

§1.4.1 クロス集計表の作成 — $l \times m$ 分割表—

3つ以上のカテゴリを含む変数を用いて、 $l \times m$ のクロス集計表による分析を行います。この例では、**race**(人種)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。人種には3つのカテゴリ、低体重出生には2つのカテゴリが含まれています。2つの変数はともにカテゴリ変数であるため、クロス集計表によって分析します。

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。

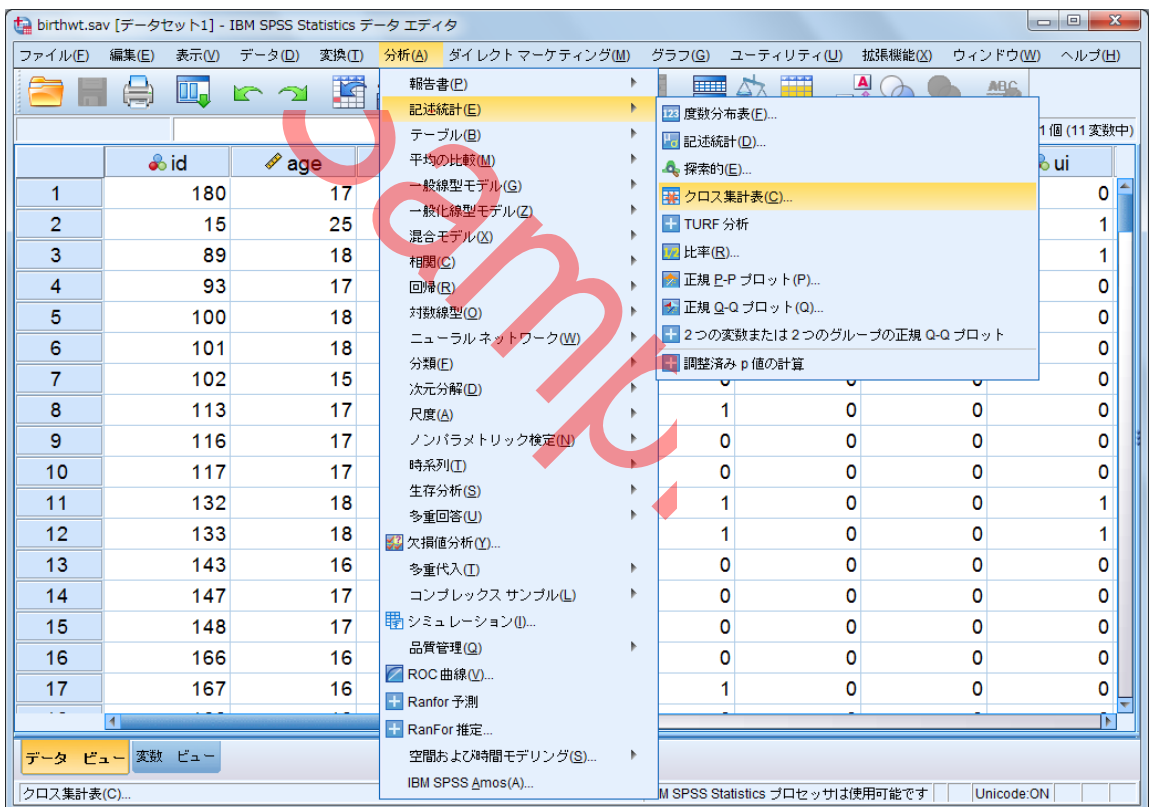


Figure1.4.1 記述統計のクロス集計表メニュー

操作手順

2. **戻す** ボタンをクリックします。
3. **race** (人種) を**行**ボックスに移動します。
4. **low** (低体重出生) を**列**ボックスに移動します。

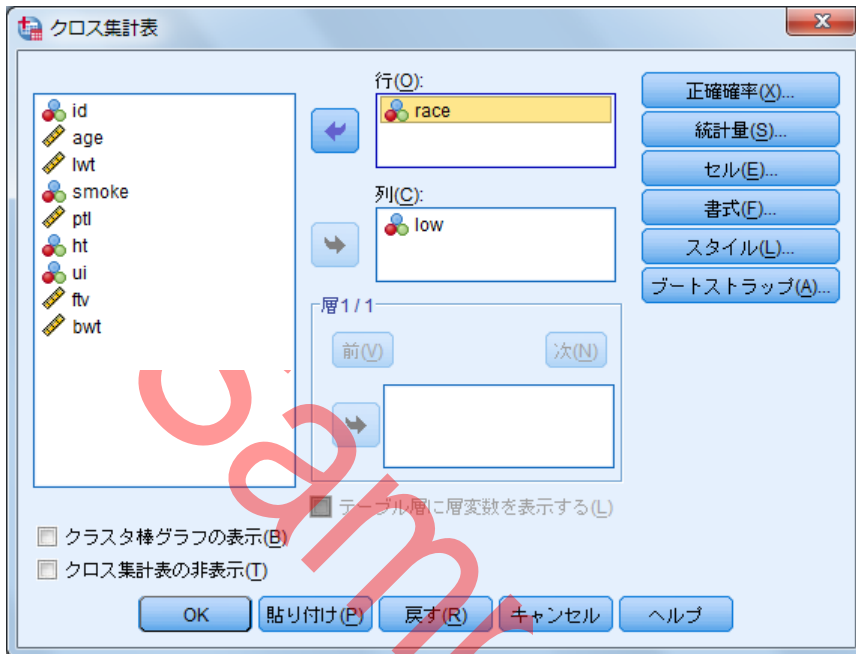


Figure1.4.2 クロス集計表の行変数と列変数の指定

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基本的にカテゴリ変数を指定します。この例では、**race** (人種) 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。

TIPS

調べたい変数が複数ある場合は、行や列のボックスに2つ以上の変数を指定することができます。その場合、行と列の変数の組合せによって、複数のクロス集計表が同時に出力されます。

操作手順

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

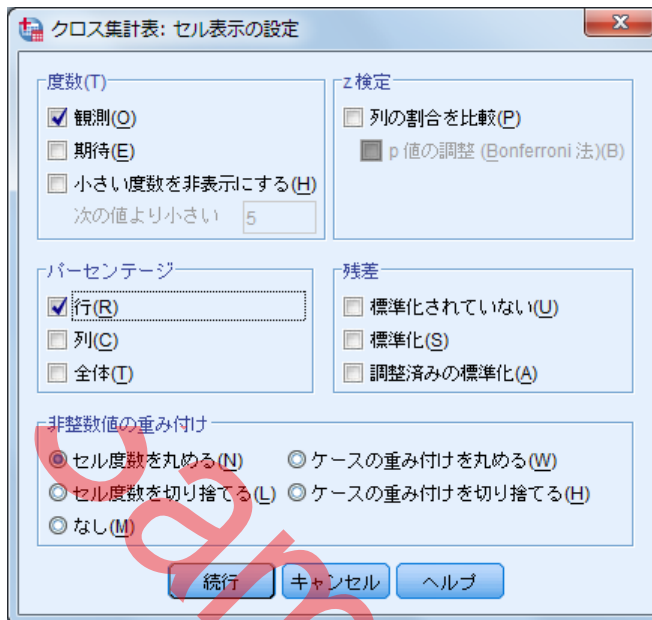


Figure1.4.4 クロス集計表のセル内容の設定

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、人種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

クロス集計表の解釈においてよく利用されるのは、行パーセントもしくは列パーセントです。体パーセントは行と列の一致を測るカッパ係数を調べる場合に用いられったりします。

操作手順

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

§1.4.2 クロス集計表の結果の解釈 — $l \times m$ 分割表 —

出力される分析結果を解釈します。

			low		
			0 2500g以上	1 2500g未満	合計
race	1 白人	度数	73	23	96
		race の %	76.0%	24.0%	100.0%
	2 黒人	度数	15	11	26
		race の %	57.7%	42.3%	100.0%
	3 その他	度数	42	25	67
		race の %	62.7%	37.3%	100.0%
合計		度数	130	59	189
		race の %	68.8%	31.2%	100.0%

Figure 1.4.6 race (人種) と low (低体重出生) のクロス集計表

この例では、行に **race** (人種) を配置し、列に **low** (低体重出生) を配置しており、3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、人種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。

白人 では、2500g未満の低体重出生に該当するのは **24.0%** です。**黒人** では、2500g未満の低体重出生に該当するのは **42.3%** であり、黒人のほうが白人より低体重出生のパーセンテージが約18%高いことが分かります。また、**その他** での低体重出生は **37.3%** であり、白人より約13%高いですが、黒人より約5%低いことが分かります。

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** を適用することができます。

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。

§1.4.3 独立性の検定の実行 — $l \times m$ 分割表—

クロス集計表の有意差検定として、独立性の検定を実行します。よく使用されるのはPearsonのカイ2乗検定とFisherの正確確率検定です。IBM SPSS Statisticsでは、 2×2 のクロス集計表においては自動的にFisherの正確確率検定を実行しますが、 $l \times m$ のクロス集計表において、Fisherの正確確率検定を実行するためには、**正確確率**ボタンを利用して検定方法を変更する必要があります。

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。

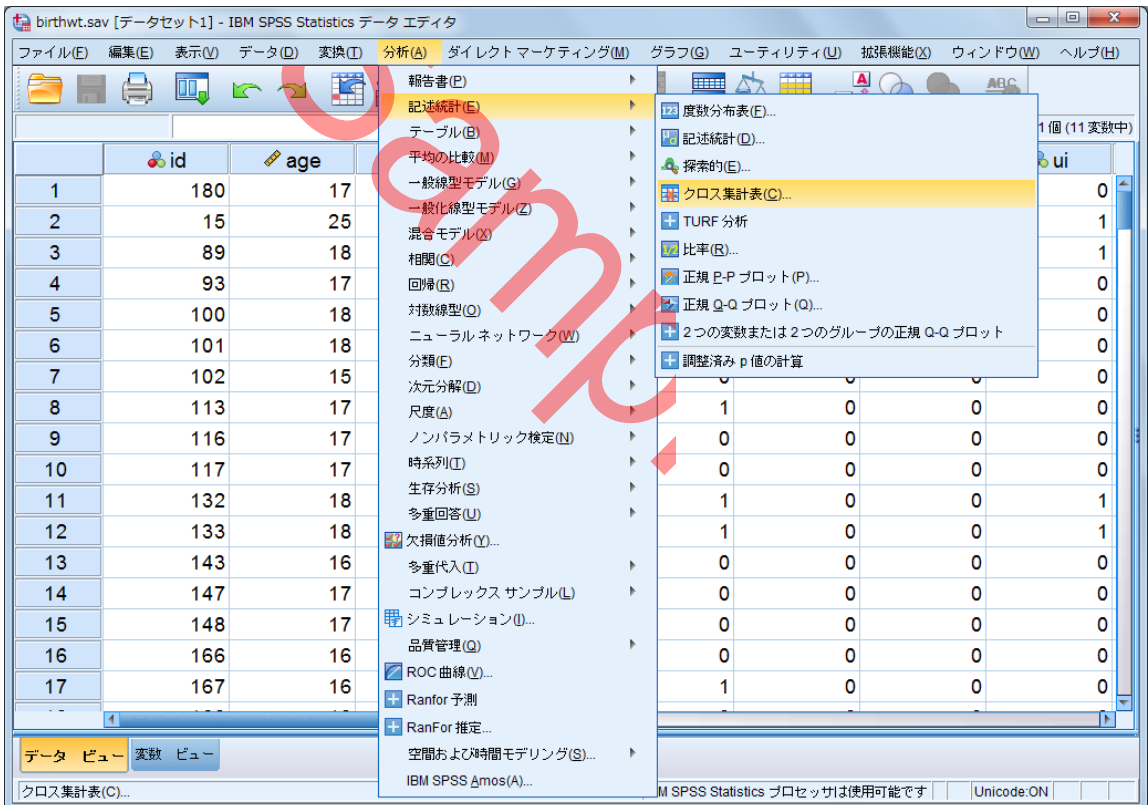


Figure1.4.7 記述統計のクロス集計表メニュー

クロス集計表ダイアログボックスが表示されます。変数の指定は完了しているため、独立性の検定を実行するための指定を行います。

操作手順

2. **統計量** ボタンをクリックします。
3. **カイ2乗** を選択します。

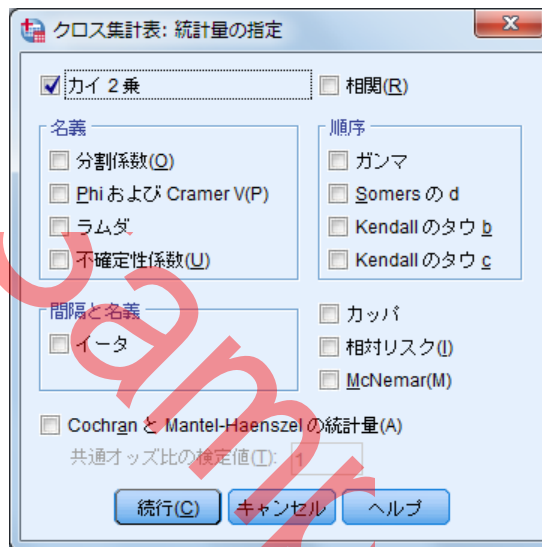


Figure1.4.8 クロス集計表の統計量の指定ダイアログボックス

独立性の検定の実行のためには、**カイ2乗** を選択します。この設定によって、Pearsonのカイ2乗検定が実行されます。

操作手順

4. **続行** ボタンをクリックします。
5. **OK** ボタンをクリックします。

§1.4.4 独立性の検定の結果の解釈 — $l \times m$ 分割表 —

独立性の検定の結果を解釈します。

カイ 2 乗検定

	値	df	漸近有意確率 (両側)
Pearson のカイ 2 乗	5.005 ^a	2	.082
尤度比	5.010	2	.082
線型と線型による連関	3.570	1	.059
有効なケースの数	189		

a. 0 セル (0.0%) は期待度数が 5 未満です。最小期待度数
は 8.12 です。

Figure 1.4.9 カイ 2 乗検定の結果

カイ 2 乗検定 テーブルに、独立性の検定の結果が出力されています。検定結果として、Pearson のカイ 2 乗の漸近有意確率 (両側) の値を確認します。

漸近有意確率 (両側) は、**0.082** ($P = 0.082 \geq 0.05$) であり、帰無仮説「行変数と列変数は独立である (関係がない)」は 5% 水準で棄却することができません。すなわち、**race** (人種) によって **low** (低体重出生) に有意な差は認められません。

POINT

仮説検定の結果が有意である場合、一定の水準で母集団にも結論できることを意味しますが、有意な結果ではない場合、その分析結果は母集団には結論できません。有意ではない場合の分析結果は、原則として標本においてのみ解釈されることを意味します。

POINT

カイ 2 乗検定の結果として、**Pearson のカイ 2 乗** の行の **漸近有意確率 (両側)** を参照します。5% 水準で検定を行う場合、この有意確率が 5% 未満 ($P < 0.05$) の場合に帰無仮説を棄却することができ、有意な関係性があると解釈することができます。

なお、重要なポイントとして、Pearsonのカイ2乗検定は、サンプルサイズ(期待度数)が十分に大きい場合に適切に使用できる検定であり、期待度数5未満のセルが含まれる場合は信頼できる検定結果ではありません。具体的には、期待度数が5未満のセルが20%を超える場合、または最小期待度数が1未満の場合は、Pearsonのカイ2乗検定の結果に基づくべきではありません。

この情報は、テーブル下部に出力されており、この例では期待度数が5未満のセルは0(0.0%)ですので、問題ありません。

a. 0 セル (0.0%) は期待度数が 5 未満です。最小期待度数は 8.12 です。

Figure1.4.10 カイ2乗検定テーブル下部のコメント

期待度数が5未満のセルが20%を超える場合、または最小期待度数が1未満の場合、Pearsonのカイ2乗検定による漸近有意確率ではなく、**Fisherの直接法**による**正確有意確率**を用いる必要があります。

TIPS

Pearsonのカイ2乗は観測度数と期待度数の差に基づいた検定ですが、**尤度比**は観測度数と期待度数の比に基づいた検定です。2つの検定結果は多くの場合一致しますが、結果が異なる場合は尤度比を優先するべきという考え方もあります。ただし、尤度比に基づく検定も漸近有意確率を計算しますので、期待度数5未満のセルが含まれる場合は、Fisherの正確確率検定を用いるべきです。

TIPS

値は χ^2 値、**df**は自由度です。自由度はクロス集計表の行数と列数から求められます。値と自由度に基づいて有意確率が算出されますので、検定結果としては基本的に有意確率を参照すれば十分です。