

Research Methods

Dr Ala Hussany and Dr Rachel Alexander

Variables:

Characteristic of an individual

A variable is anything that changes in a scientific experiment or mathematical calculation.

There are two types of variables:

Independent Variable: What you change (cause).

Dependent Variable: What changes because of the Independent Variable (effect).

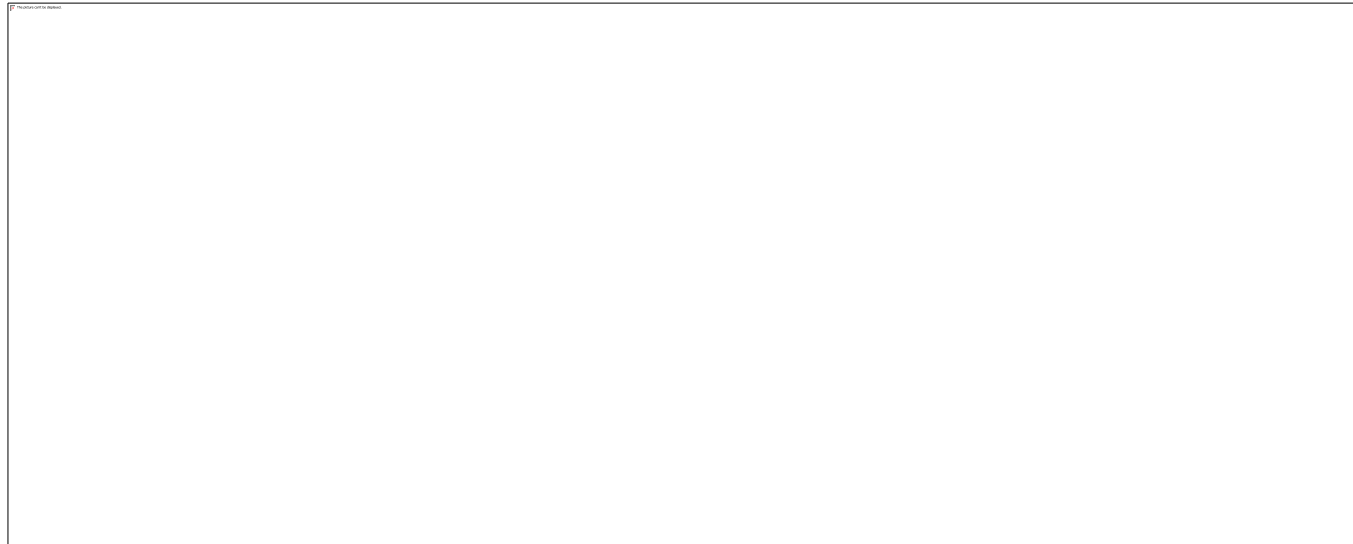
If the density of the material increased, then the velocity of the sound will increase.

If the temperature of the room is increased, then the bacteria will grow faster.

The density of material and the room temperature are independent variables.

The velocity of the sound and the bacterial grow are dependent variables.

If independent variable Then dependent variable.



The methods used to display, summarize, and analyze data depend on whether the variables are qualitative or quantitative.

Quantitative (**numerical values**): Can be added, subtracted, averaged, etc.

Qualitative (**Categorical**): An individual is placed into one of several groups or categories. These groups or categories are not **usually** numerical.

Measurable observations are called quantitative observations
Observations that cannot be measured are termed qualitative.
Qualitative observations can only be described.

age, height, gender, favorite class, speed, alive dissatisfied,
energy, stress tensile

Nominal Variable: variable that categorizes (or describes, or names) an element of a population.

Nominal level data are the simplest type of data and consist of named categories, with no implied order among the categories.

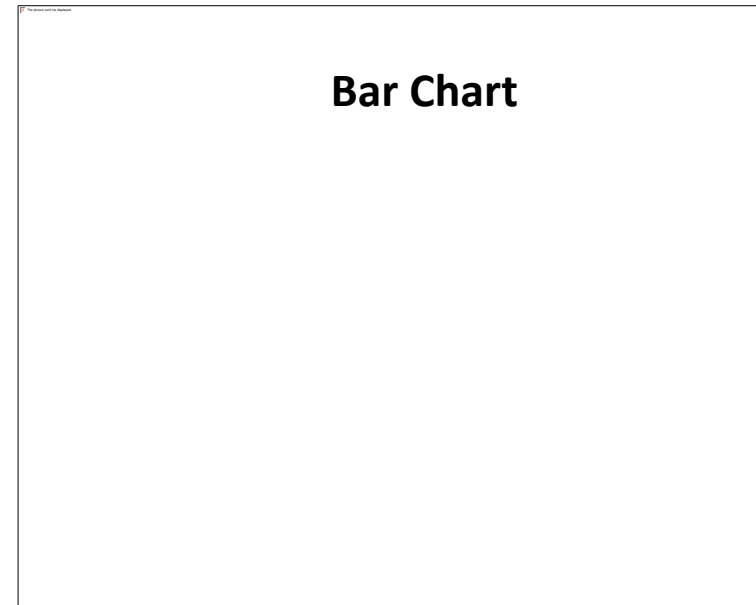
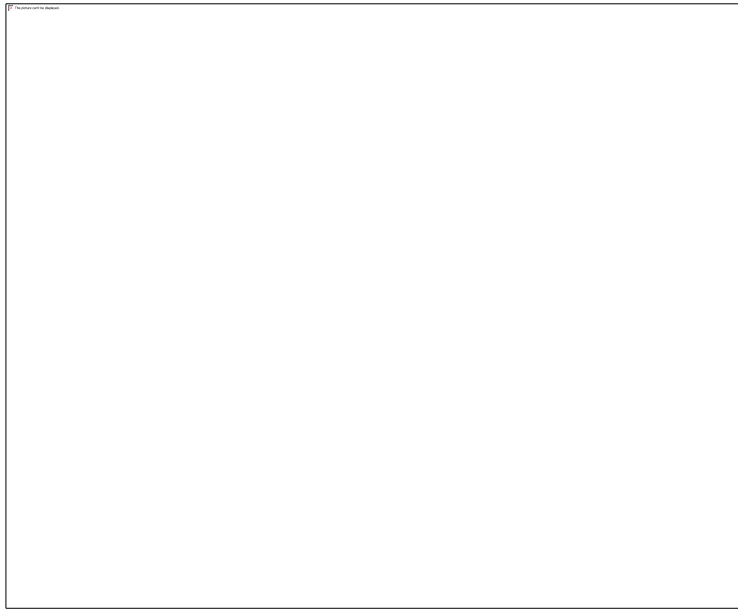


Ordinal Variable: A qualitative variable that incorporates an ordered position, or ranking.

Ordinal variables there is no origin and no unit of measurement, but there is an order

Education level (elementary, secondary, college)

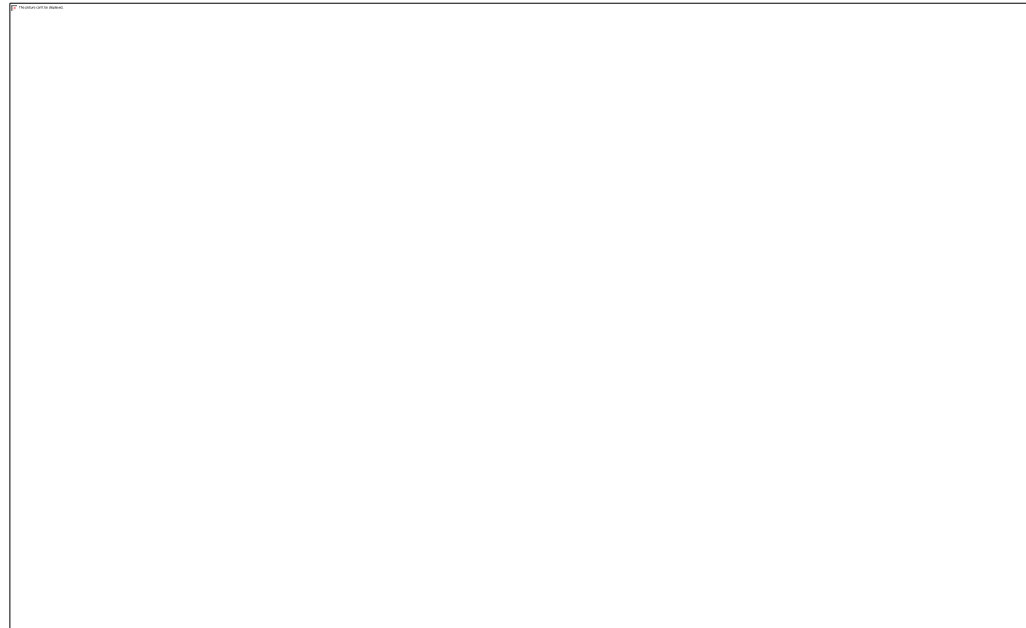
Suppose we poll 46 people on an issue. How can we exhibit their response?



Discrete Variable: A quantitative variable that can assume a countable number of values. **Intuitively**, a discrete variable can assume values corresponding to isolated points along a line interval. That is, there is a gap between any two values.



Continuous Variable: A quantitative variable that can assume an uncountable number of values. **Intuitively**, a continuous variable can assume any value along a line interval, including every possible value between any two values.



Data Frequency tables

Data frequency table refers to a way of presenting a summary of the data in such a manner that it facilitates the possibility of seeing patterns or relationships in the data.

A data frequency table shows how many times each data point (observation or outcome) occurs in a given data set.

School of Business supplied the following number of months working experience:

23	28	29	34	34	39	43	44	45	45	48	48	49	54	54
54	55	56	56	65	65	65	67	73	76	76	77	78	87	92

The first step in producing a Frequency Table is to establish the range of the data.

The second step is to decide the number of groups/intervals into which the data should be divided.

23 28 29 34 34 39 43 44 45 45 48 48 49 54 54
54 55 56 56 65 65 65 67 73 76 76 77 78 87 92

For School of Business (<25, 35, 45, 55, 65, 75, and >85)

The third step is to grouped the data into a number of intervals.
Enter possible number of experience months intervals.

<25 26-35 36-45 46-55 56-65 66-75 76-85 >85

1 4 5 7 5 2 4 2

Data Distribution

<25 26-35 36-45 46-55 56-65 66-75 76-85 >85

Resulting Frequency Counts

1 4 5 7 5 2 4 2



School of Accounting supplied the following number of months working experience:

10 12 12 16 19 20 22 23 23 23 26 28 29 32 33
34 34 41 43 43 44 45 45 54 56 56 56 65 67 76

<25 25-35 36-45 46-55 56-65 66-75 76-85 >85

10 7 6 1 4 1 1 0

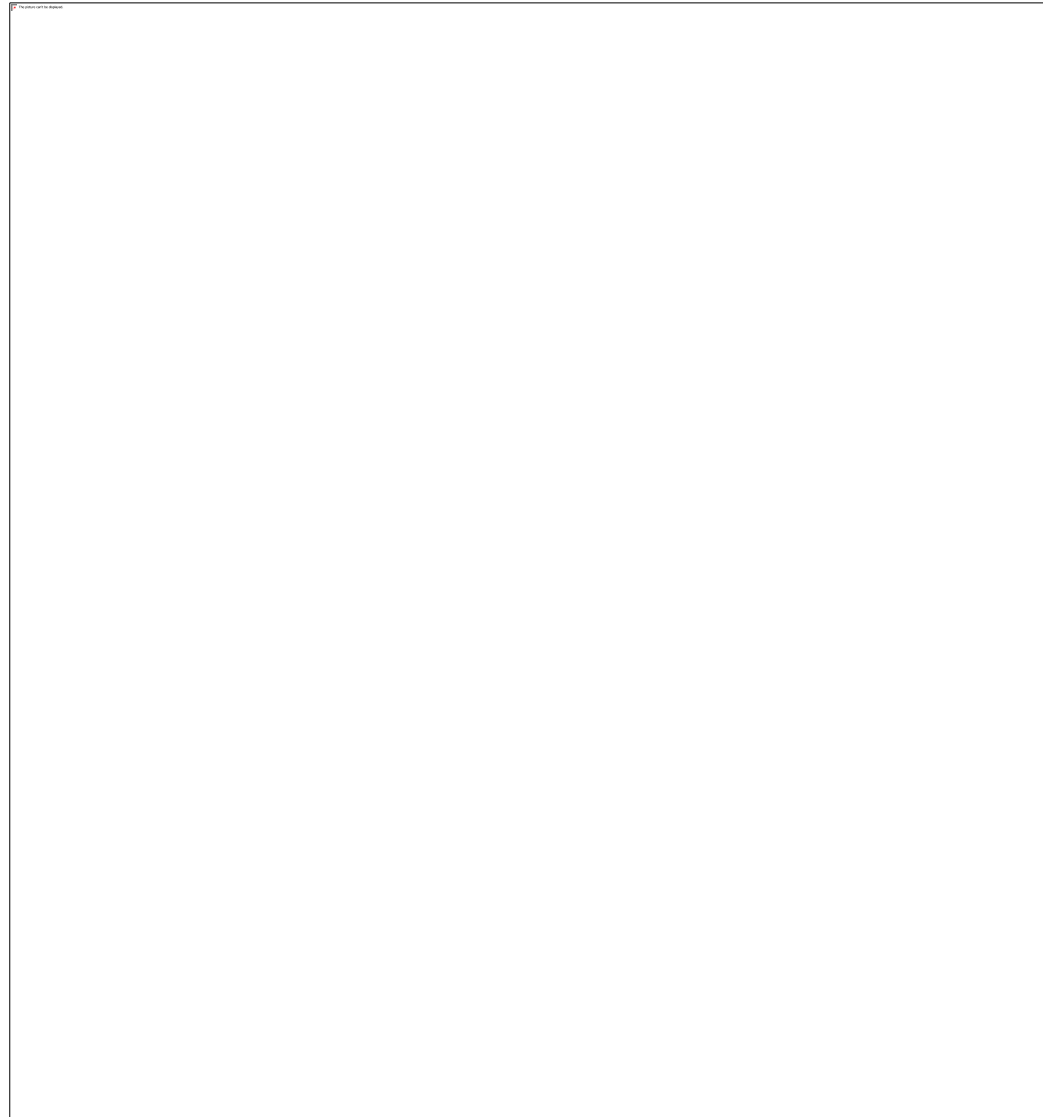


Reliability is the property of independent reproducibility of the data.

Validity is the quality of a measure being an adequate and acceptable representative of what it is supposed to represent.

*We can not have validity without reliability,
but a measure can be reliable without being
valid.*

Classic example: The scale which is consistent and able to distinguish from one person to the next but actually off by 5 pounds



In order to be valid, a test must be reliable; but reliability does not guarantee validity

Science is based on observation and measurement,
Hence,
these are imperfect.

A numerical value on the imperfection is called **an error**.

Error is defined by the webster as “*the difference between an observed or Calculated value and true value*”.

Discrepancy is the difference between two or more observations. This gives rise to **uncertainty**.

In general, the uncertainties represent the errors in the empirical results, and the procedure for estimating the errors is refer to **error analysis**.

Error = deviation of measurement from true value

Uncertainty = measure of likely error

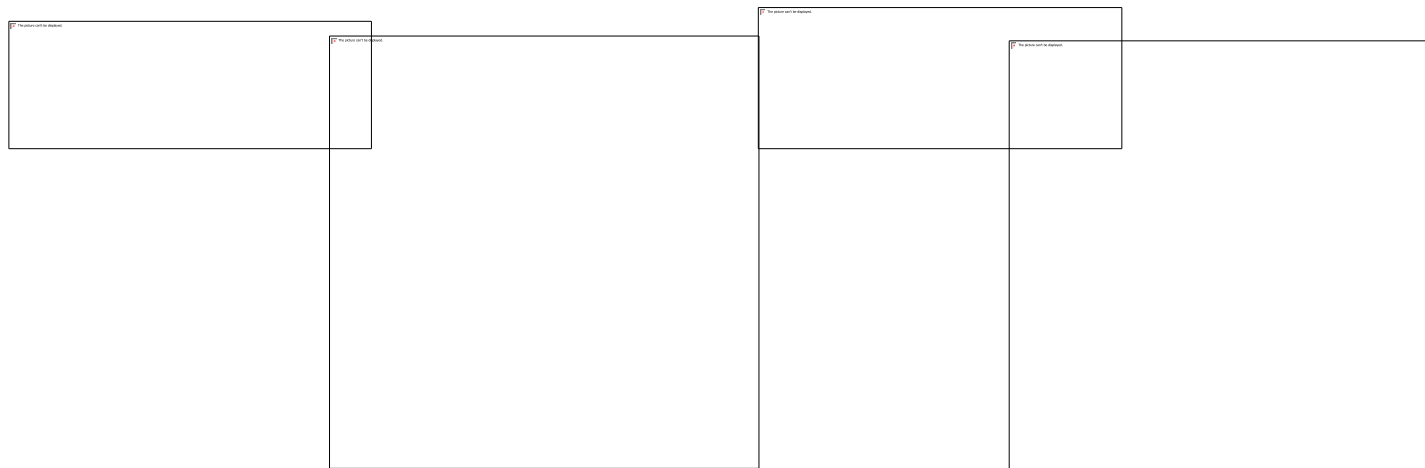
Accuracy and Precision

Accuracy is the agreement between a measured value and the true value.

(note that we may not always know the true value).

Precision also called reproducibility or repeatability, the degree to which further measurements or calculations show the same or similar results.

(A measure of the spread of data points).



**A result is
called valid if
it is both
accurate and
precise**

The precision and accuracy are limited by the instrumentation and data gathering techniques

Sources of experimental Error

Two sources of error in a measurement are

1. limitations in the sensitivity of the instruments
2. imperfections in experimental design or measurement techniques



Errors that originate from making mistake in measurement or computation called in science a **Blunder**. Error analysis does not take blunder into account.

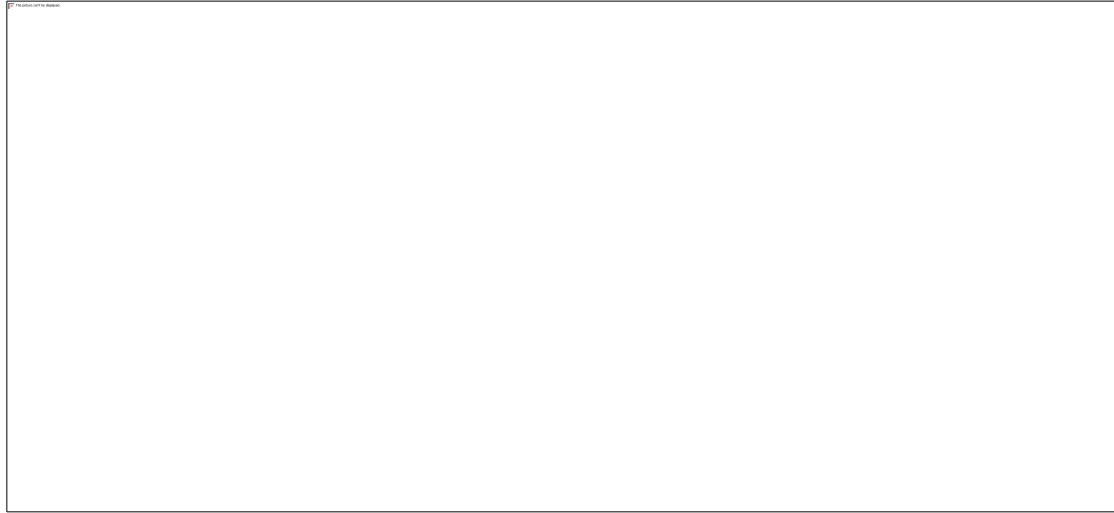
There are two types of experimental error:

1-A systematic error

This is an effect that always makes your measurement either bigger or smaller than it should be, and so distorts the results. Statistical analysis is generally not useful, but rather corrections must be made based on experimental conditions.

Measurement + Systematic Error *or* *Measurement - Systematic Error*

*Quantify
systematic shift
in measurements
away from 'true'
value*



How to minimize them?

- Careful calibration.
- Best possible techniques.

Zero Errors

A particular type of systematic error is the zero error.

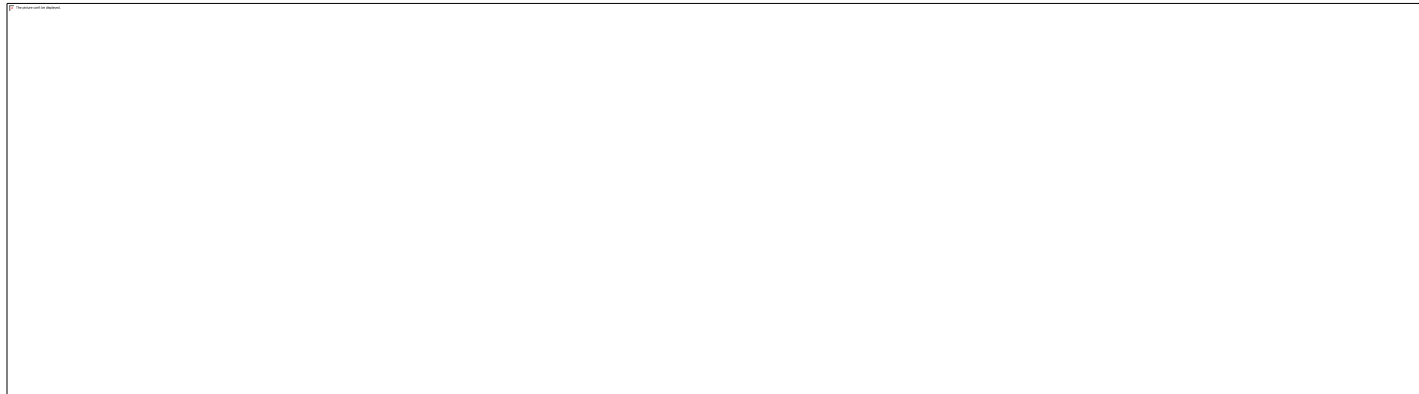
Apparatus that does not read zero even when it should

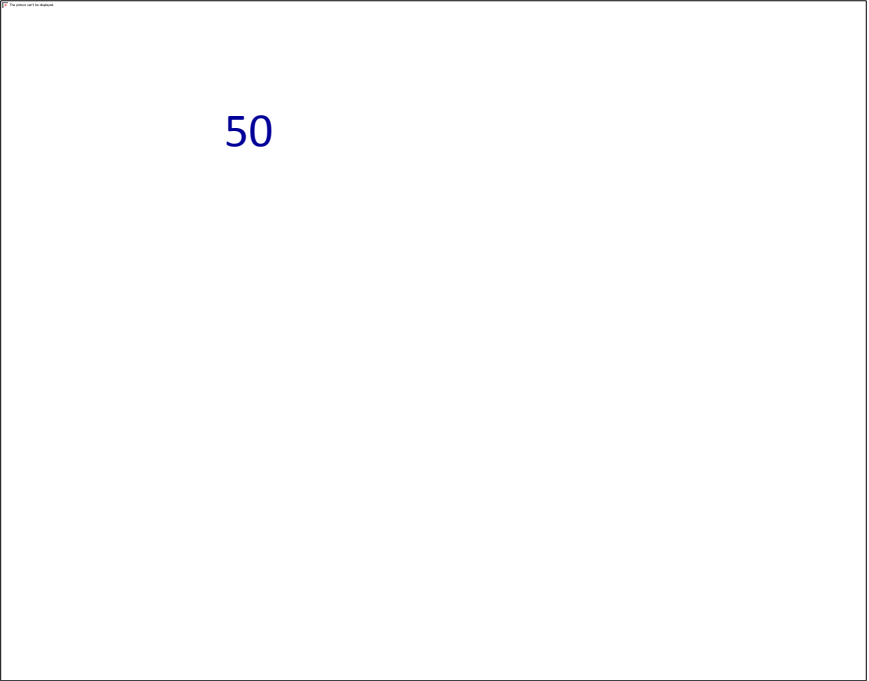
2-A random error *influences the precision of a result,*

This is due to effects that could make your measurement either larger or smaller than the real value. Requires that experiments be repeated a sufficient number of time to establish the precision of measurement.

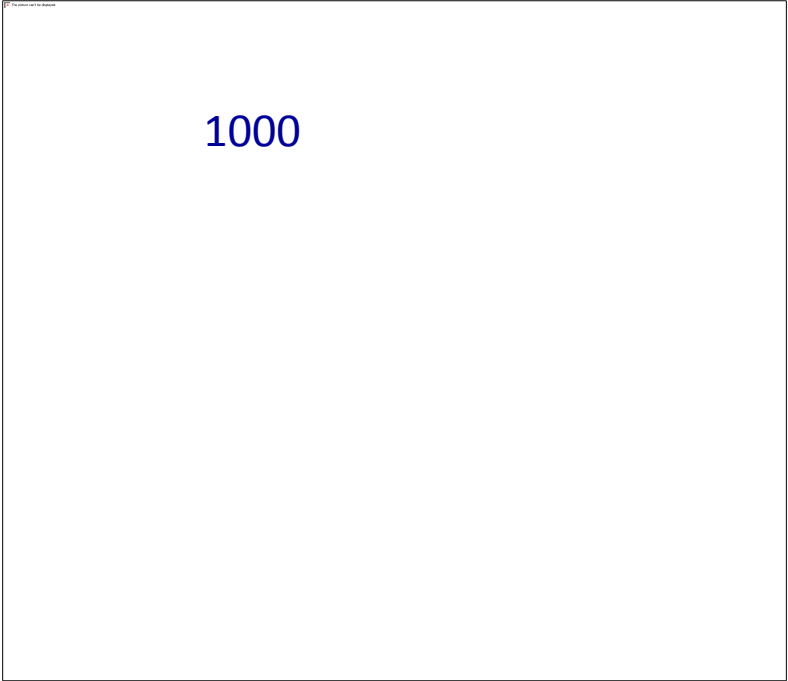
Repeated measurements result in a distribution of data indicates the presence of random error.

Best estimate of the quantity is the mean, $\bar{x}$:





50



1000

How to minimize them?

Take repeated measurements and calculate their average

Mean of measurements from large number of experiments
gives correct value for measured quantity

Measurements often approximately *Gaussian-distributed*

Dealing with Errors

Identify the errors and their magnitude.

Try to reduce the magnitude of the error by

1. Better instruments
2. Better experimental design
3. Collect a lot of data

Sources of arithmetic errors

Two sources of error

1- Round-off Error: Error created due to approximate representation of numbers

$$\frac{1}{3} \approx 0.333333$$

$$\text{Round-off Error} = \frac{1}{3} - 0.333333 = 0.0000003333$$

2-Truncation Error: the error created by approximating a mathematical procedure

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$e^{0.5} = 1 + 0.5 + \frac{0.5^2}{2!} + \frac{0.5^3}{3!} + \dots$$

$$e^{0.5} \approx 1 + 0.5 + \frac{0.5^2}{2!}$$

$$\text{Truncation error} = \frac{0.5^3}{3!} + \dots$$

Chopping

Example:

$p=3.14159265358979323846264\dots$

to be stored on a base-10 system carrying 7 significant digits.

$p=3.141592$ chopping error $e_t=0.00000065$

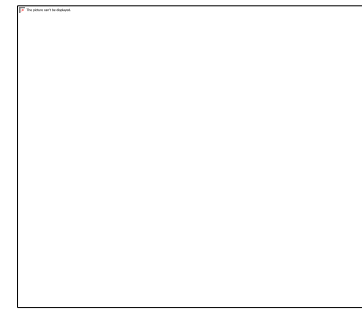
If rounded

$p=3.141593$ $e_t=0.00000035$

Round-off Errors

- Computers can represent numbers to a finite precision
- How do computers represent numbers?
- **Binary representation** of the integers and real numbers in computer memory

Mean, Median, Mode and Range



Mean: arithmetic average -sum of all observations divided by the number of observations

Mode: value that occurs most frequently in a distribution (usually the highest point of curve) may have more than one mode in a dataset

Median: value midway in the frequency distribution- half the area of curve is to right and other to left

Range: measure of dispersion about mean (maximum minus minimum)