

サンプリング

テーブルからサンプルレコードを抽出する

◇機能

THUMGY Dataにおける「サンプリング」とは、テーブルから一部のレコードをサンプルとして抽出することを指します。抽出したサンプルは、試査に使用します。

サンプリングの種類は、「レコードサンプリング」と「金額単位サンプリング」の2種類があります。さらにサンプルの選択方法を「ランダム」か「固定間隔」にするかにより、4パターンのサンプル抽出方法が用意されています。

サンプリングの種類	機能	サンプルの選択方法	説明
レコードサンプリング	レコードを基準にサンプルを抽出する	ランダム	乱数を利用して、ランダムにサンプルレコードを抽出する。
		固定間隔	指定した間隔ごとのレコードをサンプルレコードとして抽出する。
金額単位サンプリング	売上金額など、金額のフィールドの値を基準にサンプルを抽出する	ランダム	乱数を利用して、ランダムにサンプルレコードを抽出する。
		固定間隔	金額の累計が指定した間隔に達するごとにレコードをサンプルレコードとして抽出する。

金額単位サンプリングでは、「カットオフ値」を指定すると、指定した金額以上のレコードをすべてサンプルレコードとして抽出することができます。売上金額が高額なレコードを全件調査対象とする場合などに「カットオフ値」を使用します。

<サンプル抽出のイメージ>

分析対象のテーブル

No	売上番号	計上日	金額
1	11509	2022/01/05	11,200
2	11510	2022/01/05	25,000
3	11511	2022/01/05	165,000
4	11512	2022/01/06	24,480
5	11513	2022/01/06	5,400
6	11514	2022/01/08	4,500
7	11515	2022/01/09	324,000
8	11516	2022/01/09	4,200
9	11517	2022/01/09	72,000
10	11518	2022/01/10	35,200
11	11519	2022/01/10	781,600
12	11520	2022/01/11	1,352,000
13	11521	2022/01/11	1,611,350
14	11522	2022/01/13	61,200
15	11523	2022/01/14	343,800
16	11524	2022/01/16	12,000
17	11525	2022/01/17	12,000
18	11526	2022/01/19	3,600
19	11527	2022/01/20	8,840
20	11528	2022/01/20	16,380
21	11529	2022/01/22	290,000
22	11530	2022/01/24	74,100

- 「ランダム」による抽出：ランダムにレコードが抽出されます。

No	売上番号	計上日	金額
5	11513	2022/01/06	5,400
19	11527	2022/01/20	8,840
3	11511	2022/01/05	165,000
9	11517	2022/01/09	72,000
4	11512	2022/01/06	24,480

- 「固定間隔」による抽出：指定した間隔ごとにレコードが抽出されます。

- レコードサンプリングの場合：間隔値を「5」とした場合、5件ごとにレコードが抽出されます。

No	売上番号	計上日	金額
1	11509	2022/01/05	11,200
6	11514	2022/01/08	4,500
11	11519	2022/01/10	781,600
16	11524	2022/01/16	12,000
21	11529	2022/01/22	290,000

- 金額単位サンプリングの場合：金額のフィールドの累計が指定した間隔に達するごとにレコードが抽出されます。間隔値を「50万」とした場合、金額の累計が50万、100万、150万・・・に達した位置のレコードが抽出されます。カットオフ値を「100万」とすると、100万以上のレコードはすべて抽出されます。（カットオフ値の機能は、「ランダム」でも同様です。）

No.	売上番号	計上日	金額
1	11509	2022/01/05	11,200
7	11515	2022/01/09	324,000
11	11519	2022/01/10	781,600
14	11522	2022/01/13	61,200
21	11529	2022/01/22	290,000
12	11520	2022/01/11	1,352,000
13	11521	2022/01/11	1,611,350

「カットオフ値」による抽出

分析対象のテーブルのレコード件数が多い場合など、レコード全件ではなく、一部のレコードを調査対象（サンプルレコード）としてピックアップすることがあります。《サンプリング》コマンドを使用すると、ランダムにレコードを抽出することができます。

◇メニュー

《サンプリング》-《サンプリング》

◇《サンプリング》の基本的設定内容

《サンプリング》の画面は、「種類」の選択により、設定項目が異なります。下の図では、基本的設定内容について説明します。

1. テーブル名 : 操作の対象となっているテーブルが表示されます。
2. 種類 (必須) : リストからサンプリングの種類を選択します。
 - レコードサンプリング : レコードを基準にサンプルを抽出する場合に選択します。(レコードサンプリングがデフォルトで表示されます。)
 - 金額単位サンプリング : 売上金額など、金額のフィールドの値を基準にサンプルを抽出します。
3. サンプル対象 (「種類」で「金額単位サンプリング」を選択した場合は必須) : 金額単位サンプリングを行う際に、サンプル抽出の基準に使用する数値フィールドを選択します。(レコードサンプリングの場合は、表示されません。)
4. 選択方法 (必須) : リストからサンプルレコードを選択する方法を選択します。
 - ランダム : サンプルレコードをランダムに抽出する場合に選択します。(ランダムがデフォルトで表示されます。)
 - 固定間隔 : サンプルレコードを指定した間隔で抽出する場合に選択します。
5. 《サンプルサイズの計算》ボタン : 抽出すべきサンプルレコード数を計算する場合に使用します。

6. パラメータ : 「種類」、「選択方法」で選択した内容により、設定する項目が変わります。
「◇《サンプリング》のパラメータの設定内容」を参照してください。
7. 保存先（必須） : 抽出されたサンプルレコードを保存する新規テーブルの名前を入力します。
 - テーブル名の長さは、64文字までです。
 - ファイルの保存先は、プロジェクトフォルダの「04_table」フォルダです。
 - 《保存先》ボタンで、保存するフォルダを変更できます。ただし、使用可能なフォルダは、プロジェクトフォルダの「04_table」フォルダおよびそのサブフォルダのみです。
8. 《実行》ボタン : 実行します。
9. 《キャンセル》ボタン : 実行せずに画面を閉じます。

◇《サンプリング》のパラメータの設定内容

パラメータの設定内容は、サンプリングの種類、選択方法の組み合わせにより異なります。ここでは、組み合わせごとの設定内容について説明します。

- (1)レコードサンプリング - ランダム の場合
- (2)レコードサンプリング - 固定間隔 の場合
- (3)金額単位サンプリング - ランダム の場合
- (4)金額単位サンプリング - 固定間隔 の場合

※《サンプルサイズの計算》ボタンは、「◇サンプルサイズの計算」を参照してください。

(1)レコードサンプリング - ランダム の場合

1. サンプルサイズ（必須）： サンプルとして抽出するレコードの件数を数値で指定します。
《サンプルサイズの計算》ボタンでサンプルサイズを算出することもできます。
2. シード値： サンプルレコードの抽出に使用される乱数を生成するための任意の数値を入力します。
空欄にした場合は、実行の都度、異なるサンプルレコードが抽出されます。分析の再実行の際に、同じサンプルレコードを抽出する必要がある場合は、「シード値」の指定は必須です。

(2)レコードサンプリング - 固定間隔 の場合

サンプル

テーブル名 : 売上データ2017

種類 レコードサンプリング

選択方法 固定間隔

サンプルサイズの計算

パラメータ

間隔値 1

最初の値 2

保存先

キャンセル 実行

1. 間隔値（必須） : サンプルレコードを抽出する間隔を数値で指定します。
《サンプルサイズの計算》ボタンで間隔を算出することもできます。
2. 最初の値 : 何レコード目からサンプルを抽出するかを1～9の数値で指定します。空欄にした場合は、1～9の値がランダムに使用されるため、実行の都度、異なるサンプルレコードが抽出されます。分析の再実行の際に、同じサンプルレコードを抽出する必要がある場合は、「最初の値」の指定は必須です。

(3)金額単位サンプリング - ランダム の場合

サンプル

テーブル名 : 売掛金残高データ2018

種類

サンプル対象

選択方法 サンプルサイズの計算

パラメータ

サンプルサイズ 1

シード値 2

母集団 計算 3

カットオフ値 3 4

保存先

キャンセル 実行

1. サンプルサイズ（必須） : サンプルとして抽出するレコードの件数を数値で指定します。
《サンプルサイズの計算》ボタンでサンプルサイズを算出することもできます。
2. シード値 : サプリングをする前にランダムにレコードを並び替えたうえで、サンプルレコードを抽出します。「シード値」には、乱数を生成するための任意の数値を入力します。
「シード値」を空欄にした場合は、実行の都度、異なるサンプルレコードが抽出されます。
分析の再実行で、同じサンプルレコードを抽出する必要がある場合は、「シード値」の指定は必須です。
3. 母集団（必須） : 「サンプル対象」で指定した数値フィールドの合計値の絶対値を数値で指定します。
《計算》ボタンをクリックすると、自動で入力されます。
4. カットオフ値 : カットオフ値に金額（数値）を指定すると、「サンプル対象」のフィールドの値がカットオフ値以上のレコードは、すべてサンプルレコードとして抽出されます。

(4)金額単位サンプリング - 固定間隔 の場合

サンプル

テーブル名 : 売掛金残高201812

種類

サンプル対象

選択方法 サンプルサイズの計算

パラメータ

間隔値 1

最初の値 2

カットオフ値 3

シード値 4

保存先

キャンセル 実行

1. 間隔値（必須） : サンプルレコードを抽出する間隔を数値で指定します。
《サンプルサイズの計算》ボタンで間隔を算出することもできます。
2. 最初の値 : 何レコード目からサンプルを抽出するかを1～9の数値で指定します。空欄にした場合は、1～9の値がランダムに使用されるため、実行の都度、異なるサンプルレコードが抽出されます。分析の再実行の際に、同じサンプルレコードを抽出する必要がある場合は、「最初の値」の指定は必須です。
3. カットオフ値 : カットオフ値に金額（数値）を指定すると、「サンプル対象」のフィールドの値がカットオフ値以上のレコードは、すべてサンプルレコードとして抽出されます。
4. シード値 : サンプリングをする前にランダムにレコードを並び替えたうえで、固定間隔ごとのレコードを抽出します。「シード値」には、乱数を生成するための任意の数値を入力します。「シード値」を空欄にした場合は、実行の都度、異なるサンプルレコードが抽出されます。分析の再実行で、同じサンプルレコードを抽出する必要がある場合は、「シード値」の指定は必須です。

◇サンプルサイズの計算

《サンプルサイズの計算》ボタンで、適正なサンプル数を算出することができます。《サンプルサイズの計算》ボタンで設定する項目は、サンプリングの種類（レコードサンプリング/金額単位サンプリング）により異なります。

（参考）サンプルサイズの計算の考え方

内部統制のサンプル件数は25件のサンプルをテストするという考え方がありますが、このサンプル件数は、二項分布、信頼度、許容逸脱率、予想逸脱率から統計的に算出できます。具体的には、信頼度が90%、許容逸脱率が9%、予想誤謬率が0%の場合、統計的にはサンプル数は25件になります。

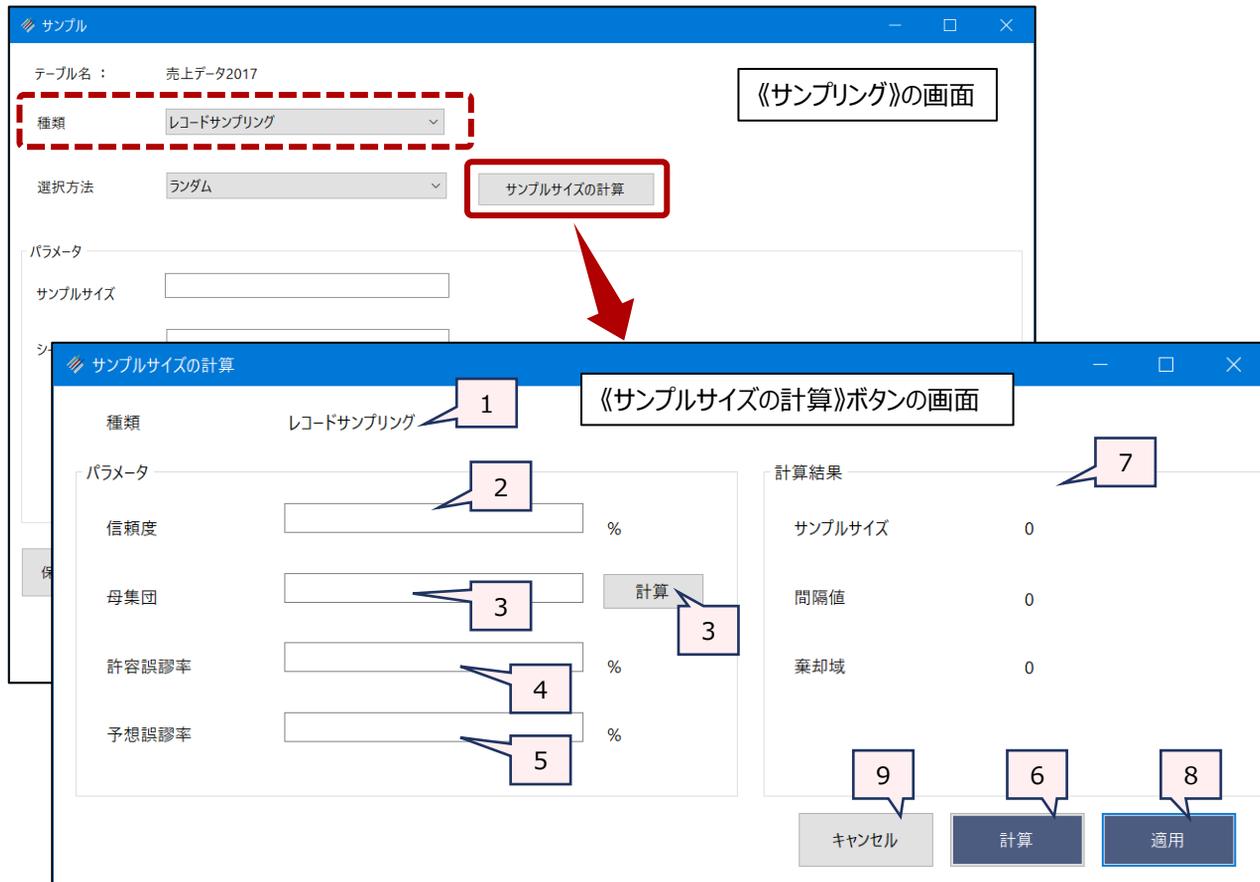
詳細は、日本公認会計士協会 監査・保証基準委員会発行の「財務報告内部統制監査基準報告書第1号 財務報告に係る内部統制の監査」の「付録2 統計的サンプル数の例示」を参照してください。

THUMGY Dataのサンプリング機能を使うと、信頼度、許容逸脱率、予想逸脱率を入力するだけで、統計的にテストが必要なサンプル件数が算出できます。

操作上の注意

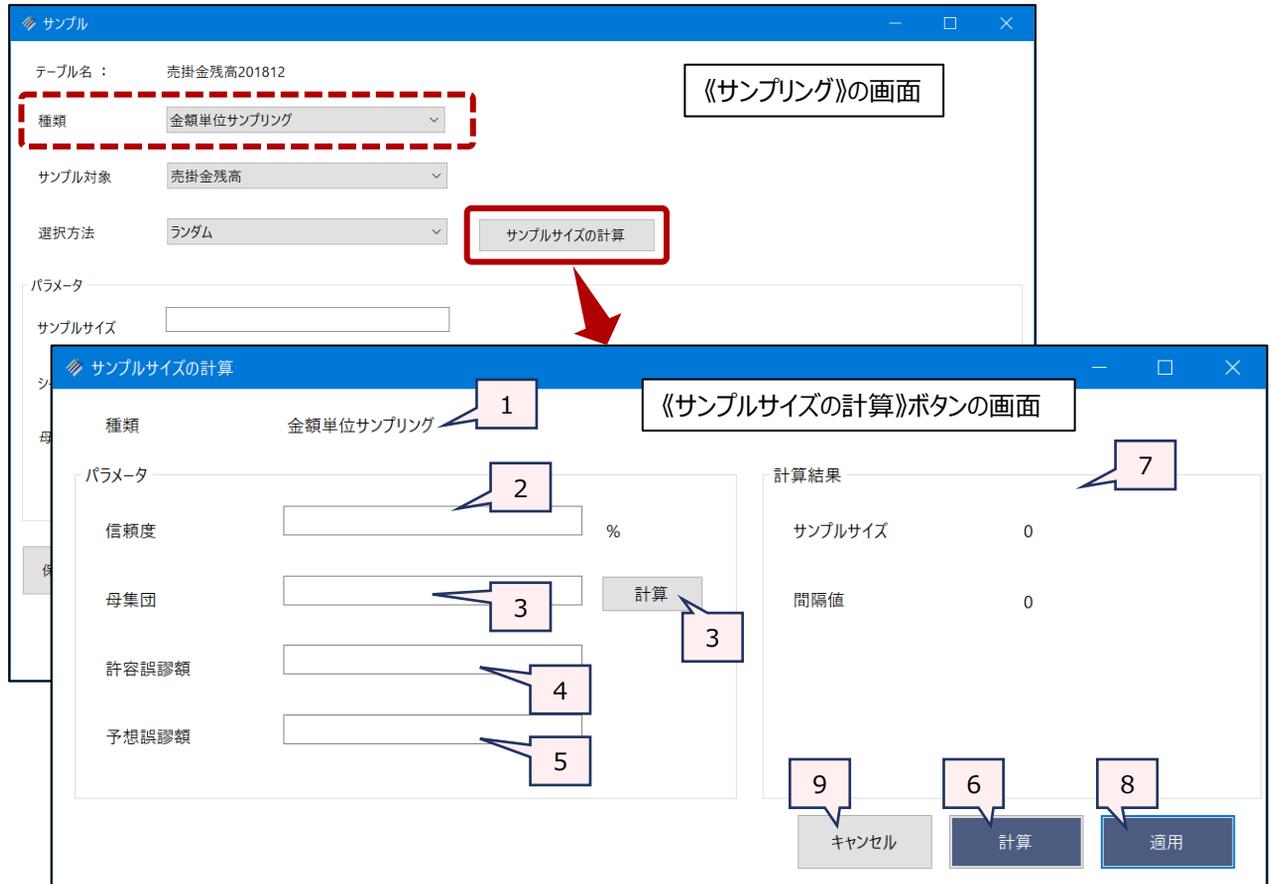
《サンプルサイズの計算》ボタンの画面で入力した内容は、《キャンセル》ボタンまたは《適用》ボタンで画面を閉じると削除されます。再度《サンプルサイズの計算》ボタンをクリックしても表示されません。入力した内容を確認したい場合は、ログで確認してください。

● レコードサンプリングの場合



1. 種類 : 選択しているサンプリングの種類が表示されます。
2. 信頼度 (必須) : 信頼度をパーセンテージで入力します。「90%」の場合は、「90」と入力します。
3. 母集団 (必須) : 対象のテーブルのレコード件数を数値で入力します。
《計算》ボタンをクリックすると、レコード件数が自動で入力されます。
4. 許容誤謬率 (必須) : 信頼度の水準を考慮して、受け入れることができる誤謬率を入力します。
5. 予想誤謬率 (必須) : 母集団の中に存在すると予想される誤謬率を入力します。
6. 《計算》ボタン : サンプルサイズの計算を実行します。
7. 計算結果 : 《計算》ボタンのクリックにより、計算結果が表示されます。
 - サンプルサイズ
 - 間隔値
 - 棄却域
8. 《適用》ボタン : サンプルサイズの計算結果を《サンプリング》の設定画面に反映します。
 - 「選択方法」がランダムの場合：計算結果の「サンプルサイズ」が《サンプリング》の「サンプルサイズ」に入力されます。
 - 「選択方法」が固定間隔の場合：計算結果の「間隔値」が《サンプリング》の「間隔値」に入力されます。
9. 《キャンセル》ボタン : 適用せずに画面を閉じます。

● 金額単位サンプリングの場合



1. 種類 : 選択しているサンプリングの種類が表示されます。
2. 信頼度 (必須) : 信頼度をパーセンテージで入力します。「90%」の場合は、「90」と入力します。
3. 母集団 (必須) : 《サンプリング》の「サンプル対象」で指定したフィールドの合計の絶対値を入力します。《計算》ボタンをクリックすると、自動で入力されます。
4. 許容誤謬額 (必須) : 信頼度の水準を考慮して、受け入れることができる誤謬額を入力します。
5. 予想誤謬額 (必須) : 母集団の中に存在すると予想される誤謬額を入力します。
6. 《計算》ボタン : サンプルサイズの計算を実行します。
7. 計算結果 : 《計算》ボタンのクリックにより、計算結果が表示されます。
 - サンプルサイズ
 - 間隔値
8. 《適用》ボタン : サンプルサイズの計算結果を《サンプリング》の設定画面に反映します。
 - 「選択方法」がランダムの場合 : 計算結果の「サンプルサイズ」が《サンプリング》の「サンプルサイズ」に入力されます。
 - 「選択方法」が固定間隔の場合 : 計算結果の「間隔値」が《サンプリング》の「間隔値」に入力されます。
9. 《キャンセル》ボタン : 適用せずに画面を閉じます。

◇操作の手順

(例) レコードサンプリングの固定間隔で、サンプルレコードを抽出する

- ・対象テーブルのレコード件数：336件
 - ・抽出するサンプル数：《サンプルサイズの計算》ボタンで算出する
 - ・信頼度：90%
 - ・許容誤謬率：10%
 - ・予想誤謬率：0%
- ※《サンプルサイズの計算》に用いる数値は、社内のルール等で定められた方法で算出済みとします。

操作は、以下のステップで行います。

操作のステップ

(1) テーブルの選択 → (2) メニューの選択 → (3) サンプリングの設定 → (4) サンプルサイズの計算 → (5) パラメータの設定 → (6) テーブルの保存

(1) テーブルの選択

ナビゲータからサンプルレコードを抽出したいテーブルを開きます。

(2) メニューの選択

メニューバーの《サンプリング》をクリックし、《サンプリング》をクリックします。

(3) サンプリングの設定

「種類」で「レコードサンプリング」、「選択方法」で「固定間隔」を選択します。

サンプル

テーブル名： 売上データ2017

種類： レコードサンプリング

選択方法： 固定間隔

サンプルサイズの計算

パラメータ

間隔値

最初の値

保存先

キャンセル 実行

(4) サンプルサイズの計算

- ①《サンプルサイズの計算》ボタンをクリックします。
- ②「サンプルサイズの計算」画面のパラメータを入力します。
 - ・信頼度：90
 - ・母集団：336 ※「母集団」の入力欄の右の《計算》ボタンをクリックすると、自動で入力されます。
 - ・許容誤謬率：10
 - ・予想誤謬率：0

サンプルサイズの計算	
種類	レコードサンプリング
パラメータ	
信頼度	90 %
母集団	336 <input type="button" value="計算"/>
許容誤謬率	10 %
予想誤謬率	0 %
計算結果	
サンプルサイズ	0
間隔値	0
棄却域	0
<input type="button" value="キャンセル"/> <input type="button" value="計算"/> <input type="button" value="適用"/>	

- ③画面右下の《計算》ボタンをクリックし、表示された「計算結果」を確認します。
ここでは、「サンプルサイズ」が24、「間隔値」が「14.0」。「棄却域」が0と表示されたものとします。

サンプルサイズの計算	
種類	レコードサンプリング
パラメータ	
信頼度	90 %
母集団	336 <input type="button" value="計算"/>
許容誤謬率	10 %
予想誤謬率	0 %
計算結果	
サンプルサイズ	24
間隔値	14.0
棄却域	0
<input type="button" value="キャンセル"/> <input type="button" value="計算"/> <input type="button" value="適用"/>	

- ④《適用》ボタンをクリックします。

(5) パラメータの設定

「間隔値」に《サンプルサイズの計算》ボタンで計算した間隔値（「14.0」）を入力します。《適用》ボタンを使用した場合、自動で入力されます。

「最初の値」を入力します。ここでは、「3」とします。

The screenshot shows a dialog box titled 'サンプル' (Sample) with the following fields and controls:

- テーブル名: 売上データ2017
- 種類: レコードサンプリング (dropdown menu)
- 選択方法: 固定間隔 (dropdown menu)
- ボタン: サンプルサイズの計算
- パラメータセクション:
 - 間隔値: 14.0 (text input)
 - 最初の値: 3 (text input)
- 保存先: (empty text input)
- ボタン: キャンセル, 実行

(6) テーブルの保存

「保存先」に抽出したサンプルレコードを書き出すテーブルのテーブル名を入力し、《実行》ボタンをクリックします。

- テーブル名は、64文字までです。
- テーブルは、拡張子「.hdf5」のファイルです。
- ファイルの保存場所は、プロジェクトフォルダ内の「04_table」フォルダです。

The screenshot shows the same dialog box as in step 5, but with the '保存先' (Save to) field populated with the text '売上データ2017_サンプルレコード'.

- テーブル名: 売上データ2017
- 種類: レコードサンプリング (dropdown menu)
- 選択方法: 固定間隔 (dropdown menu)
- ボタン: サンプルサイズの計算
- パラメータセクション:
 - 間隔値: 14.0 (text input)
 - 最初の値: 3 (text input)
- 保存先: 売上データ2017_サンプルレコード
- ボタン: キャンセル, 実行

【結果】

《サンプルサイズの計算》で、「サンプルサイズ」が24と計算されていたため、24レコードのサンプルレコードが抽出されました。

サンプリングの種類、サンプルの選択方法、パラメータの値によっては、サンプルサイズより多い、または少ないレコード件数が抽出される場合もあります。

	売上番号	計上日	商品No	数量	仕入単価	販売単価	金額	営業担当者	入力担当者	入力日時
1	11511	2017/01/01	30103006	55	734	3,000	165,000	0052	0058	2017/01/01 07:37:42
2	11525	2017/01/13	30103002	15	263	800	12,000	0052	0058	2017/01/13 11:01:38
3	11539	2017/02/09	30104006	58	439	1,450	84,100	0045	0047	2017/02/09 07:41:33
4	11554	2017/02/22	30402001	245	384	980	240,100	0024	0047	2017/02/22 14:47:46
5	11566	2017/03/10	30105002	1	161	680	680	0024	0047	2017/03/10 07:42:11
6	11580	2017/03/28	30603001	6	342	680	4,080	0052	0058	2017/03/28 13:28:35
7	11594	2017/04/07	30201004	15	769	2,400	36,000	0052	0058	2017/04/07 12:14:24
8	11608	2017/04/24	30201002	372	432	800	297,600	0045	0047	2017/04/25 16:31:41
9	11622	2017/05/09	30104002	1	1,170	4,500	4,500	0024	0058	2017/05/09 14:43:30
10	11636	2017/05/22	30201001	2	99	400	800	0052	0058	2017/05/23 07:38:54
11	11650	2017/06/09	30103005	4	533	2,400	9,600	0024	0058	2017/06/09 15:49:31
12	11664	2017/06/27	30301004	1	191	600	600	0052	0058	2017/06/27 12:20:37
13	11678	2017/07/10	30104006	4	439	1,450	5,800	0045	0047	2017/07/11 09:50:46
14	11692	2017/07/31	30301004	5	191	600	3,000	0024	0036	2017/08/01 16:33:32
15	11706	2017/08/18	30104006	871	439	1,450	1,262,950	0052	0058	2017/08/18 08:47:48
16	11720	2017/09/02	30202002	513	741	2,500	1,282,500	0021	0036	1900/01/01 00:00:00
17	11734	2017/09/14	30602001	417	363	780	325,260	0045	0047	2017/09/14 16:32:23
18	11748	2017/10/08	30603003	44	196	680	29,920	0052	0058	2017/10/08 16:31:33
19	11762	2017/10/26	30103006	21	734	3,000	63,000	0052	0058	2017/10/26 11:08:35
20	11776	2017/10/11	30103004	15	266	800	12,000	0045	0047	2017/11/10 15:55:39
21	11790	2017/11/23	30104002	2	1,170	4,500	9,000	0052	0058	2017/11/23 14:44:37
22	11804	2017/12/04	30105002	794	161	680	539,920	0010	0036	2017/12/05 11:12:15
23	11819	2017/12/14	30104002	6	1,170	4,500	27,000	0024	0036	2017/12/14 08:45:28
24	11834	2017/12/26	30604001	513	1,563	5,000	2,565,000	0024	0036	2017/12/26 11:05:49

◇操作ログ

●《サンプルサイズの計算》ボタンの操作ログ

《サンプルサイズの計算》ボタンを使用した場合は、《サンプリング》のログの前に、《サンプルサイズの計算》のログが記録されます。

SIZE 種類 CONFIDENCE “数値” POPULATION “数値” PERMISSIBLE “数値” PREDICTION “数値” TO SCREEN

※「種類」は、選択したサンプリングの種類により、ログが変わります。

種類	ログ
レコードサンプリング	RECORD
金額単位サンプリング	MONETARY

※以下の項目は、指定した数値が記録されます。(例)「信頼度」を「90」とした場合：CONFIDENCE "90"

CONFIDENCE (信頼度)

POPULATION (母集団)

PERMISSIBLE (許容誤謬率/許容誤謬額)

PREDICTION (予想誤謬率/予想誤謬額)

●《サンプリング》の操作ログ

(1)レコードサンプリング - ランダム の場合

SAMPLE ON RECORD NUMBER “数値” POPULATION "" RANDOM “数値” TO “新規テーブル名” OPEN

※「POPULATION」の数値は空欄 ("") で記述されます。

(2)レコードサンプリング - 固定間隔 の場合

SAMPLE ON RECORD INTERVAL “数値” FIXED “数値” TO “新規テーブル名” OPEN

(3)金額単位サンプリング - ランダム の場合

SAMPLE ON MUS [サンプル対象フィールド] NUMBER “数値” POPULATION “数値” CUTOFF “数値” RANDOM “数値” TO “新規テーブル名” OPEN

(4)金額単位サンプリング - 固定間隔 の場合

SAMPLE ON MUS [サンプル対象フィールド] INTERVAL “数値” FIXED “数値” CUTOFF “数値” RANDOM “数値” TO “新規テーブル名” OPEN

※以下の項目は、指定した数値が記録されます。(例)「サンプルサイズ」を「24」とした場合：NUMBER “24”

NUMBER (サンプルサイズ)

FIXED (最初の値) ※空欄にした場合は、記述されません。

POPULATION (母集団)

INTERVAL (間隔値)

CUTOFF (カットオフ値) ※空欄にした場合は、記述されません。

RANDOM (シード値) ※空欄にした場合は、記述されません。