

ヒストグラム統計の分布（インデックス列の要素の偏り）の表示 列データの要素の偏りを表しているヒストグラム統計の作成方法

列の値の種類数の確認

```
select table_name, column_name, num_distinct from dba_tab_columns
where owner = 'KOZUE'
and table_name = 'EMP'
and column_name = 'DEPTNO' ;
```

TABLE_NAME	COLUMN_NAME	NUM_DISTINCT	
EMP	DEPTNO	5	← 種類数

もしくは、

```
select count (distinct (列名)) from テーブル名;
```

ヒストグラム作成のメリット

インデックスが作成してある列でも列の値の分布に偏りがあり、インデックス使用が効率が悪い場合には、全件 **Read** を行うように実行計画を立てる

ヒストグラム統計の取得

```
begin
  dbms_stats.gather_table_stats (
    ownname => 'スキーマ名',
    tabname => 'テーブル名',
    method_opt => 'for columns (列名 , 列名) size 10'
  ) ;
end ;
/
```

もしくは、

```
analyze table テーブル名 compute statistics for columns 列名 size 分類数
, 列名 size 分類数 , . . . . ;
```

```
analyze table テーブル名 compute statistics for all columns ;
(すべての列に対して、ヒストグラム統計を作成)
```

ヒストグラム統計の状態確認

```
select table_name, column_name, num_distinct, num_buckets
from dba_tab_colstats
where owner = 'KOZUE'
and table_name = 'EMP' ;
```

TABLE_NAME	COLUMN_NAME	NUM_DISTINCT	NUM_BUCKETS
EMP	EMPNO	4	4
EMP	DEPTNO	256	20

列名 列の値の種類数 ヒストグラムの区分数

列の値の種類数 = ヒストグラムの区分数 \times 値ベース・ヒストグラム
列の値の種類数 > ヒストグラムの区分数 \times 高さベース・ヒストグラム

ヒストグラムの分布状態表示

【値ベース・ヒストグラム】

```
col column_name format a50
select table_name, column_name, endpoint_number, endpoint_value
from dba_histograms
where owner = 'スキーマ名'
and table_name = 'テーブル名'
and column_name = '列名'
order by endpoint_number ;
```

TABLE_NAME	COLUMN_NAME,	ENDPOINT_NUMBER	ENDPOINT_VULUE
		2500	10
		5000	20
		7500	30
		9500	40
		↑	↑

この列値までのレコード件数累計

列値

よって、この列に含まれるのは、「この列までの累計」 — 「前の列までの累計」

$$9500 - 7500 = 2000 \text{ (件)}$$

【高さベース・ヒストグラム】

```
col column_name format a50
select table_name, column_name, endpoint_number, endpoint_value
from dba_histograms
where owner = 'スキーマ名'
and table_name = 'テーブル名'
and column_name = '列名'
order by endpoint_number ;
```

TABLE_NAME	COLUMN_NAME,	ENDPOINT_NUMBER	ENDPOINT_VULUE
	最小値を格納する跳ぶ別なバケット→	0	100
		1	1000
		2	3000
		5	3500
		↑	↑

バケット番号

このバケットに入る最大列値

※ この例では、列値 3500 のレコードは、バケット 3~5 に入る

バケット 3 の最大値 3500

バケット 4 の最大値 3500

バケット 5 の最大値 3500

※ ここで、バケット 6 の ENDPOINT が 4500 であれば、バケット 6 には 3500~4500 の値が一様に分布していると考え