



CONTENTS

- Numerical methods: average and spread
- Boxplots
- Definition of the standard deviation
- Empirical rules
- Use of standard deviation
- The standard error



DESCRIPTIVE STATISTICS

- To cope with inherent variability: no 2 values measured in nature are identical.
- From <http://www.guinnessworldrecords.com/>
(***Guinness World Records***, known until 2000 as ***The Guinness Book of Records***)
- **Tallest man - living**
Xi Shun (China, b. 1951), measured 7 ft 8.95 in (2 m 36.1 cm) as a result of six measurements taken on January 15, 2005, at Chifeng City Hospital, Inner Mongolia, China.



VARIABILITY

Opening sentence from:

**A concise summary
of the International System of Units, the SI**

www.bipm.org/utils/common/pdf/si_summary_en.pdf

*“Metrology is the science of measurement,
embracing all measurements,
made at a known level of uncertainty,
in any field of human activity.”*



VARIABILITY

DC TO AC POWER INVERTOR 150W DC12V to AC230V A16FG Instruction Manual

Please read user manual before use.

USEFUL APPLICATIONS

RUN NOTEBOOK COMPUTERS, RADIOS, SMALL TVS,
VCRS, LAMPS, FANS, FAX, ETC.

SPECIFICATION

INPUT VOLTAGE RANGE : 10-15V DC

INPUT FULL LOAD CURRENT : 15A

STANDBY INPUT CURRENT : < 0.5A

OUTPUT VOLTAGE (AC) : 230V \pm 5%

OUTPUT WAVEFORM : MODIFY SINEWAVE

OUTPUT FREQUENCY : 50Hz

CONTINUOUS OUTPUT POWER : 150W (20 minutes)

PEAK OUTPUT POWER : 450W

EFFICIENCY : 90%

BATTERY LOW SHUTDOWN : 10 \pm 0.5V DC(WILL SHUTDOWN WITHIN 6 SECOND)

BATTERY LOW PRE-ALARM : 11 \pm 0.5V DC

THERMAL PROTECT : 60 \pm 5 $^{\circ}$ C (MICROCONTROLLER)



DESCRIPTIVE STATISTICS

- Once we have measured a sample, how do we describe or summarise what the sample is like?
- What is the 'average' value?
- How far apart are these values spread?
- There are graphical and numerical methods



NUMERICAL METHODS

- of description -

- **AVERAGE**
- Mean → the arithmetic average
- Median → the value ranked mid way

- **SPREAD - VARIABILITY**
- Range → highest value - lowest value
- Interquartile range → in terms of ranks:
3rd quarter - 1st quarter = the middle half



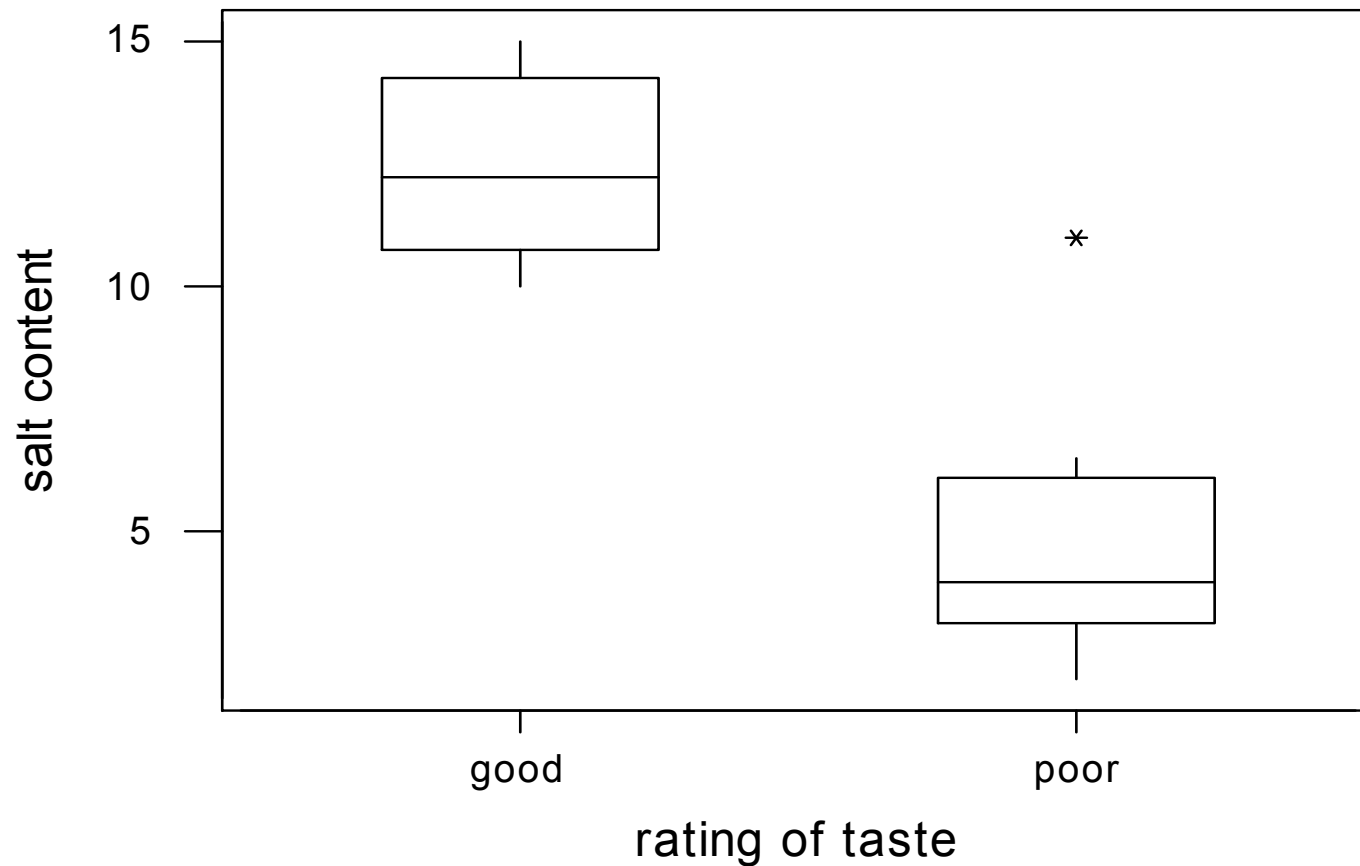
Graphical method BOXPLOT

- Uses the median and interquartile range: robust to unusually extreme values
- Emphasizes where the middle 50% of records lies by drawing a box
- Flags presence of extreme values - 'outliers'
- Depends on package but usually symbol is a * if record $< Q1 - 1.5IR$ or $> Q3 + 1.5IR$



BOXPLOT: quick visual comparison

In space between different samples...





STANDARD DEVIATION: the 'Statistical' measure of variability

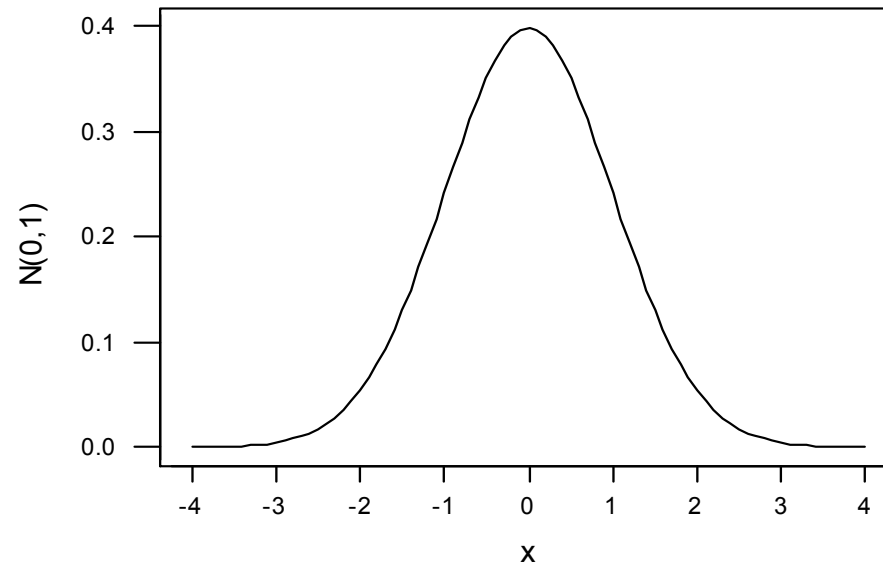
- 'Ranges' easy to find **BUT** make no use of distance between records, as they use ranks
- Variance $\sigma^2 =$ average square distance of records from the mean
- Standard deviation $\sigma = \sqrt{\text{variance}}$
 \Rightarrow thus same units as the original records
- Also coefficient of variation
 $CV = (\sigma/\text{mean}) * 100\%$



How do statisticians use σ ?

- If we deal with a so called ‘normally’ distributed variable, empirical rules hold:

% of records	Found within
68	$\mu \pm 0.97\sigma$
95	$\mu \pm 1.96\sigma$
99	$\mu \pm 2.58\sigma$



Notice symmetry

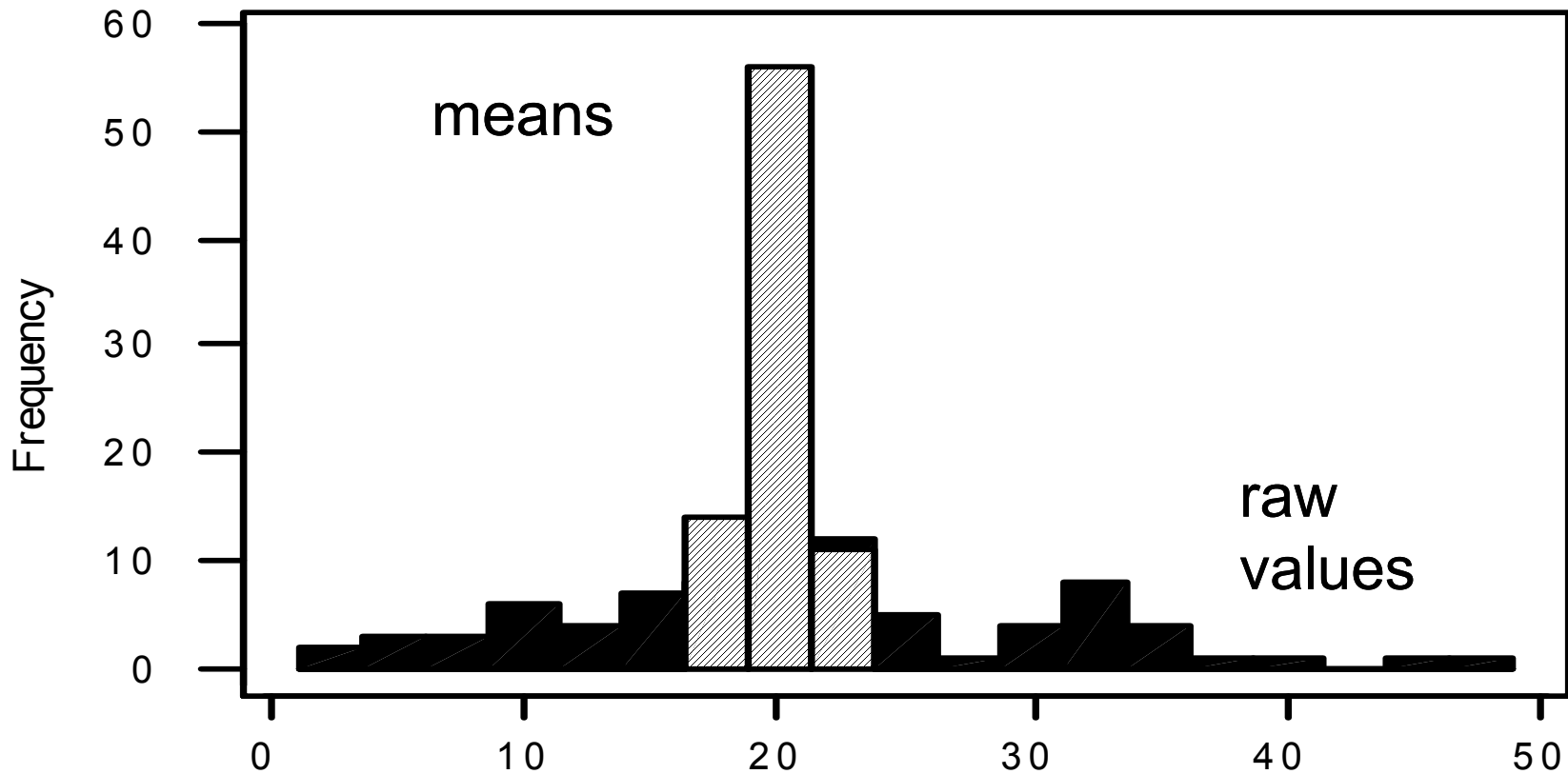


Stdev & standard error

- Draw a sample of raw values $x_1 x_2 \dots x_n$
- From the same population, draw many samples of same size and note the means $\bar{x}_1 \bar{x}_2 \dots \bar{x}_k$
- Superimpose histograms:
- Intuitively, spread of $\bar{x} < \text{spread of } X$



Stdev & standard error



thus $\sigma(\bar{x}) < \sigma(X)$

call $\sigma(\bar{x})$ standard error



The standard error of the mean

- σ is the spread of the raw records while s.e.(m) is the spread of the means
- So s.e.(m) is to \bar{x} what σ is to X

In fact

$$\text{s.e.}(m) = \frac{\sigma}{\sqrt{n}}$$

