Learning Objectives

Monday, March 2, 2020 9:39 AM

Learning objectives

1. Understand the difference between direct and indirect observations
2. Know the types and sources of errors made in surveying observations
3. Define accuracy and precision
4. Know how to adjust for systematic errors
5. Know how to calculate and examine residuals
6. Know how to calculate common statistics such as the mean, variance, standard deviation and confidence intervals
7. Understand how error can propagate and how to calculate the errors of a sum, of a series, of a product and of a mean

Homework Assignment

Do the following problems from the textbook:

3.1 (5 points)
3.2 (5 points)
3.5 (10 points)
3.7 (10 points)

3.12	For problem 3.7 determine the range within observations should fall (a) 90 percent of the time, (b) 95 percent of the time. (10 points)

3.17 (10 points)
3.18 (10 points)
3.21 (10 points)
3.23 (10 points)

Supplemental Problem 1 (20 points)

What is the area of a rectangular field and its estimated error for the following recorded values:

(a) 243.89 ± 0.05 ft, by 208.65 ± 0.04 ft

(b) 660.23 ± 0.012 ft by 1425.67 ± 0.013 ft

(c) 304.206 ± 0.005 m, by 758.234 ± 0.006 m

Introduction

Monday, March 2, 2020 9:39 AM

- **Making observations** (measurements), and subsequent computations and analyses using them, are **fundamental tasks of surveyors**.
- **Good observations** require a combination of **human skill** and **precision mechanical equipment** applied with the utmost judgment.
- No matter how carefully made, **observations are never exact** and will always contain errors.
- Geomatics engineers (surveyors) whose work must be performed to exacting standards should therefore t**horoughly understand the different kinds of errors**, their sources and expected magnitudes under varying conditions, and their manner of propagation.
- **Surveyors must be capable of assessing the magnitudes of errors** in their observations so that either their acceptability can be verified or, if necessary, new ones made.
- **Computers and sophisticated software are tools now commonly used** by surveyors to plan measurement projects, design measurement systems, investigate, and **distribute observational errors after results have been obtained.**

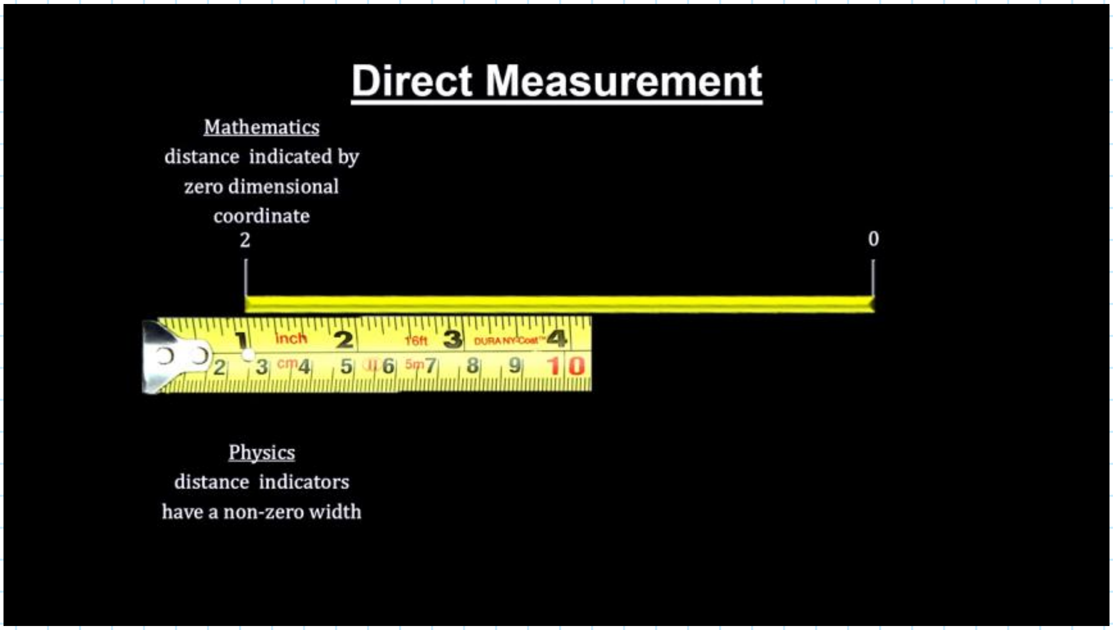
From <https://online.vitalsource.com/#/books/9781292060675/cfi/65!/4/4@0:31.6>

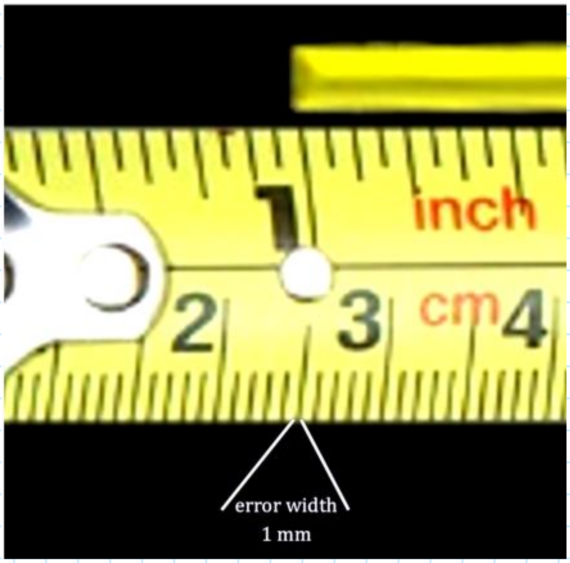
https://www.snopes.com/fact-check/misaligned-bridge-photo/

Direct Observation - Distance

Monday, March 2, 2020 9:39 AM

Direct Measurement





https://www.youtube.com/watch?v=wOjpeXIq1eI

Direct Observation - Distance

Monday, March 2, 2020 9:39 AM

Topcon GM-50

Maximum slope distance display

(Except for tracking)

Using prism or reflective sheet target	9,600.000 m (31,490 ft)
Reflectorless	1,200.000 m (3,930 ft)

(Tracking)

Using prism or reflective sheet target	1,280.000 m (4,200 ft)
Reflectorless	768.000 m (2,520ft)

Distance unit	m/ft/inch (selectable)

Accuracy (D: measurement distance; Unit: mm) (Under normal atmospheric conditions[*1])

(Using prism)[*3]

Fine measurement:	(1.5 + 2 ppm X D) mm[*9 *11]
Rapid measurement:	(5 + 2 ppm X D) mm

(Using reflective sheet target)[*4]

Fine measurement:	(2 + 2 ppm X D) mm
Rapid measurement:	(5 + 2 ppm X D) mm

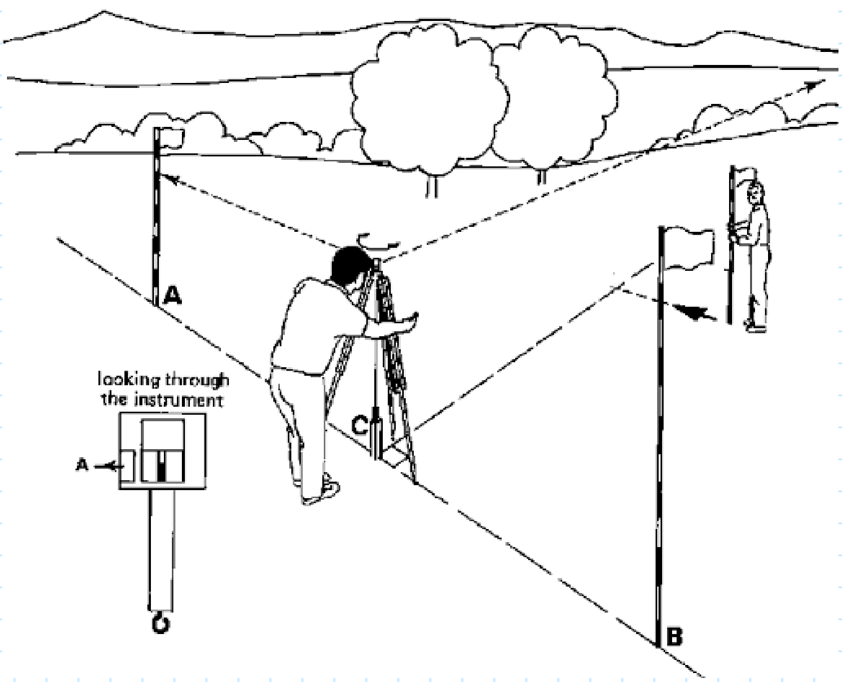
(Reflectorless (White))[*6]

Fine measurement:	(2 + 2 ppm X D) mm (0.3 to 200 m)[*10]
	(5 + 10 ppm X D) mm (over 200 to 350 m)
	(10 + 10 ppm X D) mm (over 350 to 500 m)

Direct Observation - Angle

Monday, March 2, 2020 9:39 AM

Direct Measurement



Topcon GM-50

Angle measurement

Horizontal and Vertical circles type	Rotary absolute encoder
Detecting	
GM-52:	2 sides
GM-55:	1 side
Angle units	Degree/Gon/Mil (selectable)
Minimum display	1" (0.0002 gon/0.005 mil)/5" (0.0010 gon/0.02 mil) (selectable)
Accuracy	
GM-52:	2" (0.0006 gon/0.010 mil)
GM-55:	5" (0.0015 gon/0.025 mil)
(ISO 17123-3 : 2001)	
Measuring time	0.5 sec or less
Collimation compensation	On/Off (selectable)
Measuring mode	
Horizontal angle:	Right/Left (selectable)
Vertical angle:	Zenith/Horizontal/Horizontal ±90° /% (selectable)

Direct Observation - Height

Monday, March 2, 2020 9:39 AM

mode)

Accuracy:*1 (with SOKKIA staff BGS40/50 or BIS20/30)
Height Standard deviation for 1 km of double-run

SDL30:

Electronic Measurement

0.6mm (0.03in.) (with BIS20/30)
1.0mm (0.04in.) (with BGS40/50/50G)
1.2mm (0.047in.) (with BAS55)

Visual Measurement

1.0mm (0.04in.) (with BGS40/50/50G)
1.5mm (0.06in.) (with BAS55)

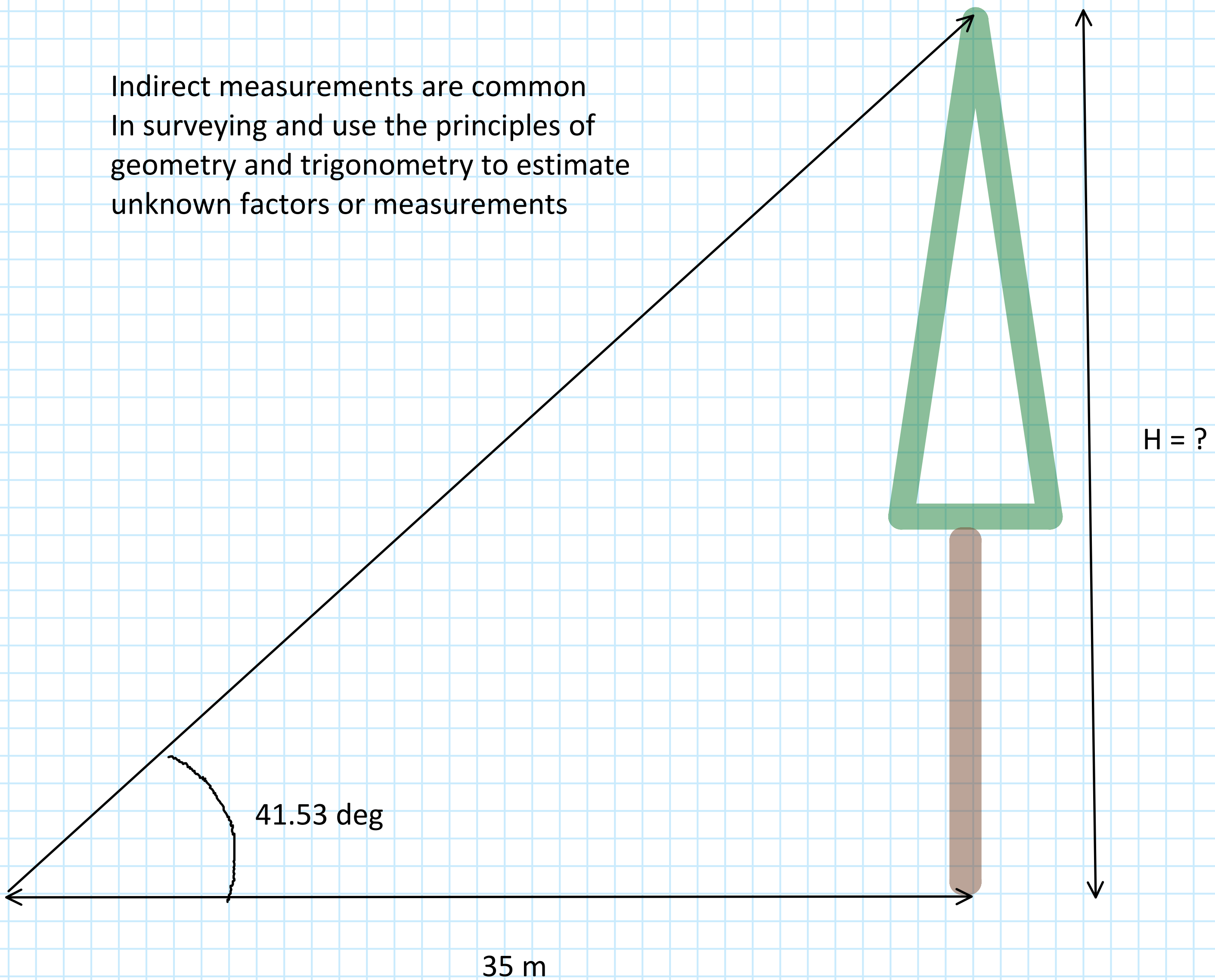
2.5mm (0.06in.) (with BAS55)

Distance ±10mm (less than 10m measurement)
±(0.1% x D) (10 to 50m measurement)
±(0.2% x D) (more than 50m measurement)
(D: measured distance, unit: m)

Indirect Measurement

Monday, March 2, 2020 9:39 AM

Indirect measurements are common
In surveying and use the principles of
geometry and trigonometry to estimate
unknown factors or measurements



Bicycle Oedometer

Saturday, January 12, 2019 1:48 PM

© Steven F. Bartlett, 2019

Indirect Measurement Exercises

Saturday, January 12, 2019 1:48 PM

1. Determine how far a bicycle will travel (km) in 3 days (27 hours). The wheel has an outside diameter of 0.66 m is rotating at 73 revolutions per minute.

2. Without using a tape measure, measure the height of tree. You cannot climb the tree but must make the height measurement (estimate) from the ground (See previous page)

0.66*3.14159*73*60*27=245,206.12604

Errors in Measurement

Monday, March 2, 2020 9:39 AM

By definition, an error is the difference between an observed value for a quantity and its true value, or

$$E = X - \overline{X} \quad (3.1)$$

where E is the error in an observation,

X the observed value,

$\overline{X}$ its true value.

It can be unconditionally stated that

(1) no observation is exact,

2) every observation contains errors,

(3) the true value of an observation is never known,

(4) the exact error present is always unknown.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/66!/4/4@14.0:15.6>

Obviously, **accuracy of observations depends on the scale's division size, reliability of equipment used, and human limitations in estimating closer than about one tenth of a scale division.** As better equipment is developed, observations more closely approach their true values, but they can never be exact. **Note that observations, not counts (of cars, pennies, marbles, or other objects), are under consideration here.**

From <https://online.vitalsource.com/#/books/9781292060675/cfi/66!/4/4@14.0:15.6>

Precision, Accuracy and Systematic Error

Monday, March 2, 2020 9:39 AM

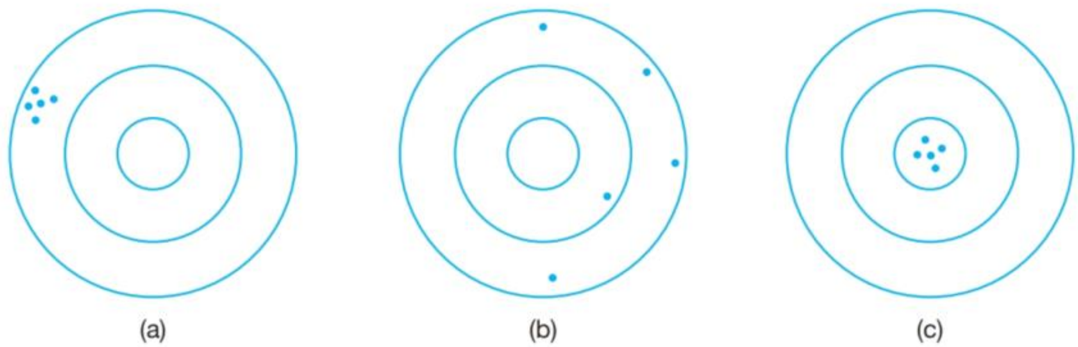
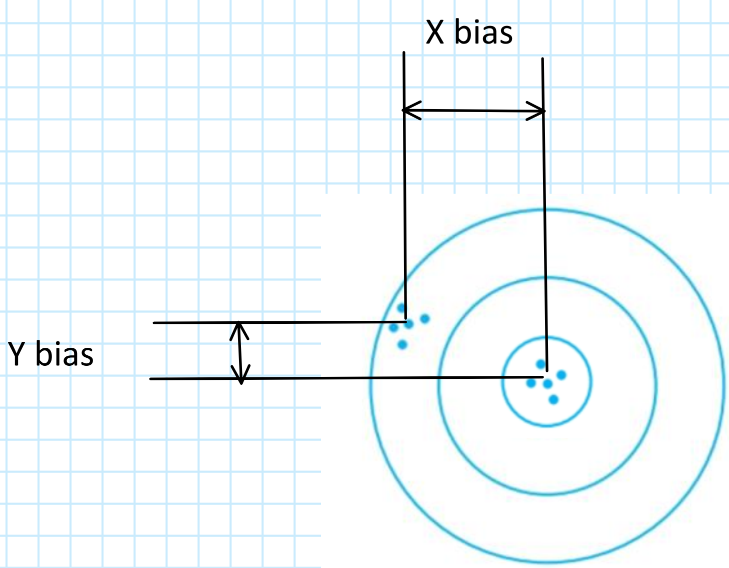


Figure 3.1 Examples of precision and accuracy. (a) Results are precise but not accurate. (b) Results are neither precise nor accurate. (c) Results are both precise and accurate.

<u>Correction for systematic bias</u>



- Note that (a) (above) has high precision but is not very accurate.

Bias and Accuracy. Accuracy is a qualitative term referring to whether there is agreement between a **measurement** made on an object and its true (target or reference) value. **Bias or systematic error** is a quantitative term describing the difference between the average of measurements made on the same object and its true value.

Accuracy

Monday, March 2, 2020 9:39 AM

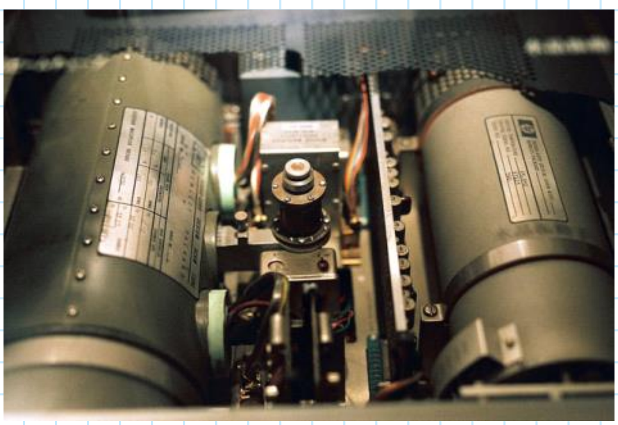
Example of High Accuracy - Atomic Clock

By measuring the oscillation of atoms, atomic clocks remain precise, but they're not perfect. They experience an error of 1 second every one-hundred million years or so.

Today, the [NIST-F1 atomic clock in Colorado]() is considered to be one of the most precise clocks in the world.

It is called a cesium fountain clock. This means the lasers – or beams – bundle the atoms into a bustling cloud, cool them off, and toss them around. What this does is slow the atoms movement so that they can be measured easier and much more precisely.

From <https://www.worldtimeserver.com/learn/how-are-atomic-clocks-so-accurate-at-keeping-time/>



https://www.worldtimeserver.com/images/content/Atomic%20Clock%20inner%20workings.jpg

How accurate is the clock?

How many seconds are in 100 million years?

60*60*24*365.25*100000000=3.15E15

Accuracy equals one part on 3.15E15

Sources of Error

Monday, March 2, 2020 9:39 AM

- **Natural errors** are caused by v**ariations in wind, temperature, humidity, atmospheric pressure, atmospheric refraction, gravity, and magnetic declination**. An example is a steel tape whose length varies with changes in temperature.

- **Instrumental errors** result from any **imperfection in the construction or adjustment of instruments and from the movement of individual parts**. For example, the graduations on a scale may not be perfectly spaced, or the scale may be warped. The effect of many instrumental errors can be reduced, or even eliminated, by adopting proper surveying procedures or applying computed corrections.

- **Personal errors** arise principally from **limitations of the human senses of sight and touch**. As an example, a small error occurs in the observed value of a horizontal angle if the vertical cross hair in a total station instrument is not aligned perfectly on the target, or if the target is the top of a rod that is being held slightly out of plumb.

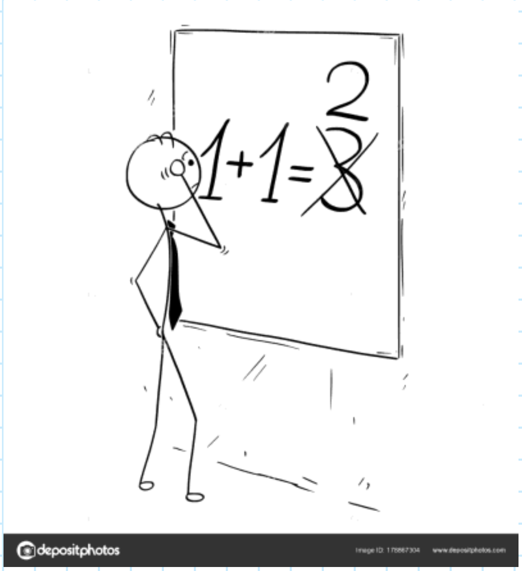
From <https://online.vitalsource.com/#/books/9781292060675/cfi/67!/4/4@0:0.00>

Types of Error

Monday, March 2, 2020 9:39 AM

Mistakes

Monday, March 2, 2020 9:39 AM



These are usually caused by **misunderstanding the problem, carelessness, fatigue, missed communication, or poor judgment**. Examples include **transposition** of numbers, such as recording 73.96 instead of the correct value of 79.36; reading an angle counterclockwise, but indicating it as a clockwise angle in the field notes; sighting the wrong target; or recording a measured distance as 682.38 instead of 862.38. **Large mistakes such as these are not considered in the succeeding discussion of errors. They must be detected by careful and systematic checking of all work, and eliminated by repeating some or all of the measurements.** It is very difficult to detect small mistakes because they merge with errors. When not exposed, these small mistakes will therefore be incorrectly treated as errors.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/66!/4/4@0:35.1>

Eliminating Mistakes and Systematic Error

Monday, March 2, 2020 9:39 AM

All field operations and office computations are g**overned by a constant effort to eliminate mistakes and systematic errors**. Of course it would be preferable if mistakes never occurred, but because humans are fallible, this is not possible. In the field, experienced observers who alertly perform their observations using standardized repetitive procedures can minimize mistakes. **Mistakes that do occur can be corrected only if discovered**. Comparing several observations of the same quantity is one of the best ways to identify mistakes. Making a common sense estimate and analysis is another. A**ssume that five observations of a line are recorded as follows: 567.91, 576.95, 567.88, 567.90, and 567.93. The second value disagrees with the others**, apparently because of a transposition of figures in reading or recording. Either casting out the doubtful value or preferably repeating the observation can eradicate this mistake. **When a mistake is detected, it is usually best to repeat the observation**. **However, if a sufficient number of other observations of the quantity are available and in agreement, as in the foregoing example, the widely divergent result may be discarded.** Serious consideration must be given to the effect on an average before discarding a value. **It is seldom safe to change a recorded number, even though there appears to be a simple transposition in figures. Tampering with physical data is always a bad practice and will certainly cause trouble, even if done infrequently.**

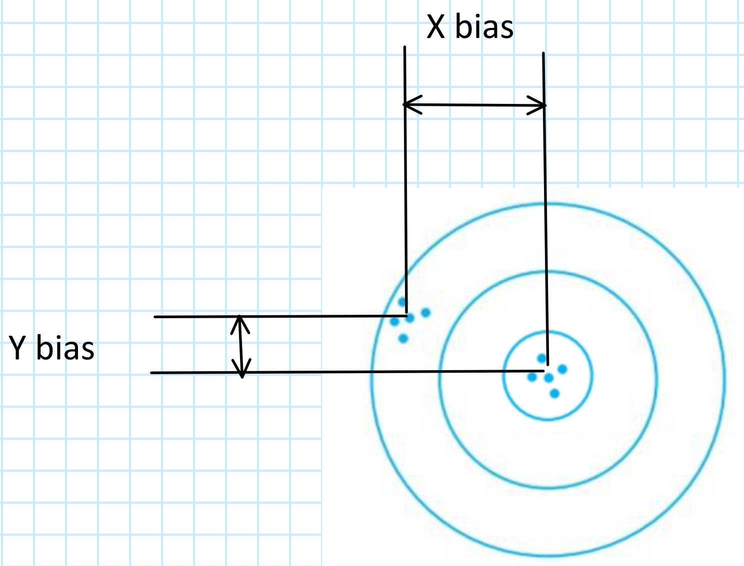
Systematic errors can be calculated and proper corrections applied to the observations. Procedures for making these corrections to all basic surveying observations are described in the chapters that follow. In some instances, it may be possible to adopt a field procedure that automatically eliminates systematic errors. For example, as explained in **Chapter 5, a leveling instrument out of adjustment causes incorrect readings, but if all backsights and foresights are made the same length, the errors cancel in differential leveling**.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/69!/4/4@0:19.5>

Systematic Error

Monday, March 2, 2020 9:39 AM

<u>Correction for systematic bias</u>



- Note that (a) (above) has high precision but is not very accurate.

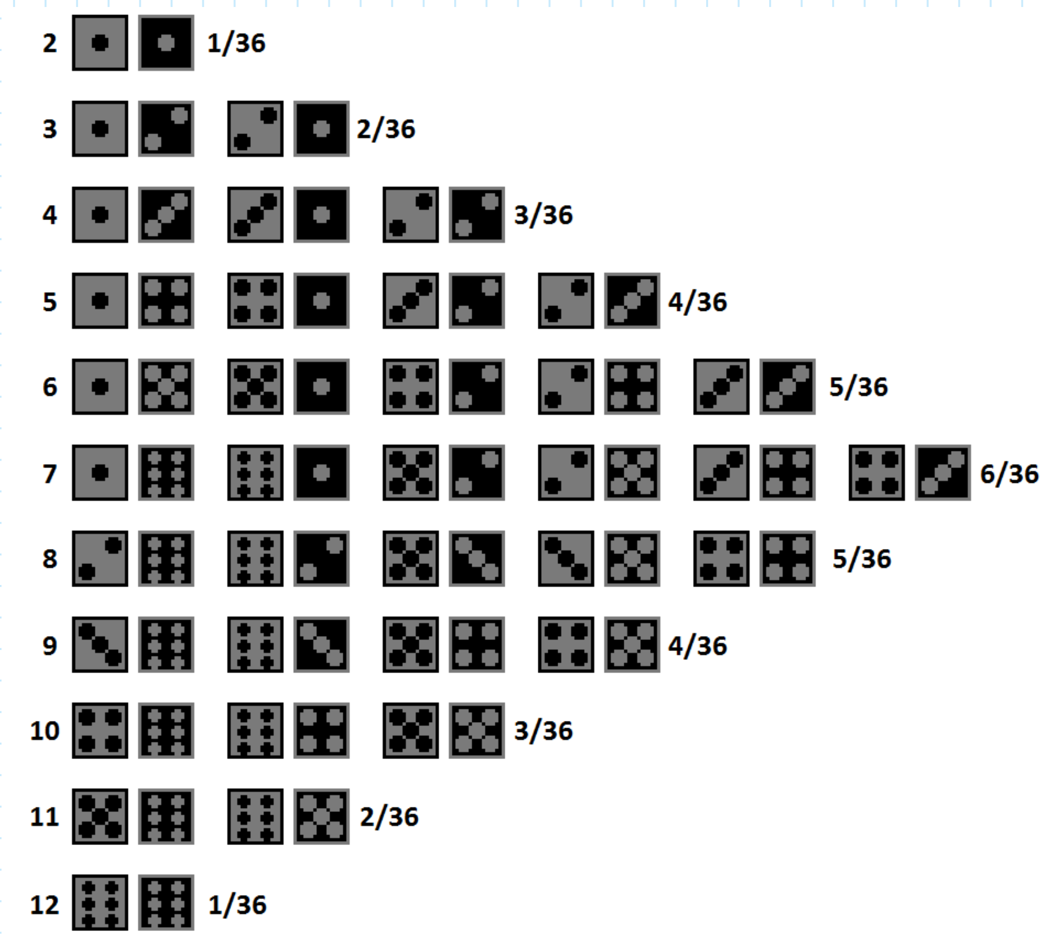
Bias and Accuracy. Accuracy is a qualitative term referring to whether there is agreement between a **measurement** made on an object and its true (target or reference) value. **Bias or systematic error** is a quantitative term describing the difference between the average of measurements made on the same object and its true value.

Probability

Monday, March 2, 2020 9:39 AM

Probability is a numerical description of how likely an event is to occur or how likely it is that a proposition is true. Probability is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility and 1 indicates certainty.[note 1][1][2] The higher the probability of an event, the more likely it is that the event will occur. A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes ("heads" and "tails") are both equally probable; the probability of "heads" equals the probability of "tails"; and since no other outcomes are possible, the probability of either "heads" or "tails" is 1/2 (which could also be written as 0.5 or 50%).

From <https://en.wikipedia.org/wiki/Probability>



The probabilities of rolling several numbers using two dice

From <https://en.wikipedia.org/wiki/Probability#/media/File:Dice_Probability.png>

Probability (cont.)

Saturday, January 12, 2019 1:48 PM

Independent events [edit]

If two events, A and B are independent then the joint probability is

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B),$$

for example, if two coins are flipped the chance of both being heads is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.[31]

Mutually exclusive events [edit]

If either event A or event B but never both occurs on a single performance of an experiment, then they are called mutually exclusive events.

If two events are mutually exclusive then the probability of **both** occurring is denoted as $P(A \cap B)$.

$$P(A \text{ and } B) = P(A \cap B) = 0$$

If two events are mutually exclusive then the probability of **either** occurring is denoted as $P(A \cup B)$.

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - 0 = P(A) + P(B)$$

For example, the chance of rolling a 1 or 2 on a six-sided die is $P(1 \text{ or } 2) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Not mutually exclusive events [edit]

If the events are not mutually exclusive then

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \text{ and } B).$$

For example, when drawing a single card at random from a regular deck of cards, the chance of getting a heart or a face card (J,Q,K) (or one that is both) is $\frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{11}{26}$, because of the 52 cards of a deck 13 are hearts, 12 are face cards, and 3 are both: here the possibilities included in the "3 that are both" are included in each of the "13 hearts" and the "12 face cards" but should only be counted once.

https://en.wikipedia.org/wiki/Probability

Summary of probabilities

Event	Probability	
A	$P(A) \in [0,1]$	
not A	$P(A^\complement) = 1 - P(A)$	
A or B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cup B) = P(A) + P(B) \qquad \text{if A and B are mutually exclusive}$	Addition Rule
A and B	$P(A \cap B) = P(A\|B)P(B) = P(B\|A)P(A)$ $P(A \cap B) = P(A)P(B) \qquad \text{if A and B are independent}$	Multiplication Rule
A given B	$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B\|A)P(A)}{P(B)}$	Bayes Theorem

For more on this topic, see

https://www.youtube.com/watch?v=94AmzeR9n2w Addition and Multiplication Rule

https://www.youtube.com/watch?v=OByl4RJxnKA Bayes Theorem

General Laws of Probability

Monday, March 2, 2020 9:39 AM

1. Small residuals (errors) occur more often than large ones; that is, they are more probable.

2. Large errors happen infrequently and are therefore less probable; for normally distributed errors, unusually large ones may be mistakes rather than random errors.

3. Positive and negative errors of the same size happen with equal frequency; that is, they are equally probable. [This enables an intuitive deduction of Equation (3.2) to be made: that is, the most probable value for a group of repeated observations, made with the same equipment and procedures, is the mean.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/75!/4/4@0:0.00>

Most Probable Value

Monday, March 2, 2020 9:39 AM

It has been stated earlier that in physical observations, the true value of any quantity is never known. However, its most probable value can be calculated if redundant observations have been made. Redundant observations are measurements in excess of the minimum needed to determine a quantity. For a single unknown, such as the length of a line that has been directly and independently observed a number of times using the same equipment and procedures,1 the first observation establishes a value for the quantity and all additional observations are redundant.

The most probable value in this case is simply the arithmetic mean, or

$$\overline{M} = \Sigma M / n$$

where $\overline{M}$ is the most probable value of the quantity,

ΣM the sum of the individual measurements M,

and n the total number of observations.

Equation (3.2) can be derived using the principle of least squares, which is based on the theory of probability.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/70!/4/4@10.0:33.5>

Residuals

Monday, March 2, 2020 9:39 AM

Having determined the most probable value of a quantity, it is possible to calculate residuals. A residual is simply the difference between the most probable value and any observed value of a quantity, which in equation form is

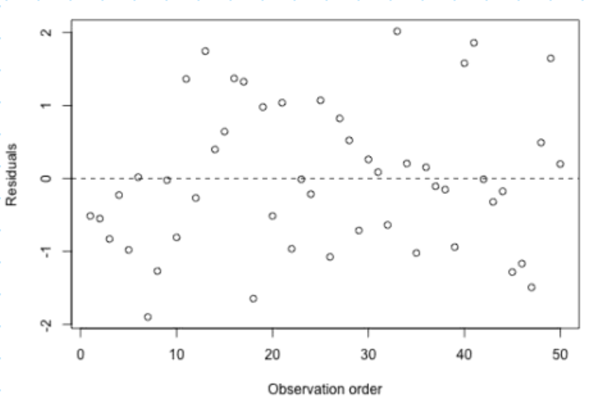
$$\nu = \overline{M} - M$$

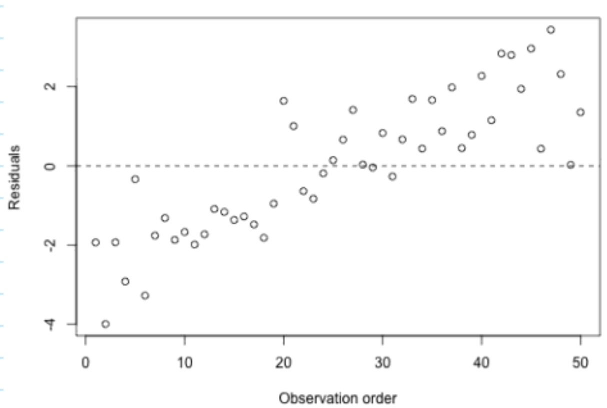
where v is the residual in any observation M, and

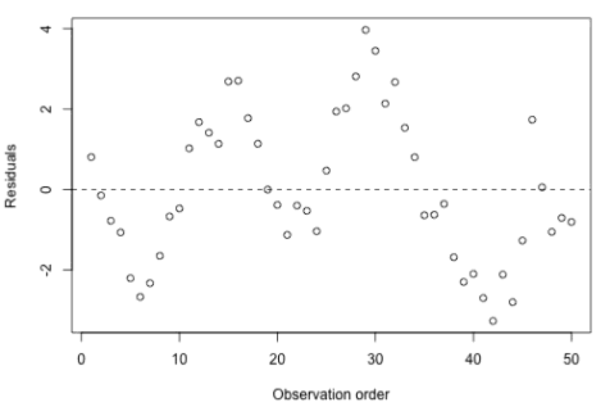
M is the most probable value for the quantity.

Residuals are theoretically identical to errors, with the exception that residuals can be calculated whereas errors cannot because true values are never known. Thus, residuals rather than errors are the values actually used in the analysis and adjustment of survey data.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/71!/4/4@0:0.00>







https://online.stat.psu.edu/stat462/node/121/

Occurrence of Random Errors

Monday, March 2, 2020 9:39 AM

To analyze the manner in which random errors occur, consider the data of Table 3.1, **which represents 100 repetitions of an angle observation made with a precise total station instrument.** Assume these observations are f**ree from mistakes and systematic errors**.

From <https://online.vitalsource.com/#/books/9781292060675/cfi/71!/4/4@0:35.3>

TABLE 3.1 ANGLE OBSERVATIONS FROM PRECISE TOTAL STATION INSTRUMENT

Observed Value (1)	No. (2)	Residual (Sec) (3)	Observed Value (1 Cont.)	No. (2. Cont.)	Residual (Sec) (3 Cont.)
27°43′19.5″	1	5.4	27°43′25.1″	3	−0.2
20.0	1	4.9	25.2	1	−0.3
20.5	1	4.4	25.4	1	−0.5
20.8	1	4.1	25.5	2	−0.6
21.2	1	3.7	25.7	3	−0.8
21.3	1	3.6	25.8	4	−0.9
21.5	1	3.4	25.9	2	−1.0
22.1	2	2.8	26.1	1	−1.2
22.3	1	2.6	26.2	2	−1.3
22.4	1	2.5	26.3	1	−1.4
22.5	2	2.4	26.5	1	−1.6
22.6	1	2.3	26.6	3	−1.7
22.8	2	2.1	26.7	1	−1.8
23.0	1	1.9	26.8	2	−1.9
23.1	2	1.8	26.9	1	−2.0
23.2	2	1.7	27.0	1	−2.1
23.3	3	1.6	27.1	3	−2.2
23.6	2	1.3	27.4	1	−2.5
23.7	2	1.2	27.5	2	−2.6
23.8	2	1.1	27.6	1	−2.7
23.9	3	1.0	27.7	2	−2.8
24.0	5	0.9	28.0	1	−3.1
24.1	3	0.8	28.6	2	−3.7
24.3	1	0.6	28.7	1	−3.8
24.5	2	0.4	29.0	1	−4.1
24.7	3	0.2	29.4	1	−4.5
24.8	3	0.1	29.7	1	−4.8
24.9	2	0.0	30.8	1	−5.9
25.0	2	−0.1		Σ = 100	

Occurrence of Random Errors (cont.)

Monday, March 2, 2020 9:39 AM

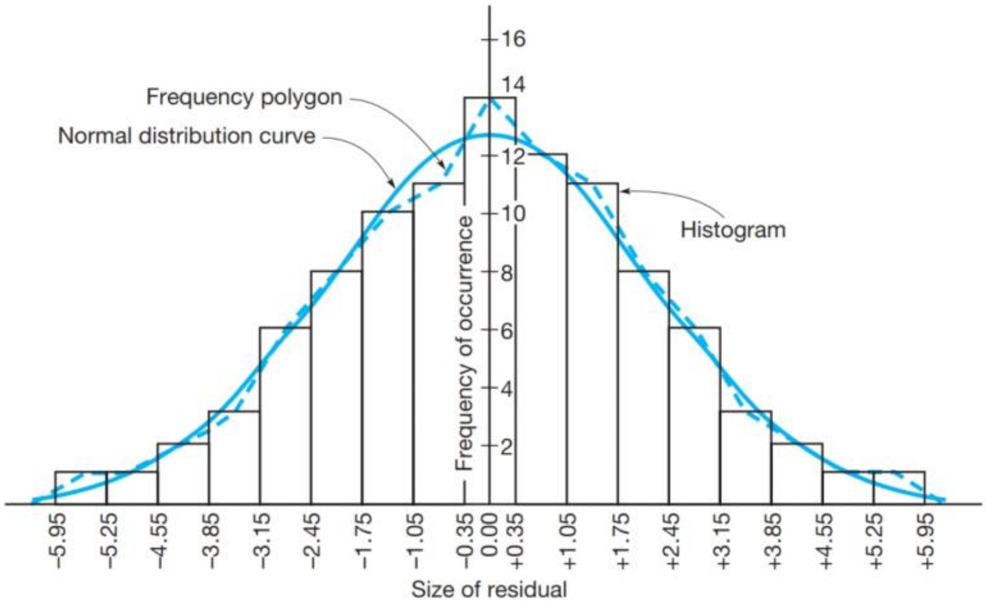


Figure 3.2 Histogram, frequency polygon, and normal distribution curve of residuals from angle measurements made with total station.

- Relative Frequency Histogram
  - If residual error is normally distributed, then a normal distribution curve can be fitted to the data (blue line)
  - In surveying, normal or nearly normally distributed errors is expected
  - Therefore normal distribution curves are used to describe the error.

Excel - Descriptive Statistics

Wednesday, March 11, 2020 3:43 AM

II. Measures of Centrality (Mean, Median, Mode)

A. Sample Mean, $\bar{x}$

The mean is a weighted sum of all possible values, with each value receiving the same weight. It is calculated from:

$$\bar{x} = (\Sigma x) / n$$ Excel Function (=AVERAGE)

where **Σ x** is sum of the all the observations and n is the number of observations (i.e., sample size).

What is the mean for the survey data strength data?

B. Median Excel Function (=MEDIAN)

The median is the value that exactly divides the cumulative frequency distribution into two equal halves.

What is the median for the survey data?

C. Mode EXCEL Function (MODE.SNG)

The mode is any <u>local</u> maximum of the frequency or relative frequency diagram. Thus, it is possible for a sample to have more than one mode.

What is the mode for the survey data?

© Steven F. Bartlett, 2013

Excel - Relative Frequency Histogram

Wednesday, March 11, 2020 3:43 AM

1. Divide the residuals into bins or classes.
2. Count the number (frequency) of residuals in each bin.
3. Plot the frequency vs the bins to using a bar chart

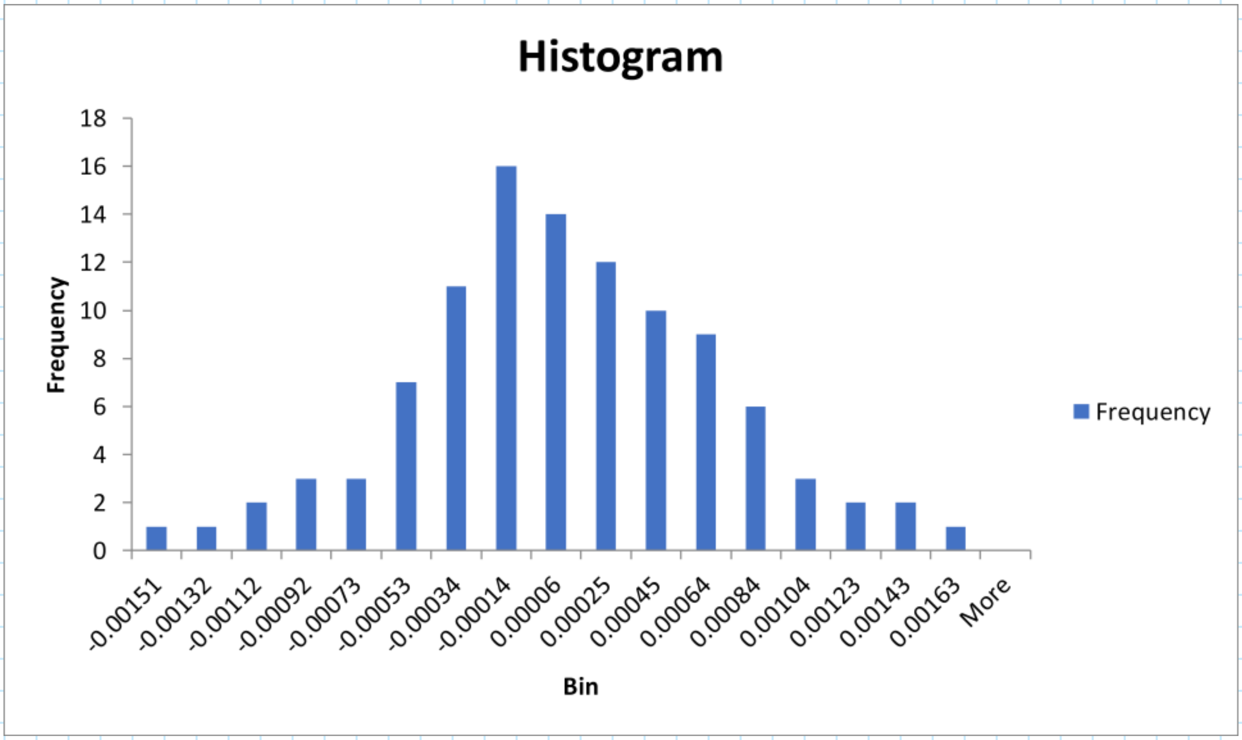
No. of Bins	Low	High	bin increment
17	-0.00151	0.00163	0.00020
	Upper Value		
1	-0.00151	*Bin*	*Frequency*
2	-0.00132	-0.00151	1
3	-0.00112	-0.00132	1
4	-0.00092	-0.00112	2
5	-0.00073	-0.00092	3
6	-0.00053	-0.00073	3
7	-0.00034	-0.00053	7
8	-0.00014	-0.00034	11
9	0.00006	-0.00014	16
10	0.00025	0.00006	14
11	0.00045	0.00025	12
12	0.00064	0.00045	10
13	0.00084	0.00064	9
14	0.00104	0.00084	6
15	0.00123	0.00104	3
16	0.00143	0.00123	2
17	0.00163	0.00143	2
		0.00163	1
		More	0

© Steven F. Bartlett, 2013

Excel - Relative Frequency Histogram (cont.)

Wednesday, March 11, 2020 3:43 AM

1. Divide the residuals into bins or classes.
2. Count the number (frequency) of residuals in each bin.
3. Plot the frequency vs the bins to using the chart option



The histogram above give us an idea about the uncertainty in the measurements. This uncertainty can be called "scatter" or "dispersion." Its formal statistical name is variance, s^2.

$$s^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}$$

The standard deviation is simply the square root of the variance

s = SQRT (s^2) EXCEL FUNCTION (=STDEV)

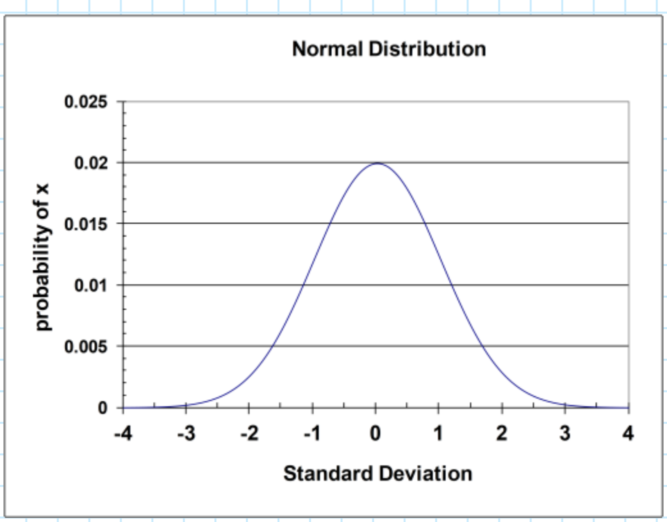
© Steven F. Bartlett, 2013

III. Continuous Probability Distribution Functions

We have introduced histograms as a way to show the distribution of data. Because the data in histograms can be divided in discrete classes or bins, these functions are *discrete distribution functions*. However, if we had numerous observations and could divide the bins into very small intervals (i.e., essentially a point), then the discrete frequency distribution function would begin to define a *continuous distribution function* (i.e., a smooth curve). Furthermore, if we could express the continuous distribution function as a *relative frequency distribution function* with the values of relative frequency function summing to 1 instead of 100 percent (i.e., the integral of the function equals 1), then we would define what is called a **continuous probability distribution function**.

A. Standard Normal Distribution

The most important and widely used continuous probability density function is the Gaussian, or normal distribution. This distribution is best recognized as a "bell-shaped" curve. Many things in the physical world are normally distributed or can be approximated by the normal distribution. This curve is also known as a "probability density function." One property of the standard normal curve (i.e., probability density function) is that the area under the curve sums to exactly 1. That is to say that the probability of z occurring between a z-score equal to negative infinity and a z-score equal to positive infinity is 1. Mathematically this is stated as: **$p(-\infty < z < \infty) = 1$**.



© Steven F. Bartlett, 2013

Excel - Standard Normal Distribution and z-score

Thursday, March 11, 2010 11:43 AM

Note that for the standard normal distribution the x-axis is plotted in terms of standard deviations or z-score and the y-axis is the probability of that z-score.

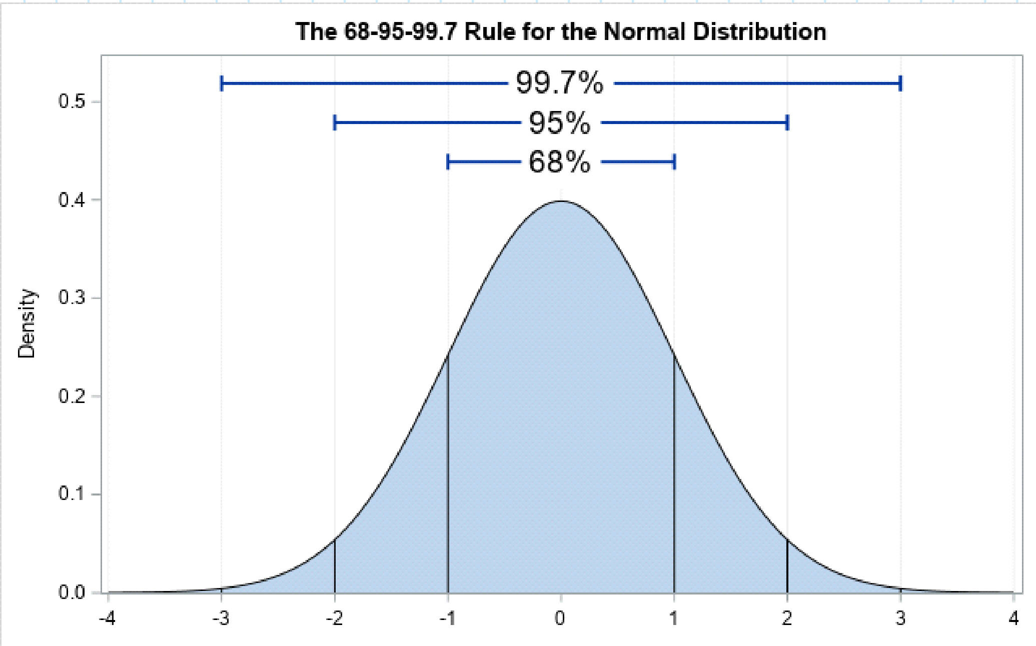
The z-score is calculated from:

$$z = (x - \bar{x}) / s$$

where x = value of interest, $\bar{x}$ is the sample mean (which is an estimate of the population mean) and s is the sample standard deviation (which is an estimate of the population standard deviation).

Note also that if x is equal to the sample mean, then z equals zero. A z-score of 1 means that the value of interest is located exactly 1 standard deviation above the mean. A z-score of -1 means that the value of interest is exactly 1 standard deviation below the mean.

If data are normally distributed then approximately 68 percent or two-thirds of the data are found between ± 1 standard deviation ($-1 < z < 1$) and 95 percent of the data are found between ± 2 standard deviations ($-2 < z < 2$) and 99.7% are found between ± 3 standard deviations

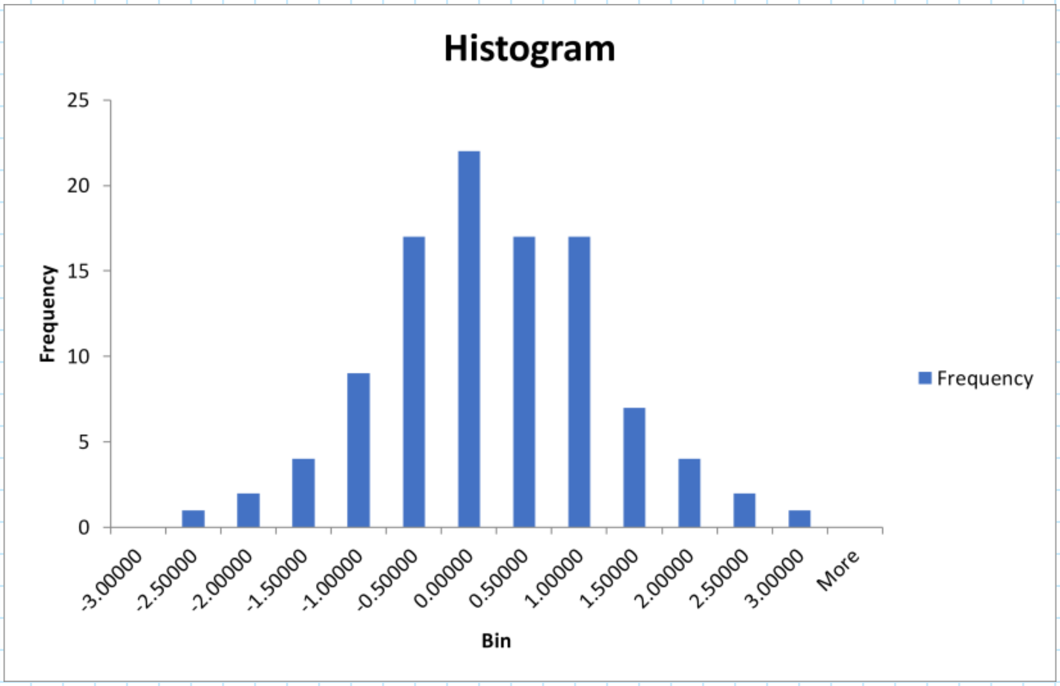


© Steven F. Bartlett, 2020

Excel - Standard Normal Distribution and z-score (cont).

Thursday, March 11, 2010 11:43 AM

No. of Bins	Low	High	bin increment
13	-3.00000	3.00000	0.50000
	Upper Value	*Bin*	*Frequency*
1	-3.00000	-3.00000	0
2	-2.50000	-2.50000	1
3	-2.00000	-2.00000	2
4	-1.50000	-1.50000	4
5	-1.00000	-1.00000	9
6	-0.50000	-0.50000	17
7	0.00000	0.00000	22
8	0.50000	0.50000	17
9	1.00000	1.00000	17
10	1.50000	1.50000	7
11	2.00000	2.00000	4
12	2.50000	2.50000	2
13	3.00000	3.00000	1



© Steven F. Bartlett, 2020

Confidence Intervals from Normal Distribution

Wednesday, March 11, 2020 3:43 AM

© Steven F. Bartlett, 2013

Error Propagation - Error of a Sum

Monday, March 2, 2020 9:39 AM

3.17 ERROR PROPAGATION

It was stated earlier that because all observations contain errors, any quantities computed from them will likewise contain errors. The process of evaluating errors in quantities computed from observed values that contain errors is called *error propagation*. The propagation of random errors in mathematical formulas can be computed using the general law of the propagation of variances. Typically in surveying (geomatics), this formula can be simplified since the observations are usually mathematically independent. For example, let $a, b, c, \ldots, n$ be observed values containing errors $E_a, E_b, E_c, \ldots, E_n$, respectively. Also let Z be a quantity derived by computation using these observed quantities in a function f, such that

$$Z = f(a, b, c, \ldots, n) \tag{3.9}$$

Then assuming that $a, b, c, \ldots, n$ are independent observations, the error in the computed quantity Z is

$$E_Z = \pm\sqrt{\left(\frac{\partial f}{\partial a}E_a\right)^2 + \left(\frac{\partial f}{\partial b}E_b\right)^2 + \left(\frac{\partial f}{\partial c}E_c\right)^2 + \cdots + \left(\frac{\partial f}{\partial n}E_n\right)^2} \tag{3.10}$$

where the terms $\partial f/\partial a, \partial f/\partial b, \partial f/\partial c, \ldots, \partial f/\partial n$ are the partial derivatives of the function f with respect to the variables $a, b, c, \ldots, n$. In the subsections that follow, specific cases of error propagation common in surveying are discussed, and examples are presented.

The partial derivatives of Z with respect to each observed quantity are $\partial Z/\partial a = \partial Z/\partial b = \partial Z/\partial c = \cdots = 1$. Substituting these partial derivatives into Equation (3.10), the following formula is obtained, which gives the propagated error in the sum of quantities, each of which contains a different random error:

$$E_{Sum} = \pm\sqrt{E_a^2 + E_b^2 + E_c^2 + \cdots} \tag{3.11}$$

where E represents any specified percentage error (such as σ, E_{50}, E_{90}, or E_{95}), and a, b, and c are the separate, independent observations.

Example 3.2

Assume that a line is observed in three sections, with the individual parts equal to (753.81, ±0.012), (1238.40, ±0.028), and (1062.95, ±0.020) ft, respectively. Determine the line's total length and its anticipated standard deviation.

Solution

Total length = 753.81 + 1238.40 + 1062.95 = 3055.16 ft.

By Equation (3.11), $E_{Sum} = \pm\sqrt{0.012^2 + 0.028^2 + 0.020^2} = \pm 0.036$ ft

Error Propagation - Error of a Series

Monday, March 2, 2020 9:39 AM

3.17.2 Error of a Series

Sometimes a series of similar quantities, such as the angles within a closed polygon, are read with each observation being in error by about the same amount. The total error in the sum of all observed quantities of such a series is called the *error of the series*, designated as E_{Series}. If the same error E in each observation is assumed and Equation (3.11) applied, the series error is

$$E_{Series} = \pm\sqrt{E^2 + E^2 + E^2 + \cdots} = \pm\sqrt{nE^2} = \pm E\sqrt{n} \qquad (3.12)$$

where E represents the error in each individual observation and n the number of observations.

This equation shows that when the same operation is repeated, random errors tend to balance out, and the resulting error of a series is proportional to the square root of the number of observations. This equation has extensive

Example 3.3

Assume that each of the interior angles in a four-sided traverse has an estimate error of $\pm 3.5''$. Determine the error in the sum of the four interior angles.

Solution

By Equation (3.12), the error in the sum of the angles is

$$E_{Series} = \pm E\sqrt{n} = \pm 3.5''\sqrt{4} = \pm 7''$$

Example 3.4

The error in sum of the interior angles of a quadrilateral must be within $\pm 10''$. Determine how accurately each of the four angles must be observed to ensure that the error will not exceed the permissible limit.

Solution

Since by Equation (3.12), $E_{Series} = \pm E\sqrt{n}$ and $n = 4$, the allowable error E in each angle is

$$E = \pm\frac{E_{Series}}{\sqrt{n}} \pm \frac{10''}{\sqrt{4}} = \pm 5''$$

3.17.3 Error of a Products

The equation for propagation of errors in the product AB, where E_a and E_b, are the respective errors in A and B, is

$$E_{prod} = \pm\sqrt{A^2E_b^2 + B^2E_a^2} \tag{3.13}$$

The physical significance of the error propagation formula for a product is illustrated in Figure 3.6, where A and B are shown to be observed sides of a rectangular parcel of land with errors E_a and E_b, respectively. The product AB is the parcel area. In Equation (3.13), $\sqrt{A^2E_b^2} = AE_b$, represents area within either of the longer (horizontal) crosshatched bars and is the error caused by either $-E_b$ or $+E_b$. The term $\sqrt{B^2E_a^2} = BE_a$ is represented by the area within the shorter (vertical) crosshatched bars, which is the error resulting from either $-E_a$ or $+E_a$.

Example 3.6

For the rectangular lot illustrated in Figure 3.6, observations of sides A and B with their 95% errors are (252.46, ±0.053) and (605.08, ±0.072) ft, respectively. Calculate the parcel area and the estimated error in the area.

Solution

$$\text{Area} = 252.46 \times 605.08 = 152{,}760 \text{ ft}^2$$

By Equation (3.13),

$$\sigma = \pm\sqrt{(252.46)^2(0.072)^2 + (605.08)^2(0.053)^2} = \pm 36.9 \text{ ft}^2$$

Example 3.6 can also be used to demonstrate the validity of one of the rules of significant figures in computation. The computed area is actually 152,758.4968 ft^2. However, the rule for significant figures in multiplication (see Section 2.4) states that there cannot be more significant figures in

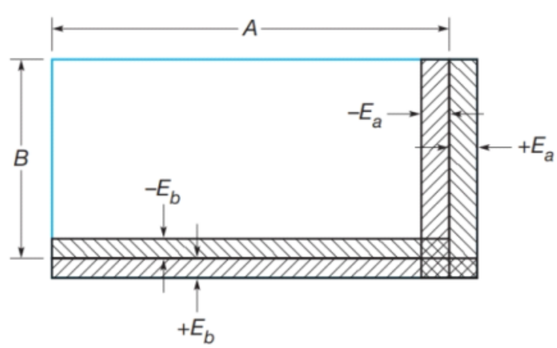


Figure 3.6
Error of area.

3.17.4 Error of the Mean

Equation (3.2) stated that the most probable value of a group of repeated observations of equal weight is the arithmetic mean. Since the mean is computed from individual observed values, each of which contains an error, the mean is also subject to error. By applying Equation (3.12), it is possible to find the error for the sum of a series of observations where each one has the same error. Since the sum divided by the number of observations gives the mean, the error of the mean is found by the relation

$$E_m = \frac{E_{series}}{n}$$

Substituting Equation (3.12) for E_{series}

$$E_m = \frac{E\sqrt{n}}{n} = \frac{E}{\sqrt{n}} \tag{3.14}$$

where E is the specified percentage error of a single observation, E_m the corresponding percentage error of the mean, and n the number of observations.

The error of the mean at any percentage probability can be determined and applied to all criteria that have been developed. For example, the standard deviation of the mean, $(E_{68})_m$ or σ_m is

$$(E_{68})_m = \sigma_m = \frac{\sigma}{\sqrt{n}} = \pm\sqrt{\frac{\Sigma v^2}{n(n-1)}} \tag{3.15a}$$

and the 90% and 95% errors of the mean are

$$(E_{90})_m = \frac{E_{90}}{\sqrt{n}} = \pm 1.6449\sqrt{\frac{\Sigma v^2}{n(n-1)}} \tag{3.15b}$$

$$(E_{95})_m = \frac{E_{95}}{\sqrt{n}} = \pm 1.9599\sqrt{\frac{\Sigma v^2}{n(n-1)}}. \tag{3.15c}$$

These equations show that *the error of the mean varies inversely as the square root of the number of repetitions*. Thus, to double the accuracy—that is, to reduce the error by one half—four times as many observations must be made.

Conditional Adjustment

Monday, March 2, 2020 9:39 AM

Weights of Observations

Monday, March 2, 2020 9:39 AM

3.20 WEIGHTS OF OBSERVATIONS

It is evident that some observations are more precise than others because of better equipment, improved techniques, and superior field conditions. In making adjustments, it is consequently desirable to assign *relative weights* to individual observations. It can logically be concluded that if an observation is very precise, it will have a small standard deviation or variance, and thus should be weighted more heavily (held closer to its observed value) in an adjustment than an observation of lower precision. From this reasoning, it is deduced that weights of observations should bear an inverse relationship to precision. In fact, it can be shown that relative weights are inversely proportional to variances, or

$$W_a \propto \frac{1}{\sigma_a^2} \tag{3.16}$$

where W_a is the weight of an observation a, which has a variance of σ_a^2. Thus, the higher the precision (the smaller the variance), the larger should be the relative weight of the observed value being adjusted. In some cases, variances are unknown originally, and weights must be assigned to observed values based on estimates of their relative precision. If a quantity is observed repeatedly and the individual observations have varying weights, the weighted mean can be computed from the expression

$$\overline{M}_W = \frac{\Sigma WM}{\Sigma W} \tag{3.17}$$

where $\overline{M}_W$ is the weighted mean, ΣWM the sum of the individual weights times their corresponding observations, and ΣW the sum of the weights.

Example 3.8

Suppose four observations of a distance are recorded as 482.16, 482.17, 482.20, and 482.18 and given weights of 1, 2, 2, and 4, respectively, by the surveyor. Determine the weighted mean.

Solution

By Equation (3.17)

$$\overline{M}_W = \frac{482.16 + 482.17(2) + 482.20(2) + 482.14(4)}{1 + 2 + 2 + 4} = 482.16 \text{ ft}$$

Least Squares Adjustment - Regression Analysis

Monday, March 2, 2020 9:39 AM

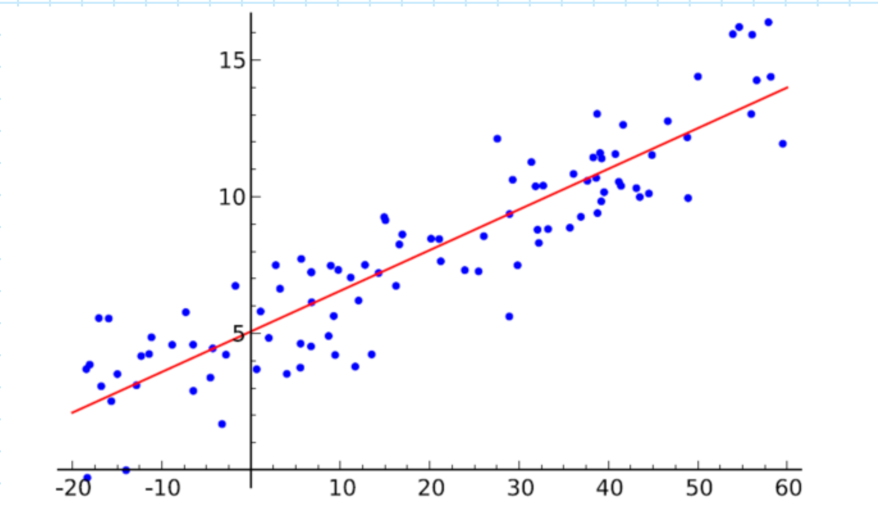
Excel - Linear Regression

Thursday, March 11, 2010 11:43 AM

In statistics, **simple linear regression** is the least squares estimator of a linear regression model with a single explanatory variable. In other words, simple linear regression fits a straight line through the set of *n* points in such a way that makes the sum of squared *residuals* of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible.
The adjective *simple* refers to the fact that this regression is one of the simplest in statistics. The slope of the fitted line is equal to the correlation between *y* and *x*corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that it passes through the center of mass (*x*, *y*) of the data points.

Other regression methods besides the simple ordinary least squares (OLS) also exist (see linear regression model). In particular, when one wants to do regression by eye, people usually tend to draw a slightly steeper line, closer to the one produced by the total least squares method. This occurs because it is more natural for one's mind to consider the orthogonal distances from the observations to the regression line, rather than the vertical ones as OLS method does

Pasted from <http://en.wikipedia.org/wiki/Simple_linear_regression>



Data showing a linear trend or correlation between x and y

© Steven F. Bartlett, 2013

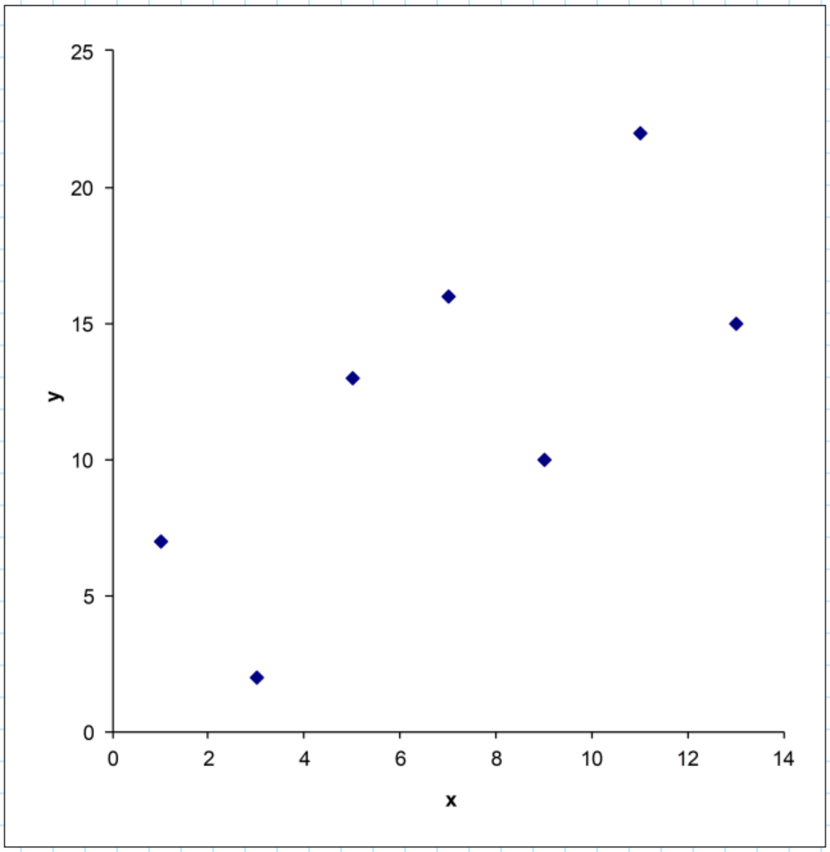
Excel - Trendline Feature to Fit Data

Thursday, March 11, 2010 11:43 AM

x	y
1	7
7	16
11	22
5	13
9	10
13	15
3	2

1. Initially the data must be in two adjacent columns are shown here. The x variable is called the independent variable and the y variable is called the dependent variable. In statistical terms, we want to use the independent variable to predict the dependent variable, assuming that some relationship or correlation exists.

2. Plot the data x versus y to see if a relationship exits, see below.

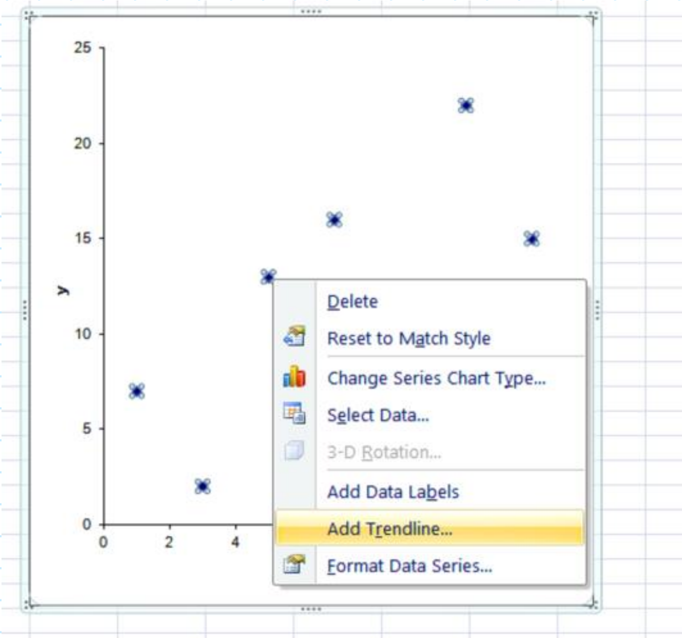


© Steven F. Bartlett, 2013

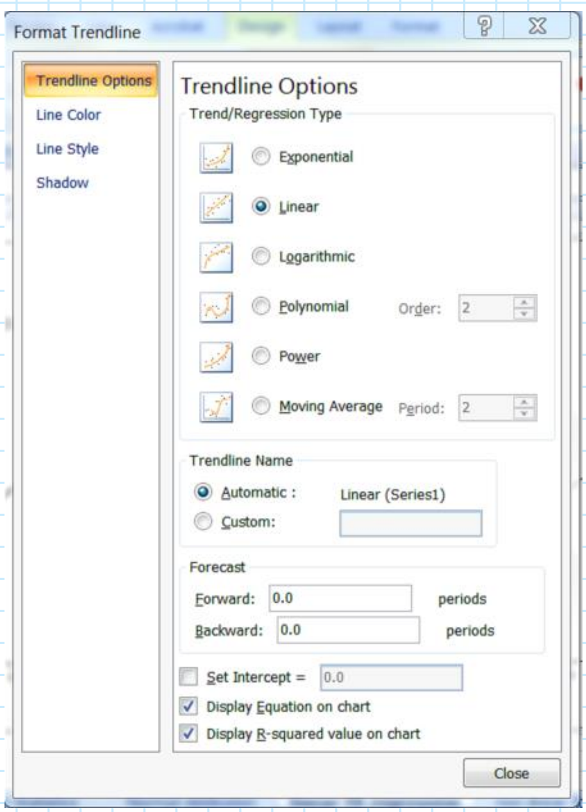
Excel - Trendline (cont.)

Thursday, March 11, 2010 11:43 AM

3. On the data plot, select the data series and right click to bring up the menu shown below.



4. Select Add Trendline, which will bring up the following menu. Select the option shown in the menu below.

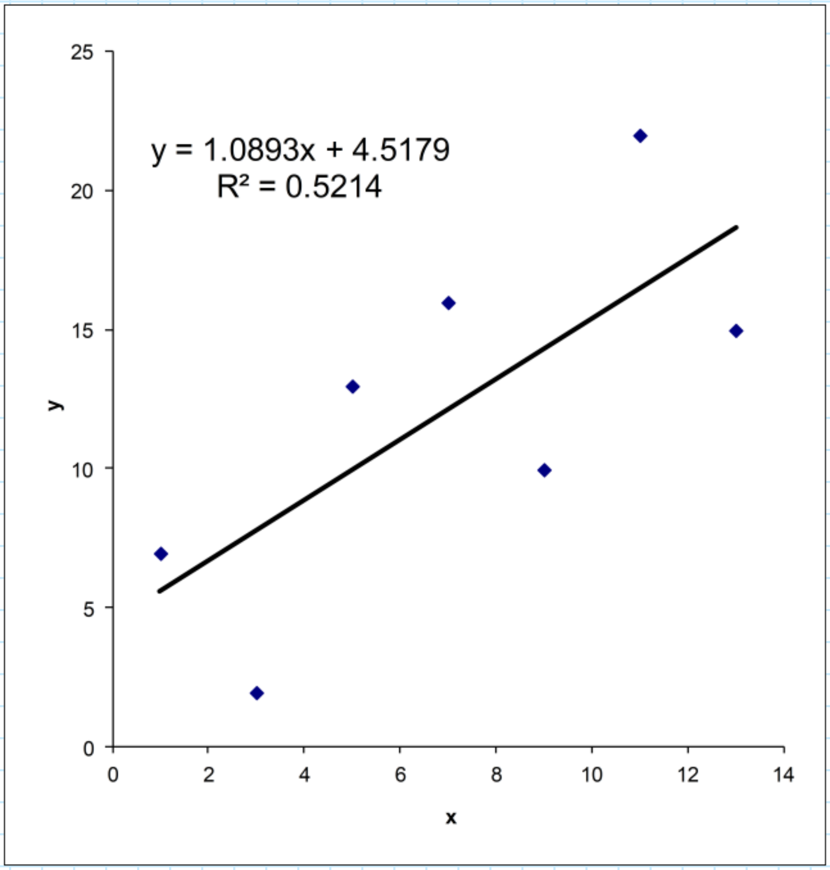


© Steven F. Bartlett, 2013

Excel - Trendline (cont.)

Thursday, March 11, 2010 11:43 AM

Your finished plot should look like this:



The intercept for this best fit line is 4.5179 and the slope of the line is 1.0893.

The R^2 value is called the coefficient of determination and 0.5214 means that 52.14 percent of the variability in y is being explained by x.

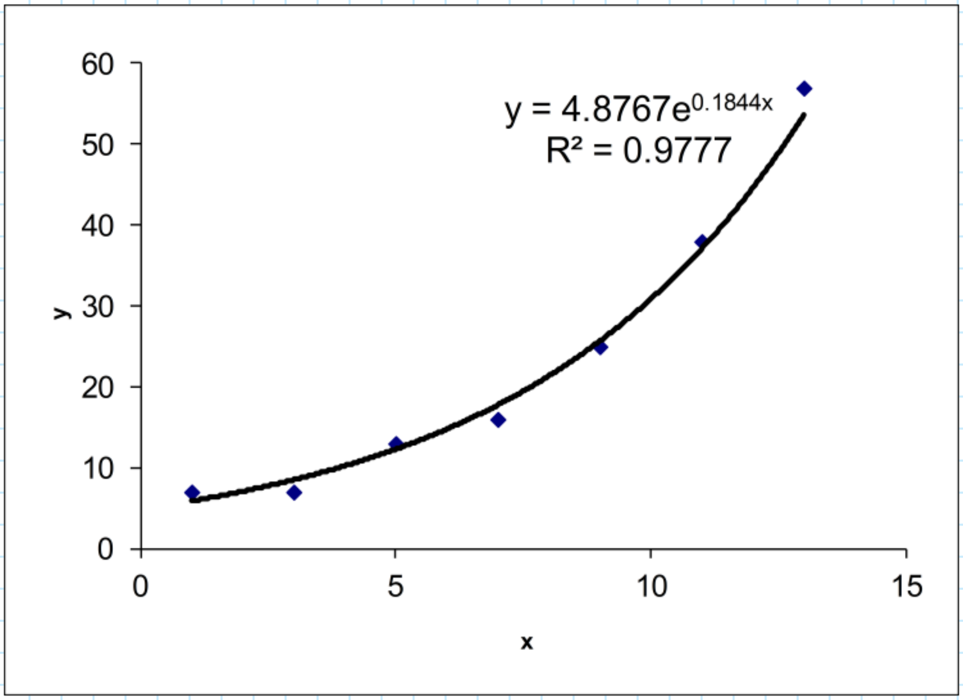
© Steven F. Bartlett, 2013

Excel - Nonlinear Regression

Thursday, March 11, 2010 11:43 AM

Regression fitting does not have to be applied only to linear trends. Below is an example of a nonlinear fit of data.

x	y
1	7
7	16
11	38
5	13
9	25
13	57
3	7



Note that this nonlinear form can be linearized. This is done by:

ln y = ln 4.8767 + 0.1844 x or in general terms, ln y = ln bo + b1 x.

The next page shows general transformation to linear forms.

© Steven F. Bartlett, 2013

Transformations and Matrix Operations for Linear Regression

Thursday, March 11, 2010 11:43 AM

If the x or y data can transformed to make the data have a linear trend appearance, then matrix operations can be used to solve for the slope and intercept of the line.

Transformations Useful for Regression Analysis (i.e., line fitting)

Non-linear Model		**Linearized Form**
Exponential $y = b_0 e^{b1x}$		$\ln y = \ln b_0 + b_1 x$
Power $y = b_0 x^{b1}$		$\ln y = \ln b_0 + b_1 \ln x$
Logarithmic $y = b_0 + b_1 \ln x$	(already linear form)	$y = b_0 + b_1 \ln x$
Polynominal $y = b_0 + b_1 x_1 + b_2 x_1^2 \ldots$	(already linear form)	$y = b_0 + b_1 x_1 + b_2 x_1^2 \ldots$

If the nonlinear data can be transformed to a linear relationship, then it is possible to use matrix operations to determine the intercept, bo, and the slope, b1 of the linearized form. This resulting fitted line will be a least squares best fit of the data. This is the process that Excel uses in fitting trendlines.

Matrix Operation Used for Regression Analysis

Least squares fit or solution to a set of x – y data pairs is given by:

$$\underline{\mathbf{B}} = (\mathbf{X'X})^{-1}(\mathbf{X'}\underline{\mathbf{Y}})$$

B is a matrix that contains the intercept and partial slopes (i.e., partial derivatives) that you want to estimate (e.g. b_o, $b_1 \ldots b_p$). For a two parameter model, the value of b_o is the intercept and b_1 is the slope of the line.

X is the matrix of independent variables (i.e., variables that you have measured to try to predict the response, Y).

Y is the matrix of dependent variables (i.e., response variable that is predicted by X).

© Steven F. Bartlett, 2013

Blank

Monday, March 2, 2020 9:39 AM