Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

Module 5
SQL Analytic Functions

Lesson 5: Functions for Ratio Comparisons

Lesson Objectives

- Understand concepts about cumulative distribution functions
- Write SELECT statements using functions for ratio comparisons
- Reflect about the importance of ratio comparisons

Motivation

- Ratio comparisons common in business intelligence
- More precise than rankings
- Contribution ratios
  - Part of a whole
  - Share of total sales for each division
- Distribution ratios
  - Size of subsets compared to a population
  - Threshold for top 5% of unit sales

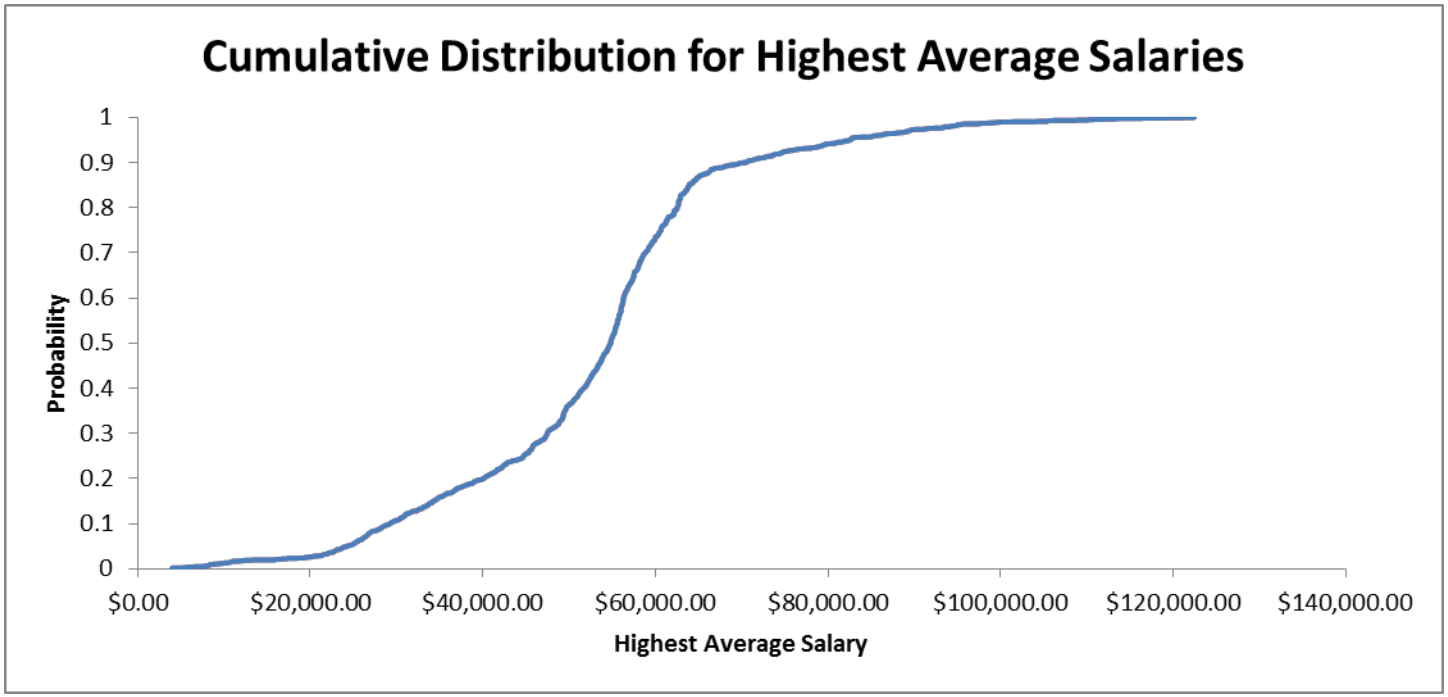
Ratio Comparison Functions

Ratio_To_Report Example

- Contribution ratio on sum of dollar sales by year and customer city
- Partition on year
- Order result by year and descending sum of sales

```
SELECT TimeYear, CustCity, SUM(SalesDollar) AS SumSales,
       RATIO_TO_REPORT(SUM(SalesDollar))
         OVER (PARTITION BY TimeYear) AS SumSalesRatio
 FROM SSCustomer, SSSales, SSTimeDim
 WHERE SSSales.CustID = SSCustomer.CustId
   AND SSSales.TimeNo = SSTimeDim.TimeNo
 GROUP BY TimeYear, CustCity
 ORDER BY TimeYear, SUM(SalesDollar) DESC;
```


Cumulative Distribution



Cume_Dist	Percent_Rank
• (rows preceding inclusive) / *N* • Value range: > 0 to 1 • Cume_Dist(54,950) = 0.50987 (801/1571)	• (rank-1) / (*N*-1) • Value range: ≥ 0 to 1 • Percent_Rank(54,950) = 0.50955 (800/1570)

Cumulative Distribution Example

- Cumulative distribution functions on item unit price
- Display item name, rank, percent rank, row number, and cumulative distribution

```
SELECT ItemName, ItemUnitPrice,
  RANK() OVER (ORDER BY ItemUnitPrice) As RankUnitPrice,
  PERCENT_RANK()
    OVER (ORDER BY ItemUnitPrice) As PercentRankUnitPrice,
  ROW_NUMBER()
    OVER (ORDER BY ItemUnitPrice) As RowNumUnitPrice,
  CUME_DIST()
    OVER (ORDER BY ItemUnitPrice) As CumDistUnitPrice
 FROM SSItem;
```


Example with Equal Values

- Cumulative distribution functions on sum of sales units by customer name
- Display customer name, rank, percent rank, row number, and cumulative distribution

```
SELECT CustName, SUM(SalesUnits) AS SumSalesUnits,
  RANK() OVER (ORDER BY SUM(SalesUnits) ) AS RankSalesUnits,
  PERCENT_RANK() OVER (ORDER BY SUM(SalesUnits) )
   AS PerRankSalesUnits,
  ROW_NUMBER()
    OVER (ORDER BY SUM(SalesUnits)) As RowNumSalesUnits,
  CUME_DIST() OVER (ORDER BY SUM(SalesUnits) ) AS CumDistSalesUnits
 FROM SSSales, SSCustomer
 WHERE SSSales.CUSTID = SSCustomer.CUSTID
 GROUP BY CustName;
```


Top Performers Example

- Cumulative distribution function on item unit price
- Only display top 30% of items with largest unit prices
- Display item name, item brand, item unit price, and cumulative distribution

```
SELECT ItemName, ItemBrand, ItemUnitPrice, CumDistUnitPrice
 FROM ( SELECT ItemId, ItemName, ItemBrand, ItemUnitPrice,
          CUME_DIST()
  OVER ( ORDER BY ItemUnitPrice ) As CumDistUnitPrice
  FROM SSItem ) X
 WHERE CumDistUnitPrice >= 0.7;
```


Additional Problems I

- Example 5
  - Cumulative distribution (Cume_Dist) of dollar sales in Colorado (CO)
  - Remove duplicates
  - Display dollar sales and cumulative distribution
- Example 6
  - Top performing (30%) customer zip codes by year on sum of dollar sales
  - Use either cumulative distribution function
  - Partition by year
  - Display year, store zip code, sum of dollar sales, and cumulative distribution
  - Order by year and cumulative distribution

Additional Problems II

- Example 7
  - Contribution ratio on sum of 2021 unit sales by month and item brand
  - Partition on month
  - Display month, item brand, sum of unit sales, and contribution ratio
  - Order result by month and descending sum of unit sales

Summary

- Support common ratio comparisons in business intelligence
- RATIO_TO_REPORT for contributions of additive columns to total
- CUME_DIST and PERCENT_RANK for cumulative distributions