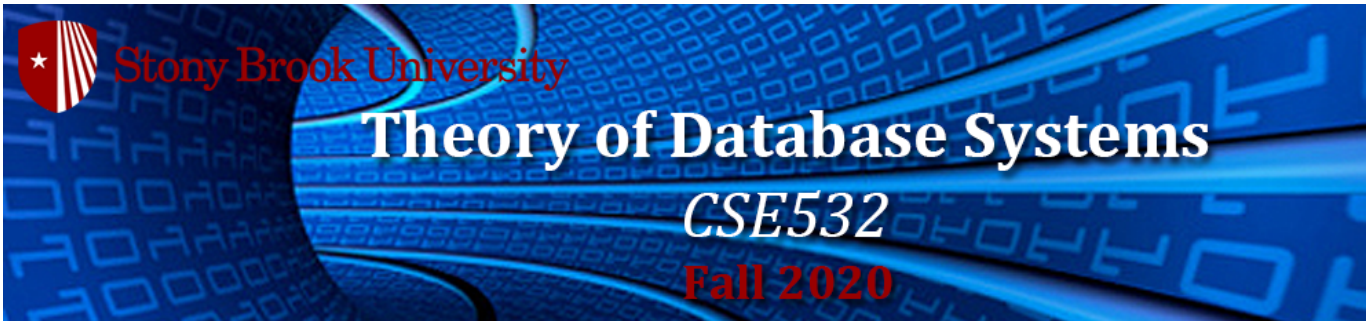


Home

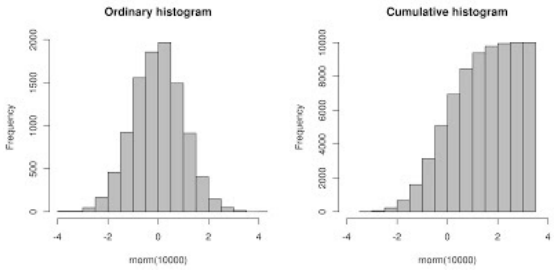
Home
Syllabus
Schedule
Homework
Textbooks
Examples
Instructor
Resources

Homework >

Homework2: Compute Cumulative Histogram

In this homework, you will implement a stored procedure to compute the cumulative histogram for employee's salaries from the employee table in DB2 sample database (*employee* table). If you want to create the table by yourself, you need to create a table employee with this schema and load the sample data into the database first. You can find examples on stored procedures and JDBC.

A histogram is the probability distribution of a continuous variable, represented by frequencies of the variables falling into a bin. A **cumulative histogram** (used in this homework) is a mapping that counts the cumulative number of observations in all of the bins up to the specified bin.



In this homework, your cumulative histogram program takes an initial value (inclusive), an end value (exclusive), and the number of bins, e.g., cumulativehistogram(start, end, number), and it returns a set of bins with bin number, cumulative frequency until this bin, bin's start value, and bin's end value, i.e., (binnum, frequency, binstart, binend). For example:

binnum	cumulativefrequency	binstart	binend
1	2	35000	40000
2	5	40000	45000

Q1 (5 points). You will write a stored procedure in SQL/PL to implement it. Please consider performance in your implementation. For example, multiple queries on the same table will involve multiple scans, and could lead to bad performance. An aggregate function (e.g., count) will normally lead to a scan of the table.

When the stored procedure is called, the cumulative histogram is returned and stored in a predefined table. Alternatively, you can return the result as an array. Refer to "Exchange data using arrays in SQL PL".

Q2 (5 points). Implement the same function with another approach, e.g., JDBC, Java Stored Procedure, or User Defined Function in PL/SQL. Again, you should consider only querying the table once to generate the result to maximize performance.

Homework 2 Submission

To test your code, please use lower bound 30,000, upper bound 170,000, and interval 20,000 to generate a sample result.

e.g., run the query "call gen_salary_cumulativehistogram(30000, 170000, 7)"

Please zip your SQL code sp.sql, the code file in alternative approach (e.g., CumulativeHistogram.java), sample result (a file or screenshot), and a readme file on how to run your programs.

Please go to blackboard, and submit it under homework 2.

homework1sampledata.txt (5k)	Fusheng Wang, Sep 28, 2020, 10:32 AM	v.1	
homework1tableschema.txt (0k)	Fusheng Wang, Sep 28, 2020, 10:32 AM	v.1	