Check-in 6

Code

AUTHOR
Maxim Dokukin

Remember, **follow the instructions below and use R Markdown to create a pdf document with your code and answers to the following questions on Gradescope.** You may find a template file by clicking "Code" in the top right corner of this page.

1. The following code downloads two sequences of 256 digits. One of them was generated by asking ChatGPT to generate to "sample 1000 random digits from 0 to 9 with replacement" (I only got 256). The other was generated using the R command `sample(0:9, 256, replace = T)`. Which is from R and which is from ChatGPT? Explain and justify your answer.

```
A <- read.csv("https://math167r-s24.github.io/static/digits-a.cs
B <- read.csv("https://math167r-s24.github.io/static/digits-b.cs
```

```
vector <- A$x

print(vector)
```

```
  [1] 6 2 4 8 1 3 7 0 9 5 3 7 1 4 0 6 9 2 8 5 2 9 1 7 3 0 4 8 5
6 0 5 8 4 2 1 7
 [38] 9 6 3 1 3 7 5 4 8 6 2 9 0 4 6 0 3 5 9 2 7 1 8 9 0 5 2 6 7
8 3 4 1 7 8 3 6
 [75] 9 4 1 5 0 2 8 1 2 9 7 5 3 6 4 0 5 4 6 0 8 2 9 1 3 7 3 7 1
4 0 6 9 2 8 5 2
[112] 9 1 7 3 0 4 8 5 6 0 5 8 4 2 1 7 9 6 3 1 3 7 5 4 8 6 2 9 0
4 6 0 3 5 9 2 7
[149] 1 8 9 0 5 2 6 7 8 3 4 1 7 8 3 6 9 4 1 5 0 2 8 1 2 9 7 5 3
6 4 0 5 4 6 0 8
[186] 2 9 1 3 7 3 7 1 4 0 6 9 2 8 5 2 9 1 7 3 0 4 8 5 6 0 5 8 4
2 1 7 9 6 3 1 3
[223] 7 5 4 8 6 2 9 0 4 6 0 3 5 9 2 7 1 8 9 0 5 2 6 7 8 3 4 1 7
8 3 6 9 4
```

```
print(mean(vector))
```

```
[1] 4.539062
```

```
print(table(vector))
```

```
vector
 0  1  2  3  4  5  6  7  8  9
25 25 25 26 26 25 26 26 26 26
```

```
group <- (seq_along(vector)-1) %/% 10
sliced_vectors <- split(vector, group)
print(sliced_vectors)
```

```
$`0`
 [1] 6 2 4 8 1 3 7 0 9 5

$`1`
 [1] 3 7 1 4 0 6 9 2 8 5

$`2`
 [1] 2 9 1 7 3 0 4 8 5 6

$`3`
 [1] 0 5 8 4 2 1 7 9 6 3

$`4`
 [1] 1 3 7 5 4 8 6 2 9 0

$`5`
 [1] 4 6 0 3 5 9 2 7 1 8

$`6`
 [1] 9 0 5 2 6 7 8 3 4 1

$`7`
 [1] 7 8 3 6 9 4 1 5 0 2

$`8`
 [1] 8 1 2 9 7 5 3 6 4 0

$`9`
 [1] 5 4 6 0 8 2 9 1 3 7

$`10`
```

```
 [1] 3 7 1 4 0 6 9 2 8 5

$`11`
 [1] 2 9 1 7 3 0 4 8 5 6

$`12`
 [1] 0 5 8 4 2 1 7 9 6 3

$`13`
 [1] 1 3 7 5 4 8 6 2 9 0

$`14`
 [1] 4 6 0 3 5 9 2 7 1 8

$`15`
 [1] 9 0 5 2 6 7 8 3 4 1

$`16`
 [1] 7 8 3 6 9 4 1 5 0 2

$`17`
 [1] 8 1 2 9 7 5 3 6 4 0

$`18`
 [1] 5 4 6 0 8 2 9 1 3 7

$`19`
 [1] 3 7 1 4 0 6 9 2 8 5

$`20`
 [1] 2 9 1 7 3 0 4 8 5 6

$`21`
 [1] 0 5 8 4 2 1 7 9 6 3

$`22`
 [1] 1 3 7 5 4 8 6 2 9 0

$`23`
 [1] 4 6 0 3 5 9 2 7 1 8

$`24`
 [1] 9 0 5 2 6 7 8 3 4 1

$`25`
```

```
[1] 7 8 3 6 9 4
```

```
vector <- B$x

print(vector)
```

```
  [1] 8 5 9 7 1 9 4 9 4 9 1 9 7 4 9 2 7 7 1 2 9 9 1 7 2 9 9 5 1
9 1 8 0 4 7 3 9
 [38] 8 1 8 1 7 5 3 1 0 4 7 7 6 1 0 6 7 9 6 5 9 0 8 9 6 3 6 6 0
7 4 9 6 3 7 5 1
 [75] 6 0 0 2 3 3 5 7 2 5 4 2 2 4 6 4 5 8 8 8 4 9 0 0 7 5 7 3 3
9 8 9 4 3 5 6 0
[112] 8 4 6 2 7 3 5 0 8 3 0 9 8 1 8 1 8 8 2 9 5 5 5 2 6 3 9 6 6
9 8 5 6 4 7 2 2
[149] 8 9 7 9 9 2 2 8 0 2 4 0 5 6 3 6 9 2 7 1 6 6 4 3 7 2 2 5 6
4 3 2 6 3 4 7 9
[186] 4 1 4 8 2 7 0 6 5 7 4 1 3 6 6 6 5 7 7 2 9 0 8 7 6 5 9 6 3
8 0 9 2 1 8 8 2
[223] 3 4 5 5 2 6 6 8 0 7 5 5 1 7 7 2 9 7 3 7 2 2 8 0 0 4 7 5 7
0 0 2 6 4
```

```
print(mean(vector))
```

```
[1] 4.882812
```

```
print(table(vector))
```

```
vector
 0  1  2  3  4  5  6  7  8  9
22 18 28 20 23 25 30 33 25 32
```

```
group <- (seq_along(vector)-1) %/% 10
sliced_vectors <- split(vector, group)
print(sliced_vectors)
```

```
$`0`
 [1] 8 5 9 7 1 9 4 9 4 9

$`1`
 [1] 1 9 7 4 9 2 7 7 1 2
```

```
$`2`
 [1] 9 9 1 7 2 9 9 5 1 9

$`3`
 [1] 1 8 0 4 7 3 9 8 1 8

$`4`
 [1] 1 7 5 3 1 0 4 7 7 6

$`5`
 [1] 1 0 6 7 9 6 5 9 0 8

$`6`
 [1] 9 6 3 6 6 0 7 4 9 6

$`7`
 [1] 3 7 5 1 6 0 0 2 3 3

$`8`
 [1] 5 7 2 5 4 2 2 4 6 4

$`9`
 [1] 5 8 8 8 4 9 0 0 7 5

$`10`
 [1] 7 3 3 9 8 9 4 3 5 6

$`11`
 [1] 0 8 4 6 2 7 3 5 0 8

$`12`
 [1] 3 0 9 8 1 8 1 8 8 2

$`13`
 [1] 9 5 5 5 2 6 3 9 6 6

$`14`
 [1] 9 8 5 6 4 7 2 2 8 9

$`15`
 [1] 7 9 9 2 2 8 0 2 4 0

$`16`
 [1] 5 6 3 6 9 2 7 1 6 6
```

```
$`17`
 [1] 4 3 7 2 2 5 6 4 3 2

$`18`
 [1] 6 3 4 7 9 4 1 4 8 2

$`19`
 [1] 7 0 6 5 7 4 1 3 6 6

$`20`
 [1] 6 5 7 7 2 9 0 8 7 6

$`21`
 [1] 5 9 6 3 8 0 9 2 1 8

$`22`
 [1] 8 2 3 4 5 5 2 6 6 8

$`23`
 [1] 0 7 5 5 1 7 7 2 9 7

$`24`
 [1] 3 7 2 2 8 0 0 4 7 5

$`25`
[1] 7 0 0 2 6 4
```

```r
vector <- sample(0:9, 256, replace = T)

print(vector)
```

```
  [1] 0 7 3 8 4 4 8 1 9 2 9 7 1 7 5 2 1 8 1 6 1 8 2 1 1 0 0 7 8
5 3 0 0 3 5 3 8
 [38] 2 9 5 0 9 8 9 8 3 8 3 7 4 9 7 1 3 4 2 6 2 7 2 8 8 0 7 6 3
0 8 9 1 8 5 5 0
 [75] 0 9 1 8 3 8 5 4 1 1 4 4 5 2 9 2 7 6 6 0 1 1 6 7 2 4 7 4 4
8 6 5 4 2 1 9 3
[112] 2 1 9 3 2 5 2 0 2 9 2 8 0 9 4 7 9 6 0 1 1 7 7 7 6 0 1 4 5
5 3 0 5 3 7 1 0
[149] 2 5 6 5 6 5 0 8 1 1 4 4 3 9 5 9 0 6 7 7 8 0 2 2 9 9 7 4 8
6 1 0 7 7 1 6 8
[186] 5 1 7 2 2 3 1 3 3 0 0 4 1 1 2 8 0 7 0 3 8 4 4 4 7 4 6 6 0
4 6 7 2 1 3 4 8
```

```
[223] 7 4 5 5 8 4 8 2 0 9 9 6 8 8 4 2 9 5 6 1 0 1 1 4 4 5 3 6 2
8 0 8 5 4
```

```
print(mean(vector))
```

```
[1] 4.320312
```

```
print(table(vector))
```

```
vector
 0  1  2  3  4  5  6  7  8  9
29 32 26 20 29 23 20 26 30 21
```

```
group <- (seq_along(vector)-1) %/% 10
sliced_vectors <- split(vector, group)
print(sliced_vectors)
```

```
$`0`
 [1] 0 7 3 8 4 4 8 1 9 2

$`1`
 [1] 9 7 1 7 5 2 1 8 1 6

$`2`
 [1] 1 8 2 1 1 0 0 7 8 5

$`3`
 [1] 3 0 0 3 5 3 8 2 9 5

$`4`
 [1] 0 9 8 9 8 3 8 3 7 4

$`5`
 [1] 9 7 1 3 4 2 6 2 7 2

$`6`
 [1] 8 8 0 7 6 3 0 8 9 1

$`7`
 [1] 8 5 5 0 0 9 1 8 3 8

$`8`
 [1] 5 4 1 1 4 4 5 2 9 2
```

```
$`9`
 [1] 7 6 6 0 1 1 6 7 2 4

$`10`
 [1] 7 4 4 8 6 5 4 2 1 9

$`11`
 [1] 3 2 1 9 3 2 5 2 0 2

$`12`
 [1] 9 2 8 0 9 4 7 9 6 0

$`13`
 [1] 1 1 7 7 7 6 0 1 4 5

$`14`
 [1] 5 3 0 5 3 7 1 0 2 5

$`15`
 [1] 6 5 6 5 0 8 1 1 4 4

$`16`
 [1] 3 9 5 9 0 6 7 7 8 0

$`17`
 [1] 2 2 9 9 7 4 8 6 1 0

$`18`
 [1] 7 7 1 6 8 5 1 7 2 2

$`19`
 [1] 3 1 3 3 0 0 4 1 1 2

$`20`
 [1] 8 0 7 0 3 8 4 4 4 7

$`21`
 [1] 4 6 6 0 4 6 7 2 1 3

$`22`
 [1] 4 8 7 4 5 5 8 4 8 2

$`23`
 [1] 0 9 9 6 8 8 4 2 9 5
```

```
$`24`
 [1] 6 1 0 1 1 4 4 5 3 6

$`25`
[1] 2 8 0 8 5 4
```

Conclusion

- Vector A has a perfect distribution of digits Every digits appears 25-26 times. This is suspicious.
- Every sequence of 10 numbers, starting from index 1, has digits 0 to 9 without repetitions, which is statistically unlikely to be a natural random distribution. It seems as in sample 'A' ChatGPT generated data by the clusters of 10 digits (0-9).
- Another observations, is that there is almost no same digits appear in sequence in sample A, which is again statistically unlikely.

It is interesting how the biggest statistical machine alive (GPT) could not fulfill a simple statistical request.