Data Management

MDM4U

Qinghao Hu

November 4, 2025

Contents

1 Unit 1 **2**
- 1.1 Lecture 1 . . . 2
  - 1.1.1 The Fundamental or Multiplicative Counting Principle . . . 2
  - 1.1.2 Additive Counting Principle . . . 2
- 1.2 Lecture 1.2 . . . 2
- 1.3 Like Term Permutations . . . 2
- 1.4 Pascal Triangle . . . 3
- 1.5 Venn Diagrams . . . 3
- 1.6 Combination . . . 3

2 Unit 2, Probability **4**
- 2.1 Probability . . . 4
  - 2.1.1 Definitions of Probability . . . 4
  - 2.1.2 Definitions of other stupid stuffs . . . 4
  - 2.1.3 Formulas . . . 5
  - 2.1.4 Examples . . . 5
- 2.2 Dependent Events . . . 5
  - 2.2.1 Definitions . . . 5
- 2.3 Mutually Exclusive Events . . . 5
  - 2.3.1 Some boring Definitions . . . 5

3 One Variable Statistics **7**
- 3.1 Variables and Data . . . 7
  - 3.1.1 Definitions . . . 7
- 3.2 One Variable Graphs . . . 8
  - 3.2.1 Some Definitions . . . 8
- 3.3 Central Tendency . . . 10
  - 3.3.1 Definitions of Central Tendency . . . 10
- 3.4 Standard Deviation . . . 11
- 3.5 Quartiles . . . 12
  - 3.5.1 Definitions . . . 12
  - 3.5.2 Percentiles . . . 13
- 3.6 Spread Grouped Data . . . 14
  - 3.6.1 For weighted data . . . 14
- 3.7 Collect data . . . 14
  - 3.7.1 Anonymous and not be anonymous . . . 14
  - 3.7.2 Survey Questions . . . 14
  - 3.7.3 Questions should be avoided . . . 14
  - 3.7.4 Experimental vs Observational study . . . 15
- 3.8 sampling . . . 15
  - 3.8.1 Some type of sampling . . . 15
- 3.9 Bias . . . 16

4 Two variable Statistics **17**
4.1 Linear Regression 17
4.1.1 Residual 17
4.1.2 Regression 17

Chapter 4

Two variable Statistics

4.1 Linear Regression

4.1.1 Residual

Definition 4.1.1. *(Residual): It is the difference between the observed value and the value predicted by the best fit line. (Ie. how "far off" is the line, vertically, from the point)*

Points abouve the line have **positive** residuals and points below the line have negative residuals.

There are some really good sample questions on teacher's handouts. You should check it though!

4.1.2 Regression

Definition 4.1.2. *(Linear Regression): an analytic technique to determine a model that can used to describe the linear correlation between two quantitative variables.*

The linear model produced by linear regression is called a **least-squares** line.

For a least-square line, there are certain properties:

- the residuals add to 0 and the sum of the squares of the residuals is **Minimized**
- the slope is defined as

$$a = \frac{s_{xy}}{(s_x)^2}$$

- the y-intercept is defined as

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