**工程統計學**

**授課教授：林遠見**

**題目：通過字數統計對小說進行分析**

系級：工學院學士班三  
學號：109305521  
姓名:林義軒

日期：2023/6/

**目錄**

[摘要 3](#_Toc137912281)

[第一章 前言 4](#_Toc137912282)

[1.1 研究動機與目的 4](#_Toc137912283)

[1.2 研究方式 4](#_Toc137912284)

[1.3研究內容及資料來源 4](#_Toc137912285)

[第二章 實驗設計 5](#_Toc137912286)

[第三章 實驗結果 6](#_Toc137912287)

[第四章 結論 8](#_Toc137912288)

# 摘要

# 前言

## 1.1 研究動機與目的

本次實驗主要的目的是要探討能否透過統計的手段，以小說內文為分析對象，藉此觀察不同作品間的差異。

舉凡是文字組成的作品，理論上都可以作為此次探討的主體。但是若是這樣，題目便會過於寬泛，導致主題失焦。而由於我個人喜歡閱讀小說，因此我想以自身的喜好作為本次研究的主題。

會想做這個題目，主要是因為我個人喜歡閱讀小說，因此我想以自身的喜好作為本次研究的主題。閱讀不同的作品，會帶來截然不同的閱讀體驗。而我想深入探討的就是這個「體驗」的部份，因為體驗本身是一個相當主觀的概念，而在經歷的一學期統計學的學習後，我希望能透過統計的手段，去量化這個抽象的概念。

## 1.2 研究方式

## 1.3研究內容及資料來源

# 實驗設計

多樣性=(字種類數/字數)\*1000

# 實驗結果

四大小說母體分析

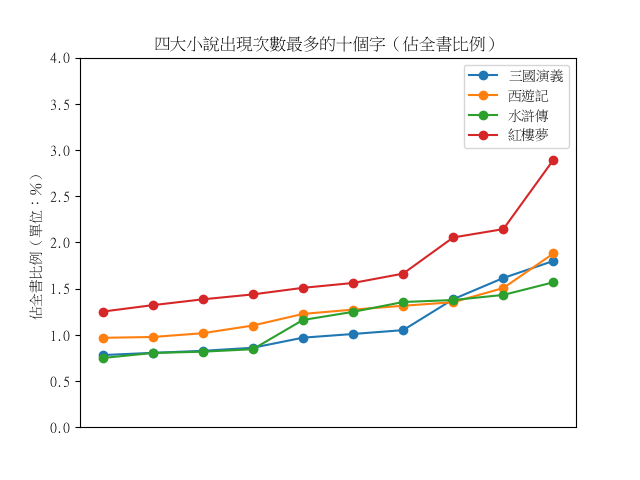
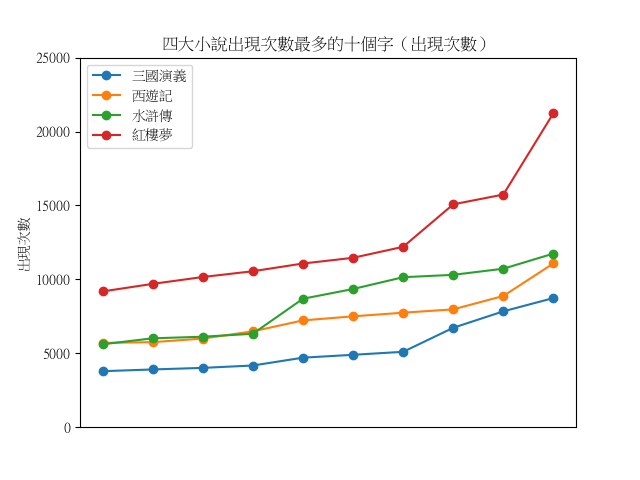
根據實驗設計中所述，下列是對四大古典小說進行母體分析的結果，其中多樣性為四捨五入到小數點後第二位之值。

|  |  |  |  |
| --- | --- | --- | --- |
| 小說 | 字數 | 字種類數 | 多樣性 |
| 三國演義 | 485394 | 4120 | 8.49 |
| 水滸傳 | 748791 | 4501 | 6.01 |
| 紅樓夢 | 734066 | 4544 | 6.19 |
| 西遊記 | 589290 | 4610 | 7.82 |

出現字數前十名

|  |  |
| --- | --- |
| 小說 | 出現字數前十名（由左至右排列） |
| 三國演義 | 曰、之、不、人、軍、兵、大、一、馬、將 |
| 水滸傳 | 了、道、一、來、人、不、是、個、上、去 |
| 紅樓夢 | 了、的、不、一、來、道、人、是、說、我 |
| 西遊記 | 道、不、一、了、那、我、是、來、他、個 |

前十名字數曲線



隨機取樣

|  |  |
| --- | --- |
| 小說 | 取樣段落 |
| 三國演義 | 1, 31, 38, 39, 44 |
| 水滸傳 | 15, 29, 59, 63, 64 |
| 紅樓夢 | 3, 6, 31, 58, 68 |
| 西遊記 | 11, 35, 42, 52, 58 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 之 | 曰 | 人 | 大 | 張 | 不 | 德 | 玄 | 軍 | 一 |
| 31 | 之 | 曰 | 公 | 關 | 不 | 一 | 人 | 大 | 將 | 兵 |
| 38 | 兵 | 曰 | 之 | 馬 | 軍 | 大 | 人 | 不 | 一 | 城 |
| 39 | 兵 | 之 | 曰 | 不 | 軍 | 大 | 明 | 馬 | 將 | 孔 |
| 44 | 兵 | 曰 | 之 | 人 | 馬 | 不 | 大 | 軍 | 將 | 一 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 15 | 武 | 了 | 人 | 來 | 松 | 一 | 道 | 那 | 不 | 大 |
| 29 | 了 | 來 | 道 | 不 | 一 | 我 | 人 | 個 | 是 | 去 |
| 59 | 道 | 一 | 兵 | 宋 | 不 | 來 | 關 | 將 | 馬 | 了 |
| 63 | 了 | 王 | 慶 | 的 | 那 | 一 | 道 | 個 | 人 | 不 |
| 64 | 一 | 了 | 那 | 人 | 來 | 王 | 慶 | 是 | 道 | 兵 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 | 一 | 不 | 了 | 人 | 之 | 的 | 是 | 有 | 來 | 玉 |
| 6 | 了 | 的 | 不 | 是 | 說 | 來 | 一 | 我 | 人 | 他 |
| 31 | 了 | 一 | 的 | 賈 | 不 | 人 | 是 | 說 | 這 | 道 |
| 58 | 了 | 的 | 不 | 人 | 說 | 來 | 是 | 道 | 這 | 玉 |
| 68 | 了 | 的 | 不 | 是 | 說 | 我 | 寶 | 道 | 一 | 人 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | 道 | 我 | 了 | 那 | 不 | 你 | 一 | 行 | 他 | 者 |
| 35 | 道 | 不 | 一 | 我 | 者 | 那 | 來 | 行 | 是 | 了 |
| 42 | 道 | 那 | 了 | 我 | 不 | 是 | 來 | 王 | 個 | 一 |
| 52 | 道 | 一 | 不 | 個 | 三 | 王 | 了 | 有 | 是 | 我 |
| 58 | 一 | 經 | 十 | 不 | 三 | 了 | 道 | 之 | 難 | 來 |

# 第四章 結論