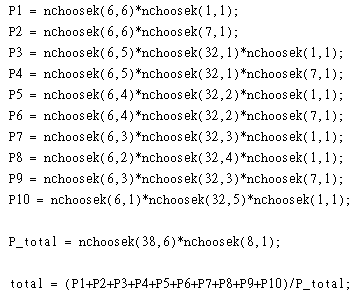
1. According to the winning rules, as shown in Figure 1, the combination of each prize can be written as:

For winning the second set, it is needed that the specific ball should be chosen, which only have 1 combination. But for not choosing the specific ball, the picked ball should be chosen from the others, which have combinations. Same for the first set, if the combination of choosing only m balls from the winning set is times choosing 6-m balls from the losing set .

The total combination of both sets are . By calculating the total probability, we can obtain:



The probability Calculated form MATLAB = 0.1178, near 0.11=

The assumption under this modeling is and iid assumption, meaning that every number has the identical probability to be chosen independently each time.

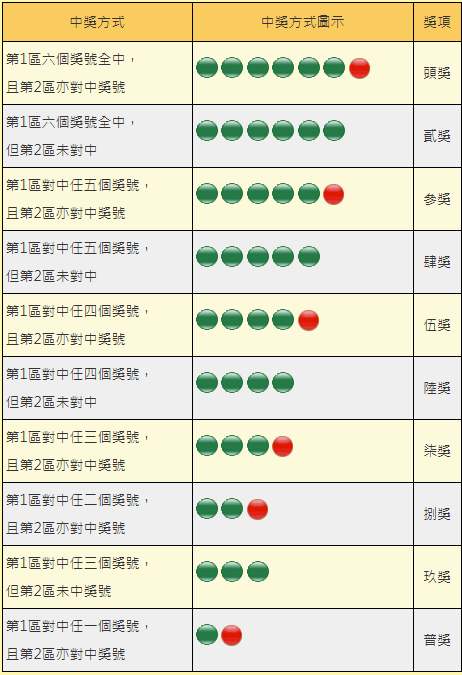


Figure 1



To check if events are independent or mutual exclusive, we first verify if the probability satisfies . If the equation is satisfied, then A and B are independent.

As we can see from the result, only event A and B are independent, both A,C and B,C are dependent.

For all the union of each two event are not empty set, neither of the two combination events are mutual exclusive.

1. From Figure 1, with the right pick from second-set number, we can have the chances to win prize 1, 3, 5, 7, 8, and 10. The condition probability is based on winning the second-set, which has a probability .

The condition probabilities of winning each prize are listed below:

1. (a) For data of first section from the last 50, 100, and 500 runs, the histogram of the number distribution is shown as Figure 2 to Figure 4.

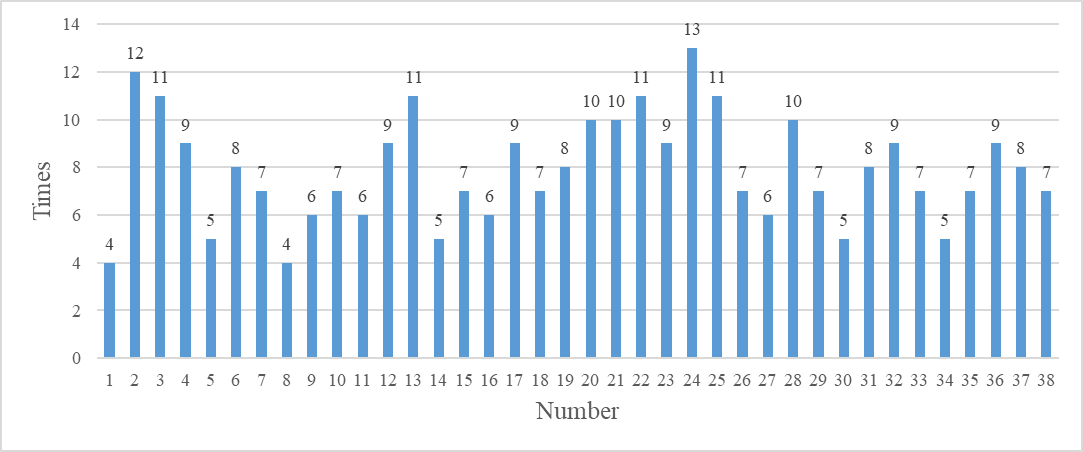


Figure histogram of 50 runs, first section

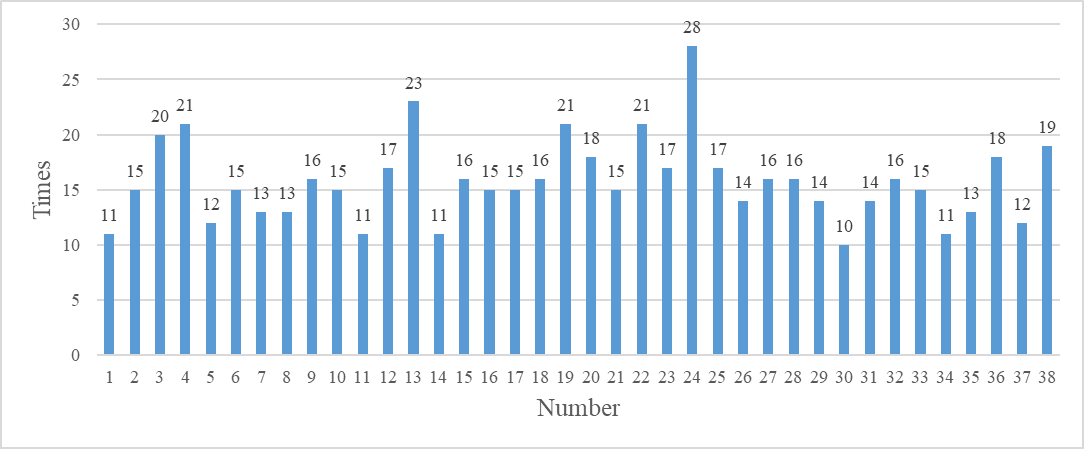


Figure histogram of 100 runs, first section

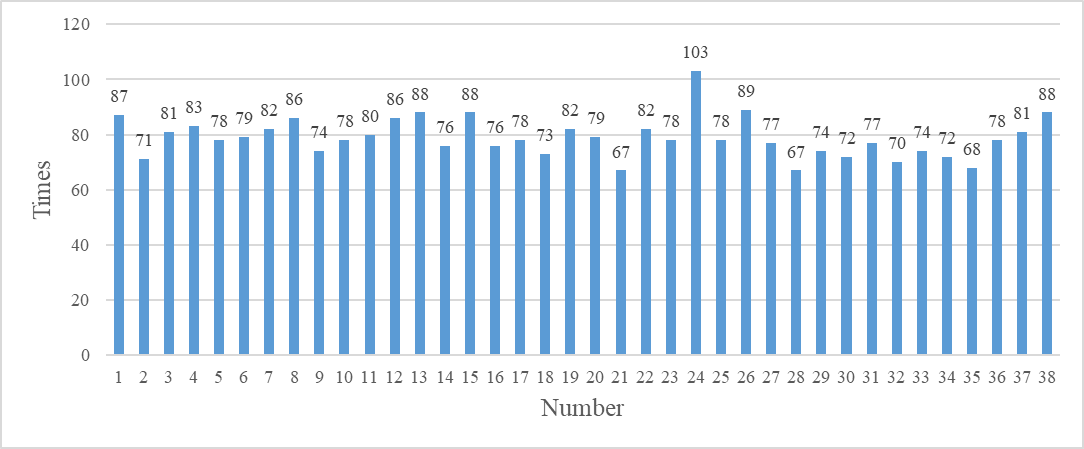


Figure histogram of 500 runs, first section

The statistic identity of the data is shown in the table below:

Table data identity for first section

|  |  |  |  |
| --- | --- | --- | --- |
|  | 50 | 100 | 500 |
| avg | 7.894737 | 15.78947 | 78.94737 |
| medium | 7.5 | 15 | 78 |
| max | 13 | 28 | 103 |
| min | 4 | 10 | 67 |
| range | 9 | 18 | 36 |
| variance | 5.069701 | 13.95448 | 52.10526 |

(b) For data of second section from the last 50, 100, and 500 runs, the histogram of the number distribution is shown as Figure 5 to Figure 7.

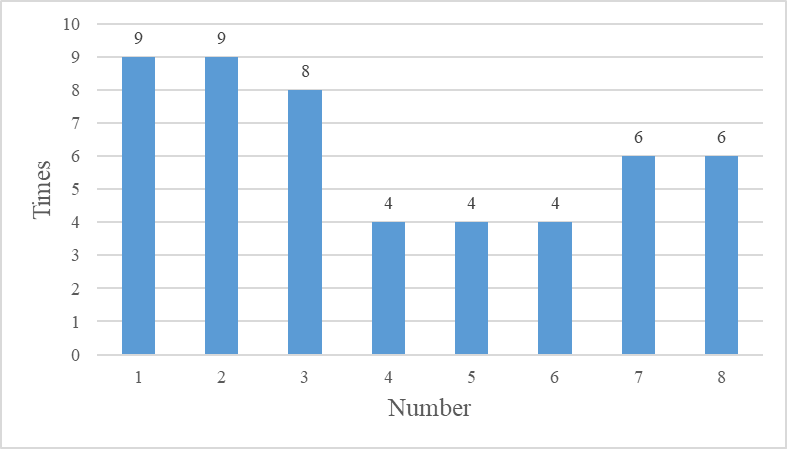


Figure histogram of 50 runs, first section

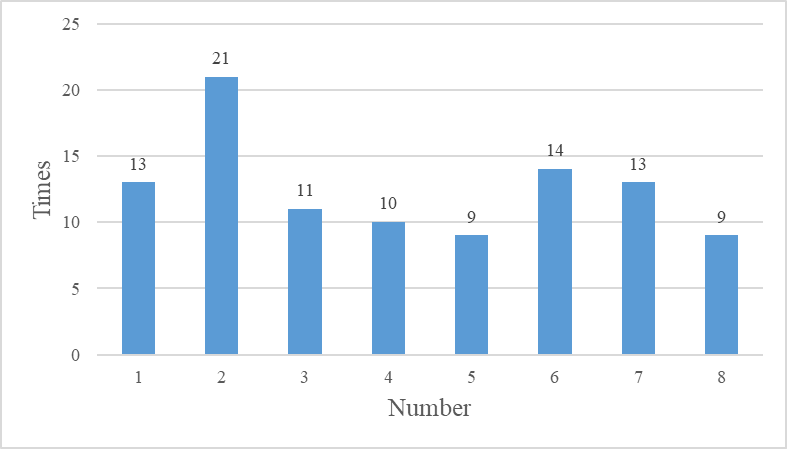


Figure histogram of 100 runs, first section

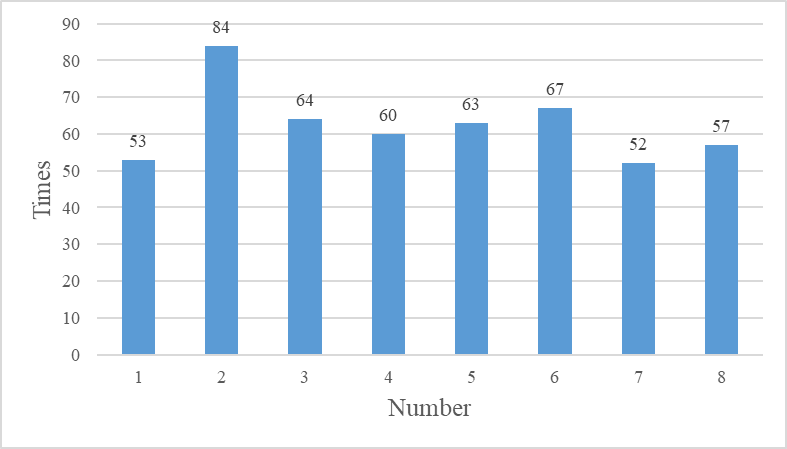


Figure histogram of 500 runs, first section

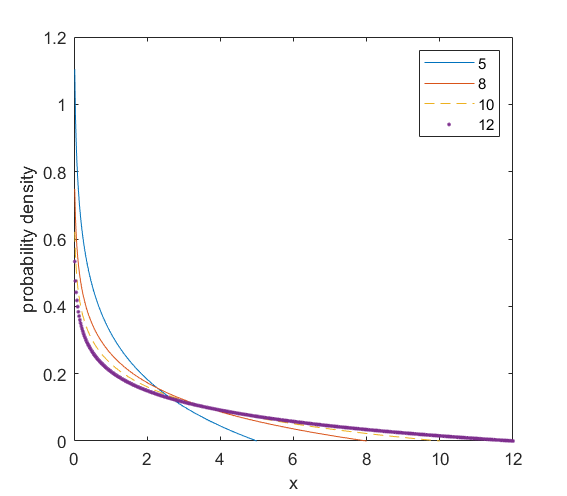
The statistic identity of the data is shown in the table below:

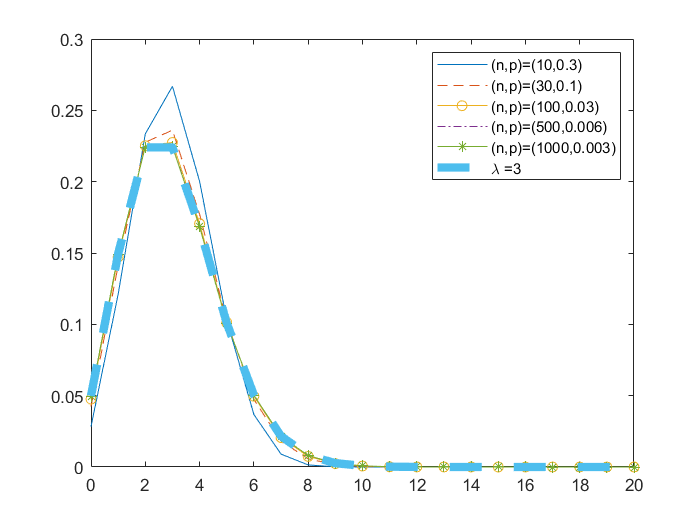
Table data identity for second section

|  |  |  |  |
| --- | --- | --- | --- |
|  | 50 | 100 | 500 |
| avg | 6.25 | 12.5 | 62.5 |
| medium | 6 | 12 | 61.5 |
| max | 9 | 21 | 84 |
| min | 4 | 9 | 52 |
| range | 5 | 12 | 32 |
| variance | 4.785714286 | 15.42857143 | 103.1428571 |

在沒有iid假設下，我們可以從條件機率中發現一個數字要連續出現的機率是較小的，從統計資料來看，一個數字要在500回的開獎中出現較多次的機率也是較小的。因此，比較在500回中出現次數較多及出現較少的數的兩種數，在計算該兩種數字在500回中的出現機率，並且下一回合要出現的機率來比較，選擇出現次數較少的數字出現機率會比較高。因此，從歷史出現最少的數字做選擇，(21,28,35,21,2)以及特別號7會是我的策略。

1. Plot



5. 

從圖可建當的值相同時，區線分布的狀態幾乎相同，大小及在x軸上的位置都是一樣的。將其對比於poisson random variable時可發現，線條也非常接近。不同處在於poisson distribution將binominal distribution進行了近似的運算，因此圖形會有些為不同，但當接相同時，圖形的走向是相同的。

6.

(a) 出現機率皆為

(b) 對於一個數字在n回合中出現的次數，期望值為。所使用的模型為平均分佈

(c) 當n = 50，出現次數的期望值為，標準差為0; 當n = 100，出現次數的期望值為，標準差為0; 當n = 500，出現次數的期望值為，標準差為0，因為每個數字的出現機率和期望值都是相同的。對比於作業一的機率分布模型，

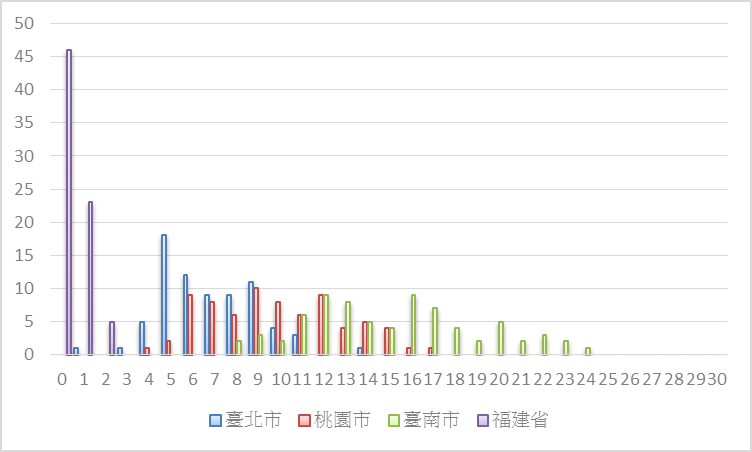
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 50 | 50理想 | 100 | 100理想 | 500 | 500理想 |
| avg | 7.894737 | 7.894737 | 15.78947 | 15.78947 | 78.94737 | 78.94737 |
| medium | 7.5 | 7.894737 | 15 | 15.78947 | 78 | 78.94737 |
| max | 13 | 7.894737 | 28 | 15.78947 | 103 | 78.94737 |
| min | 4 | 7.894737 | 10 | 15.78947 | 67 | 78.94737 |
| range | 9 | 0 | 18 | 0 | 36 | 0 |
| variance | 5.069701 | 0 | 13.95448 | 0 | 52.10526 | 0 |

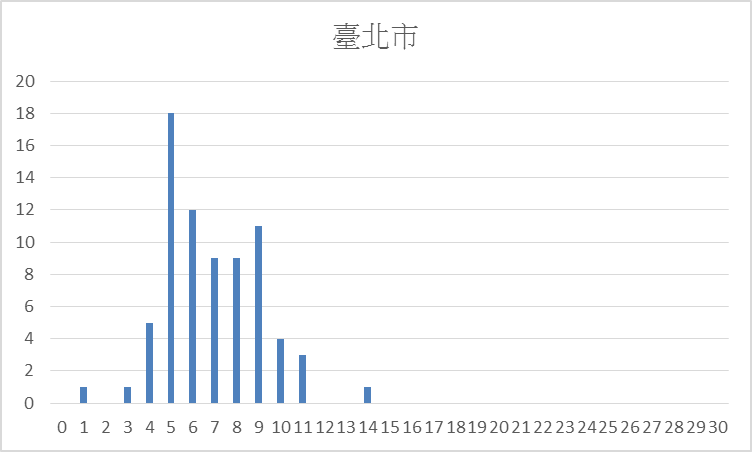
7. 每小時闖紅燈的車子數量、每個晚上在醫院小孩出生的期望值、一本書裡面的勘誤、MLB球賽中單場比賽的全壘打數量。這些事件的樣本數需要很大，發生的機率卻很小，且事件皆為可數、事件間互相獨立不相互影響。

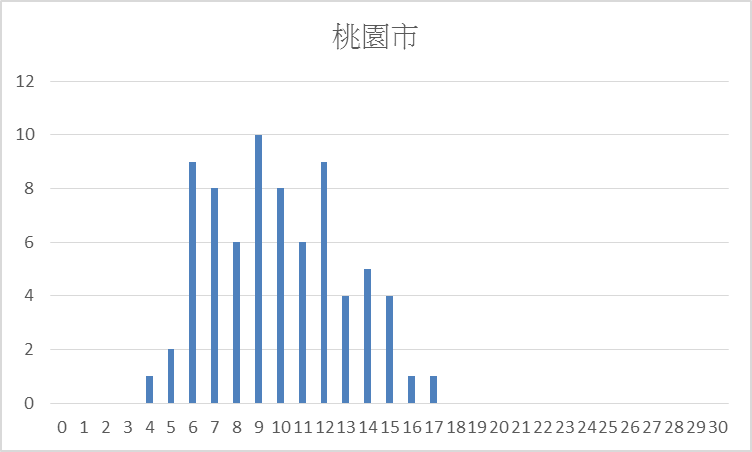
8. 計算四個城市的車禍發生次數的平均數與標準差，如下表所示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **臺北市** | **桃園市** | **臺南市** | **福建省** |
| avg | 6.851351 | 9.878378 | 15.16216 | 0.445946 |
| Var | 5.004998 | 9.286375 | 15.4254 | 0.387449 |

如表所示我們發現其平均數和標準差非常接近，且與Poisson Distribution的特性相似，因此判斷可使用Poisson Distribution來表達事件發生的機率分布。將其圖形繪製直方圖可得：







由直方圖可看出其分佈情形與Poisson Distribution相似。

