency

Variation

References

Country	Asylum seekers	$(x - \bar{x})$	$(x - \bar{x})^2$
Denmark	2.3	$2.3 - 17.1 = -14.8$	$(-14.8)^2 = 219.0$
Finland	3.6	$3.6 - 17.1 = -13.5$	$(-13.5)^2 = 182.3$
Ireland	4.3	$4.3 - 17.1 = -12.8$	$(-12.8)^2 = 163.8$
Norway	5.4	$5.4 - 17.1 = -11.7$	$(-11.7)^2 = 136.9$
Netherlands	12.4	$12.4 - 17.1 = -4.7$	$(-4.7)^2 = 22.1$
Belgium	16.0	$16.0 - 17.1 = -1.1$	$(-1.1)^2 = 1.2$
Sweden	17.5	$17.5 - 17.1 = 0.4$	$(0.4)^2 = 0.2$
Germany	28.9	$28.9 - 17.1 = 11.8$	$(11.8)^2 = 139.2$
United Kingdom	30.5	$30.5 - 17.1 = 13.4$	$(13.4)^2 = 179.6$
France	50.1	$50.1 - 17.1 = 33$	$(33)^2 = 1089$
Total	170.8	0	2133.3

$$s^2 = \frac{\sum_{i=1}^N (x - \bar{x})^2}{N} = \frac{2133.3}{10} = 213.3.$$

$$s = \sqrt{s^2} = \sqrt{213.3} = 14.6.$$

Variance and outliers



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

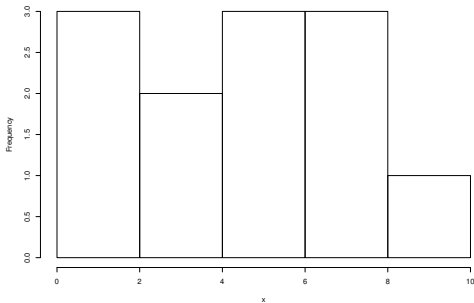
References

The variance refers to the variation in the data around the mean. It is similarly sensitive to outliers - a few extreme values in the data can significantly increase the variance.

Example summary



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---



Graphical
descriptives

Frequency
tables

Central
tendency

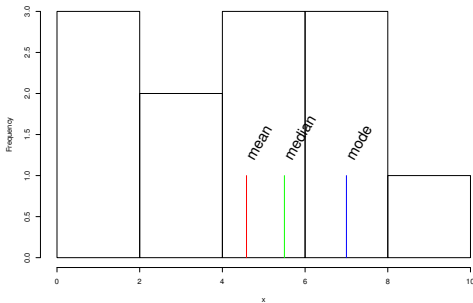
Variation

References

Example summary



6	2	3	0	7	9	6	7	5	3	7	0
---	---	---	---	---	---	---	---	---	---	---	---

Graphical
descriptivesFrequency
tablesCentral
tendency**Variation**

References



Graphical
descriptives

Frequency
tables

Central
tendency

Variation

References

Baturo, Alexander and Johan A. Elkink. 2014. "Office or Officeholder? Regime Deinstitutionalisation and Sources of Individual Political Influence." *Journal of Politics* 76(3).

Baturo, Alexander and Johan A. Elkink. 2015. "Dynamics of Regime Personalization and Patron-Client Networks in Russia, 1999–2014." *Post-Soviet Affairs* 32(1):75–98.

Healey, Joseph F. 1996. *Statistics: a tool for social research*. Wadsworth.

Marshall, M.G. and K. Jagers. 2002. "Polity IV project: political regime characteristics and transitions, 1800-2002."

URL: <http://www.bsos.umd.edu/cidcm/polity/>

Moore, David S. 2003. *The basic practice of statistics*. 3rd ed. New York: W.H. Freeman.