

Exam Notes for MDM4UI

John Oss
December 16, 2016
L^AT_EX

Abstract

Notes for all topics taught in MDM4UI (Data Management) from the winter term of the 2016-2017 school year at GCI. Please refer to the appendix for a walk-through of example questions that will probably be on the exam. Sorry for the formatting. I'm not quite done yet.

1. Permutations

1.1. *What are Permutations?*

Permutations are the arrangements of objects in a set order. For instance, how many ways can you rearrange the letters in the word "APPLE", or how many ways can you bring 4 people out of 20 friends to a party?

1.2. *Factorial Notation*

Often in permutations, we need to multiply descending consecutive numbers. Since these numbers are descending and not ascending, we can safely assume that **every factorial must be positive**. For instance, $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ can often be simplified and written as **5!**, or 5 factorial.

The general rule for factorial notation is:

$$n! = n \cdot (n - 1) \cdot (n - 2) \cdots 3 \cdot 2 \cdot 1$$

The exception to this rule is $0!$, which in this case would be $0! = 1$, which would follow the same thinking as $n^0 = 1$.

1.3. *Rule of Product and Rule of Sum*

How many different results can one get from flipping a coin 3 times? You have 2 possibilities, heads or tails, and you flip it 3 times. Essentially, you have three groups of two, or $2 \cdot 3$ different results, which turns out to be 6 different possible results. This is called the *Fundamental Rule of Product* or the *Counting Principle*

Definition 1. *The Rule of Product states that if the 1st action can be performed in n ways, and the 2nd action can be performed in m ways. They can be performed together in $n \cdot m$ ways.*

What happens if the actions cannot be performed together? Such as which car to buy? In that case we have another rule, known as the *Rule of Sum*.

Definition 2. *The Rule of Sum states that if the 1st action can be performed in n ways, and the 2nd action can be performed in m ways, and these actions cannot occur together, then there are $n + m$ ways for either of the actions to occur.*

1.4. N-Arrangements

Let's say you have 5 people, and you need to arrange them in a line. You put one person in the first spot, and now you have 4 left. Repeat this until you are left with one person, which fits into the last spot. You can express this mathematically by writing out $5!$.

This is quite similar to the Rule of Product as:

1. Each placement is a choice
2. The number of choices is reduced with each previous action.

A permutation of n objects is an arrangement of the objects in a **definite order**. This can be expressed with permutation notation, shown below where $P(n,n)$ equals the number of ways to permute n objects.

$$P(n, n) = n!$$

This equation will only work when the number of spaces to occupy and the number of objects are equal. In other cases we'll have to use an r-arrangement.

1.5. R-Arrangements

If you have more objects than places you can put the objects into, then we can use something called an r-arrangement. An r-arrangement is a permutation of n objects taken r at a time, as is shown below.

$$P(n, r) = \frac{n!}{(n - r)!}$$

Where $P(n, r)$ equals the total number of arrangements you have, divided by the number of unused objects. This is most commonly displayed on calculators as nPr . For example you have 4 subjects, but you can only study two at a time. The equation for that would be $\frac{4!}{(4-2)!} = 12$ combinations.

1.6. Permutations with Repeating Elements

All previous examples will ONLY work with non-repeating elements, such as the letters in the word "MONTREAL", or 5 different books. They will definitely NOT work for questions with repeating elements, such as the word "FIJI". For all repeating elements we'll have to follow another rule shown below.

Definition 3. *In general, the number of arrangements of n objects of which a of one kind are alike, and b of one kind are alike and so forth is given by the expression below, Where the number of permutations can be found by dividing the total number of possible combinations by the repeating elements.*

$$\frac{n!}{a!b!c! \dots}$$

1.7. Problem Solving with Permutations

There are 3 methods for solving problems with permutations. They are called: the Indirect Method, the Case Method, and Circular Arrangements.

1.7.1. Indirect Method

The indirect method involves taking all possibilities and subtracting those which are not wanted. For instance, say you have 3 people that need to sit at a table, but 2 of them don't want to sit beside each other. There is a step-by-step process that you can use for this question that is explained below.

1. Find all possible permutations (ex. $P(3,3)$),
2. Find all unwanted permutations (in this case when they are together),
3. Subtract the unwanted permutations from the wanted permutations.

This can be written as:

$$\textbf{Wanted Outcomes} = \text{Total Outcomes} - \text{Unwanted Outcomes}$$

1.7.2. Case Method

The Case Method involves breaking down the equation into manageable parts, then adding them to get the final solution. There is a step-by-step example problem in Section 1 of the appendix, along with an explanation of why, and how you manage the "breaking into parts" aspect.

1.7.3. Circular Arrangements

Circular arrangements are no longer a part of the curriculum, and as such they will not be tested on, but they are still quite useful to know in niche cases. If you have seven people seated at a table, and they all move to their rights, their positions may have moved, but their order remains the same, like a circle. Keep in mind that if there is some fixed point, the question just ends into a typical linear arrangement, with the equation simplifying to $1 \cdot (n - 1)!$.

2. Combinations

2.1. What are Combinations?

A Combination is a way of selecting objects (permuting) from a group, without the order being important. In essence, it is orderless permutations. Let ABC be a permutation. With permutations we asserted that ABC and BAC are different, as they are ordered differently. In combinations, we assert that ABC and BAC are the same, as they contain the same objects, just in a different order.

2.2. Set Theory

Set Theory involves the creation and manipulation of sets and their properties. A set is a collection of distinct objects in which their order does not matter. The objects in this set are called elements. If a set does not have any elements in it, we refer to it as a *null set*, which is shown as \emptyset or (\emptyset) . This null set is a subset of everything. If we want to describe a set containing elements, we must follow the example below.

$$X = \{a, b, c, d, e, f, \dots, y, z\}$$

If we wish to describe the number of elements in the set, we can use $n(X)$, where X represents the set, and n returns the number of elements in that set. In this case $n(X) = 26$, as set X contains all the letters of the alphabet. Sets are referred to differently when compared to other sets, as defined below.

Definition 4. *Special Properties of Sets when Compared to Other Sets*

1. *If two sets have no elements in common they are called disjoint sets.*
2. *If two sets have all elements in common they are equal.*
3. *If all of elements of A are also in B then A is a subset of B ($A \subseteq B$)*

2.2.1. Universal Set and Complement Sets

The set of all elements being considered is called the *universal set* and is always denoted by S . When referring to a set of all elements that are in the universal set, but **not in set A**, we call this the complement of A or A^c . For example:

$$S = \{1, 2, 3, 4\}, A = \{1, 3\}, A^c = \{2, 4\}$$

2.2.2. Unions and Intersections

For two sets of A and B , the **union** of A and B or $A \cup B$, is the set of all elements in A or B , not including duplicates.

$$S = \{1, 2, 3, 4\}, A = \{1, 3\}, B = \{1, 2\}, A \cup B = \{1, 2, 3\}$$

For two sets of A and B the **intersection** of A and B , $A \cap B$, is the set of all the elements that are in both set A and set B .

$$S = \{1, 2, 3, 4\}, A = \{1, 3\}, B = \{1, 2\}, A \cap B = \{1\}$$

If you are ever having trouble remembering the difference between the symbols representing the union and the intersection, remember that a union looks like a cup, and intersection looks like a cap.

2.2.3. Subsets

Subsets, are sets that are also in another set. To find every possible subset (including null set) we need to use the equation:

$$2^{n(X)}$$

In this case, $n(X)$ is the number of elements in Set X . If a question comes up, where we need the number of subsets of a given set with the length y , of set x , we'd simply use $\binom{n(x)}{y}$

2.3. Introduction to Combinations

A combination is a selection of r objects, from n distinct objects *without regard for order*. The equation for solving combinations is written below.

$$C(n, r) = \frac{n!}{(n - r)! \cdot r!}$$

This equations finds all combinations by dividing number of objects, by the unused objects $(n - r)!$ and the order of the selected objects. This can be further simplified to:

$$C(n, r) = \frac{P(n, r)}{r!}$$

If order does not matter, the process is called a *Combination*, whereas when order does matter, it is called a *Permutation*. On your calculator, the formulae for combinations will be shown as **nCr**. If written out by hand, we will use the form:

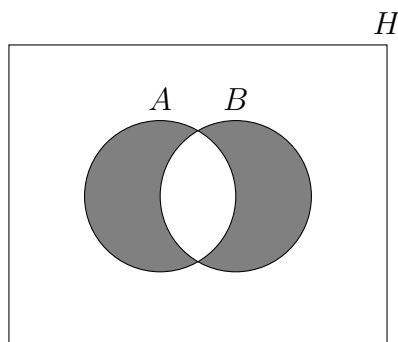
$$\binom{n}{r} = \frac{n!}{(n - r)! \cdot r!}$$

Or simply $\binom{n}{r}$.

2.4. Inclusion—Exclusion Principle

The Inclusion-exclusion principle is a method of obtaining the number of elements in finite sets. Typically used when asked for the number of combinations of two things, where there is overlap between the two.

Let A represent Set A, and let B represent Set B.



This diagram perfectly represents the overlap between the two sets. If you caught it earlier, what we're actually finding is $A \cup B$. The Inclusion—Exclusion Principle is merely a way of finding it algebraically. The image above represents the equation below.

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

What happens if we have have three events we need to find the union of though? To do that, we'll follow the same principles, and use the equation below.

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

3. Modelling with Matrices

3.1. Intro to Matrices

A matrix is a rectangular array of numbers that is used to organize data. If you are used to computer programming, you will recognize them as 2D arrays. If you are good with set notation, they are in essence, a set of sets. Matrices follow the format below:

$$X = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

Matrices have various properties, and unique terms created to describe them. For instance, looking at the matrix above, we can safely say that a is an entry, as defined below.

Definition 5. *Entry: An entry refers to a number in a matrix.*

How did we figure out the number of entries in the matrix? Simple, we can use the matrix's *dimensions* to figure that out.

Definition 6. *Dimensions: These refer to the size of the matrix, described by the number of rows and columns. The formula below is used to solve for dimensions, where D is the result, R is the rows, and C is the columns.*

$$D = R \cdot C$$

Matrix's entries are numbered in a way so that it is easy to refer to a specific entry. Capital letters are used for the names of matrices, and lowercase characters are used for the names of entries. Shown below are the entries named, with numbers instead of letters. The first character represents the y value, whereas the second represents the x value. In the equation below, a_{13} represents the entry in row 1, column 3.

$$X = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

3.2. Variations and Classifications of Matrices

Matrices can be transformed, and compared to each other, much like sets. The definitions of the variations and classifications of the matrices are listed below.

Definition 7. *Transpose Matrix:* A transpose matrix is a matrix obtained by interchanging the rows and columns. This is written as X^+ . Shown below is Matrix W , and the transposed equivalent.

$$W = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \rightarrow X^+ = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

Definition 8. *Column Matrix:* A Column Matrix is a matrix that has only one x value. Instead of rows and columns, it is merely a column. For instance, Matrix X in this case would be a Column Matrix.

$$X = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Definition 9. *Row Matrix:* A Row Matrix is a matrix that has only one y value. It is the transpose of the Column Matrix. For instance, Matrix Y in this case would be a Row Matrix.

$$Y = [a \quad b \quad c]$$

Definition 10. *Square Matrix:* A Square Matrix is a matrix in which the rows and columns are equal. For instance, Matrix Z below is a Square Matrix.

$$Z = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

3.3. Basic Operations with Matrices

Matrices may only be added or subtracted when they have the same dimensions. Adding matrices is done by adding entry by entry to the other matrix. A_{12} will be added with B_{12} and so forth. Shown below is a valid example of adding with matrices in which the entries are labeled to show proper addition:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} (a+b)_{11} & (a+b)_{12} & (a+b)_{13} \\ (a+b)_{21} & (a+b)_{22} & (a+b)_{23} \end{bmatrix}$$

They may also be multiplied by a coefficient, by multiplying the coefficient with every entry in the matrix. Shown below is a proper solution in which a matrix is multiplied by a coefficient, with entries labeled:

$$x \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix} = \begin{bmatrix} xy_{11} & xy_{12} & xy_{13} \\ xy_{21} & xy_{22} & xy_{23} \end{bmatrix}$$

3.4. Multiplying Matrices Together

Matrices may also be multiplied with other matrices, yet to do so we must follow certain rules. Let's say we have Matrix A and Matrix B as shown below. To multiply, we need to make sure the *inner dimensions* are the same. We know that Matrix A's dimensions are $3 \cdot 2$, whereas for Matrix B, they are $2 \cdot 3$. Since we know this, we can find the *inner dimensions*. The *inner dimensions*, equals the columns of Matrix A and the rows of Matrix B. In this case, that would work out to be 2 and 2. Since they are the same, we are allowed to multiply these matrices.

$$A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} B = \begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix}$$

Next, the rows of Matrix A times the columns of Matrix B, also known as the *outer dimensions*, will give us the dimensions of the resultant matrix. Since Matrix A has two rows, and Matrix B has two columns, our resultant Matrix will possess the dimensions $2 \cdot 2$. Shown below is the updated equation.

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \cdot \begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

Multiplying them is a bit strange at first, because to do it, we multiply row by row. Starting at the first row in Matrix A, we can see we have a_{11}, b_{12}, c_{13} as our values, with their locations being labeled for convenience. We must multiply these values each column in Matrix B, starting with the values u_{11}, w_{21}, y_{31} . To find entry (1,1) we can use the formula shown below, where E represents the entry, r represents a row in Matrix A, and c represents a column in Matrix B.

$$E_{11} = r_{11} \cdot c_{11} + r_{12} \cdot c_{21} + r_{13} \cdot c_{31}$$

If you want to do it just by eye, you can multiply A_{11} by B_{11} , A_{12} by B_{21} , and A_{13} by B_{31} , and add them all up. It will seem totally awful at first, but after you do a couple it'll all work out fine.

3.5. Identity Matrices

An Identity Matrix is a $n \cdot n$ size matrix in which every number along the main diagonal is a 1, and the rest are zeros. Shown below is Matrix I,

a $3 \cdot 3$ Identity Matrix.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This matrix is very special, as any square matrix multiplied by its inverse will equal an Identity Matrix. We can further formalize this assumption, by referring to a given matrix as Matrix A.

$$A \cdot A^{-1} = I$$

Its too bad we can't divide matrices, how can we ever solve for I now? Good thing we can use something called an Inverse Matrix.

3.6. Finding Inverse Matrices

Finding the inverse of large matrices is one of the most mathematically intensive processes out there. Good thing we'll only be finding the inverse of a $2 \cdot 2$ matrix instead. To find the inverse of a $2 \cdot 2$ matrix, we must use first define the matrices.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Now that we have the starting matrix, it is time to find the inverse. To find it, we must multiply the determinant, by the matrix, and swap several spots of numbers. Again, sounds odd, but we just need to follow the formula below.

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

After this step, simply multiply the determinant into the matrix, and you have found A^{-1} !

4. Probability

4.1. What is probability?

Probability is the measurement of the chance of an event occurring. If you've ever checked the weather and seen that it's a "20% chance of rain", you've seen probability in action. To start, we'll need to define some terms.

Definition 11. *Outcomes: A possible result, typically referencing all of the possible results from the situation.*

Definition 12. *Events: These typically refer to desired outcomes*

Typically when finding probability, we'll divide all possible outcomes from our desired outcomes. Thinking in terms of set notation, we'll refer to all of our desired outcomes as $n(A)$, where A is our desired outcome. $n(S)$ will refer to all of the possible outcomes, and $P(A)$ referring to the likelihood of A occurring. To generalize, we'll use the equation below.

$$P(A) = \frac{n(A)}{n(S)}$$

Since the number of desired outcomes will never be greater than the number of total outcomes, we'll always have a decimal between 0 (0%) and 1 (100%), which we can convert to a percentage (ex. $\frac{1}{4} = 0.25 = 25\%$), though typically we'll simply leave it as a reduced version of the fraction.

4.2. Mutually Exclusive Events and Non-Mutually Exclusive Events

Mutually Exclusive events are events that have zero overlap with other events. Recall the Venn Diagram of Sets in the Combinations section, and assume A and B are both events. To find the probability of either of them happening, we'll have to remember the equation:

$$n(A \cup B) = n(A) + n(B)$$

This equation finds all the ways A or B can happen, with zero overlap. The problem with this equation is that it only finds the number of wanted outcomes. Remember, to find the probability of something, we find the wanted outcomes, and divide it by all of the possible outcomes. After this, our equation will look like this:

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A) + n(B)}{n(S)}$$

Which reduces to:

$$P(A \cup B) = P(A) + P(B)$$

For non-mutually exclusive events, we must instead use a different equation, more accurate to the one we've discussed in the Sets Unit. Following the same logic we used earlier, our equation works out to:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Where $P(A \cap B)$ equals the probability of both of those events occurring.

4.3. Independent Probability

When two events occur at the same time, or one after another with no effect on each other, they are called Independent Events. If you roll a die, then select a card, this would be independent, as the outcomes don't affect each other. So solve an Independent Probability question we'll have to think in terms of sets, so the probability of A and B, can be solved with the equation below.

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

But what happens if we need to find the probability of (A or B)? To find that we'll need to use another equation.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

4.4. Markov Chains

Markov Chains are ways of predicting probability after successive iterations. Suppose that Product A has a 70% chance of repurchasing, and Product B has a 70% chance of repurchasing, how would we model this? We'd use a Markov Chain. Shown below is our probability matrix.

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

Now, we have to find the initial probability of something, which we call a **transition matrix**. We'll represent this in the form of how many people started buying A or B, (shown below). Since this is the initial time, we'll refer to it as S_0 , though you may also hear it called a **probability vector**.

$$S_0 = [0.6 \quad 0.4]$$

To find the next probability vector, or S_1 , we'll have to multiply the transition matrix by the probability vector.

$$[0.6 \quad 0.4] \cdot \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} = [0.54 \quad 0.46] = S_1$$

Much like rolling a die, the more times you repeat this process, the closer you are to the actual results. Though what if there was an easier way to do this?

4.5. Steady State Vectors

In Markov Chains, the probability vectors will eventually stop changing. A probability vector that remains unchanged upon multiplication is called the **steady state vector**. This vector will represent the long term trend of the event. To find this without repeating the Markov Chain multiplication, we must think of the problem in terms of a system of equations.

$$\begin{bmatrix} a & b \end{bmatrix} \cdot \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix}$$

To have a systems of equations from this, we must first assert that $a+b = 1$, and find the other by multiplying for one of the values. In this case, we can see that $0.7a + 0.3b = a$. Now solve this systems of equations like you have done in Grade 9.

4.6. Odds

Have you ever heard the (seemingly wrong) Roll Up The Rim odds? They claim that 1 in 5 people are winners! But what does this mean? Odds are the degree of confidence that someone has that an event will occur. To find the odds of something occurring, we must use the formula below.

$$P(A) : P(A)'$$

This equation represents a ratio of the probability of A happening, against the probability of A **not** happening. Likewise with the odds of something **not occurring**, we can use the reverse of this formula.

$$P(A)' : P(A)$$

What happens if we want to win a bet though? You've probably heard of betting odds before. To do that, we'll use the equation below, where X will equal the amount won.

$$\frac{P(A)}{P(A)'} = \frac{\text{Amount Bet}}{x}$$

After this, isolate X, which will give you the following equation.

$$X = \text{Amount Bet} \cdot \frac{P(A)'}{P(A)}$$

5. Probability Distributions

5.1. What is a probability distribution?

A probability distribution is a method of displaying data that displays the probabilities of every possible outcome. We'll need to define a few types of distributions first though, which are uniform and non-uniform distributions.

Definition 13. *Uniform Distribution:* A distribution is called uniform when all of the outcomes are equally likely, just like a single die roll.

Definition 14. *Non-Uniform Distribution:* A distribution is called non-uniform when not all of the outcomes are equally likely, just like rolling the sum of a pair of dice.

The variable we measure when calculating these, is called the random variable, and is denoted as X . The various possibilities for X are the outcomes, denoted x . We'll typically find the probability distributions that rely on two different types of variables: discrete and continuous variables.

Definition 15. *Discrete Variables:* These are values that are typically separate from each other, and have a finite number of possibilities.

Definition 16. *Continuous Variables:* These are values that have an infinite number of possibilities, such as a length of time.

5.2. Sigma Notation & Regular Probability Distributions

Sigma Notation, or \sum , may look terrifying, but it simply means the sum of a lot of numbers. We use this when finding an expected value of something, which we'll denote as $E(x)$. To find this, we'll use the equation below.

$$E(x) = \sum_{i=1}^n x_i P(x_i)$$

To break it down, the $i = 1$ is the start, the n is the stop, x_i is the outcome, and $P(x_i)$ is the possibility of that outcome. To show this in action, let's suppose that in a game, you'll get the amount of points you'll roll on a die. To find that, we'll need to use find $E(x)$. We know that the probability of rolling each number is $\frac{1}{6}$, and we know all the outcomes, so let's fill out the equation. First, we'll need to set our limits though.

$$E(x) = \sum_{i=1}^6$$

This means that we must repeat the $x_i P(x_i)$ starting at one, going to six, which would be six times. Now we'll expand that below.

$$E(x) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

We can also simplify this, as we know it is a uniform distribution. Shown below is the updated equation for this uniform distribution.

$$E(x) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6)$$

5.3. Binomial Distribution

Remember binomials from combinations? They look like this $\binom{x}{y}$, and come in handy when working with distributions. We have to mention something called a Bernoulli Trial first.

Definition 17. *Bernoulli Trials: This refers to repeated, independent trials that are measured in terms of success, or failure (or other Boolean outcomes).*

The probability distribution of the number of successes in Bernoulli Trials is called a binomial distribution, and is found using the equation below.

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

The equation solves for the probability of exactly x successful trials out of n total trials. In this equation, p represents the probability of success, and q represents the probability of failure. To find the expected value for a binomial distribution, we'll resort to using the formula below.

$$E(x) = n \cdot p$$

5.4. Geometric Distributions

A geometric distribution is found when an experiment is repeated just **until the first success**. It differs from binomial distributions, as the number of trials is not known at the start. Just before the success, there is something called a "waiting period".

Definition 18. *Waiting Period: This is the number of failures before the first and only success.*

If you have 8 failures before your success, then you have a waiting period of 8. To solve for the probability of success after x failures, we can use the formula below. Remember, x represents the waiting period.

$$P(x) = q^x p^1$$

In this equation, the variable p **must have a max of one** as you can never have more than one success in a Geometric Distribution. To find your expected waiting period, you must solve for $E(x)$, which you can do by using the formula below.

$$E(x) = \frac{q}{p}$$

5.5. Hyper—Geometric Distribution

Hyper-Geometric Distribution are quite similar to Geometric Distributions, yet they are dependent, rather than independent. Repeated sampling without replacement exhibits this type of process. To find the probability of this, we must use the equation below, where n represents the total number of choices, r represents the total number of items chosen, a represents the number of successes to choose from, and x represents the successes chosen.

$$P(x) = \frac{\binom{a}{x} \cdot \binom{n-a}{r-x}}{\binom{n}{r}}$$

We can also find the number of expected successes chosen by using the formula below, using the same variables as in the equation above.

$$E(x) = r \cdot \frac{a}{n}$$

6. Single Variable Statistics

6.1. Intro to Statistics

This is a massive shift compared to the rest of the stuff we've done so far. It is a lot of (relatively simple) graphing, and data management, and also includes a lot of spreadsheet work (which I may or may not include). Using a frequency table will help you, as well as attention to detail.

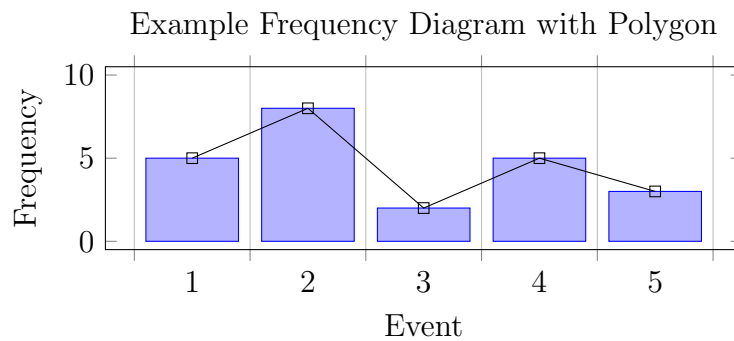
6.2. Displaying Frequency

To start, we're going to have to (re)define some stuff you have learned in earlier years. This will seem like a lot at first, but most of it is quite basic, and there are no equations to memorize.

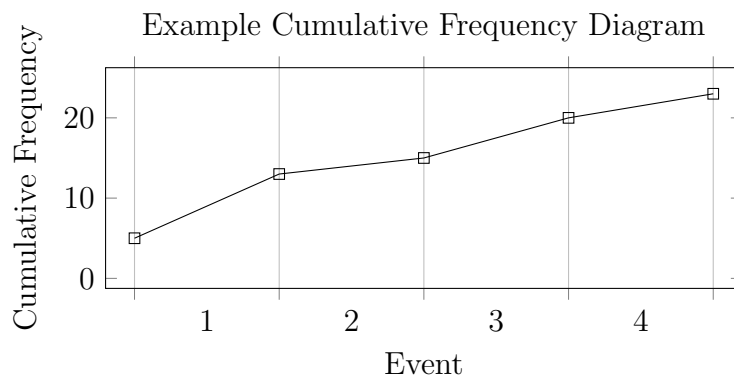
Definition 19. *Range:* A range, in relation to sets of data, is how far apart the data points are spread. It is calculated by taking the largest number, and subtracting it by the smallest.

Definition 20. *Frequency Diagram:* A frequency diagram is a graph used to show the frequency of various points, or intervals of data (ex. years, or decades).

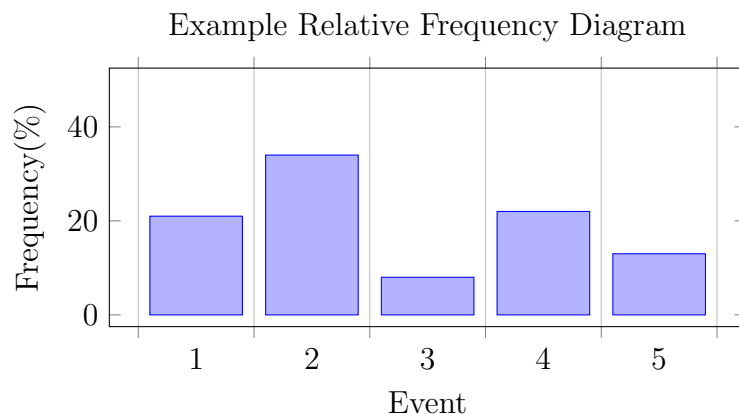
Definition 21. *Frequency Polygon:* A shape that helps the reader understand the overall shape of the diagram or distribution.



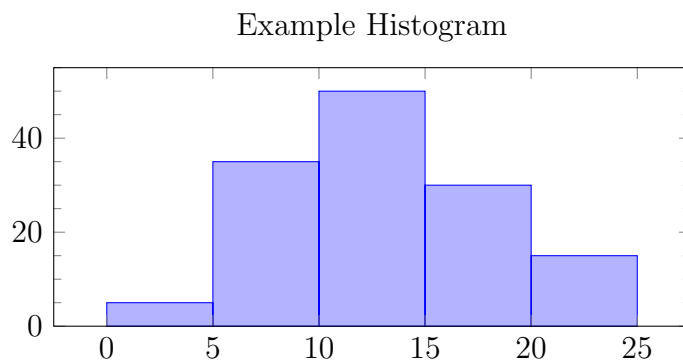
Definition 22. *Cumulative Frequency:* This refers to the running total of frequencies. Shown below is an example of once such diagram.



Definition 23. *Relative Frequency: This means the frequency expressed as a percentage, found by dividing the frequency by the total number of results.*



Definition 24. *Histogram: Essentially a bar graph, yet with the data in groupings, instead of being separate.*



All of this data and results above can be displayed and organized in something called a frequency table, as defined, and shown below.

Definition 25. *Frequency Table: This is a table that contains all of the above results, as well as tallies of data points.*

(x)	Frequency	Relative Frequency	Cumulative Frequency
1	5	$\frac{5}{23}=21\%$	5
2	8	$\frac{8}{23}=34\%$	13
3	2	$\frac{2}{23}=8\%$	15
4	5	$\frac{5}{23}=22\%$	20
5	3	$\frac{3}{23}=13\%$	23

6.3. Measures of Central Tendency

Measures of Central Tendency deal with the various numbers we use to find data points we can make assumptions based on. These three numbers we use are called the **mean, median and mode**.

Definition 26. *Mean: The mean represents the average of the data, and is found by using the equation below.*

$$\bar{x} = \frac{\sum x}{n}$$

Definition 27. *Median: The median represents the middle-most point of data, for example in a set of 5 integers, the median will be the third. To find the median of an list of data of which **n is odd numbered** you can use the equation below.*

$$\tilde{x} = \text{Number in middle of data}$$

if the equation is even numbered, you must use the average of the numbers that the median is in between. For example if you have a list of 8 numbers, the median would be the average of 4 and 5. The equation for this can be seen below.

$$\tilde{x} = \frac{\text{Left Number} + \text{Right Number}}{2}$$

Definition 28. *Mode: The mode represents the data point that has been repeated the most often. Can be found by merely looking at the highest repeated data set.*

6.4. Measures of Spread

A measure of spread is a measurement that describes how the data is distributed, and how widely it fluctuates. To measure these, we divided the spread of the data into something called *quartiles*. A quartile, just like its name implies, is merely a quarter of the data. The first Quartile (referred to henceforth as Q1) is the median of the **FIRST HALF** of the data, Q2 is the middle (or median), and Q3 is the median of the **LAST HALF** of the data. If we want to find the most relevant data, we'll typically look at the data between Q1 and Q3, this range of data is called the Inter Quartile Range (IQR).

Definition 29. *Inter Quartile Range: This is the middle-most data, and a grouping you can easily make assumptions from. To find this, we use the equation below.*

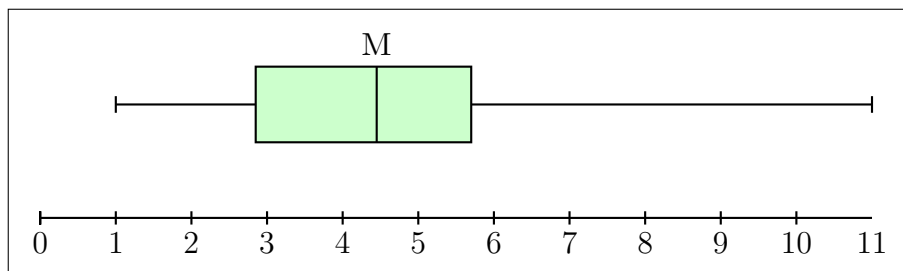
$$IQR = Q3 - Q1$$

Using this, we can also find other data to make assumptions off of, such as the Semi-Inter Quartile Range.

Definition 30. *Semi Inter Quartile Range: This is $\frac{1}{2}$ of the difference of the IQR, and therefore is less affected by outliers. Can also be referred to as the SIQR, and is found by the equation below.*

$$SIQR = \frac{IQR}{2}$$

We can display these using something called a **box and whisker plot**. This will show Q1, Q2, and Q3, along with the highest and lowest datum. An example of this is shown below.



6.5. Intro to Variance

To calculate the variance, we must first calculate the deviations. A deviation is the difference between an individual value in a data set and the mean of that set. To calculate the deviation, we'll need to use the equation below.

$$\text{Deviation} = x - \bar{x}$$

We use different kinds of deviations for various reasons, such as Mean Deviations (md) and Standard Deviations (σ). They are defined below, though most typically we will use the Standard Deviation. Most of these can be calculated much easier if you use the *STAT* mode on your calculator, but the equations will be below if your calculator doesn't support that.

Definition 31. *Mean Deviation:* The mean deviation is the mean of all the distances to the mean. To calculate this, we'll use the formula below to solve for md , which represents the mean deviation.

$$md = \frac{\sum |x - \bar{x}|}{n}$$

Definition 32. *Standard Deviation:* This is more widely used, and differs from the mean deviation in the fact that it uses square roots and exponents rather than the absolute value function. Will always be bigger than the mean deviation, and is shown as (σ) .

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Once we know the deviation, we can find the variance of various numbers.

Definition 33. *Variance:* This number represents the average of the squares of the deviations. You can find this by simply squaring the Standard Deviation (giving you σ^2), or by using the equation below.

$$Variance = \frac{\sum (x - \bar{x})^2}{n}$$

6.6. Z—Scores

A Z-Score is the number of Standard Deviations that a datum is from the mean. We can find this using the equation below.

$$z = \frac{x - \bar{x}}{\sigma}$$

If $z < 0$ the element is less than the mean, if $z = 0$, the element is equal to the mean, and if $z > 0$, the number is greater than the mean.

6.7. Deviations with Intervals Data

When data is grouped into intervals, the formula for mean deviation and standard deviation change. Mean Deviation changes to the equation below.

$$md = \frac{\sum f|x - \bar{x}|}{n}$$

Standard Deviation also changes into the formula below.

$$\sigma = \sqrt{\frac{\sum f(x - \bar{x})^2}{n}}$$

In both of these equations, f represents the frequency of the interval.

6.8. Chebyshev's Theorem

Chebyshev's Theorem states that given a set of data with a mean (designated \bar{x}) and a Standard Deviation (σ), then **at least** $x\%$ of the will be within k standard deviations of the mean. We find this using the equation below.

$$x = (1 - \frac{1}{k^2})\%$$

6.9. Gathering Data

This subsection is entirely definitions, and as such is quite lengthy. It is tested on, but I believe it is multiple choice. As such this is semi-important if any research or actual studies are ever done in your class, or in the real world.

Definition 34. *Population:* Refers to individuals who belong to a group being studied.

Definition 35. *Sample:* A subset of a population

Definition 36. *Simple Random Sample:* Every member of the population has an equal chance of being selected.

Definition 37. *Systematic Sample:* Selected by iterating through the population and selecting members at regular intervals.

Definition 38. *Stratified Sample:* A group of population members with similar characteristics is called a stratum. A stratified sample has the same proportion of members from each stratum.

Definition 39. *Cluster Sample:* When certain groups are representative of the entire population, this takes a random selection of such groups.

Definition 40. *Multistage Sample:* Uses several levels of random sampling, randomly sample from groups within groups.

Definition 41. *Voluntary Response Sample:* Researcher invites any member of population to participate. Results skewed as people who choose to participate not always representative of the population.

Definition 42. *Convenience Sample:* Easily accessible sample of the population.

Definition 43. *Bias:* Any factor that favors certain outcomes or responses

Definition 44. *Sampling Bias: Occurs when the sampling frame does not reflect the characteristics of the population.*

Definition 45. *Non-response Bias: Occurs when particular groups are under-represented because they choose to not participate.*

Definition 46. *Measurement Bias: Occurs when the data collection method consistently misrepresents the characteristics of a population.*

Definition 47. *Response Bias: Occurs when the participants in a survey deliberately give false/misleading answers.*

7. Two-Variable-Stats

7.1. Differences compared to Single Variable Stats

Two variable is used quite widely, and can be used to measure the effects of various things, rather than just comparing frequency as we have previously done. We can compare both x and y values, giving us a lot more freedom, but a lot more complex graphing.

7.2. Linear Correlation

Linear correlation is used to find relationships between two variables. We plot or graph this to easily visualize the results, and relationships. We often use this to discern if the variations in one variable (independent) affect the other (dependent). This correlation will be either known as a positive correlation, or a negative correlation.

Definition 48. *Positive Correlation: A correlation is positive if the dependent variable increases when the independent variable is increased.*

Definition 49. *Negative Correlation: A correlation is negative if the dependent variable decreases when the independent variable is increased.*

These correlations also have various **weights**, and we can use an equation to find that; but first we must define what a weight is.

Definition 50. *Relationships: These relationships show how close the line of best fit is to the points. These can be either defined as weak, moderate, and strong, and can be found by using the correlation coefficient.*

Oh no! What's the correlation coefficient? Earlier, we said we can find the weight using an equation, and this equation is known as the **correlation coefficient**.

Definition 51. *Correlation Coefficient:* This is a measurement of how strong the correlation is between two variables. Typically designated as r , we can find it by using the equation below.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

This correlation is always between -1 and 1. If it is positive, it is a positive correlation, and the opposite is true if it is a negative. You can classify a relationship type by using this number, and these following inequalities.

- Strong correlation when: $\frac{2}{3} < r < 1$
- Moderate correlation when: $\frac{1}{3} < r < \frac{2}{3}$
- Weak correlation when: $0 < r < \frac{1}{3}$

If the correlation coefficient (henceforth referred to as r) is not what you'd expect, you can always improve the accuracy by increasing the number of data points. Keep in mind that you also need to follow one of the commandments of statistics:

A STRONG CORRELATION DOES NOT NECESSARY IMPLY CAUSATION.

7.3. Linear Regression

Regression (in general) is defined as "an analytical technique for determining a relationship between a dependant, and an independent variable". When two variables *appear* to have a linear correlation, you can find the line of best fit. When the correlation is strong (recall: when $\frac{2}{3} < r < 1$), it is easy to estimate this line, but when any other time it is nigh impossible. Luckily, mathematicians don't like doing difficult things, so we have an equation for this line.

We begin with $y = mx + b$, the standard linear formula. For a , we use the following equation:

$$a = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

Then for b , we'll use the next equation.

$$b = \bar{y} - a\bar{x}$$

7.4. Other Types of Regressions

Unfortunately, not all relationships are linear, in fact most are exponential, or polynomial. We can define certain growths and decays as various types defined below. Please keep in mind they are almost all done on the graphing calculator, and as such do not need to be memorized. If you have taken MHF4UI or MHF3UI, you will almost certainly know what most of these are.

Definition 52. *Exponential Growth/Decay: This has an y_{int} of a , and an equation of:*

$$y = ab^x$$

Definition 53. *Linear Growth/Decay: This has a y_{int} of b , and an equation of:*

$$y = ax + b$$

Definition 54. *Quadratic Growth/Decay: This has a y_{int} of c , and an equation of:*

$$y = ax^2 + bx + c$$

Definition 55. *Cubic Growth/Decay: This has a y_{int} of d , and an equation of:*

$$y = ax^3 + bx^2 + cx + d$$

Definition 56. *Power Growth/Decay: This has a y_{int} of 0 , and an equation of:*

$$y = ax^b$$

7.5. Selecting a Regression

To select a regression, we must take into account various factors.

- r^2 : Coefficient of determination
- Shape of the data.
- Starting Value.
- Ending Value.

7.6. Cause and Effect Relationships

The main reason correlations are useful to us is for finding evidence of a cause/effect relationship. **Remember: A strong correlation doesn't prove causation.** There are various types and degrees of causal relationships between variables.

Definition 57. *Cause and Effect Relationships: Exists when a change in X produces a guaranteed change in Y . Physical processes display this type of relationship.*

Definition 58. *Common Cause Factor:*

Definition 59. *Reverse Cause and Effect Relationship:*

Definition 60. *Accidental Relationships:*

Definition 61. *Presumed Relationships:*