

Your Name:

Imagine you have a relation $R(A, B)$, with a total of 100,000 tuples. For each of the following scenarios, estimate the number of tuples in the result using the available information. Make uniformity and independence assumptions as needed. The information is NOT cumulative (i.e., the information listed in (1) does not apply to (2))

(1) Information: A is a primary key; Query: $A = 10$

Since A is primary key, only one tuple will satisfy

(2) Information: B is an integer, $\min(B) = 501$, $\max(B) = 1000$; Query: $B = 600$

There are 500 distinct values of B in that range, so assuming uniformity: $100,000/500 = 200$

(3) Information: B is an integer, $\min(B) = 501$, $\max(B) = 1000$, No. of distinct values of B = 50
Query: $B = 600$

Number of distinct values is explicitly given, so; $100,000/50 = 2000$

(4) Information: B is an integer, $\min(B) = 501$, $\max(B) = 1000$; Query: $601 \leq B \leq 700$

This is similar to (2). $100 * 100,000/500 = 20,000$

(5) Information: No. of distinct values of B = 1000, No. of distinct values of A = 500
Query: $B = 500 \text{ AND } A = 100$

$100,000/1000 = 100$ tuples satisfy the first one, and of those, only $100/500$ satisfy the second one, so the answer is basically 0 (or 1 if you round up)

Same as above, but now we have two relations: $R(A, B)$, and $S(B, C)$: $|R| = 10,000$, $|S| = 1,000$

Our query in all cases is to estimate the size of $R \text{ NATURAL JOIN } S$.

(6) Information: B is a primary key for S

Each tuple of R matches at most one tuple of S, so the result estimate = 10,000

(7) Information: No. of distinct values of B = 100 (in both relations)

Each tuple of R matches about $1000/100$ tuples of S, so we get: $10000 * 10 = 100,000$ tuples in the result (estimated)