

Statistics of One Variable

Statistics is a mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data. Statistical methods can be used to summarize or describe a collection of data; this is called **descriptive statistics**. In addition, patterns in the data may be modeled in a way that accounts for randomness and uncertainty in the observations, to draw inferences about the process or population being studied; this is called **inferential statistics**. Both descriptive and inferential statistics can be considered part of **applied statistics**. There is also a discipline of **mathematical statistics**, which is concerned with the theoretical basis of the subject. The word *statistics* is also the plural of *statistic* (singular), which refers to the result of applying a statistical algorithm to a set of data, as in employment statistics, accident statistics, etc.

Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analysing data (singular is datum), as well as drawing valid conclusions and making reasonable inferences on the basis of such analysis. One can collect data on the entire target group (or a group of theoretical infinite size) called the **population** or on a small part of the group called a **sample**. The variables used in listing or describing the data can be **discrete** (i.e. they assume any of a prescribed finite set of values called the domain of the variable) or **continuous** (i.e. they can theoretically assume any value between two given values which determine the range of the domain).

e.g. the number of people attending this session is a discrete variable
 the height of the people at this session is a continuous variable

The concept of a variable can be extended to non-numerical entities.

In **rounding off** data when compiling statistics, we use a special rule for numbers in which the key (and final) digit is 5. The reason for using this approach is due to the fact that the number 12.65 is equidistant from 12.6 and 12.7. If we always rounded such data points up to the next higher number (12.7 in this case) we would tend to introduce cumulative rounding errors into subsequent statistical calculations. Thus the accepted practice is to round such numbers to the nearest even number preceding the 5. For example, 8.75 is rounded to 8.8 while 8.25 is rounded to 8.2. Any listing or computation involving measured data should also be accompanied by error limits (e.g. ± 0.01).

Data can also be categorized as **qualitative** or **quantitative**. The definitions of these terms vary somewhat, but usually non-numerical data is qualitative and numerical data is quantitative.

Raw data are collected data that have not been organized numerically. An **array** is an arrangement of raw numerical data in ascending (usually) or descending order of magnitude. Large masses of raw data are often divided into classes or categories. The number of individuals belonging to each class is called the **frequency**. A tabular arrangement of data by classes is called a **frequency distribution table**. Data organized or summarized in this manner are called grouped data. It is generally not possible to determine the values for all of the data points when they are organized in this way. Each class in such a table is defined as a **class interval** (e.g. 51-60). The end numbers of the interval are called its **class limits** (here 51 is called the lower class limit and 60 the upper class limit). A class interval in which there is no (theoretical) upper or lower limit is called an **open class limit**.

The **class width** of an interval is given by the difference between the upper and lower limits (also called the class size or strength or interval). The **class mark** is the midpoint of the interval (given by the sum of the upper and lower limits, divided by two). You will wish to consult additional definitions for frequency polygons, frequency curves, cumulative distributions and histograms, as well as the rules (more precisely, the accepted practices) governing their formation and presentation.

Data sets involving single-variables are most commonly analysed by means of **measures of central tendency** (or measures of position) and by **measures of spread** (also called measures of dispersion or variation). Data can also be described by measures of **skewness** (the degree of asymmetry) and of **kurtosis** (the degree of peakedness) but these are not elaborated here. Data sets that relate two or more variables are analysed by investigating relationships between those variables.

Measures of Central Tendency

An **average** is a value which is typical or representative of the set of data. These tend to lie centrally within a set of data that has been arranged according to magnitude. Some common measures of central tendency are:

The **arithmetic mean**, defined by: $\bar{X} = \frac{\sum_{j=1}^N X_j}{N}$ where X denotes the data set, X_j denotes an

individual datum and N denotes the number of data in the set (called the index when used in sigma notation). This formula is used for raw data. The corresponding formula for the **mean for grouped**

data is: $\bar{X} = \frac{\sum fx}{N}$ where f is the frequency of the interval and x is the mid-value (class mark) of the interval. The **index** relates the value of a variable to an accepted (or arbitrary) base level or time (e.g. Consumer Price Index).

The **weighted arithmetic mean** is defined by: $\bar{X} = \frac{\sum wX}{\sum w}$ where w denotes the weighting factor.

A **moving average** is one statistical technique used to analyze time series data. Moving averages are used to smooth out short-term fluctuations, thus highlighting longer-term trends or cycles. The moving average used in secondary school mathematics studies is really the **simple moving average** (or SMA), which calculated the unweighted mean of the previous n data points. It is given by:

$$SMA = \frac{x_m + x_{m-1} + x_{m-2} + \dots + x_{m-n+1}}{n} .$$

In order to reduce the lag in simple moving

averages, technicians often use **exponential moving averages** (also called exponentially weighted moving averages). EMA's reduce the lag by giving more weight to recent values relative to older ones. The weighting applied to the recent values depends on the specified period of the moving average. The shorter the EMA's period, the more weight that will be applied to the most recent value.

The **median** is the middle value of a set of ranked (in order of size) data or the arithmetic mean of two middle values. There are two formulas for the computation of the median, depending on whether the size of your sample is even or odd. If n (the number of observations in your sample) is odd, select $(n+1)/2$ observation. If n is even, select the midpoint between the $n/2$ and $n/2+1$ observation. The **mode** is the value(s) which occur with the greatest frequency. The mode may not be unique and it may not exist.

The **geometric mean** is defined by: $G = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_N}$

The **harmonic mean** is defined by: $H = \frac{N}{\sum \frac{1}{X}}$. We have the relationship that $H \leq G \leq \bar{X}$.

Measures of position include the measures of central tendency as well as other point locations away from the centre of the data distribution. The values that divide a set of ranked data into four equal parts are called the first, second and third **quartiles**, and these are denoted by Q_1 , Q_2 and Q_3 where Q_2 is equal to the median. **Deciles** and **percentiles** divide the data into ten and 100 equal parts respectively. When data are organized into a frequency distribution, all values falling within a given class interval are considered as coincident with the class mark (mid-value) of that interval.

Measures of Dispersion

The degree to which numerical data tend to spread about an average value is called the **variation** or **dispersion** of the data. The most common measure of spread used is the **standard deviation**, but several others are listed here. (See the attached Excel sheets for determining the standard deviation for given data sets.) Other types of measures of spread include:

The **range** – the difference between the largest and smallest numbers in the set

The **mean deviation**: $M.D. = \overline{|X - \bar{X}|} = \frac{\sum_{j=1}^N |X_j - \bar{X}|}{N}$ in which \bar{X} is the arithmetic mean. The mean

deviation for grouped data is given by: $\frac{\sum_{j=1}^N f_j |X_j - \bar{X}|}{N}$ where the X_j 's represent class marks and the

f_j 's are the corresponding frequencies. The **interquartile range** (usually abbreviated as IQR) is given by $IQR = Q_3 - Q_1$. The **semi-interquartile** range, is $SIQR = \frac{Q_3 - Q_1}{2}$.

The **population standard deviation** of a set of N numbers $X_1, X_2, X_3, \dots, X_N$, is denoted by σ and is

defined by: $\sigma = \sqrt{\frac{\sum_{j=1}^N (X_j - \bar{X})^2}{N}}$ and in an alternate form for easier calculations by the formula:

$\sigma = \sqrt{\frac{\sum_{j=1}^N x_j^2}{N} - \left(\frac{\sum_{j=1}^N X_j}{N}\right)^2}$. The formula for the **sample standard deviation** is identical to these

except that the denominator contains $N - 1$ instead of N . This factor $N - 1$ is considered to represent a better estimate of the standard deviation for samples of size $N < 30$. This formula is usually denoted by s instead of σ . However, you should also be aware that many statistics texts reverse the use of the symbols s and σ in these formulas. Most calculators use σ_n and σ_{n-1} for the population and sample S.D., respectively. It should also be noted that for samples sizes greater than 30, the values for both versions of the standard deviation are very close. In general, a data set with N members is said to have $N - 1$ degrees of freedom. For a large data set (i.e. a large population), the numerical difference between the results using either $N - 1$ or N is insignificant. The formula for the standard deviation for

grouped data is: $\sigma = \sqrt{\frac{\sum_{j=1}^N f_j (X_j - \bar{X})^2}{N}}$ in which the f_j 's are the frequency of the data in each of the

classes, the X_j 's are the class marks and \bar{X} represents the mean for the grouped data.

The **variance** of a set of data is defined as the square of the standard deviation and denoted by σ^2 or s^2 . For normal distributions, 68.27% of the data lie between $\bar{X} - \sigma$ and $\bar{X} + \sigma$ (one standard deviation on each side of the mean). As well, 95.45% of the data lie between the mean and two standard deviations and 99.73% between the mean and three standard deviations. It is taken as another “rule of thumb” that any set with more than 30 data points closely approximates that of normally distributed data.

For binomial distributions, p is the probability that an event will occur in any single trial and q is the probability that it will not occur. Here, p is called the probability of a “success” and q that of a “failure” where $p = 1 - q$. For **binomial distribution**, we have the following: μ (the mean) = Np , σ^2 (the variance) = Npq and $\sigma = \sqrt{Npq}$. It is necessary that N be large and neither p nor q be close to zero for these approximations to hold. In practice, the criteria used here is that **both** Np and Nq must exceed 5 for these approximations to be acceptable.

Elementary Sampling Theory

Samples of size $N > 30$ are considered large enough that we can use normal curve formulas and approximations for analysis. It is important that **confidence intervals** be stated for samples of all sizes. Students who carry out surveys or polls should include this interval as part of their analysis. The following terms are briefly defined here. The **population parameters** are the statistical descriptions of the population. Generally, we will want to infer these from the sample chosen. Samples are used to estimate the measures of the entire population when it is too time-consuming or expensive to determine the population attributes directly (in a **census**). Sample data can be used to obtain **point estimates** (estimates stated as a single number) or **interval estimates** (an interval within which the parameter will lie). Far and away, the most common point estimate obtained from a sample is the population mean, although we sometimes are interested in the median, a quartile, a percentile or some other point estimate. Because samples only provide estimates and not exact determinations, we must express limits on the reliability of the measures predicted (such as $\pm 5\%$, or within 2 standard deviations). In all cases here, I have assumed that the sample is taken without replacement. Be aware that all of the formulas given below are multiplied by the factor $\sqrt{\frac{N_p - 1}{N_p}}$ if replacement is allowed.

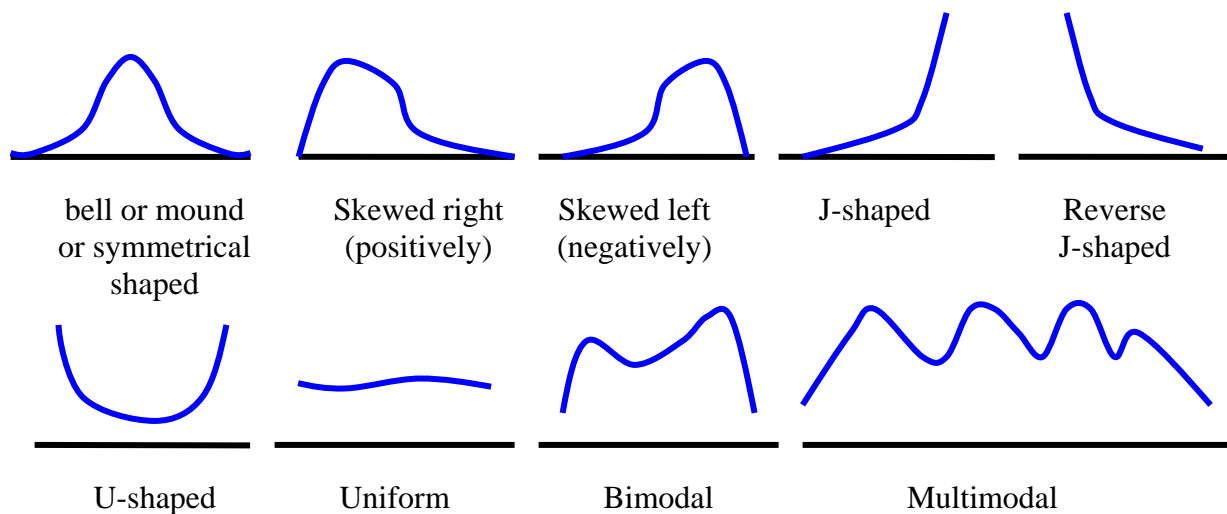
The degree of confidence denotes the probability the interval will actually contain the quantity that is being estimated from the sample. For sample sizes where $N > 30$, we automatically assume that 95% of the \bar{X} 's will lie within $\bar{X} \pm 1.960 \frac{\sigma}{\sqrt{N}}$ (\bar{X} is the sample mean and σ the sample standard deviation,

which is the estimate for the population standard deviation). The value of $1.960 \frac{\sigma}{\sqrt{N}}$ is called the **probable error**. This accounts for the common phrase “results are accurate 19 times out of 20” when poll findings are released. We have $z = \frac{\bar{X} - \mu}{\sigma / \sqrt{N}}$, where z is the z -score from a standard normal

distribution table and μ is the mean for the population. The factor $\frac{\sigma}{\sqrt{N}}$ is called the **standard error for the mean**. A significance level of 95% is the most common interval applied by statisticians, but any arbitrary number of z -scores can be selected. Tests involving life-threatening situations (e.g. testing parachutes or new drugs) are often required to employ significance intervals as high as 99.7% by government regulation. The table at the bottom of the last page of this document provides the corresponding z -scores for the most commonly used confidence intervals. It will be required to allow us to answer many of the examples given below.

When the sample size is less than 30, we base confidence intervals for μ on a distribution which is quite similar to normal distribution – the **Student’s *t*-distribution**. It is called this because William Gosset, the man who first developed this method of interpreting data, was forbidden from publishing anything by his employer, Guinness. Gosset secretly published his work under the name “Student” in order to avoid detection. Care must be taken to avoid **bias** in sampling techniques. Sampling bias (or selection bias or selection effect) is the error of distorting a statistical analysis due to the methodology of how the samples are collected. Typically this causes measures of statistical significance to appear much stronger than they are, but it is also possible to cause completely illusory artifacts.

One-variable data can be illustrated using a wide variety of graphing techniques. The most common of these are pie charts (circle graphs), bar graphs (multiple, horizontal and vertical) line graphs, area plots, histograms, frequency polygons and curves as well as box-and-whisker plots. The diagrams below illustrate the most common appearances of frequency curves, but the names are somewhat arbitrary.



A second important branch of statistics is the statistics of **two variables**. Here we are more interested in finding the relationships between two (or more) quantities, than we are in measuring their attributes (although that is also required). This topic will be left until a later presentation.

In secondary (and elementary) school mathematics of data management, we usually are concerned with: **Data collection**, **data presentation** (data **organization** in tables or charts, and data **illustration** by means of graphs) and **data analysis** (the computation of statistical measures and determining inferences, predictions or conclusions from these measures). Data that has been collected directly by the researcher (through samples, questionnaires, tests, measurements, etc.) is called **primary data**. Data which has been collected elsewhere, but used for analysis by the researcher, is called **secondary data**. Secondary data can be used for analysis which goes beyond the purpose for which it was originally compiled, or even can be used to contradict the findings of the original work in which it appeared, but the original source must always be credited whenever secondary data is used.

The study of the statistics of one variable requires a preliminary understanding of counting theory (permutations and combinations) and probability theory. Hopefully, these are kept to an introductory level in the Data Management course. The Ministry of Education likely intended that the more challenging combinatorial applications are left to the Geometry and Discrete Mathematics grade 12 course (but that is also about to be substantially revised in the near future). A brief outline of the curriculum content specified by the Ministry of Education for the grade 12 Data Management course is provided on the third last page of this document.

The questions which begin on this page will constitute the **start of the presentation on Saturday**. It is not necessary to understand (or even bring) the material outlined in the pages above, to follow the session planned (but it might help). We will begin by analyzing and presenting the solutions to the problems given here, with a view as to how they might be explained to secondary school students taking Data Management. Quite likely, these students were not the most proficient students in their previous encounters with the secondary school mathematics curriculum.

All of the solutions for the problems below will be presented on Saturday. You are free to work them out in advance if you like. Our emphasis will really focus on what those solutions mean in terms of the context of the problems offered. You will need a scientific calculator and the tables on the last two pages of this document, or you a graphing calculator (the TI-83 is well-suited for this purpose).

- 1.) The data in the table below represents the heights of the mathematics teachers who have attended my workshops over the last two years.

Height (in cm)	Tally	Frequency	Midpoint	Relative Frequency	Cumulative Frequency
141-145		2	143	0.025	2
146-150		6	148	0.075	8
151-155		9	153	0.1125	17
156-160		12	158	0.15	29
161-165		15	163	0.1875	44
166-170		14	168	0.175	58
171-175		10	173	0.125	68
176-180		7	178	0.0875	75
181-185		5	183	0.0625	80

- a) Find the mean and (sample) standard deviation for the data set.
- b) Construct a histogram, frequency polygon and frequency curve for this data.
- 2.) The data given below represents all of the 18-hole round scores accumulated by Tiger Woods in his last 15 stroke-play tournaments. (Source: The Professional Golfers Association of America.

18-hole Total	63	64	65	66	67	68	69	70	71	72	73	74	75	76
Frequency	2	3	2	8	9	6	5	2	7	7	1	1	1	2



- a) Find the mean and standard deviation for these scores.
- b) Find the probability that Woods will get at least one 18-hole round below 59.
- c) A typical year consists of 50 rounds for Woods. Find the probability that he will get one round below 59 out of 50 rounds played.
- d) Do you think this statistical estimate is too low, too high or about right for real-life conditions?
- 3.) A lazy student studied only part of the content necessary for a final exam in Biology. The exam consisted of 100 multiple choice questions, each with four possible answer selections. The student was absolutely confident of his answers on 20 of the 100 questions (the part of the course he actually studied) and simply guessed at all of the remaining questions.
- a) Compare the probability of getting exactly 30 questions correct (out of 100) by guessing at all answers, using i) a binomial probability distribution, and ii) normal curve approximations
- b) Find the probability that he passed the test (50% is considered to be a pass).

- 4.) Use the heights in the sample of mathematics teachers given in question 1.) above, to predict the average height of a mathematics teacher in Ontario. (Use an appropriate confidence level.)
- 5.) A mathematics teacher at Brand X High School has developed a pedagogical method which she claims produces better student scores on the grade 9 EQAO tests. In the past, the students at her school have achieved a mean score of 2.94 with a standard deviation of 0.42. Using her new technique with 30 students this year, she produced an average EQAO score of 3.08. Did her method produce a significant improvement in the EQAO test scores of her students?

Grade 12 Data Management

Of the three new grade 12 (1998) mathematics courses designed for university-bound students, this contains the most features that were not a part of the three mathematics courses at the OAC level. The course was designed for students bound for programs in the social sciences, life sciences, business, law, medicine, pharmacy, **and** students who would someday consider postgraduate studies in any field. For a variety of reasons, this purpose has not been well received or understood by many grade 12 students with the Near North Board (and most other boards as well).

One reason was that teachers of Finite Mathematics were often assigned to teach the course (and it was not envisioned as a near replacement for that course at all). A second reason is that many mathematics teachers do not have a strong background in statistics but rather in analysis. The course was also designed to utilize many computer and graphing calculator features that are quite new to teachers. Finally, as with most of the secondary school mathematics texts, the text books available for the Data Management are of very poor quality, resulting in the fact that teachers would be required to supplement the text with a variety of material from many other sources. Add to this, the fact that most Ontario universities use the Advanced Functions and Introductory Calculus course as a means of “sorting” through applications from secondary school students for most programs – even those that have absolutely no requirement for Calculus.

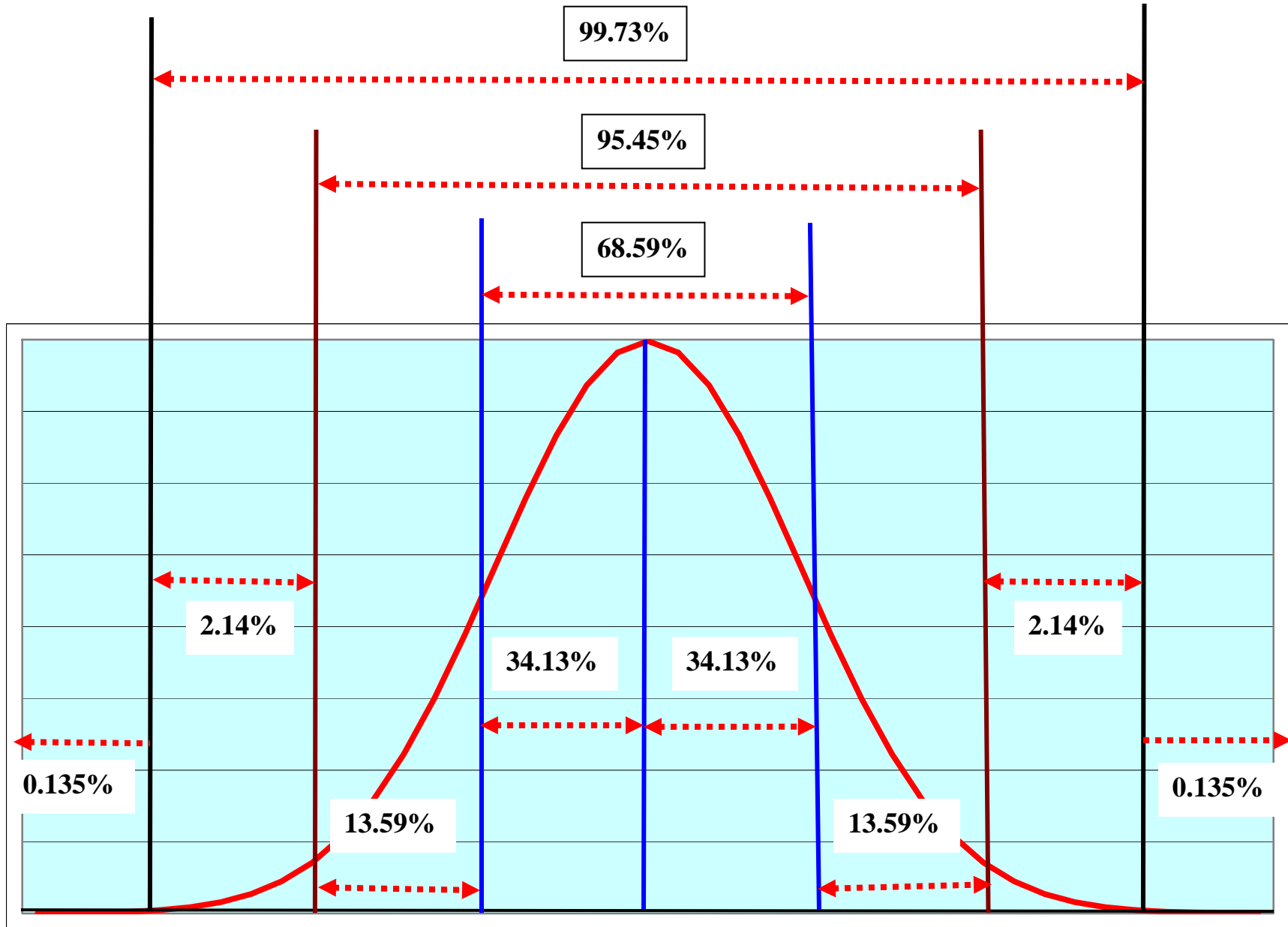
We can perform a valuable service for our students who contemplate obtaining a university degree (40% of them) by steering them into selecting the Data Management course (in addition to Calculus, if that course is mandated). This will require a significant change on the part of the teachers of Data Management. They will have to aim the curriculum material and level of difficulty at students who have traditionally struggled with mathematics. A high failure or drop-out rate among Data Management students will quickly signal to students that they should not opt for that course (even though it is against their best interest).

The Data Management Course curriculum content (as envisioned by the Ministry) consists of:

- Search for and locate sources of data concerning a wide range of subjects
- Create data bases that allow for the manipulation and retrieval of data
- Use diagrams (tree, graph, optimum path and network) to investigate outcomes
- Carry out simple operations with matrices and use these to solve problems
- Solve counting problems using diagrams, formulas and Pascal’s Triangle
- Represent, interpret and solve various probability distributions
- Design, carry out and analyze simulations (particularly with respect to probability)
- Understand and describe sampling techniques and bias
- Compute and interpret statistics of one variable
- Demonstrate an understanding of the normal curve and its applications
- Demonstrate an understanding of the relationships involving two variables
- Asses the validity and conclusions made form statistics in public sources
- Carry out the culminating project or investigation described for this course

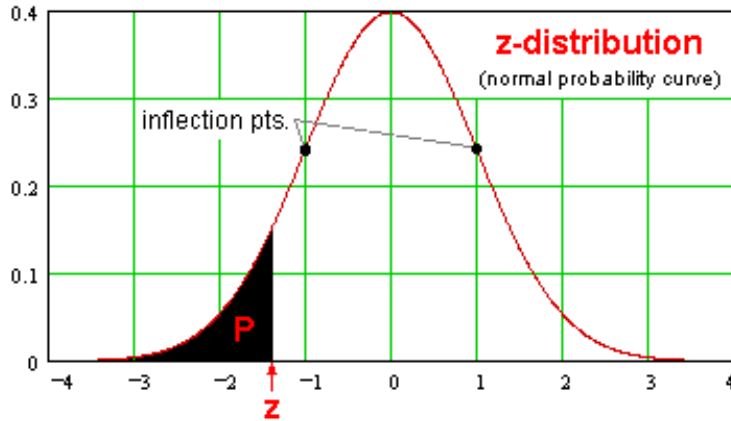
Clearly, this course requires a distinct approach to teaching mathematics to secondary students.

Percentages of data within selected standard deviations



Standard Normal Probabilities: (The table is based on the area **P** under the standard normal probability curve, below the respective **z**-statistic.)

z-distribution



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00103	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02067	0.02018	0.01970	0.01923	0.01876	0.01831

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04456	0.04363	0.04272	0.04181	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06425	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07214	0.07078	0.06944	0.06811
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09852
-1.1	0.13566	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1.0	0.15865	0.15625	0.15386	0.15150	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17878	0.17618	0.17361	0.17105	0.16853	0.16602	0.16354	0.16109
-0.8	0.21185	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.23269	0.22965	0.22663	0.22363	0.22065	0.21769	0.21476
-0.6	0.27425	0.27093	0.26763	0.26434	0.26108	0.25784	0.25462	0.25143	0.24825	0.24509
-0.5	0.30853	0.30502	0.30153	0.29805	0.29460	0.29116	0.28774	0.28434	0.28095	0.27759
-0.4	0.34457	0.34090	0.33724	0.33359	0.32997	0.32635	0.32276	0.31917	0.31561	0.31206
-0.3	0.38209	0.37828	0.37448	0.37070	0.36692	0.36317	0.35942	0.35569	0.35197	0.34826
-0.2	0.42074	0.41683	0.41293	0.40904	0.40516	0.40129	0.39743	0.39358	0.38974	0.38590
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43250	0.42857	0.42465
-0.0	0.50000	0.49601	0.49202	0.48803	0.48404	0.48006	0.47607	0.47209	0.46811	0.46414

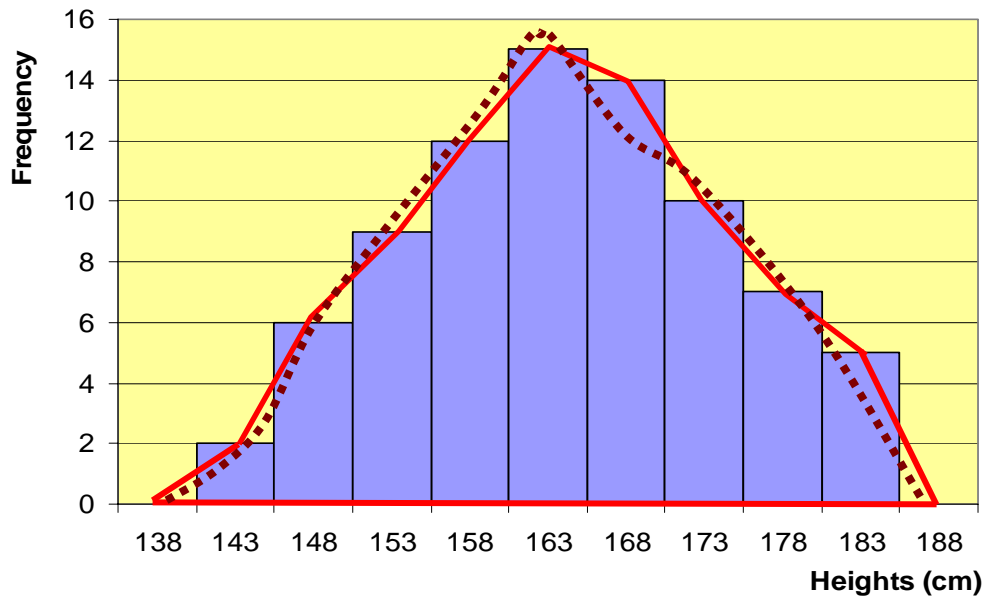
This table is useful for calculations with Z-scores greater than 2 (or less than -2).

Simply use 1 minus the decimal given for positive Z-scores.

Level of Significance α	0.10	0.05	0.01	0.005
Critical values of z for one-tailed tests	1.280	1.645	2.33	2.876
Critical values of z for two-tailed tests	1.645	1.960	2.576	3.08

Confidence Level	99.73%	99%	98%	95.45%	95%	90%	80%	68.27%
z_c	3.00	2.58	2.33	2.00	1.96	1.645	1.28	1.00

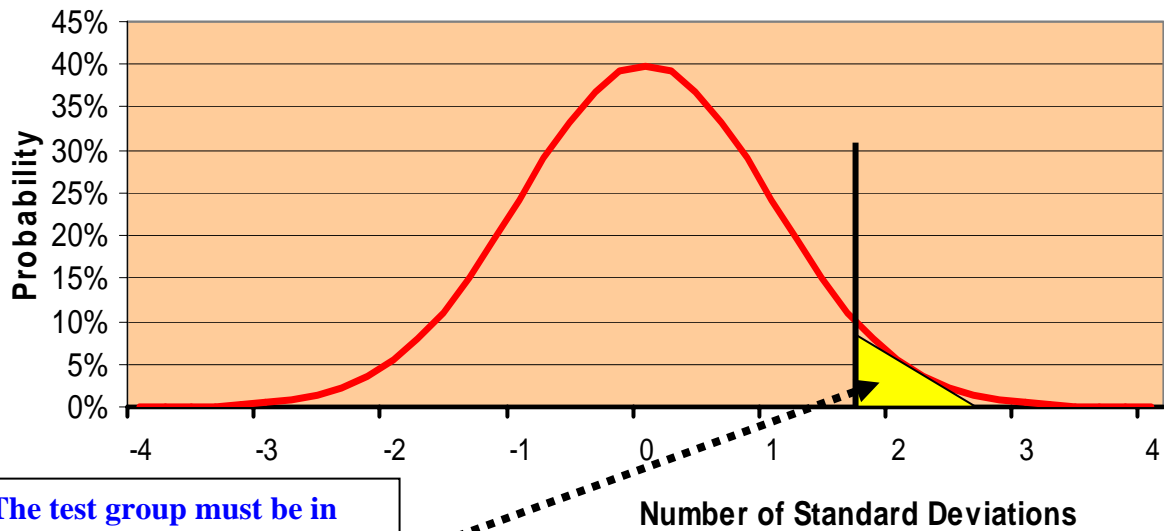
Mathematics Teacher Heights



Frequency Polygon ———

Frequency Curve ·····

Standard Normal Curve



The test group must be in this region of the normal curve (to the right of the arbitrary “confidence limit” in order to be considered statistically significant”

Statistics of One-variable Handout Sheet

