John Graunt

1620 - 1674
John Graunt

- 1620 - 1674
- “Haberdasher of small-wares” (buttons, needles, and such)
John Graunt

- 1620 - 1674
- “Haberdasher of small-wares” (buttons, needles, and such)
- Elected Fellow of the Royal Society in 1662 at the special request of Charles II
Weekly accounts, issued by parish clerks, of all deaths, along with their causes, and Christenings in the parish for the week.
Bills of Mortality

- Weekly accounts, issued by parish clerks, of all deaths, along with their causes, and Christenings in the parish for the week
- Graunt is the first to recognize the wealth of information, useful for both the state and for business, contained in these bills.
Bills of Mortality

- Weekly accounts, issued by parish clerks, of all deaths, along with their causes, and Christenings in the parish for the week.
- Graunt is the first to recognize the wealth of information, useful for both the state and for business, contained in these bills.
- Graunt (page 1421): “Now having (I know not by what accident) engaged my thoughts upon the *Bills of Mortality*, and so far succeeded therein, as to have reduced several great confused *Volumes* into a few perspicuous *Tables*, and abridged such *Observations* as naturally flowed from them, into a few succinct *Paragraphs* . . . ”
Graunt (page 1435): “I conclude, That a clear knowledge of all these particulars, and many more, whereat I have shot but at rovers, is necessary in order to good, certain, and easie Government, and even to balance Parties, and factions both in Church and State.”

These data of the state became known as statistics.
Graunt (page 1435): “I conclude, That a clear knowledge of all these particulars, and many more, whereat I have shot but at rovers, is necessary in order to good, certain, and easie Government, and even to balance Parties, and factions both in *Church* and *State.*”

See reasons on page 1434, and questions which may be answered on page 1433.
Graunt (page 1435): “I conclude, That a clear knowledge of all these particulars, and many more, whereat I have shot but at rovers, is necessary in order to good, certain, and easie Government, and even to balance Parties, and factions both in Church and State.”

- See reasons on page 1434, and questions which may be answered on page 1433.
- These data of the state became known as statistics.
Page 1429: Since few starve, wouldn’t it be “better for the State to keep them?”
Examples

- Page 1429: Since few starve, wouldn’t it be “better for the State to keep them?”

- Page 1430: There are some causes of death about which “there be daily talk,” but little effect.
Edmond Halley

- 1658 - 1744

Pushed Newton to complete and publish his *Philosophiae naturalis principia mathematica*.

Studied comets, and, in particular, predicted the time of return for the comet we now know as Halley's comet.

His tables of mortality rates, based on the birth and death records of Breslaw, provided the first firm data for calculating insurance and annuity rates.
Edmond Halley

- 1658 - 1744
- Pushed Newton to complete and publish his *Philosophiae naturalis principia mathematica*

Studied comets, and, in particular, predicted the time of return for the comet we now know as Halley's comet.

His tables of mortality rates, based on the birth and death records of Breslaw, provided the first firm data for calculating insurance and annuity rates.
Edmond Halley

- 1658 - 1744
- Pushed Newton to complete and publish his *Philosophiae naturalis principia mathematica*
- Studied comets, and, in particular, predicted the time of return for the comet we now know as Halley’s comet.
Edmond Halley

- 1658 - 1744
- Pushed Newton to complete and publish his *Philosophiae naturalis principia mathematica*
- Studied comets, and, in particular, predicted the time of return for the comet we now know as Halley’s comet.
- His tables of mortality rates, based on the birth and death records of Breslaw, provided the first firm data for calculating insurance and annuity rates.
On average

- Although we cannot predict if a given individual will die during the year, or contract a certain disease, we can predict on average how many people of his or her age will die, or contract that disease.
On average

Although we cannot predict if a given individual will die during the year, or contract a certain disease, we can predict on average how many people of his or her age will die, or contract that disease.

Similarly, although we cannot predict exactly the yield of a given field, we can say how much a field of this type should produce on average.
Some statistics

- Given a list of data, the *mean* is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.
Some statistics

- Given a list of data, the *mean* is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.
- **Example**: Given the data 5, 6, 13, 14, 3, 3, 3, 4, and 12, the mean is
  \[
  \frac{5 + 6 + 13 + 14 + 3 + 3 + 3 + 4 + 12}{9} = \frac{63}{9} = 7.
  \]
Some statistics

- Given a list of data, the *mean* is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.

  Example: Given the data 5, 6, 13, 14, 3, 3, 3, 4, and 12, the mean is
  \[
  \frac{5 + 6 + 13 + 14 + 3 + 3 + 3 + 4 + 12}{9} = \frac{63}{9} = 7.
  \]

- The *median* of a list of data is the middle value when the data are listed in ascending order.
Some statistics

- Given a list of data, the mean is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.

- Example: Given the data 5, 6, 13, 14, 3, 3, 3, 4, and 12, the mean is
  \[
  \frac{5 + 6 + 13 + 14 + 3 + 3 + 3 + 4 + 12}{9} = \frac{63}{9} = 7.
  \]

- The median of a list of data is the middle value when the data are listed in ascending order.
  - Note: there is a unique middle value for an odd number of data values, but two middle values for an even number of data values.
Some statistics

- Given a list of data, the *mean* is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.
- Example: Given the data 5, 6, 13, 14, 3, 3, 3, 4, and 12, the mean is
  \[
  \frac{5 + 6 + 13 + 14 + 3 + 3 + 3 + 4 + 12}{9} = \frac{63}{9} = 7.
  \]
- The *median* of a list of data is the middle value when the data are listed in ascending order.
  - Note: there is a unique middle value for an odd number of data values, but two middle values for an even number of data values.
  - In the latter case, the average of the two middle values is taken as the median.
Some statistics

- Given a list of data, the *mean* is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.
- Example: Given the data 5, 6, 13, 14, 3, 3, 3, 4, and 12, the mean is
  \[
  \frac{5 + 6 + 13 + 14 + 3 + 3 + 3 + 4 + 12}{9} = \frac{63}{9} = 7.
  \]
- The *median* of a list of data is the middle value when the data are listed in ascending order.
  - Note: there is a unique middle value for an odd number of data values, but two middle values for an even number of data values.
  - In the latter case, the average of the two middle values is taken as the median.
- Example: The previous data listed in order are 3, 3, 3, 4, 5, 6, 12, 13, and 14, so the median value is 5.
Some statistics

- Given a list of data, the **mean** is the arithmetic average of the data, that is, the sum of the data divided by the number of data values.

- **Example:** Given the data 5, 6, 13, 14, 3, 3, 3, 4, and 12, the mean is
  \[
  \frac{5 + 6 + 13 + 14 + 3 + 3 + 3 + 4 + 12}{9} = \frac{63}{9} = 7.
  \]

- The **median** of a list of data is the middle value when the data are listed in ascending order.

  - **Note:** there is a unique middle value for an odd number of data values, but two middle values for an even number of data values.
  
  - In the latter case, the average of the two middle values is taken as the median.

- **Example:** The previous data listed in order are 3, 3, 3, 4, 5, 6, 12, 13, and 14, so the median value is 5.

- **Example:** The median of 4, 8, 9, 13, 14, 22 is
  \[
  \frac{9 + 13}{2} = 11.
  \]
Some statistics (cont’d)

- The *mode* of a set of data is the value which occurs most frequently.
Some statistics (cont’d)

- The *mode* of a set of data is the value which occurs most frequently.
- Example: The mode of the data 5, 6, 13, 14, 3, 3, 3, 4, and 12 is 3.

Suppose a company has 100 employees with a salary of $30,000 per year, 20 employees who make $50,000 per year, 5 employees who make $100,000 per year, and one employee who makes $5,000,000 per year.

Then the mean salary is 

\[
(100 \times 30,000) + (20 \times 50,000) + (5 \times 100,000) + 5,000,000 = 9,500,000
\]

$75,000 per year, the median salary is $30,000 per year, and the mode is also $30,000 per year.

What is the “average” salary in this company?
Some statistics (cont’d)

▶ The *mode* of a set of data is the value which occurs most frequently.
▶ Example: The mode of the data 5, 6, 13, 14, 3, 3, 3, 4, and 12 is 3.
▶ Example
Some statistics (cont’d)

- The *mode* of a set of data is the value which occurs most frequently.
- Example: The mode of the data 5, 6, 13, 14, 3, 3, 3, 4, and 12 is 3.
- Example
  - Suppose a company has 100 employees with a salary of $30,000 per year, 20 employees who make $50,000 per year, 5 employees who make $100,000 per year, and one employee who makes $5,000,000 per year.
Some statistics (cont’d)

- The *mode* of a set a data is the value which occurs most frequently.

- Example: The mode of the data 5, 6, 13, 14, 3, 3, 3, 4, and 12 is 3.

- Example

  - Suppose a company has 100 employees with a salary of $30,000 per year, 20 employees who make $50,000 per year, 5 employees who make $100,000 per year, and one employee who makes $5,000,000 per year.

  - Then the mean salary is

    \[
    \frac{(100 \times 30,000) + (20 \times 50,000) + (5 \times 100,000) + 5,000,000}{126} = \frac{9,500,000}{126} = $75,397 \text{ per year},
    \]

    the median salary is $30,000 per year, and the mode is also $30,000 per year.
Some statistics (cont’d)

- The mode of a set of data is the value which occurs most frequently.
- Example: The mode of the data 5, 6, 13, 14, 3, 3, 3, 4, and 12 is 3.
- Example
  - Suppose a company has 100 employees with a salary of $30,000 per year, 20 employees who make $50,000 per year, 5 employees who make $100,000 per year, and one employee who makes $5,000,000 per year.
  - Then the mean salary is
    \[
    \frac{(100 \times 30,000) + (20 \times 50,000) + (5 \times 100,000) + 5,000,000}{126} = \frac{9,500,000}{126} = \$75,397 \text{ per year,}
    \]
  - the median salary is $30,000 per year, and the mode is also $30,000 per year.
- What is the “average” salary in this company?
Note: If the data are symmetrically distributed, then the median and the mean will be close to each other, but if the data are not symmetrically distributed they can be very different.
Note: If the data are symmetrically distributed, then the median and the mean will be close to each other, but if the data are not symmetrically distributed they can be very different.

- In particular, like in the last example, a few very large data values will affect the mean but not the median.
Some statistics (cont’d)

- Note: If the data are symmetrically distributed, then the median and the mean will be close to each other, but if the data are not symmetrically distributed they can be very different.
  - In particular, like in the last example, a few very large data values will affect the mean but not the median.
  - The result is that for economic data like incomes or housing prices, the mean is often much larger than the median.
Note: If the data are symmetrically distributed, then the median and the mean will be close to each other, but if the data are not symmetrically distributed they can be very different.

- In particular, like in the last example, a few very large data values will affect the mean but not the median.
- The result is that for economic data like incomes or housing prices, the mean is often much larger than the median.
- In such cases, the median is more indicative of the average than is the mean.
Problems

1. In 1798 Henry Cavendish repeated an experiment for measuring the density of the earth 23 times. His results were

   5.36  5.62  5.27  5.46  5.53  5.57
   5.29  5.29  5.39  5.30  5.10  5.79
   5.58  5.44  5.42  5.75  5.34  5.63
   5.65  5.34  5.47  5.68  5.85

   a. Find the mean of this data.
   b. Find the median of this data.
   c. Find the mode of this data.
2. The number of home runs hit by the American League home run leaders for the years 1972 to 1991 are as follows: 37, 32, 32, 36, 32, 39, 46, 45, 41, 22, 39, 39, 43, 40, 40, 49, 42, 36, 51, 44.

   a. Find the mean of this data.
   b. Find the median of this data.
   c. Find the mode of this data.
   d. One of the numbers in this data set appears to be inconsistent with the other values. Remove this value and recompute the mean, median, and mode for the remaining data. Can you think of an explanation for the unusual value?

3. Suppose you read in one newspaper that the average salary of an NBA basketball player is $1,000,000 and you read in another newspaper that the average salary of an NBA basketball player is $4,000,000. Which one of these numbers is the mean salary and which one is the median salary?