

평균, 편차, 분산, 표준편차

이번 장에서는 SQLite에서 자체적으로 지원하지 않는 함수를 설명하는데, 표준편차를 그 예로 든다. 이 장의 내용이 너무 어렵다면 건너뛰었다가 나중에 읽어봐도 좋다.

통계량의 정의

우선 용어 정의부터 알아보자.

이 장에서 ‘평균’은 ‘산술 평균’을 가리킨다. 설명은 [위키백과](#)를 참조하자.

편차, 분산, 표준편차는 [위키백과](#)에 다음과 같이 설명되어 있다.

편차(deviation)는 관측값에서 평균 또는 중앙값을 뺀 것이다.

분산(variance)은 관측값에서 평균을 뺀 값을 제곱하고, 그것을 모두 더한 후 전체 개수로 나눠서 구한다. 즉, 차이값의 제곱의 평균이다. 관측값에서 평균을 뺀 값인 편차를 모두 더하면 0이 나오므로 제곱해서 더한다.

표준 편차(standard deviation)는 분산을 제곱근한 것이다. 제곱해서 값이 부풀려진 분산을 제곱근해서 다시 원래 크기로 만들어준다.

평균

평균을 구하는 방법은 [12장](#)에서 이미 다뤘다. 그런데 평균의 대상이 되는 개별 값과 평균을 표에 나란히 나타내려면 어떻게 해야 할까?

다음 예는 `Person` 테이블에서 `Height`의 개별 값을 나열하며, 평균값도 함께 보여준다. 질의문을 보기 전에 결과부터 먼저 살펴보자. 질의문을 보기 전에 직접 시도해보면 더 좋다.

이름	키	평균
헤리	166.8	167.025
소진	167	167.025
유라	170.3	167.025
민아	164	167.025

키의 평균값은 `167.025`로 단일하므로 모든 행에 똑같이 나타난다. 12장에서 구한 평균값과 차이가 나는 것은 연습 문제 12.1에서 값이 바뀌었기 때문이다.

그러면 위의 결과가 어떻게 해서 나왔는지 질의문을 살펴보자.

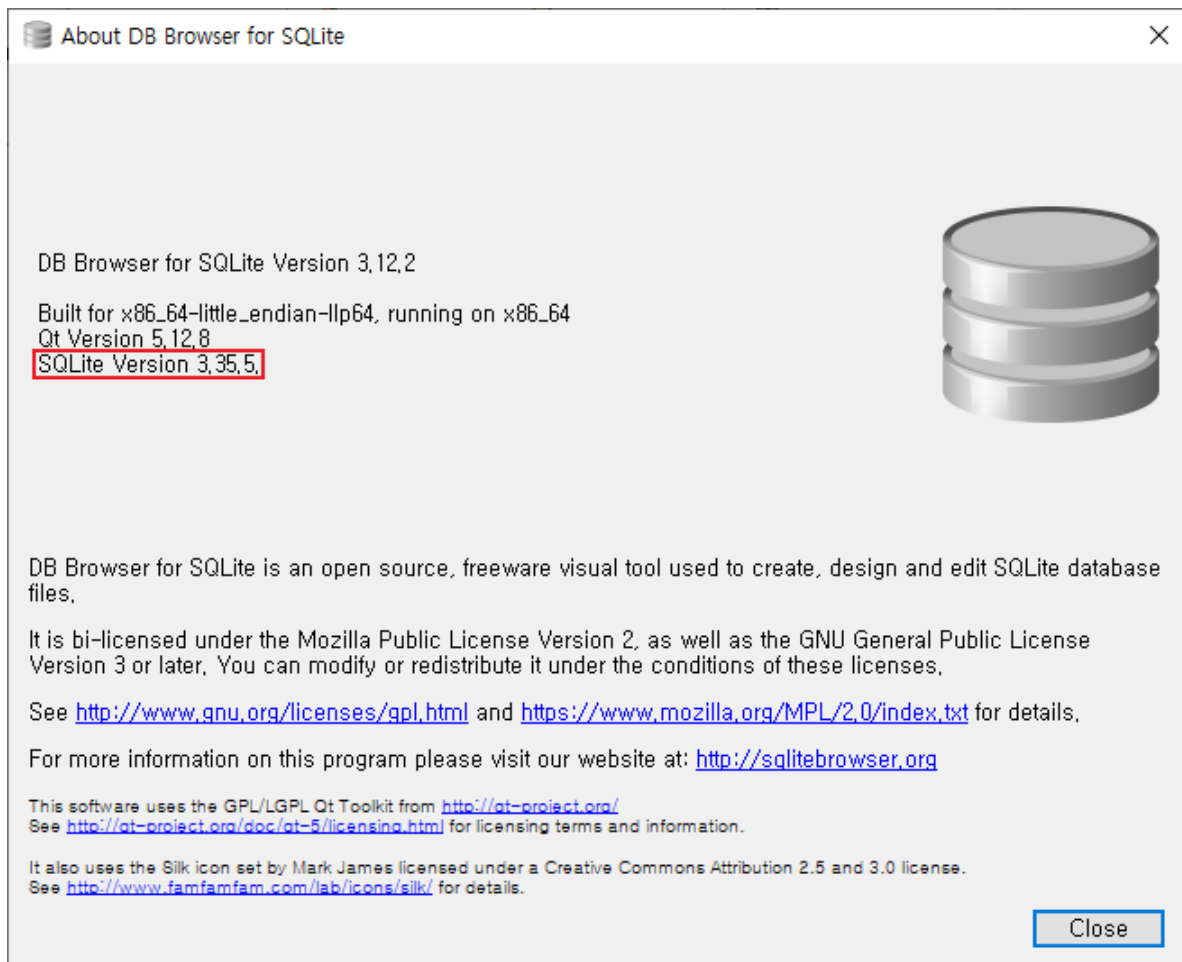
```
SELECT
    Name AS 이름,
    Height AS 키,
    avg(Height) over () AS 평균
FROM Person
```

여기서는 집계 함수로 계산한 결과를 행에 나타내려고 `over ()` 구문을 사용했다.

tip

💡 SQLite 버전 확인

`over ()` 는 SQLite 3.25 버전 이후에서 작동한다. 현재 사용하는 SQLite 버전을 확인하려면 **Help→About** 메뉴를 선택한다.



또는 다음 문장을 실행해도 된다.

```
select sqlite_version();
```

sqlite_version()
3.35.5

`over ()` 를 포함한 질의를 실행할 때 `syntax error`가 발생한다면 위의 방법으로 SQLite 버전을 확인해보고, 버전이 낮다면 DB Browser for SQLite의 최신 버전을 다운로드해 설치하자.

편차

Person 테이블에서 Height 의 평균값과의 편차를 구해보자. 앞에서와 마찬가지로 over () 를 사용한다.

```
SELECT
  Name AS 이름,
  Height AS 키,
  avg(Height) over () AS 평균,
  round(Height - avg(Height) over (), 3) AS 편차
FROM Person
```

이름	키	평균	편차
헤리	166.8	167.025	-0.225
소진	167	167.025	-0.025
유라	170.3	167.025	3.275
민아	164	167.025	-3.025

분산에 대한 설명에서 '편차를 모두 더하면 0이 나오므로 제공해서 더한다'고 했는데, 그렇다면 편차를 제공하지 않은 채로 그대로 더하면 정말로 0이 나오는지 확인해보자.

```
SELECT sum(편차)
FROM (
  SELECT
    round(avg(Height) over () - height, 3) AS 편차
  FROM Person
)
```

sum(편차)
0.0

분산

이번에는 분산을 계산해보자.

```
SELECT avg(편차*편차) AS 분산
FROM (
  SELECT
    round(avg(Height) over () - height, 3) AS 편차
  FROM Person
)
```

분산
4.981875

표준 편차

이제 분산의 제곱근을 계산하면 표준편차도 구할 수 있다. 그런데 이것은 생각보다 복잡한 문제다. SQLite는 제곱근을 구하는 함수를 자체적으로 제공하지 않는다!

SQLite에서 자체적으로 제공하지 않는 함수

SQLite는 자체적으로 제공하는 함수가 많지 않은 대신, 사용자가 함수를 직접 만들어 쓸 수 있게 되어 있다. 함수를 만드는 방법이 궁금하면 [이 글](#)을 읽어 보라.

다행히 제곱근과 표준편차 같이 일반적인 함수는 다른 사람들이 만들어둔 것을 얻어와서 사용할 수 있다. SQLite 웹사이트의 [Contributed Files](#) 페이지에 소개되어 있다. 이것들을 이용하려면 [익스텐션을 로드](#)해야 하는데, 로딩 가능한 파일을 만들려면 소스 코드를 컴파일해야 하며, 이를 위해서는 컴파일러가 필요하다. 윈도우에 마이크로소프트 빌드 도구를 설치하고 사용하는 방법은 [이 문서](#)를 참조한다.

파이썬에서 SQLite를 사용한다면 [이 방법](#)도 검토해 보라.

SQLite 대신 PostgreSQL이나 SQL Server 등 다른 데이터베이스를 사용하는 것이 나은 경우도 있을 것이다