# Basic Statistics using SQL

*MD Arshad Ahmad*
*15 Years+ Experience in Data Science*
*Mentored 100+ people*

Decoding Data Science

# Variance and Standard Deviation

Code

```
select
  month_of_year,
  SUM(units_sold),
  ROUND( AVG(units_sold),2),
  ROUND(VAR_POP(units_sold), 2),
  ROUND(STDDEV_POP(units_sold), 2)
FROM
   store_sales
GROUP BY
    Month_of_year;
```

Output

Data Output | Explain | Messages | Notifications

| | month_of_year integer | sum bigint | round numeric | round numeric | round numeric |
|---|---|---|---|---|---|
| 1 | 12 | 18613 | 600.42 | 53431.34 | 231.15 |
| 2 | 3 | 16134 | 520.45 | 13782.12 | 117.40 |
| 3 | 11 | 16073 | 535.77 | 28680.25 | 169.35 |
| 4 | 8 | 15816 | 510.19 | 12227.38 | 110.58 |

# Continuous Percentiles

```
3  SELECT
4      PERCENTILE_CONT(0.95) WITHIN GROUP (order by revenue) as pct_95c_rev,
5      PERCENTILE_DISC(0.95) WITHIN GROUP (order by revenue) as pct_95d_rev
6  FROM
7      store_sales;
8
9
```

Data Output | Explain | Messages | Notifications

| | pct_95c_rev double precision | pct_95d_rev integer |
|---|---|---|
| 1 | 9189.6 | 9192 |

Decoding Data Science

# Discrete Percentiles

```
36  SELECT
37     PERCENTILE_DISC(0.50) WITHIN GROUP (ORDER BY revenue) as pct_50_rev,
38     PERCENTILE_DISC(0.60) WITHIN GROUP (ORDER BY revenue) as pct_60_rev,
39         PERCENTILE_DISC(0.90) WITHIN GROUP (ORDER BY revenue) as pct_90_rev,
40         PERCENTILE_DISC(0.95) WITHIN GROUP (ORDER BY revenue) as pct_95_rev FROM
41    store_sales;
42
```

Data Output | Explain | Messages | Notifications

| | pct_50_rev integer | pct_60_rev integer | pct_90_rev integer | pct_95_rev integer |
|---|---|---|---|---|
| 1 | 5856 | 6360 | 8604 | 9192 |

Decoding Data Science

# Correlation Coefficients

```
12
13  SELECT   CORR(units_sold, employee_shifts) as corr_units_shift ,CORR(units_sold, month_of_year) as corr_unit_moy  FROM   store_sales
14
```

Data Output | Explain | Messages | Notifications

| | corr_units_shift<br>double precision | corr_unit_moy<br>double precision |
|---|---|---|
| 1 | 0.5595361395476149 | 0.1281421944964 |

Decoding Data Science

# Mode

```
12  SELECT   month_of_year,   MODE() WITHIN GROUP (ORDER BY employee_shifts) as emp_mode
13  FROM   store_sales GROUP BY   month_of_year
14
```

Data Output | Explain | Messages | Notifications

| | month_of_year integer | emp_mode integer |
|---|---|---|
| 1 | 1 | 3 |
| 2 | 2 | 4 |
| 3 | 3 | 4 |
| 4 | 4 | 4 |
| 5 | 5 | 5 |
| 6 | 6 | 5 |

Decoding Data Science

# Row Number

```
11
12  SELECT   sale_date,   month_of_year,   units_sold,   ROW_NUMBER () OVER (ORDER BY units_sold)
13  FROM   store_sales ORDER BY   sale_date
14
```

Data Output | Explain | Messages | Notifications

| | sale_date date | month_of_year integer | units_sold integer | row_number bigint |
|---|---|---|---|---|
| 1 | 2017-01-01 | 1 | 241 | 41 |
| 2 | 2017-01-02 | 1 | 205 | 32 |
| 3 | 2017-01-03 | 1 | 200 | 28 |
| 4 | 2017-01-04 | 1 | 458 | 158 |

Decoding Data Science

# Intercepts

```
11  SELECT REGR_INTERCEPT(units_sold,employee_shifts) as intercept1 ,REGR_INTERCEPT(employee_shifts, units_sold) as intercept1 FROM store_sales
12
13
```

Data Output | Explain | Messages | Notifications

| | intercept1<br>double precision | intercept1<br>double precision |
|---|---|---|
| 1 | 63.08848077595483 | 2.6458185064135895 |

Decoding Data Science

# Intercepts

```
11  SELECT REGR_INTERCEPT(units_sold,employee_shifts) as intercept1 ,REGR_INTERCEPT(employee_shifts, units_sold) as intercept1 FROM store_sales
12
13
```

Data Output | Explain | Messages | Notifications

| | intercept1<br>double precision | intercept1<br>double precision |
|---|---|---|
| 1 | 63.08848077595483 | 2.6458185064135895 |

Decoding Data Science

# Thank You!
# To know more Get In Touch!

# KicK start your Data Science Career

Book Mentoring Session

Analytics Blog

Decoding Data Science