Q1. 一家陶瓷工廠專門生產陶瓷碗，每個生產批次中，工廠會將陶瓷碗組成五個一組的套裝。由於陶瓷碗在生產和包裝過程中可能會發生破損，因此工廠管理層非常關注產品的破損率。

最近，工廠接到了一些客戶反映，收到的套裝中有破損的碗。為了確保產品質量，工廠決定進行質量檢查。他們從最近生產的批次中隨機抽查了 1000 組（每組包含 5 個碗），並記錄下每組中破損碗的數量。

假設每個碗破損的概率是固定的，且每個碗是否破損是獨立事件。基於這些檢查數據，工廠希望估計每個碗破損的概率 p。

Q1-1 請根據這個檢查情境，推測每組（5 個碗）的破損數量應該符合什麼樣的機率分布？

Q1-2 對應期望值怎麼推論?

Q1-3 請推測 p 為多少?

Q1-4 請推測變異數?

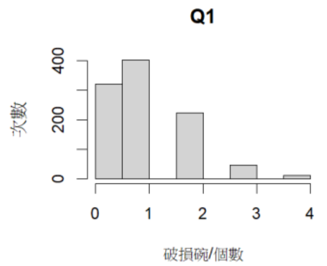
Q1-5 破損 3(含 3) 個碗以上的機率是多少?

Ans1-1

[點開數據] 可觀察到裡面都是正整數，因此可以猜測此題是離散型。且裡面含有不只是 0 & 1 的資料，由此可排除 Bernoulli distribution。

[觀看題目] 題目中提到碗破損是獨立事件、每個碗破損的概率都一樣，大概就可以猜測這是一個 $X \sim Binomial(n = 5,\ p)$。

[匯入 R 查看]

<table>
<tr><td><pre>
> summary(X_1$x)
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
  0.000   0.000   1.000   1.026   2.000   4.000
</pre></td><td>由此 summary 可再次確定他是離散型。</td></tr>
<tr><td>
</td><td>此 histogram 有點像是偏向一邊的鐘型曲線，可以更加確定他是 binomial distribution，也可以看出右邊比較像是尾巴，可以推估他為右偏，這可以由計算去驗證他。</td></tr>
<tr><td><pre>
> # [[[[[[[[[[[[偏度]]]]]]]]]]]]
> mean((X_1$x - mean(X_1$x))^3)/ ((var(X_1$x)^ (0.5)) ^ 3)
[1] 0.6517447
>
> # [[[[[[[[[[[[峰度]]]]]]]]]]]]
> (mean((X_1$x - mean(X_1$x))^4)/ ((var(X_1$x)^ (0.5)) ^ 4)) - 3
[1] 0.0482033
</pre></td><td>這裡計算出偏度是正的，確實為右偏的。由峰度可以看的出來他的圖形是接近常態分配的。且他是正的所以若要對比常態分配資料是相對不集中的。</td></tr>
</table>

binomial distribution #

Ans1-2

由上面的 summary 表可以看出 mean 為 1.026。#

Ans1-3

$\mu = n \times p \;=>\; 1.026 = 5 * p \;=>\; p = \frac{1.026}{5} \;=>\; p = 0.2052$ #

Ans1-4

$Var(X) = n \times p \times (1 - p) \;=>\; Var(X) = 5 \times 0.2052 \times (1 - 0.2052) \;=>\; Var(X) \approx 0.8155$#

Ans1-5

```
> n = 5
> p = 0.2052
> 1 - pbinom(2, n, p) # 預設是 <= 的機率
[1] 0.06199152
```

$P(X \geq 3) = 1 - P(X \leq 2) \approx 0.0620$#

Q2 一家大型醫院的急診室每天 24 小時運作，接收各種緊急情況的患者。為了更好地管理急診室的資源和人員配置，醫院的管理層希望了解在高峰期（例如下午 2 點到下午 3 點）每小時平均接收多少名急診患者。

過去一段時間的記錄顯示，在這一小時內的患者到達數量是隨機的，並且每小時內到達的患者數與前後小時無關。管理層認為，在這一小時內到達的患者數可能符合某種機率分布。

為了驗證這一假設，醫院在過去的 30 天內記錄了每天這一小時內到達的急診患者數量。根據這些數據，醫院希望：

1. 了解每小時到達的急診患者數量的平均值是多少。
2. 利用這些數據推斷未來某個小時內可能接收到的患者數量。

Q2-1 請根據這個檢查情境，每天這一小時內到達的急診患者數量應該符合什麼樣的機率分布？

Q2-2 對應期望值 u 怎麼推論?

Q2-3 根據 Q2-1 請推測變異數?

Q2-4 這小時內沒有患者到達的概率是多少？

Q2-5 請問資料的左偏還是右偏?

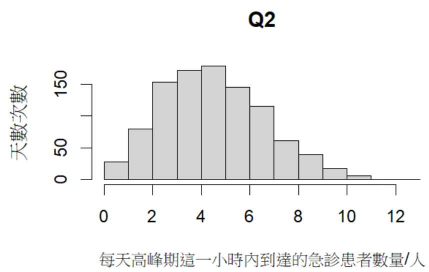
Q2-6 請問資料跟常態分佈比容易出現極端值嗎?

Ans2-1

[點開數據] 可觀察到裡面都是正整數，因此可以猜測此題是離散型。

[觀看題目] 針對計次的資料，且不受空間影響的可以推估他為 Poisson distribution。

[匯入 R 查看]

<table>
<tr>
<td><pre>> summary(X_2$x)
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
  0.000   3.000   5.000   5.026   6.000 13.000
> var(X_2$x)
[1] 4.577902</pre></td>
<td>先前我們判斷他是 Poisson distribution，這裡可以看出 mean、median 和 var 都很接近 5，因此他不會是超幾何。</td>
</tr>
<tr>
<td>
</td>
<td>從這個圖中可以看出他是有峰度，中間凸起來的這種圖形就會是 Poisson distribution。</td>
</tr>
</table>

Ans2-2

由上面的 summary 表可以看出 mean 為 5.026。#

Ans2-3

Poisson distribution 中，期望值和變異數一樣都為 λ，因此也為 5.026。#

Ans2-4

$$P(X=0) = \frac{(e^{-5.026} * 5.026^{0})}{1!} \;=>\; P(X=0) = e^{-5.026} \;=>\; P(X=0) \approx 0.0066 \#$$

Ans2-5

```
> #  [[[[[[[[[[[ 偏度 ]]]]]]]]]]]]]
> mean((X_2$x - mean(X_2$x))^3 )/ ((var(X_2$x)^ (0.5)) ^ 3)
[1] 0.374814
```

其實從上面的 histogram 就可以看出來是右邊。這裡計算出偏度是正的，確定為右偏的。

Ans2-6

```
> # [[[[[[[[[[[ 峰度 ]]]]]]]]]]]]]
> (mean((X_2$x - mean(X_2$x))^4 )/ ((var(X_2$x)^ (0.5)) ^ 4)) - 3
[1] -0.117092
```

由峰度很接近 0 可以看的出來他的圖形是接近常態分配的。且他是負的表示大部分資料落在四倍標準差中間，所以若要對比常態分配資料是相對集中的，較不容易出現極端值。

Q3

一家大型電信公司收集了過去一年中設備故障之間的時間間隔數據（以小時計），希望能夠了解這些時間間隔的分布，以便更好地預測未來的故障發生時間和計劃維修。數據顯示，每次設備故障之間的時間間隔是隨機的，且與之前的故障無關。

Q3-1: 假設你計算了這些時間間隔數據的平均值，如果數據來自於某個特定分布，這個平均值能幫助你推測數據來自什麼分布嗎？請說明你的推理過程。

Q3-2: 根據 Q3-1 你能直接知道資料的變異數嗎?

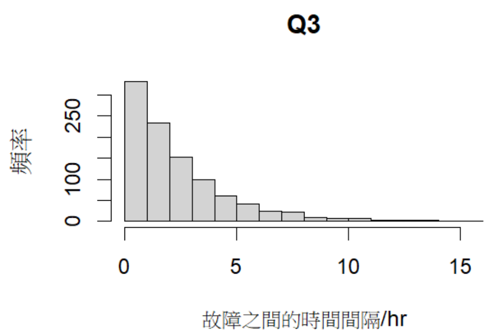
Q3-3: 假設你在觀察設備的運行時發現它已經正常運行了 2 小時還沒有故障。請問未來 1 小時內發生故障的概率是多少？這個概率會因為設備已經運行了 2 小時而改變嗎？請解釋你的推理。

Ans3-1

[點開數據] 可觀察到裡面都是有小數點的，因此可以猜測此題是連續型。

[觀看題目] 針對隨機發生故障，並且不帶記憶的連續型分配，可以推測為 exponential distribution。

[匯入 R 查看]

<table>
<tr><td><pre>> summary(X_3$x)
    Min. 1st Qu.   Median   Mean 3rd Qu.   Max.
0.008327 0.735066 1.715893 2.392672 3.326353 15.396534</pre></td><td>這裡可以看出他是連續型的，median 和 mean 是不一樣的，所以他是偏的。</td></tr>
<tr><td>
</td><td>像這種下滑的圖形可以猜測為 exponential distribution 或是 geometric distribution，因為它們都是偏健忘的分配。但上面判定是連續型，因此為 exponential distribution。</td></tr>
<tr><td><pre>> var(X_3$x)
[1] 5.445363
> var(X_3$x)^(0.5)
[1] 2.33353
> sd(X_3$x)
[1] 2.33353</pre></td><td>這裡看到變異數開根號就是期望值了，可以更加確定此題目為 exponential distribution。</td></tr>
</table>

Ans3-2

可以，就會約為 R 中計算的 5.4454#

Ans3-3

$$Var(X) = \frac{1}{\lambda^2} \quad => \lambda = \frac{1}{\sqrt{Var(X)}} \quad => \lambda = \frac{1}{\sqrt{5.4454}} \quad => \lambda \approx 0.4285$$

$$\mathrm{F}(1) = 1 - \mathrm{e}^{-\lambda} \quad => \mathrm{F}(1) = 1 - \mathrm{e}^{-0.4285} \quad => \mathrm{F}(1) \approx 0.3485 \ \#$$

題目已經有提到他是健忘的，因此他的概率跟運轉了 2 小時沒有關係，不管運轉幾個小時都會是上面的算法。

Q4 一家遊戲公司正在開發一款新的抽獎遊戲。玩家每次抽獎可以獲得一定數量的遊戲幣，這個數量是隨機決定的。

Q4-1 以資料 Q41，應該符合什麼樣的機率分布？

Q4-2 根據 Q4-1 期望值 u 怎麼推論?

Q4-3 根據 Q4-1 請推測變異數?

Q4-4 請問資料的左偏還是右偏?

Q4-6 請問資料跟常態分佈比容易出現極端值嗎?

Q4-7 以資料 Q42，應該符合什麼樣的機率分布？

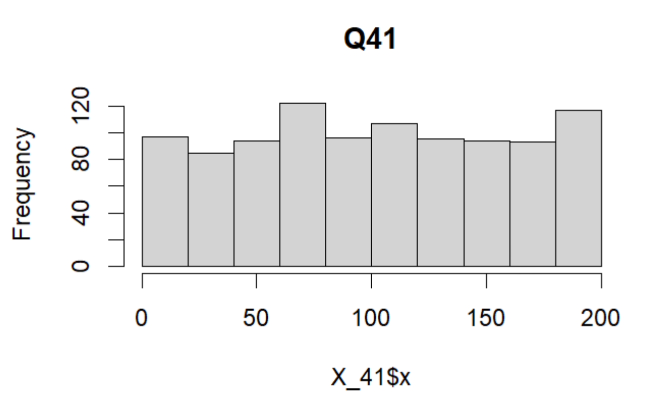
Q4-8 請問資料跟常態分佈比容易出現極端值嗎?

Ans4-1

[點開數據] 可觀察到裡面都是有小數點的，因此可以猜測此題是連續型。

[觀看題目] 看到抽獎遊戲、戳戳樂這種第一個就會想到是 Uniform distribution。

[匯入 R 查看]

<table>
<tr><td><pre>> summary(X_41$x)
   Min. 1st Qu.   Median   Mean 3rd Qu.   Max.
 0.2076 54.1666 102.2982 101.5423 151.3675 199.6708</pre></td><td>看到這個都可以大概知道 他是連續型的資料。</td></tr>
<tr><td>
</td><td>平平的看不到山峰在哪裡，因此可以更確定他是 Uniform distribution。#</td></tr>
</table>

Ans4-2

從上面的 summary 可得最大值為 199.6708、最小值為 0.2076

$$\mu = \frac{a+b}{2} \quad => \mu = \frac{199.6708 + 0.2076}{2} \quad => \mu = 99.9392\#$$

Ans4-3

$$\sigma^2 = \frac{(b-a)^2}{12} \quad => \sigma^2 = \frac{(199.6708 - 0.2076)^2}{12} \quad => \sigma^2 \approx 3315.4640\#$$

Ans4-4

```
> #  [[[[[[[[[[[ 偏度 ]]]]]]]]]]]]]
> mean((X_41$x - mean(X_41$x))^3 )/ ((var(X_41$x)^ (0.5)) ^ 3)
[1] -0.002635993
```

這裡計算出偏度是負的，大約為左偏。

Ans4-6

```
> # [[[[[[[[[[[ 峰度 ]]]]]]]]]]]]]
> (mean((X_41$x - mean(X_41$x))^4 )/ ((var(X_41$x)^ (0.5)) ^ 4)) - 3
[1] -1.166115
```

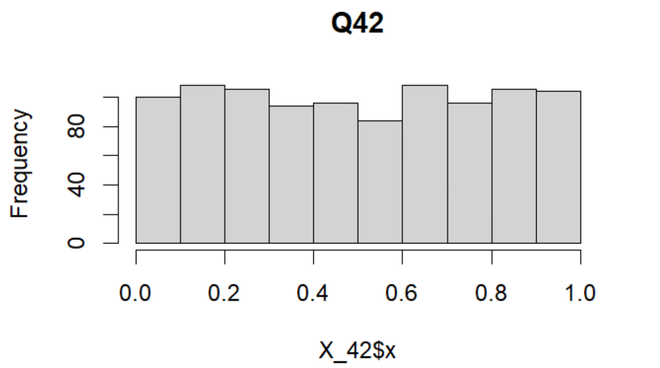
他是負的表示大部分資料落在四倍標準差中間，所以若要對比常態分配資料是相對集中的，較不容易出現極端值。

Ans4-7

[點開數據] 可觀察到裡面都是有小數點的，因此可以猜測此題是連續型。

[觀看題目] 看到抽獎遊戲、戳戳樂這種第一個就會想到是 Uniform distribution。

[匯入 R 查看]

<table>
<tr><td><pre>> summary(X_42$x)
     Min.   1st Qu.   Median   Mean   3rd Qu.   Max.
0.0004059 0.2346421 0.4928910 0.4997408 0.7562594 0.9993939</pre></td><td>看到這個都可以大概知道 他是連續型的資料。</td></tr>
<tr><td>
</td><td>平平的看不到山峰在哪裡，因此可以更確定他是 Uniform distribution。#</td></tr>
</table>

但從資料的上下介大約為 0 跟 1 的時候，可以稍微猜測資料是被正規化後的結果。但不是 Q41 的正規化結果，很好奇為什麼這題會多加一個資料。

Ans4-8

```
> # [[[[[[[[[[[ 峰度 ]]]]]]]]]]]]]
> (mean((X_42$x - mean(X_42$x))^4 )/ ((var(X_42$x)^ (0.5)) ^ 4)) - 3
[1] -1.246413
```

他是負的表示大部分資料落在四倍標準差中間，所以若要對比常態分配資料是相對集中的，較不容易出現極端值。

給定的 Joint PMF

定義隨機變數 X_1 和 X_2 的聯合機率質量函數如下：

X1/ X2	**1**	**2**
1	0.1	0.2
2	0.3	0.1
3	0.1	0.2

問題 1: 邊際 PMF

請計算 X_1 和 X_2 的邊際機率質量函數。

問題 2: 條件 PMF

請計算以下條件機率質量函數：

1. $P(X_1 = x_1 \mid X_2 = 1)$
2. $P(X_2 = x_2 \mid X_1 = 2)$

問題 3: X_1 的期望值

問題 4: $E[X_1 \mid X_2 = 2]$ 的期望值

問題 5: X_1 的變異數

問題 6: Var(X1|X2=2)。

X1/ X2	1	2	
1	0.1	0.2	0.3
2	0.3	0.1	0.4
3	0.1	0.2	0.3
	0.5	0.5	

Ans5-1

- X_1的邊際機率

  $P(X_1 = 1) = 0.3\#$

  $P(X_2 = 1) = 0.4\#$

  $P(X_3 = 1) = 0.3\#$

- X_2的邊際機率

  $P(X_1 = 1) = 0.3\#$

  $P(X_2 = 1) = 0.4\#$

  $P(X_3 = 1) = 0.3\#$

Ans5-2

- $P(\, X_1 = x_1 \mid X_2 = 1\,)$

  $P(\, X_1 = 1 \mid X_2 = 1\,) = \frac{0.1}{0.5} = 0.2\#$

  $P(\, X_1 = 2 \mid X_2 = 1\,) = \frac{0.3}{0.5} = 0.6\#$

  $P(\, X_1 = 3 \mid X_2 = 1\,) = \frac{0.1}{0.5} = 0.2\#$

- $P(X_2 = x_2 \mid X_1 = 2)$

$P(X_2 = 1 \mid X_1 = 2) = \frac{0.3}{0.4} = 0.75\#$

$P(X_2 = 2 \mid X_1 = 2) = \frac{0.1}{0.4} = 0.25\#$

Ans5-3

$E[X_1] = 1 \times P(X_1 = 1) + 2 \times P(X_1 = 2) + 3 \times P(X_1 = 3)$

$= 1 \times 0.3 + 2 \times 0.4 + 3 \times 0.3 = 2\ \#$

Ans5-4

$\boldsymbol{E[X_1 \mid X_2 = 2]}$

$= 1 \times P(X_1 = 1 \mid X_2 = 2) + 2 \times P(X_1 = 2 \mid X_2 = 2)$
$+ 3 \times P(X_1 = 3 \mid X_2 = 2)$

$= 1 \times 0.4 + 2 \times 0.2 + 3 \times 0.4 = 2\#$

Ans5-5

$\boldsymbol{Var(X_1) = (1^2 \times 0.3 + 2^2 \times 0.4 + 3^2 \times 0.3) - 2^2 = 0.6\#}$

Ans5-6

$Var(X_1 \mid X_2 = 2) = (\boldsymbol{1^2 \times 0.4 + 2^2 \times 0.2 + 3^2 \times 0.4) - 2^2 = 0.8\#}$ **-**