UNIVERSITY of WASHINGTON

We thank you for your time spent taking this survey.
Your response has been recorded.

24/30 **80.0%**

Question 1: **1/1**

In an experiment, the depth of the lake was measured every day for approximately two years, and the historical average depth was subtracted from this value. This means that a measurement of zero is the same as the historical average value. Figure 1 is a histogram representing the collection of measurements. During the experiment, for approximately how many days was the lake at a depth above the historical average, but not more than 1 in above the average?

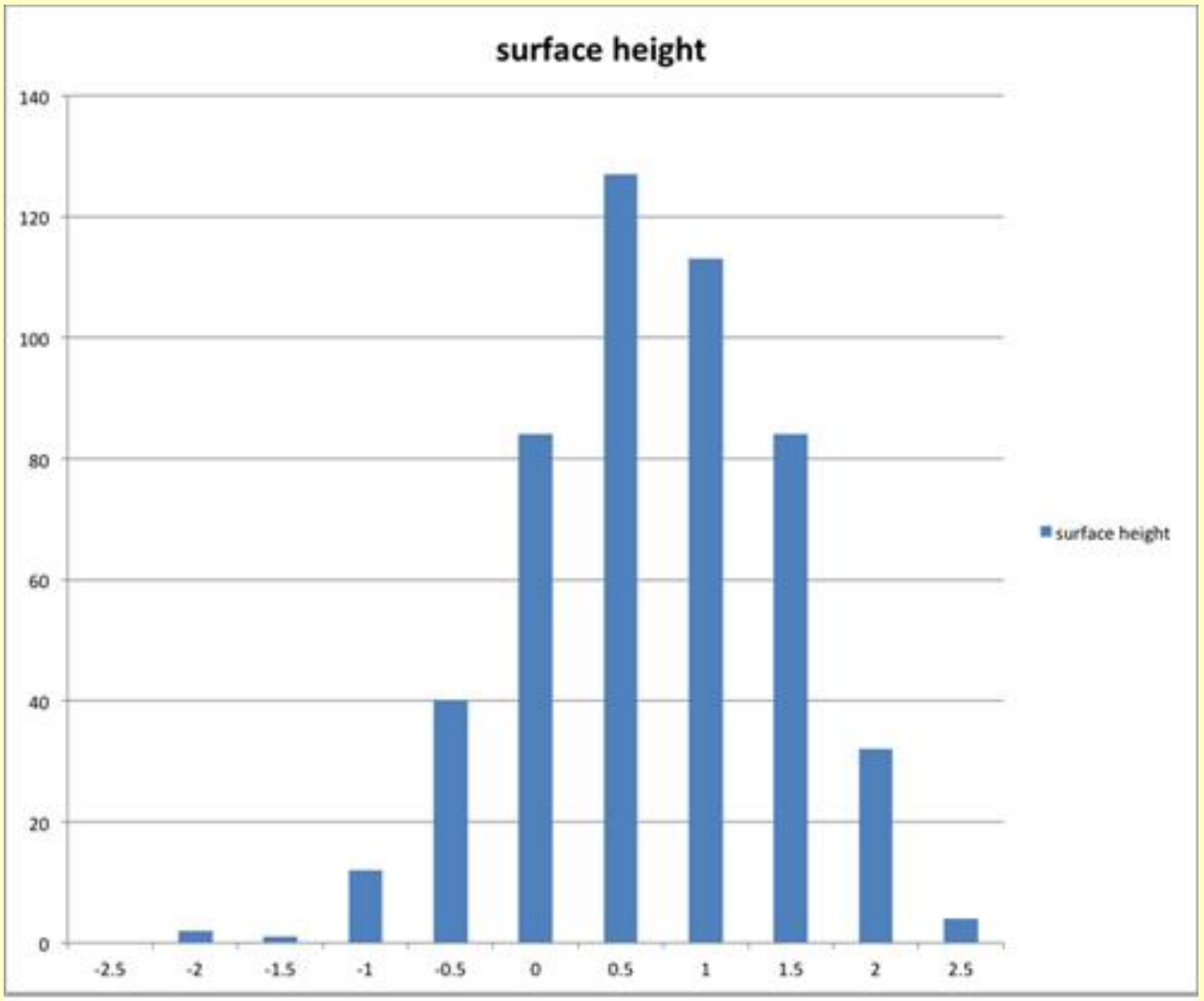

Figure 1

0 to 100

100 to 200

✓ **200 to 300**

300 to 400

Question 2: 1/1
The cumulative distribution function for the lake depth experiment of Problem 1 is shown in Figure 2. During the experiment, what was the probability that the lake was deeper than the historical average when measured?

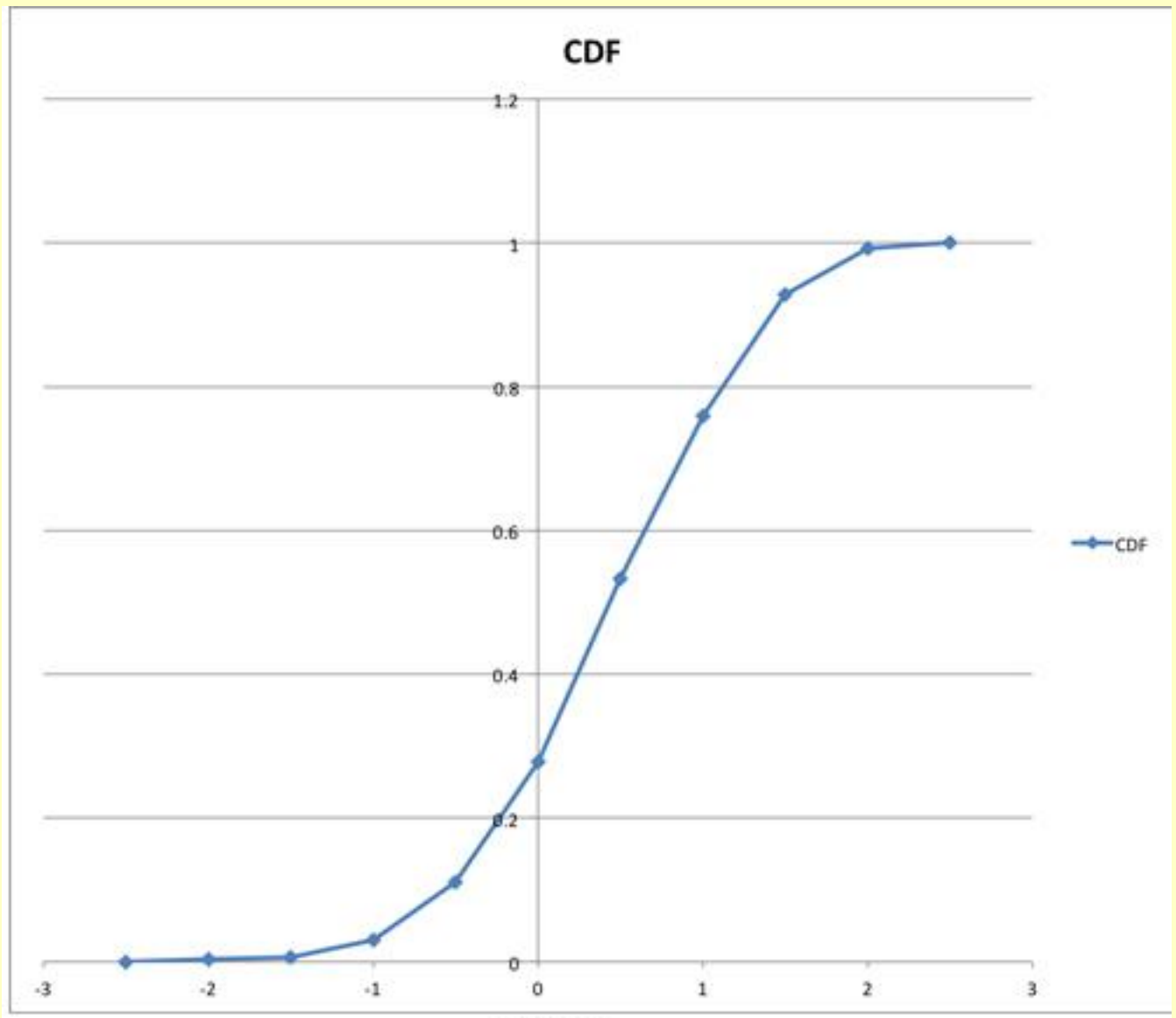

Figure 2

Less than 20%

20% - 40%

40% - 60%

✔ **Greater than 60%**

Question 3: 1/1
For a weighted die, the probability of each outcome is located below. Which of the following correctly separates the results into quartiles?

value	probability
1	3/10
2	1/5
3	1/10
4	1/5
5	3/20
6	1/20

✓ **{1}{2,3}{4,5,6}**

{1,2}{3,4}{5,6}

{1}{2}{3}{4,5}{6}

{1}{2}{3,4,5,6}

Question 4:

1/1

What is the value of $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

✓

Question 5:
What is the mean value of the following set {80, 125, 140, 85}?

1/1

78

112.4

✓ **107.5**

132

Question 6:
What is the standard deviation of the set {1, 2, 3, 4, 5}?

1/1

3

✓ **√2**

√15

4

Question 7:
If a fair die is rolled six times, which result is more likely?

1/1

a) 4, 4, 4, 4, 4, 4
b) 3, 1, 1, 4, 2, 5

(a)

(b)

✓ **Both are equally likely**

Cannot be determined

Question 8:
If a fair coin is flipped three times, what is the probability of observing the sequence: heads, heads, tails?

1/1

1/3

✓ **1/8**

3/8

1/4

Question 9:
Which of the following is the least likely observation?

0/1

Normal distribution, mean 0, standard deviation 1, observed value 1

✗ **Normal distribution, mean 0, standard deviation 2, observed value 3**

Normal distribution, mean 1, standard deviation 1, observed value -1

Normal distribution, mean 0, standard deviation 3, observed value 4

Question 10:

1/1

Let $M = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$, $N = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

M · N is:

✓

Question 11:
Four hexadecimal digits represent the range from '0000' to 'FFFF'. How many bits are necessary to represent eight hexadecimal digits?

1/1

16

✓ **32**

64

128

Question 12: 1/1
The series 1+2+3+4+5+6+...+998+999+1000 is equivalent to:

- ✓ **(1+1000)*500**
- 1000*(500+1)
- 1000*500/2
- None of the above

Question 13: 0/1
Different sorting algorithms may have very different average time complexities when sorting integer numbers. Using Big-O notation, what is the expected time complexity of, respectively, [1] Bubble sort, [2] QuickSort, [3] Mergesort, [4] Bucket sort (aka Cont sort or Distribution sort)?

- [1] O(n^2), [2] O(n Log(n)), [3] O(log(n)), [4] O(n Log(n))
- [1] O(n^2), [2] O(n Log(n)), [3] O(n Log(n)), [4] O(n)
- [1] O(n Log (n)), [2] O(n Log(n)), [3] O(log(n)), [4] O(n^2)
- ✗ **None of the above**

Question 14: 1/1
Djikstra's Algorithm is typically used for:

- finding the overlap between two arbitrary strings.
- finding the elements that intersect between two sets in linear time.
- ✓ **finding the shortest path in a graph with non-negative edges.**
- finding the most likely ancestor between two leaves in a binary tree.

Question 15: 1/1
Given a sorted array with N numbers, what assertion is true about Binary Search?

- Binary Search can find an arbitrary element in the array in O(N log(N)) time
- Binary Search can find an arbitrary element in the array in O(N) time
- ✓ **Binary Search can find an arbitrary element in the array in O(log(N) time**
- Linear Search can find an arbitrary element in the array in O(N log(N)) time

Question 16: 1/1
Dictionary lookups are often implemented using a Hash Table. What can be said about Hash Tables storing N key-value pairs?

- hash Tables need O(N^2) storage to accomplish fast search times.
- hash Tables cannot be implemented with Linked Lists.

✓ **hash Tables have O(1) average search time.**

hash Tables keep values internally in sorted order.

Question 17: 1/1
Divide and Conquer algorithms break down a problem recursively into a larger number of smaller problems that can be solved in a very simple way. Given the following sample programs, what statements can you make about whether they represent a divide and conquer algorithm?

Program A:

```
def programA(inputTree):
  if (inputTree.isLeaf()):
    return inputTree.value()
  else:
    return programA(inputTree.left()) + programA(inputTree.right())
```

Program B:

```
def programB(inputTree):
  sum = 0
  for x in inputTree:
    sum = sum + x
  return sum
```

Program C:

```
def programC(inputTree):
  return helper(0, inputTree)
def helper (sumSoFar, inputTree):
  if (inputTree.isLeaf()):
    return sumSoFar
  else:
    return helper(helper(sumSoFar, inputTree.left()), inputTree.right())
```

all three compute the same answer and thus are all equivalent

programs A and C are divide and conquer but program B is iterative

✓ **program A is divide and conquer, program B is iterative, program C is recursive, but not divide and conquer**

none of the programs use divide and conquer methods

Question 18: 1/1
A function to traverse a binary search tree data structure "in-order" (aka in-order traversal) can be accomplished by this sequence of operations:

(1) visit the root, (2) traverse the left subtree, (3) traverse the right subtree.

(1) visit the root, (2) traverse the right subtree, (3) traverse the left subtree.

(1) traverse the left subtree, (2) traverse the right subtree, (3) visit the root.

✓ **(1) traverse the left subtree, (2) visit the root, (3) traverse the right subtree.**

Question 19: 1/1
A Heap is a tree-based data structure frequently used:

to find the top K largest elements of a set.

to find the top K smallest elements in a set.

to find the median value of the elements in a set.

✓ **all of the above.**

Question 20:
What does the function Zombie(N) below do?

1/1

```
int Zombie(int N){
If (N<1)
   return 1;
Else
return N*Zombie(N-1);
}
```

calculates the square root of N

✓ **calculates N factorial**

calculates 2 to the power of N (that is, 2^N)

calculates the Nth Fibonacci series number

Question 21:
A row in a relational database always contains:

0/1

a map of arbitrary keys to arbitrary values.

a set of values for columns whose names and types are defined by the database schema.

an index of all data in the database.

either a primary key or a foreign key.

Question 22:
In a SQL database a BLOB is used:

1/1

✓ **to store binary data whose precise encoding and meaning is unknown to the database server.**

to move data to and from the disk.

when the database grows too large to fit in memory.

only as a hack, and should be avoided at all costs.

Question 23:
MiniCorp wants to track how their stores are doing by region. In order to do so, they added two new tables to their SQL database. The first table is called REGIONS and stores the name and ID of each region. The second is called STOREREGIONS and indicates which stores are in which regions.

1/1

You are beginning to explore these tables and you run the following queries with the results shown:

SELECT r.name, count(s.store_id) FROM storeregions s, regions r WHERE s.region_id = r.regionid GROUP BY s.region_id;

North 156

South 591
East 302
Southwest 208
Northwest 396

SELECT COUNT(DISTINCT store_id) FROM storeregions;

956

What can we conclude?

There are more stores in Oregon and Washington than in Michigan.

MiniCorp is growing most rapidly in the south.

✓ **Some stores are in more than one region.**

There are other regions where there are no stores.

Question 24:
A full table scan:

0/1

is used to detect viruses in the database.

is a critical step in verifying data integrity on the disk.

✗ **uses more memory than any other database operation.**

can often be avoided by adding an appropriate index.

Question 25:
Compared to an inner join, an outer join:

0/1

is never a better option.

✗ **will always produce a larger result.**

will always produce a smaller result.

none of the above is true.

Question 26:
Which of the following <u>is not</u> true of SQL databases?

1/1

they can model many-to-many relationships.

✓ **they can only be used effectively by writing regular expressions.**

they normally support unique key constraints.

stored procedures run on the server.

Question 27:
In a SQL database table, NULL values:

1/1

are forbidden.

✓ **can be disallowed for some columns with the appropriate schema definition.**

indicate that data has not yet been loaded but you should check back later.

should rarely be present because they indicate an error condition.

Question 28:
In SQL, a WHERE clause:

1/1

✓ **can cause the result set of the query to be empty.**

can only be used with a GROUP BY clause.

will prevent the query planner from relying on an index.

will force the query planner to rely on an index.

Question 29:
Consider the query:

0/1

SELECT price FROM products WHERE product_id = 2387436;

We run this query on two SQL databases, one from vendor A and one from vendor B. Both systems typically execute the query in 20 milliseconds on a database of ten thousand products. If we increase the number of the products in the database to ten million, the query takes 10 seconds on the system from vendor A and 25 milliseconds on the system from vendor B.

Which of the following is our best next step?

Select vendor B. Vendor A clearly cannot handle the larger data set.

✗ **Make sure that the server running A has as much memory as the server running B.**

Make sure we have the same indices on our tables in both systems.

Make sure that the disks in the two systems are either exactly the same make and model or at least comparable models from different manufacturers.

Question 30:
Sharding is a technique for:

1/1

slicing data in reports.

✓ **distributing large data sets across more than one machine.**

compressing data to use disk space more efficiently.

locating inflection points in large data sets.

UNIVERSITY of WASHINGTON

Data Science Assessment Quiz

This assessment includes 30 multiple-choice questions divided into three sections: Statistics & Linear Algebra, Programming, and Databases & SQL. Applicants must earn a total score of at least 18 out of 30, with a minimum score of 6 out of 10 in each section.

Once you start the assessment, you must complete it within 90 minutes. The timer does not display on the page, so **please make sure to set your own timer** to track how much time has elapsed. After each section, you will be able to see your score, as well as your total score at the end. Please use the same email address on both your assessment and your

application.

Applicant Information

Name (Last, First)

Email

→