Quiz for PhD applicants

Saif HadjSassi

20 July 2020

# I (Q1-Q2) Use of Python

**Q1. (10 Marks) During a literature review, Lili found urinary peptides related to diabetes type I (listed in db1) and urinary peptides related to diabetes type II (listed in db2).**

Use your knowledge in programming, extract the peptides that were **only** related to diabetes type I **but not** to diabetes type II. (If possible, hold these peptides in a list-like structure with the name db1s)

db1:'e00001','e00002','e00020','e00030','e00100','e00444','e01009','e09000','e10000','e100001'  
db2:'e00301','e00002','e00020','e00030','e00101','e00400','e01000','e09000','e10000','e100001','e20002','e30003'

**Q2. (10 Marks) Lili called the peptides that you extracted in Q1 as “diabetes I-specific peptides”, or in short, “db1s peptides”. She learned that the name of the db1s peptides must be unified before entering in a SQL database, such that it is a 9-digit positive integer:**

|  |  |
| --- | --- |
| old name | new name |
| e00001 | 99900001 |
| … | … |
| e99999 | 99999999 |

Use your knowledge in programming, help Lili to convert the names of the db1s-peptides. (If possible, hold these peptides in a list-like structure with the name db1s)

# II (Q3-Q5) Biostatistics

**Q3. (10 Marks) Lili received 100 urine samples from a local hospital. For each db1s peptide, Lili measured its abundance in four *randomly-selected samples* from the 100 samples, and calculated the *mean* of the four measurements. However, her supervisor thought that her analysis in 4 samples was not so informative and adviced her to measure the abundances of the same peptides in all 100 samples. Assume that for the 4-sample analysis, the mean abundance of peptide A was (unitless), where 20 was the standard deviation.**

1. (5 Marks) Can you estimate the mean abundance of peptide A in the 100-sample analysis? Will it be smaller or larger than 100?
2. (5 Marks) Can you estimate the standard deviation of peptide A abudance in the 100-sample analysis? Will it be smaller or larger than 20?

**Q4. (10 Marks) Lili just received the age the 100 patients and plotted its probability density function. What can you say about the distribution of age? It is likely a:**

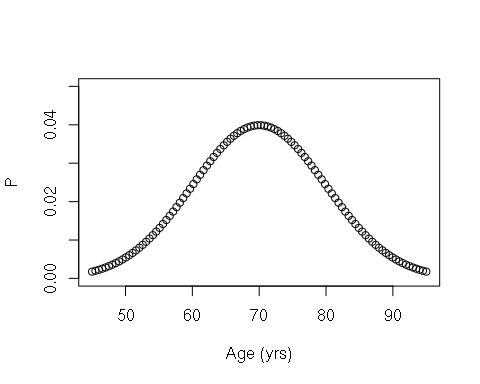
1. Binomial distribution

2. Hypergeometrical distribution

3. Standard normal distribution

4. Uniform distribution

5. None of the above



**Q5. (10 Marks) Lili went on and investigated the correlation of peptide A to age using linear regression. Her null hypothesis is that peptide A is NOT correlated to age. The linear regressor from the scikit-learn package reported, that the p-value of peptide A was smaller than 0.05.**

1. (5 Marks) What can you say about the result? Can you accept the null hypothesis?

2. (5 Marks) What is the potential “danger” of the interpretation, when she tested a large number of candidate peptides? (*Hint*: multiple testing)

# III (Q6-Q10) Algorithm

* Unless specified, do not use any built-in functions or imported libary
* Unless specified, answers without codes will NOT be marked
* Please try your best to provide implementable codes; if not, show your pseudocodes
* Codes will be evaluated based on efficiency

**Q6. (10 Marks) For a given array arr1 with 7 numbers (**7.358375,14.349093,6.055756,10.220360,15.376650,8.105909,8.516646**), give a subarray whose element x is smaller than 14 but at least 10 for all x.**

**Q7. (10 Marks) Find the maximum of the given array arr1 with 7 numbers (**7.358375,14.349093,6.055756,10.220360,15.376650,8.105909,8.516646**).**

# do not use the built-in max(x) or its equivalent

**Q8. (10 Marks) Calculate the factorial of 15.**

Remarks:

# do not use the built-in factorial(n) or its equivalent

**Q9. (10 Marks) A Fibonacci number is defined as the sum of its two preceding terms, that is: . By definition, the first two Fibonacci numbers are and . The beginning of the Fibonacci series is:**

**Give the value of .**

**Q10. (10 Marks) On a rectangular grid, an ant wants to walk from the North-west vertex A, to the South-east vertex B (run the below chrunk to see). It can only walk either east or south-ward, with the length of a step equals to 1.**

1. (5 Marks) How many possible routes can it take? (*Please give your answer directly, no codes are required*)

2. (5 Marks) List all the possible routes it can take. You may give the results in any readable and consistent form you prefer. For example, the route with 5 steps East followed by 3 steps South can be expressed as:

\* a string 'EEEESSS'

\* a string '1111000'

\* an integer

\* a list [1,1,1,1,0,0,0]

\* a tuple ('1','1','1','1','0','0','0')

\* …

