

IBM SPSS Statistics による 統計解析

【初級編 A】 クロス集計表と平均値の検定



1

クロス集計表と独立性の検定

クロス集計表 cross table は、2つの質的変数(カテゴリ変数)の関係性を分析する統計手法です。2つの変数を、クロス集計表の行と列に配置して、それぞれのカテゴリの**度数**や**%**によってその差を比較することができます。また、**仮説検定**を実行することによって、クロス集計表に見られる差や関係性が有意な結果として認められるかを調べることができます。

Contents

- クロス集計表の概要
- 独立変数と従属変数
- 仮説検定
- Pearsonのカイ2乗検定
- Fisherの正確確率検定
- サンプルデータの確認
- クロス集計表による分析の実行— 2×2 分割表—
- クロス集計表による分析の実行— $l \times m$ 分割表—
- 残差分析

Keyword

クロス集計表 / 分割表 / 仮説検定 / Pearsonのカイ2乗検定 / Fisherの正確確率検定 / 漸近有意確率 / 正確有意確率 / 観測度数 / 期待度数 / 残差 / 調整済標準化残差 /

§1.1.1 クロス集計表の概要

クロス集計表 cross table は、2つの質的変数(カテゴリ変数)の関係性を分析する統計手法です。**分割表** contingency table とも呼ばれます。

質的変数の例としては、**性別**(男性/女性)、**喫煙の有無**(あり/なし)、**疾患の有無**(あり/なし)、**年代**(20代/30代/40代)などが挙げられます。これらの質的変数を、クロス集計表の行と列に配置して、それぞれのカテゴリの**度数**や**%**を出力して、その差を比較することができます。

		疾患		合計	
		なし	あり		
喫煙	なし	度数	86	29	115
		%	74.8%	25.2%	100.0%
	あり	度数	44	30	74
		%	59.5%	40.5%	100.0%
合計		度数	130	59	189
		%	68.8%	31.2%	100.0%

Figure1.1.1 喫煙ごとの疾患の有無をあらわすクロス集計表(分割表)の例

上記の例では、喫煙と疾患の関係性について、行(喫煙)ごとの列(疾患)に対するパーセンテージが計算されています。疾患ありに注目すると、喫煙なし群でのその比率は25.2%、喫煙あり群では40.5%であり、喫煙している方が疾患ありのパーセンテージが高いことが分かります。

なお、クロス集計表は集めたデータ(標本/サンプル)に基づいて集計される記述統計の方法です。クロス集計表にみられる差が有意差として認められるかどうかは、**仮説検定**による**有意確率 P value**を調べる必要があります。

POINT

クロス集計表は、**質的変数**(カテゴリ変数)の関係性をパーセンテージに基づいて分析する統計手法です。

POINT

クロス集計表は、標本(サンプル)のデータに基づいて集計される、**記述統計**の方法です。

POINT

カテゴリを2つだけ持つ変数に基づく2行2列で構成されるクロス集計表は、 2×2 クロス集計表または 2×2 分割表と呼ばれます。

POINT

カテゴリを l 個持つ変数と、 m 個持つ変数で構成されるクロス集計表は、 $l \times m$ クロス集計表または $l \times m$ 分割表と呼ばれます。

Copyright

§1.3.1 クロス集計表の作成 —2×2分割表—

はじめに、カテゴリを2つだけ含む質的変数(カテゴリ変数)を用いて、2×2のクロス集計表による分析を行います。

この例では、**smoke**(喫煙の有無)によって**low**(低体重出生)に差が認められるかどうかを分析します。2つの変数はそれぞれ、喫煙のあり/なし、低体重児出生のあり/なしをあらわす変数です。

操作手順

1. 分析メニュー > 記述統計 > クロス集計表を選択します。

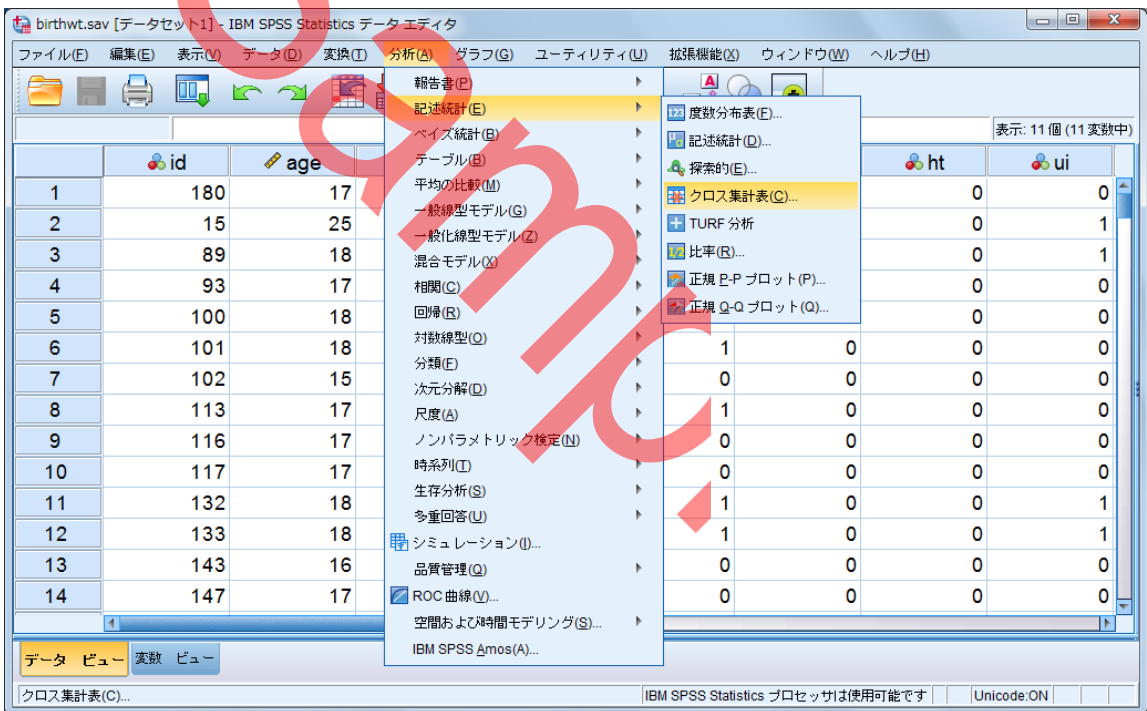


Figure1.3.1 記述統計のクロス集計表メニュー

操作手順

2. **smoke** (喫煙の有無) を**行**ボックスに移動します。
3. **low** (低体重出生) を**列**ボックスに移動します。

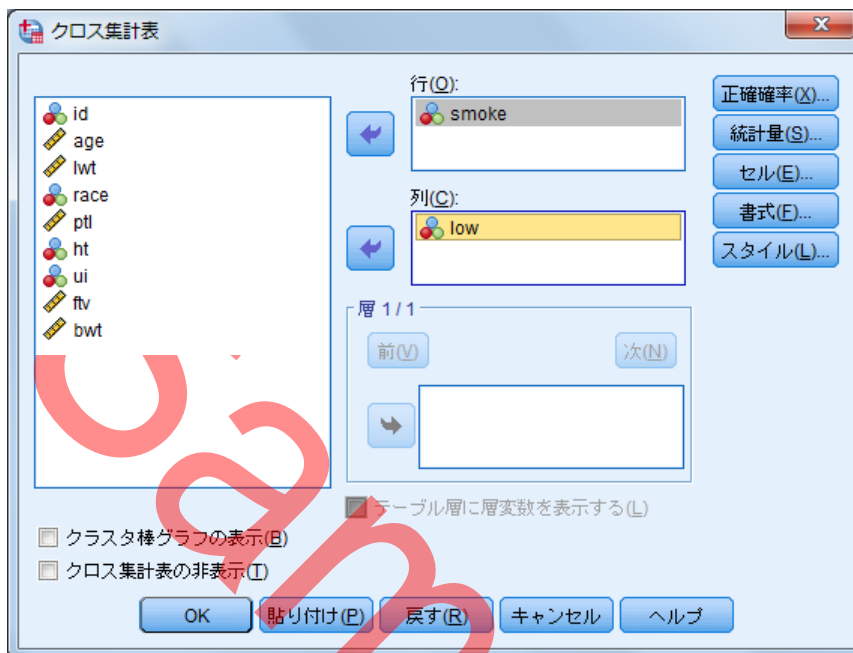


Figure1.3.2 クロス集計表の行変数と列変数の指定

クロス集計表の、行変数と列変数を指定します。これらのボックスには質的変数(カテゴリ変数)を指定します。

なお、変数間に説明関係を仮定する場合は、説明する変数(独立変数)を行に置き、説明される変数(従属変数)を列に置くという考え方があります。この例では、**smoke** (喫煙の有無)によって**low** (低体重出生)に影響があるかどうかを調べます。

次に、クロス集計表に出力するパーセンテージの設定を行います。

POINT

クロス集計表の**行**に説明変数(独立変数)、**列**に目的変数(従属変数)を配置します。これらは、結果の解釈の読みやすさの観点から逆に配置されることもあります。

操作手順

4. **セル**ボタンをクリックします。
5. パーセンテージの**行**を選択します。

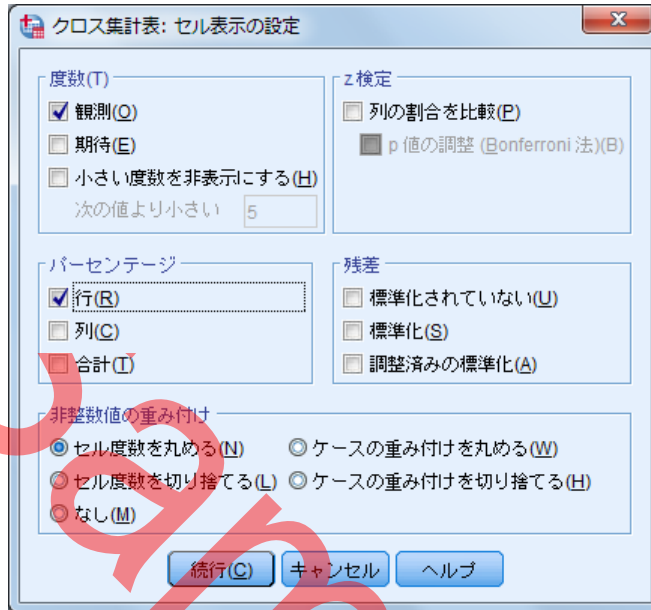


Figure1.3.4 クロス集計表のセル内容の設定

クロス集計表の結果の解釈は、度数 n ではなくパーセンテージ % によって行います。パーセンテージの出力は3種類の指定を行うことができ、それぞれパーセンテージ計算の分母を指定します。

パーセンテージ

行	行ごとの列に対するパーセンテージを出力します。 行の合計が各セルのパーセンテージ計算の分母になります。
列	列ごとの行に対するパーセンテージを出力します。 列の合計が各セルのパーセンテージ計算の分母になります。
全体	全体の合計に対するパーセンテージを出力します。 総合計が各セルのパーセンテージ計算の分母になります。

基本的な考え方として、行に説明変数(独立変数)、列に目的変数(従属変数)を配置した場合は、行パーセンテージを出力します。これによって、要因ごとの結果に対するパーセンテージを示すことができます。

ここでは、行パーセントを指定していますので、喫煙ごとの低体重出生のパーセンテージを調べることになります。

TIPS

行パーセントや列パーセントを全て出力すると、結果のクロス集計表の解釈がしにくくなるので、いずれか1つのパーセンテージを出力すると良いでしょう。

操作手順

6. **続行**ボタンをクリックします。
7. **OK**ボタンをクリックします。

§1.3.2 クロス集計表の結果の解釈 – 2×2分割表 –

出力される分析結果を解釈します。

処理したケースの要約

	ケース					
	有効数		欠損		合計	
	度数	パーセント	度数	パーセント	度数	パーセント
喫煙の有無 * 低体重出生	189	100.0%	0	0.0%	189	100.0%

Figure1.3.4 ケース処理の要約テーブル

ケース処理の要約テーブルには、分析に使用されたケース数と欠損値が要約されます。欠損値が含まれるケースは分析から自動的に除外されます。この例では、欠損値は含まれておらず、分析に使用されたケース数は有効数から**189**であることが分かります。

TIPS

2つの変数の組合せにおいて欠損値となるケースは分析から自動的に除外されます。例えば、アンケートにおける未回答ケースをクロス集計表に含めたい場合は、この値をシステム欠損値(データ未入力)とするのではなく、任意のコードを入力しておく必要があります。

喫煙の有無と低体重出生のクロス表

		低体重出生		合計	
		0 2500g以上	1 2500g未満		
喫煙の有無	0 なし	度数	86	29	115
		喫煙の有無の%	74.8%	25.2%	100.0%
	1 あり	度数	44	30	74
		喫煙の有無の%	59.5%	40.5%	100.0%
合計		度数	130	59	189
		喫煙の有無の%	68.8%	31.2%	100.0%

Figure1.3.5 smoke(喫煙の有無)とlow(低体重出生)のクロス集計表

クロス表テーブルは、分析結果の解釈の中心となるテーブルです。この例では、行に**smoke**(喫煙の有無)を配置し、列に**low**(低体重出生)を配置している2行2列のクロス集計表です。行パーセンテージを出力しているため、喫煙ごとの低体重出生のパーセンテージとして解釈することができます。

喫煙なしでは、2500g未満の低体重出生に該当するのは25.2%です。一方、喫煙ありでは、2500g未満の低体重出生に該当するのは40.5%であり、喫煙あり群のほうが約15%低体重出生のパーセンテージが高いことが分かります。つまり、母親の喫煙によって低体重出生になりやすいことが示唆されます。

次に、このような差が標本における偶然の差ではなく、母集団全体にも結論できる有意差かどうかを確認します。クロス集計表の有意差を調べる検定として、**独立性の検定**を適用します。

POINT

クロス集計表による分析は、標本データに基づく**記述統計**です。分析結果が有意かどうかを確認するために適用される独立性の検定は、標本から母集団を推測するための**推測統計**です。