Research Data on Anime Dataset Regarding Statistics

Melvin Vazquez Andujar

April 26, 2024

Table of Contents

[Abstract: 3](#_Toc165096544)

[Chapter 2 3](#_Toc165096545)

[Theorem 2.1 5](#_Toc165096546)

[Theorem 2.2 5](#_Toc165096547)

[Theorem 2.3 5](#_Toc165096548)

[Theorem 2.4 5](#_Toc165096549)

[Theorem 2.5 5](#_Toc165096550)

[Theorem 2.6 5](#_Toc165096551)

[Theorem 2.7 5](#_Toc165096552)

[Theorem 2.8 5](#_Toc165096553)

[Theorem 2.9 5](#_Toc165096554)

[Definition 2.6 5](#_Toc165096555)

[Definition 2.7 6](#_Toc165096556)

[Definition 2.8 6](#_Toc165096557)

[Definition 2.9 6](#_Toc165096558)

[Definition 2.10 6](#_Toc165096559)

[Definition 2.11 6](#_Toc165096560)

[Chapter 3 6](#_Toc165096561)

[Theorem 3.1 6](#_Toc165096562)

[Theorem 3.2 6](#_Toc165096563)

[Theorem 3.3 6](#_Toc165096564)

[Theorem 3.4 6](#_Toc165096565)

[Theorem 3.5 6](#_Toc165096566)

[Theorem 3.6 6](#_Toc165096567)

[Theorem 3.7 6](#_Toc165096568)

[Theorem 3.8 6](#_Toc165096569)

[Theorem 3.9 6](#_Toc165096570)

[Theorem 3.10 6](#_Toc165096571)

[Theorem 3.11 6](#_Toc165096572)

[Theorem 3.14 6](#_Toc165096573)

[Definition 3.2 6](#_Toc165096574)

[Definition 3.3 7](#_Toc165096575)

[Definition 3.4 7](#_Toc165096576)

[Definition 3.5 7](#_Toc165096577)

[Definition 3.6 7](#_Toc165096578)

[Definition 3.7 7](#_Toc165096579)

[Definition 3.8 7](#_Toc165096580)

[Definition 3.9 7](#_Toc165096581)

[Definition 3.10 7](#_Toc165096582)

[Definition 3.11 7](#_Toc165096583)

[Chapter 4 7](#_Toc165096584)

[Theorem 4.1 7](#_Toc165096585)

[Theorem 4.2 7](#_Toc165096586)

[Theorem 4.3 7](#_Toc165096587)

[Theorem 4.4 7](#_Toc165096588)

[Theorem 4.5 7](#_Toc165096589)

[Theorem 4.6 7](#_Toc165096590)

[Theorem 4.8 7](#_Toc165096591)

[Theorem 4.10 7](#_Toc165096592)

[Definition 4.1 7](#_Toc165096593)

[Definition 4.2 8](#_Toc165096594)

[Definition 4.3 8](#_Toc165096595)

[Definition 4.5 8](#_Toc165096596)

[Definition 4.6 8](#_Toc165096597)

[Definition 4.9 8](#_Toc165096598)

[Definition 4.11 8](#_Toc165096599)

# Abstract:

This report will be based on a data set found from a website called Kaggle on Anime. The data found in the CSV will be used to get results from inputting that data into formulas from my Probability and Statistics class. The results will give a deeper understanding of the data set and will help better understand the dataset.

# Chapter 2

For chapter 2 there is not many equations that can be used for the data set, however, when looking at the data we see that each category has multiple options when looking at each anime. For example, when we look at a single anime, we also see the source that it came from, the studio that animated it, and which company licensed the anime. We can make a sample space that every anime has three options when it comes to source, either it is anime original or it is directly based on the source material, which would be the manga, or it is based on the light novel. So an example would be   
  
S = {O}, S = {M}, S = {L} – O = Original M = Manga L = Light Novel

From those three options we know we have a one-third chance of it being one of those. Additionally, this list can become even more complicated as you add more options and more categories like genre, type of anime, studios, etc.

For Theorem 2.2 that deals with Permutation, we can use example 2.8 as a guide on how to come up with a similar problem for this dataset. A way we can do it is putting 100 anime inside of a hat and drawing 6 different anime shows from that hat. By doing that we can use the permutation formula because we now have the n, which is 100 and we have the r, which is the amount drawn.

That is the total number of sample points from that problem alone.

Another formula we can use is the Combinations formula and this can be used in correlation with chances of picking 3 anime shows from a set of 30. We can use this to find the number of combinations that can occur when choosing 3 anime shows from a set of 30 books.

There are 4060 combinations that can occur when picking 3 anime shows from a set of 30.

When using conditional probability, we can say that P(A) = .33 since there is a 1/3 chance of an anime coming from one of 3 different sources, P(B) can be .33 that there is only 1/3 chance of the anime being either a movie, tv show, or an ova. Since these two events are independent of each other, the chances of an anime being both ova and manga is 1/9 since P(A) is independent from P(B). This can be found in definition 2.10

### Theorem 2.1

### Theorem 2.2

### Theorem 2.3

### Theorem 2.4

### Theorem 2.5

### Theorem 2.6

### Theorem 2.7

### Theorem 2.8

### Theorem 2.9

### Definition 2.6

### Definition 2.7

### Definition 2.8

### Definition 2.9

### Definition 2.10

### Definition 2.11

# Chapter 3

### Theorem 3.1

### Theorem 3.2

### Theorem 3.3

### Theorem 3.4

### Theorem 3.5

### Theorem 3.6

### Theorem 3.7

### Theorem 3.8

### Theorem 3.9

### Theorem 3.10

### Theorem 3.11

### Theorem 3.14

### Definition 3.2

### Definition 3.3

### Definition 3.4

### Definition 3.5

### Definition 3.6

### Definition 3.7

### Definition 3.8

### Definition 3.9

### Definition 3.10

### Definition 3.11

# Chapter 4

### Theorem 4.1

### Theorem 4.2

### Theorem 4.3

### Theorem 4.4

### Theorem 4.5

### Theorem 4.6

### Theorem 4.8

### Theorem 4.10

### Definition 4.1

### Definition 4.2

### Definition 4.3

### Definition 4.5

### Definition 4.6

### Definition 4.9

### Definition 4.11

[Anime Dataset 2023 (kaggle.com)](https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset?resource=download&select=anime-dataset-2023.csv)