

CSC 217 Midterm Review:

Descriptive Statistics

What is Descriptive Statistics?

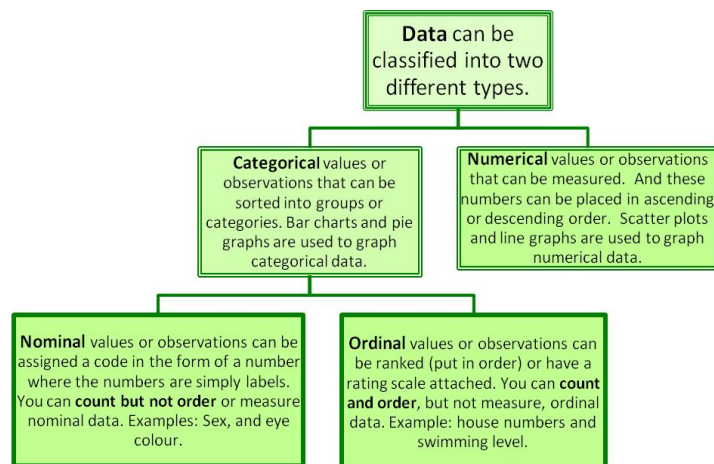
Descriptive Statistics involves describing the data **without** drawing conclusions

Categorical Data (QUALITATIVE):

Categorical Data. Categorical variables represent types of data which may be divided into groups. Examples of categorical variables are race, sex, age group, and educational level.

What types of categorical data are there?

There are two types of categorical variable, **nominal and ordinal**.



A **nominal variable** has no intrinsic ordering to its categories. For example, gender is a categorical variable having two categories (male and female) with **no intrinsic ordering** to the categories.

An **ordinal variable** has a clear ordering. For example, temperature as a variable with three *orderly* categories (low, medium and high).

Measures of Central Tendency:

What are some common measures of central tendency? When are each of them useful? When are each of them less useful?

Arithmetic mean or simply, **mean**

the sum of all measurements divided by the number of observations in the data set.

BEST: When data is normally distributed.

LESS USEFUL: When the data is skewed

Median

the middle value that separates the higher half from the lower half of the data set. The median and the mode are the only measures of central tendency that can be used for ordinal data, in which values are ranked relative to each other but are not measured absolutely.

BEST: The median is usually preferred to other measures of central tendency when your data set is skewed. (Has outliers)

LESS USEFUL: Normally distributed mean is preferred over median

Mode

the most frequent value in the data set. This is the only central tendency measure that can be used with nominal data, which have purely qualitative category assignments.

BEST: Only when dealing with nominal data

LESS USEFUL: With data that isn't nominal

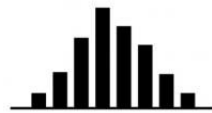
Measures of Spread

What are some common measures of spread? How are they found?

	<p>Measures of spread describe how similar or varied the set of observed values are for a particular variable (data item).</p> <p>Measures of spread include the range, quartiles and the interquartile range, variance and standard deviation.</p> <p>The range is the difference between the smallest value and the largest value in a dataset.</p> <p>Quartiles divide an ordered dataset into four equal parts, and refer to the values of the point <i>between</i> the quarters. A dataset may also be divided into quintiles (five equal parts) or deciles (ten equal parts).</p> <ul style="list-style-type: none"> - The lower quartile (Q1) is the point between the lowest 25% of values and the highest 75% of values. It is also called the 25th percentile. - The second quartile (Q2) is the middle of the data set. It is also called the 50th percentile, or the median. - The upper quartile (Q3) is the point between the lowest 75% and highest 25% of values. It is also called the 75th percentile. <p>The interquartile range (IQR) is the difference between the upper (Q3) and lower (Q1) quartiles, and describes the middle 50% of values when ordered from lowest to highest. The IQR is often seen as a better measure of spread than the range as <i>it is not affected by outliers</i>.</p> <p>$IQR = Q3 - Q1$</p> <p>The variance and the standard deviation are measures of the spread of the data around the mean. They summarise how close each observed data value is to the mean value.</p>
--	--

	<p>The population Variance σ^2 (pronounced <i>sigma squared</i>) of a discrete set of numbers is expressed by the following formula:</p> $\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$ <p>where:</p> <p>X_i represents the <i>i</i>th unit, starting from the first observation to the last</p> <p>μ represents the population mean</p> <p>N represents the number of units in the population</p> <p>The Variance of a sample s^2 (pronounced <i>s squared</i>) is expressed by a slightly different formula:</p> $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ <p>where:</p> <p>x_i represents the <i>i</i>th unit, starting from the first observation to the last</p> <p>\bar{x} represents the sample mean</p> <p>n represents the number of units in the sample</p> <p>The standard deviation is the square root of the variance. The standard deviation for a population is represented by σ, and the standard deviation for a sample is represented by s.</p>
<p>Data Visualization:</p>	<p>Histograms</p> <p>What does a histogram show?</p> <p>A histogram is a plot that lets you discover, and show, the underlying frequency distribution (shape) of a set of continuous data.</p> <p>**Histograms can only be used for numerical data.</p> <p>What are the benefits of a histogram? What are the drawbacks?</p> <p>A histogram is great for large sets of data because they can be grouped within the intervals. Change in intervals of a histogram completely changes the way that it looks, and therefore the way it is perceived.</p>

Types of Histograms



Bell Shaped:
The normal pattern



Double Peaked: Suggests two distributions



Skewed: Look for other processes in the tail



Truncated: Look for reasons for sharp end of distribution or pattern



Ragged Plateau: No single clear process or pattern

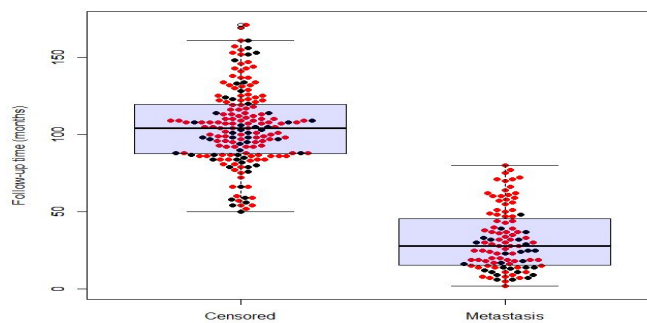
Swarm Plots:

What does a swarm plot show?

A representation of the distribution of values (illustrates frequency)

What are the benefits of a swarm plot? What are the drawbacks?

It gives a better representation of the distribution of values, but it does not scale well to large numbers of observations



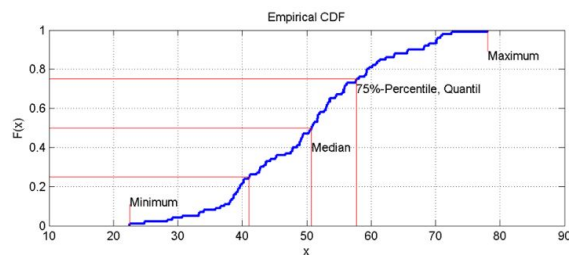
ECDF

What does an ECDF show?

An ECDF is an estimator of the Cumulative Distribution Function. The ECDF essentially allows you to plot a feature of your data in order from least to greatest and see the whole feature as if is distributed across the data set.

What are the benefits of an ECDF? What are the drawbacks?

Key values and features like minimum, maximum, median, quantiles, percentiles, etc. can be directly read from the diagram. Can't use to find mean



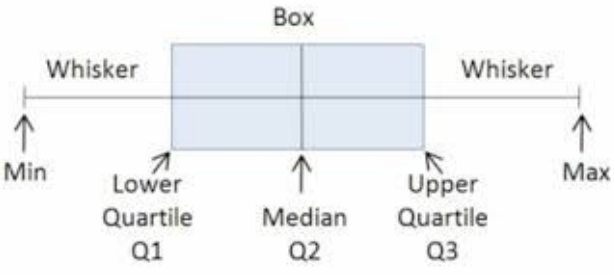
Box Plots

What does a boxplot show?

This type of graph is used to show the shape of the distribution, its central value, and its variability.

What are the benefits of a boxplot? What are the drawbacks?

Box plots display a range and distribution of data. Also, box plots show outliers, and they show some skew-ness and symmetry in the graph. But cannot clearly identify the original data, and they can only be used for numerical data.

	
Probability	<p><u>What is the difference between statistics and probability?</u></p> <p>Probability deals with <u>predicting</u> the likelihood of future events, while statistics involves the <u>analysis</u> of the frequency of past events.</p> <ul style="list-style-type: none"> • Probability is primarily a <u>theoretical branch of mathematics</u>, which studies the consequences of mathematical definitions. Statistics is primarily an <u>applied branch of mathematics</u>, which tries to make sense of observations in the real world. <p><u>What is a sample space?</u></p> <p>https://www.probablisticworld.com/delving-into-sample-spaces/</p> <p>Set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.</p> <p><u>What is an event?</u></p> <p>The event space contains all <i>sets of outcomes</i>; all subsets of the sample space.</p> <p><u>What is a probability function?</u></p>

A **probability distribution** is a table or an equation that links each outcome of a **statistical** experiment with its **probability** of occurrence.

What is the intersection of two sets?

The union of two sets contains all the elements contained in either set (or both sets).

What is the union of two sets?

The intersection of two sets contains only the elements that are in both sets.

What is the complement of a set?

The complement of a set A contains everything that is not in the set A.

_____ NOTATION _____

$\cup \rightarrow$ UNION (in both E and F)

EF , called the **intersection** of E and F, to consist of all outcomes that are in both E and F.

If $EF = \emptyset$, implying that E and F cannot both occur, then E and F are said to be **mutually exclusive**.

Venn diagram and The Algebra of Events:

A graphical representation of events that is very useful for illustrating logical relations among them is the Venn diagram

Commutative law	$E \cup F = F \cup E$	$EF = FE$
Associative law	$(E \cup F) \cup G = E \cup (F \cup G)$	$(EF)G = E(FG)$
Distributive law	$(E \cup F)G = EG \cup FG$	$EF \cup G = (E \cup G)(F \cup G)$

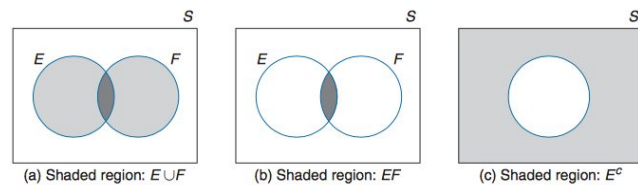
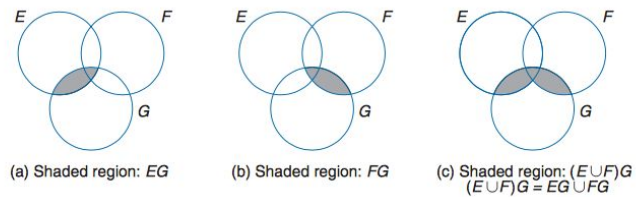
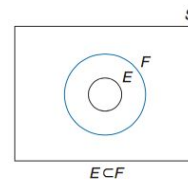


FIGURE 3.1 Venn diagrams.



The following useful relationship between the three basic operations of forming unions, intersections, and complements of events is known as *DeMorgan's laws*.

$$(E \cup F)^c = E^c F^c$$

$$(EF)^c = E^c \cup F^c$$

What is the Inclusion-Exclusion principle?

Inclusion–exclusion principle

In combinatorics, the **inclusion–exclusion principle** (also known as the **sieve principle**) is an equation relating the sizes of two sets and their union. It states that if A and B are two (finite) sets, then

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

The meaning of the statement is that the number of elements in the union of the two sets is the sum of the elements in each set, respectively, minus the number of elements that are in both. Similarly, for three sets A , B and C ,

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

What is disjoint?

In mathematics, two sets are said to be **disjoint sets** if they have no element in common. Equivalently, two disjoint sets are sets whose intersection is the empty set.

For example, $\{1, 2, 3\}$ and $\{4, 5, 6\}$ are *disjoint sets*, while $\{1, 2, 3\}$ and $\{3, 4, 5\}$ are not disjoint.

What is the rule of product?

<https://brilliant.org/wiki/rule-of-product/>

<https://brilliant.org/wiki/rule-of-sum-and-rule-of-product-problem-solving/>

The rule of product states that if there are n ways of doing something, and m ways of doing another thing after that, then there are $n*m$ ways to perform both of these actions. In other words, when choosing an option for m and an option for n , there are $m*n$ different ways to do both actions.

Rule of Sum

If there are n ways to arrange something, and there are m ways to arrange something else, and these arrangements cannot both happen, then the number of ways to arrange either of those things is $n+m$

Permutations and Combinations

What is a permutation?

A permutation is an arrangement of objects with regard to order.

What is a combination?

A combination is an arrangement of objects without regard to order.

What is the difference between the two?

Permutations are for lists (order matters) and combinations are for groups (order doesn't matter).

Do you know when to use one or the other?

<https://brilliant.org/wiki/permutations/>

<https://brilliant.org/wiki/combinations/>

Combination = Choice w/ order does not matter

Permutation = Arrangement w/ order matters

FORMULA:

PERMUTATIONS

Order Matters
Repetition Allowed

$$\text{Possibilities} = n^r$$

Order Matters
Repetition Not Allowed

$$\text{Possibilities} = \frac{n!}{(n-r)!}$$

COMBINATIONS

Order Doesn't Matter
Repetition Allowed

$$\text{Possibilities} = \frac{n!}{r!(n-r)!}$$

Order Doesn't Matter
Repetition Not Allowed

$$\text{Possibilities} = \frac{(n+r-1)!}{r!(n-1)!}$$

Sampling

What is sampling with replacement?

When we sample with replacement, the two sample values are independent. Practically, this means that what we get on the first one doesn't affect what we get on the second.

What is sampling without replacement?

In sampling without replacement, the two sample values aren't independent. Practically, this means that what we

got on the for the first one affects what we can get for the second one.

Conditional Probability

<https://brilliant.org/wiki/conditional-probability/>

What is conditional probability?

The ***conditional probability*** of an event B is the probability that the event will occur given the knowledge that an event A has already occurred. This probability is written $P(B|A)$, notation for the *probability of B given A* . In the case where events A and B are *independent* (where event A has no effect on the probability of event B), the conditional probability of event B given event A is simply the probability of event B , that is $P(B)$.

If events A and B are not independent, then the probability of the *intersection of A and B* (the probability that both events occur) is defined by

$$P(A \text{ and } B) = P(A)P(B|A).$$

From this definition, the conditional probability $P(B|A)$ is easily obtained by dividing by $P(A)$:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

What is the multiplication rule?

Rule of Multiplication

If events A and B come from the same sample space, the probability that both A and B occur is equal to the

probability the event A occurs times the probability that B occurs, given that A has occurred.

$$P(A \cap B) = P(A) P(B|A)$$

What is the law of total probability?

In probability theory, the **law (or formula) of total probability** is a fundamental rule relating marginal probabilities to conditional probabilities. It expresses the total probability of an outcome which can be realized via several distinct events

The summation can be interpreted as a [weighted average](#), and consequently the marginal probability, $\Pr(A)$, is sometimes called "average probability";^[2] "overall probability" is sometimes used in less formal writings.^[3]

The law of total probability can also be stated for conditional probabilities. Taking the B_n as above, and assuming C is an event [independent](#) with any of the B_n :

$$\Pr(A | C) = \sum_n \Pr(A | C \cap B_n) \Pr(B_n | C) = \sum_n \Pr(A | C \cap B_n) \Pr(B_n)$$

When are two events independent? Can you identify whether two events are independent or not?

Independent Events:

Two events E and F are said to be independent if Equation 3.8.1 holds. Two events E and F that are not independent are said to be dependent.

Independent Events

The outcome of one event **does not** affect the outcome of the other.

If A and B are independent events then the probability of both occurring is

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

Dependent Events

The outcome of one event affects the outcome of the other.

If A and B are dependent events then the probability of both occurring is

$$P(A \text{ and } B) = P(A) \times P(B|A)$$

Probability of B given A

Bayes' Theorem :

<https://brilliant.org/wiki/bayes-theorem/>

Bayes' theorem is stated mathematically as the following equation:^[2]

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

where A and B are events and $P(B) \neq 0$.

- $P(A | B)$ is a **conditional probability**; the likelihood of event A occurring given that B is true.
- $P(B | A)$ is also a conditional probability; the likelihood of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other; this is known as the **marginal probability**.

When can it be particularly useful?

The **theorem** provides a way to revise existing predictions or theories given new or additional evidence. In finance, **Bayes' theorem can be used** to rate the risk of lending money to potential borrowers.

What is a True Positive? What is a False Positive? What is a True Negative? What is a False Negative?

A **true positive** is an outcome where the model *correctly* predicts the *positive* class. Similarly, a **true negative** is an outcome where the model *correctly* predicts the *negative* class.

A **false positive** is an outcome where the model *incorrectly* predicts the *positive* class. And a **false negative** is an outcome where the model *incorrectly* predicts the *negative* class.

Classification Metrics

What is Accuracy

$$Accuracy = \frac{truepositives + truenegatives}{totalexamples}$$

?

Accuracy can immediately tell us whether a model is being trained correctly and how it may perform generally.

	<p>What is Precision?</p> <p>When the model predicts positive, how often is it correct?</p> $Precision = \frac{truepositives}{truepositives + falsepositives}$ <p>Precision helps when the costs of false positives are high.</p> <p>What is Recall?</p> $Recall = \frac{truepositives}{truepositives + falsenegatives}$ <p>Recall helps when the cost of false negatives is high.</p>
<p>Bernoulli Distributions</p>	<p>What is a Bernoulli distribution?</p> <p>https://brilliant.org/wiki/bernoulli-distribution/</p> <p>The Bernoulli distribution essentially models a single trial of flipping a weighted coin. It is the probability distribution of a random variable taking on only two values, 1("success") and 0("failure") with complementary probabilities P and 1-P, respectively. The Bernoulli distribution therefore describes events having exactly two outcomes, which are ubiquitous in real life.</p> <p>A <u>Bernoulli trial</u>, or Bernoulli experiment, is an experiment satisfying two key properties:</p> <ul style="list-style-type: none"> • There are exactly two complementary outcomes, success and failure.

	<ul style="list-style-type: none"> • The probability of success is the same every time the experiment is repeated.
--	---

Intuitively, it describes a single experiment having two outcomes: success ("1") occurring with probability p , and failure ("0") occurring with probability $1 - p$. It describes a single trial of a [Bernoulli experiment](#).

A closed form of the probability density function of Bernoulli distribution is $P(x) = p^x(1 - p)^{1-x}$.

One can represent the Bernoulli distribution graphically as follows:

Binomial distribution	<p>https://brilliant.org/wiki/binomial-distribution/#formal-definition</p> <p>The binomial distribution is, in essence, the probability distribution of the number of heads resulting from flipping a weighted coin multiple times. It is useful for analyzing the results of repeated independent trials.</p> <p>A binomial experiment is a series of n Bernoulli trials, whose outcomes are independent of each other. A random variable, X, is defined as the number of successes in a binomial experiment. Finally, a binomial distribution is the probability distribution of X.</p>
-----------------------	--

Geometric Distribution	<p>https://brilliant.org/wiki/geometric-distribution/</p> <p>The geometric distribution, intuitively</p>
-------------------------------	---

	<p>speaking, is the probability distribution of the number of tails one must flip before the first head using a weighted coin. It is useful for modeling situations in which it is necessary to know how many attempts are likely necessary for success.</p> <p>A series of Bernoulli trials is conducted until a success occurs, and a random variable X is defined as either</p> <p>the number of trials in the series, or the number of failures in the series.</p> <p>Note that the geometric distribution satisfies the important property of being memoryless, meaning that if a success has not yet occurred at some given point, the probability distribution of the number of additional failures does not depend on the number of failures already observed.</p>
--	---

Finding the Geometric Distribution

For a geometric distribution with probability p of success, the probability that exactly k failures occur before the first success is

$$(1 - p)^k p.$$

This is written as $\Pr(X = k)$, denoting the probability that the random variable X is equal to k , or as $g(k; p)$, denoting the geometric distribution with parameters k and p .

Poisson distribution	<p>https://brilliant.org/wiki/poisson-distribution/</p> <p>The Poisson distribution is the discrete probability distribution of the number of events occurring in a given time period, given the average number of times the event occurs over that time period.</p> <p>Conditions for Poisson Distribution:</p> <p>An event can occur any number of times during a time period.</p>
----------------------	---

	<p>Events occur independently. In other words, if an event occurs, it does not affect the probability of another event occurring in the same time period.</p> <p>The rate of occurrence is constant; that is, the rate does not change based on time.</p> <p>The probability of an event occurring is proportional to the length of the time period. For example, it should be twice as likely for an event to occur in a 2 hour time period than it is for an event to occur in a 1 hour period.</p>
--	---

Probabilities with the Poisson Distribution

Given that a situation follows a Poisson distribution, there is a formula which allows one to calculate the probability of observing k events over a time period for any non-negative integer value of k .

Let X be the **discrete random variable** that represents the number of events observed over a given time period. Let λ be the **expected value** (average) of X . If X follows a Poisson distribution, then the probability of observing k events over the time period is

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} ,$$

where e is **Euler's number**.

THEOREM

Poisson Limit Theorem:

As n approaches infinity and p approaches 0 such that λ is a constant with $\lambda = np$, the binomial distribution with parameters n and p is approximated by a Poisson distribution with parameter λ :

$$\binom{n}{k} p^k (1 - p)^{n-k} \simeq \frac{\lambda^k e^{-\lambda}}{k!} .$$

probability mass function	Let X be a discrete random variable. Hence Range(X) is a countable set
---------------------------	--

	<p>and thus it can be written as a list like $\{x_1, x_2, x_3, \dots\}$. Thus the only values the random variable X can take are x_1, x_2, x_3, \dots</p> <p>Now the next natural question to ask is what is $P(X = x_i)$ for the various values x_1, x_2, x_3, \dots?</p>
--	---

Cumulative distribution function	<p>https://brilliant.org/wiki/continuous-random-variables-cumulative/</p> <p>The cumulative distribution function, CDF, or cumulant is a function derived from the probability density function for a continuous random variable. It gives the probability of finding the random variable at a value less than or equal to a given cutoff.</p>
---	---

discrete distribution	<p>A discrete distribution describes the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.</p> <p>With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability. Thus, a discrete probability distribution is often presented in tabular form.</p>
------------------------------	---

continuous distribution	<p>A continuous distribution describes the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.</p> <p>Probabilities of continuous random variables</p>
--------------------------------	---

	(X) are defined as the area under the curve of its PDF. Thus, only ranges of values can have a nonzero probability. The probability that a continuous random variable equals some value is always zero.
--	---

If I give you a classification scenario, can you identify which metric we want to maximize? (i.e. we want to create software that detects cheating and ensure that we minimize false positives at the cost of letting some people get away with cheating)

Continuous Distributions

	<p>What is the probability density function? A function of a continuous random variable, whose integral across an interval gives the probability that the value of the variable lies within the same interval.</p> <p>What is the fundamental difference between the probability mass function and the probability distribution function, and thus a central tenet of the difference between discrete and continuous variables, and the method in which we measure them?</p> <p>A probability mass function differs from a probability density function (pdf) in that the latter is associated with continuous rather than discrete random variables; the values of the probability density function are not probabilities as such: a pdf must be integrated over an interval to yield a probability. PMF = Discrete Variable PDF = Continuous Variable</p>
--	--

Standard Normal Distribution

<https://brilliant.org/wiki/normal-distribution/>

What are the parameters for the normal distribution?

$$Z = (X - \mu)/\sigma$$

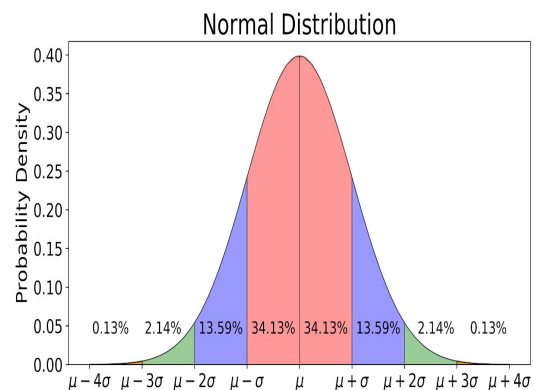
where Z is the value on the standard normal distribution, X is the value on the original distribution, μ is the mean of the original distribution, and σ is the standard deviation of the original distribution.

MAYBE important:

The empirical rule, or the 68-95-99.7 rule, states that 68% of the data modeled by a **normal distribution** falls within 1 standard deviation of the mean, 95% within 2 standard deviations, and 99.7% within 3 standard deviations.

What is the standard normal distribution?

A normal distribution with a mean of 0 and a standard deviation of 1 is called a *standard normal distribution*.



What is a Z-score? Why is it important?

	<p>1) Z-scores are expressed in terms of standard deviations from their means. Resultantly, these z-scores have a distribution with a mean of 0 and a standard deviation of 1.</p> <p>2) The standard score (more commonly referred to as a z-score) is a very useful statistic because it (a) allows us to calculate the probability of a score occurring within our normal distribution and (b) enables us to compare two scores that are from different normal distributions.</p> <p><u>What does a negative Z-score mean? What does a positive Z-score mean?</u></p> <p>A positive Z-score indicates the observed value is above the mean of all values, while a negative Z-score indicates the observed value is below the mean of all values.</p> <p><u>What is the law of large numbers?</u></p> <p>The law of large numbers, in probability and statistics, states that as a sample size grows, its mean gets closer to the average of the whole population.</p>
Center Limit Theorem	The central limit theorem is a theorem about independent random variables, which says roughly that the probability distribution of the average of independent random variables will converge to a normal distribution, as the number of observations increases.
Central Limit Theorem Importance	Central Limit Theorem and Statistical Inferences. Central Limit Theorem (CLT) is an important result in statistics, most specifically, probability theory. This theorem

	enables you to measure how much the means of various samples vary without having to use other sample means as a comparison.
--	---

Which distribution(s) can the normal distribution be an approximate for?	Normal Approximations. The normal distribution can be used as an approximation to the binomial distribution and Poisson Approximation
--	--

Why might this be useful? Why might this be not so useful anymore?	
--	--