

■ 1 章

まず、 $n = 6, a = 4$ を代入して Σ 記号の中を、それぞれの i について計算しましょう。つまり、

$$\frac{1}{n} \sum (x - a)^2 = \frac{1}{6} \{(2 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + (7 - 4)^2\}$$

です。

これは、

$$\frac{1}{6} \{4 + 1 + 0 + 1 + 4 + 9\} = \frac{1}{6} * 19 = \frac{19}{6}$$

となります。

■2章

1. 次のデータは、それぞれの尺度水準に属するでしょうか。

(a)西暦 : 間隔尺度水準

等間隔に数値が並んでいる。ただし西暦 0 年以前も歴史は存在するので、0 は絶対的の原点とは考えず、比例尺度水準ではないため。

(b)貯金額 : 比例尺度水準

等間隔に数値が並んでいる。また、貯金額が 0 であることは、貯金が「ない」ことを意味するため。

(c)居住する都道府県 : 名義尺度水準

仮に、北海道を 1、青森県を 2、...のように数値化したとしても、その数値に順序性は存在しないため。

2. 統計ソフトウェアを用いて、任意のデータを可視化してみましょう。

第 2 回目の放送教材で HAD を用いた可視化方法を解説しているため、そちらを参考にしてください。

■3章

次のデータの平均値、中央値、分散、標準偏差を計算してみましょう。

$$x_i = (1,5,0,2,6,3,10,1,4,7)$$

(a) 平均値

$$\frac{1}{10}(1+5+0+2+6+3+10+1+4+7) = \frac{39}{10} = 3.9$$

(b) 中央値

数値が小さいほうから順に並び替えると以下のようなになる。

$$x_i = (0,1,1,2,3,4,5,6,7,10)$$

データの総数が偶数（10個）であるため、小さいほうから5番目と大きいほうから5番目の数値の平均値が中央値となる。よって中央値は、 $\frac{1}{2}(3+4) = \frac{7}{2} = 3.5$

(c) 分散（ここでは標本分散のこと）

平均が3.9であることはすでに判明しているので、分散は

$$\frac{1}{10}((1-3.9)^2 + (5-3.9)^2 + (0-3.9)^2 + (2-3.9)^2 + (6-3.9)^2 + (3-3.9)^2 + (10-3.9)^2 + (1-3.9)^2 + (4-3.9)^2 + (7-3.9)^2) = 8.89$$

(d) 標準偏差（ここでは標本分散の正の平方根のこと）

分散が8.89であることはすでに判明しているので、標準偏差は

$$\sqrt{8.89} \approx 2.98$$

■4章

1. 2変数 x, y について、以下のデータが得られました。このデータについて、共分散と相関係数を算出してみましょう。

x	y
4.2	7.0
2.9	6.2
5.0	6.6
4.9	5.8
3.0	4.4

およそ以下の値となれば正解です。

共分散：0.40

相関係数：0.50

2. 表 4.5 の (a) と (b) のデータについて、 χ^2 値とクラメールの連関係数をそれぞれ算出してみましょう。

(a) のデータ

χ^2 値：0

クラメールの連関係数：0

(b) のデータ

χ^2 値：400

クラメールの連関係数：1

■5章

1. 2変数 x, y について、第4章の演習問題1のデータが得られました。このデータについて、平均、分散、そして共分散を用いて、回帰直線の切片(a)と傾き(b)を求めてみましょう。

x を説明変数、 y を目的変数とします。およそ以下の値となれば正解です。

切片(a) : 4

傾き(b) : 0.5

<参考>

x の平均 : 4

y の平均 : 6

x の分散 : 0.8

y の分散 : 0.8

共分散 : 0.4

2. 体重と50m走のタイムの相関のように、残差に注目することが興味深いケースについて考えてみましょう。

特定の正解がある問題ではありませんので、各自で考えてもらえればこの課題はOKです。

■ 6章

以下の操作で偏相関係数を計算できます。
データシートに以下のように入力します。

	A	B	C	D	E
変数名	ID	X	Y	Z	
データ読み込み	A		5	3	1
	B		2	1	4
	C		3	2	2
	D		4	6	5
	E		7	4	3

データ読み込みボタンを押して、次のように X と Y を使用変数、Z を統制変数として指定します。

	A	B	C	D	E	F
変数名	ID	X	Y	Z		
分析			選択セルを使用		統制変数を投入	
データシート						
使用変数	ID	X	Y	\$	Z	
変数情報	フィルタ	値ラベル	ラベル	@変数	コード	
1 ID						
2 X						
3 Y						
4 Z						

「分析」ボタンを押して、下のように「相関分析」をチェックします。

統計分析マクロ HAD (簡易モード) ×

データの要約	変数間の関連性
<input type="checkbox"/> 要約統計量	<input checked="" type="checkbox"/> 相関分析
<input type="checkbox"/> 箱ひげ図	<input type="checkbox"/> 順位相関分析
<input type="checkbox"/> ヒストグラム	<input type="checkbox"/> 項目分析 (α係数)
<input type="checkbox"/> 度数分布表	<input type="checkbox"/> セルごとの度数
<input type="checkbox"/> 散布図	
<input type="checkbox"/> クロス表	
<input type="checkbox"/> 群ごとの統計量	
<input type="checkbox"/> バブルチャート	
差の検定	マルチレベル分析
<input type="checkbox"/> 平均の推定 基準値 = 0	<input type="checkbox"/> 級内相関係数
<input type="checkbox"/> 平均の差の検定	<input type="checkbox"/> マルチレベル相関分析
<input checked="" type="radio"/> 対応なし <input type="radio"/> 対応あり	<input type="checkbox"/> グループごとの回帰直線
<input type="checkbox"/> 順位の差の検定 (ノンパラ検定)	<input type="checkbox"/> ペアワイス共分散行列
<input checked="" type="radio"/> 対応なし <input type="radio"/> 対応あり	
<input type="checkbox"/> 出力を上書きしない	<input type="checkbox"/> フィルタをオフにする

詳細モード グラフ設定 OK キャンセル

結果は、0.711 となります。

■7章

1.

確率変数と確率変数の実現値の違いは、確率変数は値がランダムに変わりうる対象を表しているのに対し、実現値はすでに決まった、具体的な値を指します。たとえば、1,2,3,4,5,6,の6つの数字がランダムにでる変数 X があるとします。この X のことを確率変数といいます。一方、 X について具体的な値、たとえば3が現れたとします（サイコロを振った結果などによって）。その場合、この3が実現値です。

2. 標本分布とは、標本統計量が従う確率分布です。データの分布や、母集団分布とは異なる点に注意しましょう。

データの分布は、得た標本についてヒストグラムを作ったときに見られる、データの特徴を表すものです。よって、データの分布は確率分布ではありません。これは実現値の集まりです。続いて、母集団分布は、母集団から標本抽出するとき、どのようなデータが得られるかを表す確率分布です。

そして標本分布は、標本から計算された統計量が、どのように確率的に変化するのかを表す確率分布のことです。ポイントは、標本統計量が確率変数であると考えている点です。標本統計量は標本から計算された量ですから、標本をどのように取るのかに依存する量です。統計学では、標本抽出が確率的であると考えるため、標本統計量も確率変数となるのです。そして、その分布が、標本分布となります。

■ 8 章

まず、平均値の推定値を計算しましょう。

$$\hat{\mu} = \frac{1}{6}(2 + 6 + 7 + 3 + 4 + 5) = \frac{1}{6} * 27 = 4.5$$

つづいて、標準誤差を計算します。標準誤差は、

$$\sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{4}{6}} \cong 1.22$$

です。

ここから 95%信頼区間は、

$$\mu_{\text{lower}} = 4.5 - 1.96 * 1.22 \cong 2.11$$

$$\mu_{\text{upper}} = 4.5 + 1.96 * 1.22 \cong 6.89$$

で求められ、

$$95\% \text{ CI}[2.11, 6.89]$$

となります。

■9章

ソフトウェアでの計算方法を解説します。

まずデータを読み込み、X を使用変数に指定します。

	A	B	C	D	E	F
1	変数名	ID	X			
2						
3	分析		選択セルを使用		統制変数を投入	
4						
5	データシート					
6						
7						
8	使用変数					
9		ID	X			
0						
00	変数情報	フィルタ	値ラベル	ラベル	@変数	コード
01	ID					
02						

続いて、下のように「平均値の推定」にチェックを入れます。

統計分析マクロ HAD (簡易モード) X

データの要約	変数間の関連性
<input type="checkbox"/> 要約統計量	<input type="checkbox"/> 箱ひげ図
<input type="checkbox"/> ヒストグラム	<input type="checkbox"/> 度数分布表
<input type="checkbox"/> 散布図	<input type="checkbox"/> クロス表
<input type="checkbox"/> 群ごとの統計量	<input type="checkbox"/> バブルチャート
差の検定	変数間の関連性
<input checked="" type="checkbox"/> 平均の推定 基準値 = 0	<input type="checkbox"/> 相関分析
<input type="checkbox"/> 平均の差の検定	<input type="checkbox"/> 順位相関分析
<input type="checkbox"/> 対応なし <input type="radio"/> 対応あり	<input type="checkbox"/> 項目分析 (α係数)
<input type="checkbox"/> 順位の差の検定 (ノンパラ検定)	<input type="checkbox"/> セルごとの度数
<input type="checkbox"/> 対応なし <input type="radio"/> 対応あり	マルチレベル分析
<input type="checkbox"/> 出力を上書きしない <input type="checkbox"/> フィルタをオフにする	<input type="checkbox"/> 級内相関係数
<input type="checkbox"/> 詳細モード	<input type="checkbox"/> マルチレベル相関分析
<input type="checkbox"/> グラフ設定	<input type="checkbox"/> グループごとの回帰直線
<input type="checkbox"/> OK	<input type="checkbox"/> ペアワイス共分散行列
<input type="checkbox"/> キャンセル	

OK ボタンを押すと結果が得られます。

信頼区間	
95%下限	95%上限
2.537	6.463

■10 章

1. サンプルデータを用いて、標本の相関係数から母集団の相関係数を推定し、その結果をまとめてみましょう。

サンプルデータとして、「sample_data_10_1.xlsx」という名前のエクセルファイルのデータを使用します。エクセルファイルには、「ID」「体重」「50m 走タイム」という変数名のデータがあります。これらをすべてコピーし、HAD の「データ」シートに貼りつけ、「データ読み込み」をクリックします。

「モデリング」シートに移り、「ID」「体重」「50m 走タイム」を使用変数とし、「分析」ボタンをクリックします。すると、ボックスがでてきます。

「相関分析」にチェックを入れて、「OK」をクリックすると、結果が出力されます。

「相関分析」のセクションに、相関係数が出力されます。また、「検定統計量(t 値)と有意確率」のセクションに、「95%CI」として、95%信頼区間が出力されているので、この値を報告します。

なお、サンプルデータでの相関係数の推定値は.416、95%信頼区間の下限値は.155、上限値は.622 でした。

2. サンプルデータを用いて、標本の連関係数から母集団の連関係数を推定し、その結果をまとめてみましょう。

サンプルデータとして、「sample_data_10_2.xlsx」という名前のエクセルファイルのデータを使用します。エクセルファイルには、「ID」「出身地」「好きな麺類」という変数名のデータがあります。これらをすべてコピーし、HAD の「データ」シートに貼りつけ、「データ読み込み」をクリックします。

「モデリング」シートに移り、「ID」「出身地」「好きな麺類」を使用変数とし、「分析」ボタンをクリックします。すると、ボックスがでてきます。

「クロス表」にチェックを入れて、「OK」をクリックすると、結果が出力されます。

「連関係数と独立性の検定」のセクションに、クラメール V の推定値および 95%下限と上限が出力されているので、この値を報告します。

なお、サンプルデータでのクラメール V の推定値は.316、95%信頼区間の下限値は.156、上限値は.522 でした。

■11 章

1. サンプルデータを用いて、対応のない 2 条件の場合において、標本における平均値差から母集団における平均値差を推定し、その結果をまとめてみましょう。

サンプルデータとして、「sample_data_11_1.xlsx」という名前のエクセルファイルのデータを使用します。エクセルファイルには、「ID」「値」「条件」という変数名のデータがあります。これらをすべてコピーし、HAD の「データ」シートに貼りつけ、「データ読み込み」をクリックします。

「モデリング」シートに移り、「値」「条件」を使用変数とし、「分析」ボタンをクリックします。すると、ボックスがでてきます。

「平均の差の検定」にチェックを入れて、「対応なし」を選択します。「OK」をクリックすると、結果が出力されます。

「対応のない平均値の差の検定（対応のない t 検定）」のセクションに、2 条件の差の値および 95% 下限と上限が出力されているので、この値を報告します。

なお、サンプルデータでの条件差の推定値は 12.06、95% 信頼区間の下限値は 9.67、上限値は 14.45 でした。

また、「効果量」のセクションには、「効果量 d」として、対応のない場合の標準化平均値差の推定値および 95% 下限と上限が出力されているので、この値を報告します。

なお、サンプルデータでの標準化平均値差の推定値は 1.99、95% 信頼区間の下限値は 1.51、上限値は 2.47 でした。

2. サンプルデータを用いて、対応のある 2 条件の場合において、標本における条件差の平均値から母集団における条件差の平均を推定し、その結果をまとめてみましょう。

サンプルデータとして、「sample_data_11_2.xlsx」という名前のエクセルファイルのデータを使用します。エクセルファイルには、「ID」「有条件」「無条件」という変数名のデータがあります。これらをすべてコピーし、HAD の「データ」シートに貼りつけ、「データ読み込み」をクリックします。

「モデリング」シートに移り、「有条件」「無条件」を使用変数とし、「分析」ボタンをクリックします。すると、ボックスがでてきます。「平均の差の検定」にチェックを入れて、「対応あり」を選択します。「OK」をクリックすると、結果が出力されます。

「対応のある平均値の差の検定（対応のある t 検定）」のセクションに、2 条件の差の値および 95% 下限と上限が出力されているので、この値を報告します。

なお、サンプルデータでの条件差の推定値は 12.06、95% 信頼区間の下限値は 10.46、上限値は 13.66 でした。

また、「効果量」のセクションには、「効果量 d」として、対応のある場合の標準化平均値差の推定値および 95% 下限と上限が出力されているので、この値を報告します。なお、サンプルデータでの標準化平均値差の推定値は 2.01、95% 信頼区間の下限値は 1.52、上限値は 2.50 でした。

■12章

ソフトウェアでの計算方法を解説します。

以下のようにデータを入力します。

	A	B	C	D
1	変数名	ID	X	Y
2	データ 読み込み	A		2 3
3		B		6 5
4		C		7 8
5	モデリング シート	D		3 7
6		E		4 6
7		F		5 4
8				

X と Y を使用変数に指定します。

変数名	ID	X	Y		
分析				選択セルを使用	統制変数を投入
データ シート					
使用変数	ID	X	Y		
変数情報	フィルタ	値ラベル	ラベル	@変数	コード

以下のように、「平均値の差の検定」で「対応あり」を選択します。

統計分析マクロ HAD (簡易モード) ×

データの要約

要約統計量 箱ひげ図

ヒストグラム 度数分布表

散布図 クロス表

群ごとの統計量 バブルチャート

変数間の関連性

相関分析

順位相関分析

項目分析 (α係数)

セルごとの度数

差の検定

平均の推定 基準値 = 0

平均の差の検定

対応なし 対応あり

順位の差の検定 (ノンパラ検定)

対応なし 対応あり

マルチレベル分析

級内相関係数

マルチレベル相関分析

グループごとの回帰直線

ペアワイス共分散行列

出力を上書きしない フィルタをオフにする

詳細モード
グラフ設定
OK
キャンセル

結果が出力されます。

差の検定							
水準の組	差	標準誤差	95%下限	95%上限	t値	df	p値
X - Y	-1.000	0.775	-2.991	0.991	-1.291	5	.253

■13 章

統計ソフトウェアを用いて、表 13-2 のデータに対して分散分析を実行してみましょう。

第 13 回目の放送教材で HAD を用いた分析方法を解説しているため、そちらを参考にしてください。

■14 章

統計ソフトウェアを用いて、表 14-1 のデータに対して分散分析を実行してみましょう。

第 14 回目の放送教材で HAD を用いた分析方法を解説しているため、そちらを参考にしてください。

■15 章

母平均の推定の仮定は、

- 母集団分布が正規分布である
- 無作為抽出で標本が得られている
- 対応のない平均値の差の推定の場合は、両群の分散が等しい
- 標本サイズ n の標本を得る（事前に決めておく）
- 検定の場合は、帰無仮説を事前に設定しておく

母相関係数の推定の仮定は、

- 母集団分布が2つの変数ともに正規分布である
- 無作為抽出で標本が得られている
- 標本サイズ n の標本を得る（事前に決めておく）
- 検定の場合は、帰無仮説を事前に設定しておく