

記事

[Mihoko Iijima](#) · 2020年11月6日 4m read

外れ値について

これはInterSystems FAQ サイトの記事です。

テーブルチューニングを行った際に、
フィールドに値がほとんど登録されていない (Null)

場合や、特定の値がほとんどを占める場合、その値を **[外れ値]** として除外して選択性計算を行います。
また、**外れ値が全レコードの何 % を占めているかの値**は **[外れ値の選択性]** として記録されます。

InterSystems 製品のクエリオプティマイザは、選択性数値とエクステントサイズを使用してクエリの経路を決定しますが、クラスクエリ、埋め込み SQL に使用しているクエリに外れ値が含まれる場合は、外れ値の選択性が自動的に考慮され、インデックスの使用有無を決定しています。

ダイナミック SQL、ODBC / JDBC 経由でのクエリについては、外れ値が Null である場合、自動的に外れ値の選択性が考慮されますが、Null 以外の特定の値が外れ値に検出される場合は、明示的に指示を与えるまで考慮しません。

詳細は、ドキュメント ([異常値に対する述語条件【IRIS】](#) / [異常値に対する述語条件【Caché / Ensemble】](#)) をご参照ください。

SAMPLES ネームスペースの Sample.Person テーブルの FavoriteColors には、好きな色が複数格納されていますが未登録 (NULL) の値も多く含まれています (インストール直後の状態で 200 件中 71 件が NULL)。選択性の計算で FavoriteColors を外れ値と認識した場合、[外れ値] に NULL、[外れ値の選択性] に 35.5% (71/200×100=35.5) が表示されます。

IRIS / IRIS for Health でお試しいただく場合の Sample.Person 作成方法

IRIS / IRIS for Health にはSAMPLESネームスペースはありませんので、任意のネームスペースに Sample.Person のサンプルをインポートします。

サンプルインポート方法詳細は、ドキュメント([InterSystems IRIS で使用するサンプルのダウンロード](#))、または[関連記事【サンプル \(Sample.Person\) のクラス定義ダウンロードとサンプルデータの作成について】](#)をご参照ください。

IRIS / IRIS for Health でお試しいただく場合、以降の説明に登場する SAMPLES ネームスペースは、Sample.Person クラスをインポートしたネームスペースに読み替えてお試しく下さい。

管理ポータルで確認できます。

- 1、[システムエクスプローラ] > [SQL] を開きます。
- 2、画面中央上部にある[変更]を押下し、ネームスペースをSAMPLESに切り替えます。

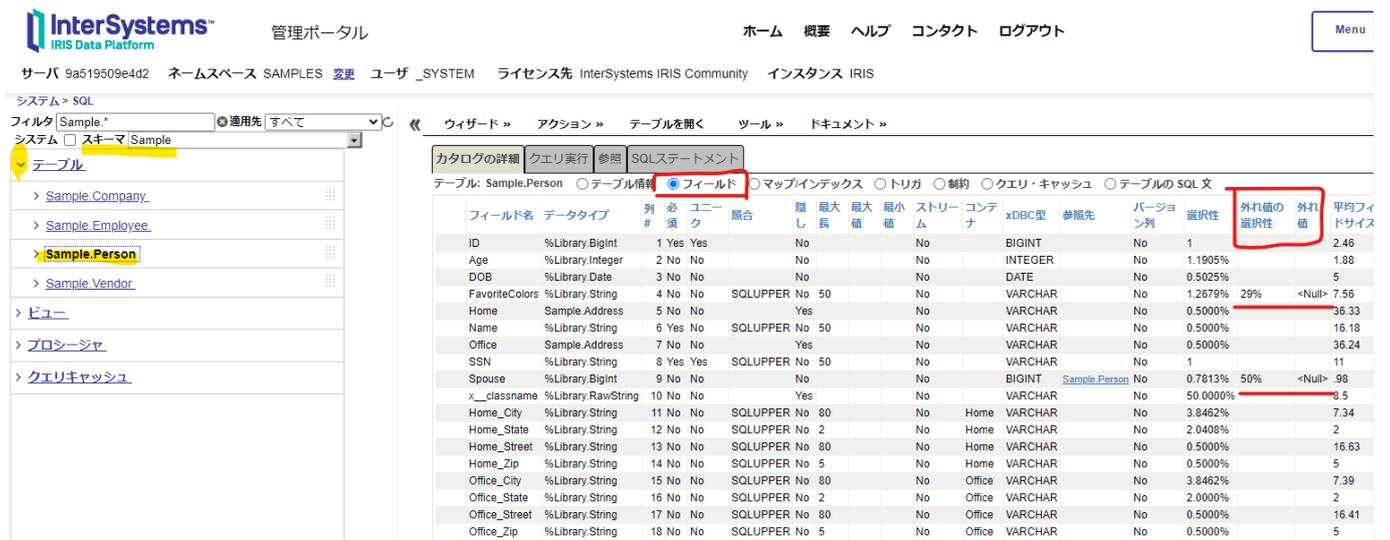
外れ値について

Published on InterSystems Developer Community (<https://community.intersystems.com>)



[スキーマ]のプルダウンから Sample を選択し、[テーブル]の > をクリックして階層を展開し、[Sample.Person] を選択します。

画面右側で [カタログ詳細] タブの [フィールド] のラジオボタンをクリックすると詳細表示が開きます。



外れ値を考慮したプランを確認するため、Sample.Person クラスの FavoriteColors に対するインデックスを追加し、インデックスを再構築します。

インデックス追加例は以下の通りです。

```
/// ???????? ?FavoriteColors????????????????????
```

```
Index FavoriteColorsIdx On (FavoriteColors(ELEMENTS), FavoriteColors(KEYS));
```

作成したインデックス再構築します。メソッドを利用する方法は以下の通りです。

```
do ##class(Sample.Person).%BuildIndices($LB("FavoriteColorsIdx"))
```

外れ値について

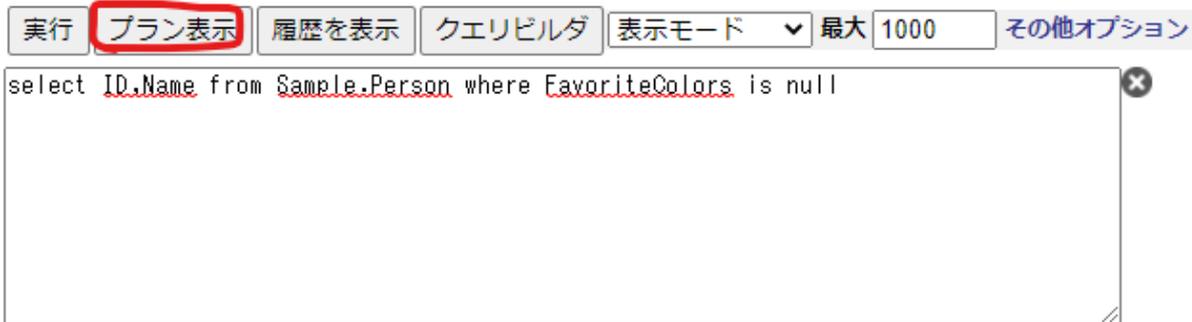
Published on InterSystems Developer Community (<https://community.intersystems.com>)

管理ポータルで再構築する方法は、以下の通りです。

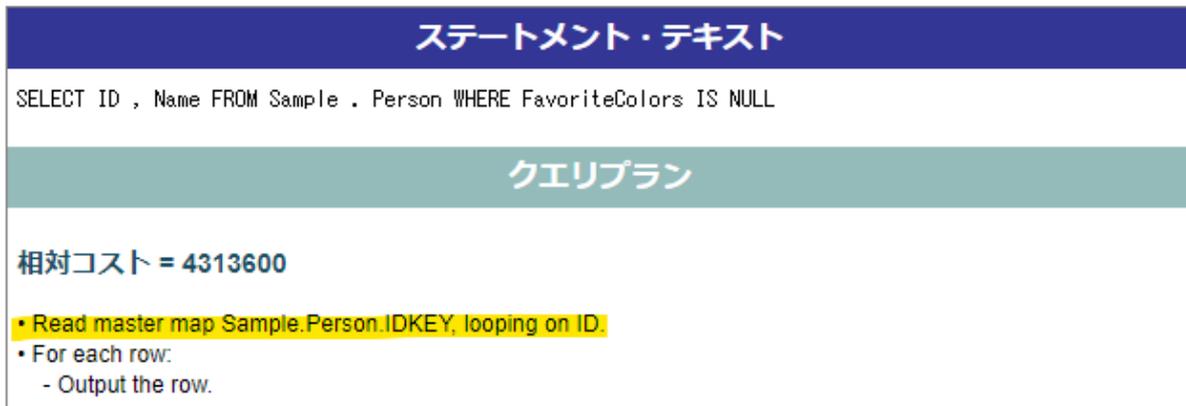


管理ポータル、またはターミナルで以下のSQLを実行し、クエリプランを確認します。

```
select ID,Name from Sample.Person where FavoriteColors is null
```



実行プランが以下に表示されます:



プラン表示より レコード全体 (Read master map) を参照したことを示しています。

where FavoriteColors is null を指定した場合、ダイナミック SQL でも外れ値 : Null を考慮したプランを選択するため、インデックス : FavoriteColorsIdx が使用されていないことがわかります。

次に、条件式をNull以外に変え実行します。

```
select ID,Name from Sample.Person where for some %element(FavoriteColors) (%value='red')
```

```
select ID,Name from Sample.Person where for some %element(FavoriteColors) (%value='red')
```

実行プランが以下に表示されます:

ステートメント・テキスト

```
SELECT ID , Name FROM Sample . Person WHERE FOR SOME %element ( FavoriteColors ) ( %VALUE = ? ) /*#OPTIONS {"DynamicSQLTypeList":"1"} */
```

クエリプラン

相対コスト = 35132

- Call module B, which populates bitmap temp-file A.
- Read bitmap temp-file A, looping on ID.
- For each row:
 - Read master map Sample.Person.IDKEY, using the given idkey value.
 - Output the row.

Module: B

- Read index map Sample.Person.FavoriteColorsIdx, using the given %SQLUPPER(Subvalue(FavoriteColors)), and looping on %EXACT(Subvalue(FavoriteColors)) and ID.
- For each row:
 - Add ID bit to bitmap temp-file A.

Module:Bを参照すると、FavoriteColorsIdxを使用していることがわかります。

外れ値についての詳細は、ドキュメント ([Selectivity と Outlier Selectivityについて【IRIS】](#) / [Selectivity と Outlier Selectivityについて【Caché / Ensemble】](#)) もご参照ください。

上記ドキュメント中の [異常値] や [Outlier] は [外れ値] と同義語です。

[#SQL](#) [#インデックス付け](#) [#ヒントとコツ](#) [#Caché](#) [#Ensemble](#) [#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)

ソースURL:

<https://jp.community.intersystems.com/post/%E5%A4%96%E3%82%8C%E5%80%A4%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6>