

	Resultado
N (movie)	240
N (person)	5000
N (movie_cast)	3179
V(year,movie)	54
Min(year, movie)	1927
Max(year, movie)	2014
V(cast_order, movie_cast)	65

Questão 1

```
SELECT * FROM movie CROSS JOIN person;
```

Resultado esperado:

R = 1200000

Resultado estimado:

$R = n(\text{movie}) * n(\text{person}) = 240 * 5000 = 1200000$

Questão 2

```
SELECT * FROM movie NATURAL JOIN movie_cast WHERE cast_order = 1 AND release_year > 2000;
```

Resultado esperado:

R = 87

Resultado estimado:

$P(\text{movie}, \text{release_year} > 2000) = 1 - P(\text{movie}, \text{release_year} \leq 2000) = 1 - (2000 - 1927 + 1) / (2014 - 1927 + 1) = 7 / 44$

$$P(\text{movie_cast}, \text{cast_order} = 1) = 1 / 65$$

$$\text{sel}(\text{PK}) = P(\text{movie}, \text{release_year} > 2000) = 7 / 44$$

$$n(\text{FK}) = n(\text{movie_cast}) * P(\text{movie_cast}, \text{cast_order} = 1) = 3179 * 1 / 65 = 3179 / 65$$

$$R = n(\text{FK}) * \text{sel}(\text{PK}) = 3179 / 65 * 7 / 44 = 2023 / 260 = 7.78$$

A diferença entre o resultado estimado e o esperado ocorre porque a estimativa presume distribuições uniformes nos registros, como anos em `movies` e `cast_order` em `movie_cast`. Na prática, isso não acontece. Por exemplo, todos os filmes têm obrigatoriamente um `cast_order = 1` antes de qualquer outro valor, gerando concentração e distorcendo a estimativa.

Questão 3

```
SELECT * FROM movie JOIN movie_cast USING (movie_id) JOIN person USING (person_id)
WHERE cast_order = 1 AND (release_year = 1990 OR release_year = 2000)
```

Resultado esperado:

$$R = 18$$

Resultado estimado:

$$P(\text{movie_cast}, \text{cast_order} = 1) = 1 / 65$$

$$P(\text{movie}, \text{release_year} = 1990 \text{ OR } \text{release_year} = 2000) = 2 / 54$$

$$\text{sel}(\text{PK}_1) = P(\text{movie}, \text{release_year} = 1990 \text{ OR } \text{release_year} = 2000) = 2 / 54$$

$$n(\text{FK}_1) = n(\text{movie_cast}) * P(\text{movie_cast}, \text{cast_order} = 1) = 3179 * 1 / 65 = 3179 / 65$$

$$\text{sel}(\text{PK}_2) = 1$$

$$n(\text{FK}_2) = n(\text{FK}_1) * \text{sel}(\text{PK}_1) = 3179 / 65 * 2 / 54 = 3179 / 1755$$

$$R = n(\text{FK}_2) * \text{sel}(\text{PK}_2) = 3179 / 1755 * 1 = 1.81$$

Assim como na questão anterior, a diferença entre o resultado estimado e o esperado ocorre porque a estimativa assume uma distribuição uniforme dos registros, o que não reflete a realidade. No caso da tabela

`movie_cast`, há uma concentração de registros com `cast_order = 1`, já que esse valor precisa existir primeiro para cada filme, distorcendo novamente a estimativa.

Questão 4

```
SELECT title FROM movie UNION ALL SELECT character_name FROM movie_cast;
```

Resultado esperado:

R = 3419

Resultado estimado:

$R = n(\text{movie}) + n(\text{movie_cast}) = 3419$

Questão 5

```
SELECT title FROM movie EXCEPT SELECT character_name FROM movie_cast
```

Resultado esperado:

R = 227

Resultado estimado:

$R = n(\text{movie}) = 240$

A estimativa não é totalmente precisa, pois assume o pior cenário, em que não existirá nenhum `character_name` igual a um `title`. Dessa forma, a estimativa estabelece um limite superior para a quantidade de registros retornados pela consulta.