Implementation of DBMS
Exercise Sheet 13
Klingemann, WS 2024 / 2025

1) Below are some statistics for four relations W, X, Y and Z. We assume in this task that a projection does not remove duplicates. We further assume that all values used in selection conditions actually appear in the corresponding instance of the relation.

W(a, b)	X(b, c)	Y(b, d)	Z(b, d)
T(W) = 100	T(X) = 200	T(Y) = 300	T(Z) = 400
V(W, a) = 20	V(X, b) = 40	V(Y, b) = 30	V(Z, b) = 40
V(W, b) = 60	V(X, c) = 100	V(Y, d) = 50	V(Z, d) = 100

Estimate the number of tuples of the relations that are the result of the following expressions:
a) $\sigma_{a=10}(W)$
b) $W \times Y$
c) $\sigma_{d>10}(Z)$
d) $W \bowtie X$
e) $\sigma_{a=1 \text{ AND } b=2}(W)$
f) $\sigma_{a=1 \text{ AND } b>2}(W)$
g) $\sigma_{b=10}(W) \bowtie X$
h) $\pi_b (\sigma_{a=20}(W))$
i) $W \bowtie X \bowtie Y$
j) $Y \bowtie Z$
k) $W \bowtie_{a=d} Z$

2) Consider a query optimizer that uses statistical data. In particular, the following information is known about an attribute A of relation R. Attribute A is of type integer. Make the best use of the given the information.

- There are 100 tuples with A values between 1 and 10. In this range, there are 8 unique A values.
- There are 200 tuples with A values between 11 and 20. In this range, there are 5 unique A values.
- There are 300 tuples with A values between 21 and 30. In this range, there are 10 unique A values.
- There are 400 tuples with A values between 31 and 40. In this range, there are 10 unique A values.

a) Consider the query $\sigma_{A=7}(R)$. How many tuples are expected in the answer, assuming values are uniformly distributed over possible V(R, A) values?
b) Consider the query $\sigma_{A=17}(R)$. How many tuples are expected in the answer, assuming values are uniformly distributed over possible domain values?
c) Consider the query $R \bowtie S$, where R has attributes R(A, B, C) and S has attributes S(A, D, E). Assume that S has the same number of tuples as R, and that the A attribute in S has the same distribution as A has in R. Assuming values are uniformly distributed over possible V(R, A) values, how many tuples are expected in the answer?