

2022.11.26

第 45 回ソーシャルワーク研究会

堀米史一

## Excel を使った基本統計ワークショップ

### 1. 尺度

統計データは 4 種類の尺度（測定値）に分けられ、尺度によっては  $1+2=3$  ではないものもある！！

- (1) 名義尺度
- (2) 順序尺度
- (3) 間隔尺度
- (4) 比率尺度

#### (1) 名義尺度

質的な相違に形式的に数値を割り振ったもの。単なるレッテルや記号として数字を用いている。→演算はできない。

例) 性別（男性=1、女性=2（男性=2、女性=1 のように逆でも可。)), 学籍番号、背番号

#### (2) 順序尺度

順序に従って数値を割り振ったもの。数値の大小には意味があるが、演算には意味がない。→等差性が保証されない。

例) 成績の順位→1 位、2 位、3 位。

意識調査（1 とてもそう思う、2 そう思う、3 そう思わない、4 全くそう思わない）

→リッカート尺度（リッカート形式）や SD（Semantic Differential）法（ここは長年議論があって学問領域によって間隔尺度でもよいとされる場合もある）

#### (3) 間隔尺度

数値の差が等しく並べられているデータであり、加減演算（足し算・引き算）が可能である。→順序の概念の他に「数値の間隔」の概念が加わる。

例) テストの点数、温度、暦の日付

#### (4) 比率尺度

数値の値にも意味があるデータで数値の原点 0 (絶対零点) が一義的に決まっている (0=無)。→四則演算が使える。

例) 長さ (身長等)、重さ (体重等)、年齢

#### エクセルを使ったワーク①

ワークシート「性別と出身地」

①=COUNTIF (範囲, 検索条件)

②ピボットテーブル

挿入 → ピボットテーブル → 範囲の選択 → ピボットテーブルのフィールドで  を入れる → テーブルを値にドラッグする

## 2.記述統計量

記述統計量・・・統計データの特徴を計算数値で表すもの

- (1) 最大値＝一連の数値の中でもっとも大きな値
- (2) 最小値＝一連の数値の中でもっとも小さな値
- (3) 標本平均 (平均値) = データの数値の和をデータ数  $n$  で除したもの
- (4) 中央値＝データを値が小さい順に並べた場合に、中央に位置するデータの値
- (5) 標本標準偏差＝各データが標本平均からどの程度乖離しているかを示す値 (平均値との差の平均値)

#### エクセルを使ったワーク②

ワークシート「テストの点数 (1)」、「テストの点数 (2)」

「平均 60 点のテストで 70 点をとった D さん」

- (1) (2) 最大値・最小値

まずは Excel の設定の確認から

データをクリック → データ分析が出てくるか確認。

※出てこない場合

ファイルをクリック オプションから アドインをクリック、設定をクリック、分析ツールに

を入れる。

データ分析が出てきたらクリック、次に基本統計量をクリック、入力範囲を選択して、出力

オプションの K 番目に大きな値 (A) と K 番目に小さな値 (M) に  を入れる。

(3) 標本平均 (平均値)

=AVERAGE (範囲)

(4) 中央値

=MEDIAN(範囲)

(5) 標本標準偏差

=STDEV.P (範囲)

### 3.2 変量の分析

(1) クロス集計

複数の変数について2つの条件 (A であって a である、A であって b である) をみたまものを集計する

	列側→	結果的なこと		
行側 ↓		A	B	
理由になること	a	1行1列目	1行2列目	1行目行計
b		2行1列目	2行2列目	2行目行計
		1列目の列計	2列目の列計	総計(合計)

例) 出身地と性別、転倒の有無と性別 等々

#### エクセルを使ったワーク③

ワークシート「性別と出身地」

挿入 → ピボットテーブル → 範囲の選択 → ピボットテーブルのフィールドで  を入れる → テ

ーブルを値にドラッグする

例) インフルエンザの予防接種と感染

	列側→	結果的なこと		
行側↓		感染した	感染しなかった	
理由になること	予防接種を受けた	a	b	a+b
	予防接種を受けてない	c	d	c+d
		a+c	b+d	a+b+c+d

(2)  $\chi^2$ 検定

データから得られた実測値と理論的に期待される期待値の相違を表す統計量。期待値（理論値）と実測値（観測値）のズレを $\chi^2$ 検定で測る。

=CHISQ.TEST（実測値，期待値）

帰無仮説は「A と B では差がない（同じ）」と立て、帰無仮説を有意水準によって検定し、仮説が棄却された場合には対立仮説を採用する。対立仮説は「A と B では差があるのではないか」と立てる。一般的に有意確率（P）の値が 0.05 より大きい場合は帰無仮説を採用し、0.05 未満の場合は対立仮説を採用する。

0.05 未満→帰無仮説を棄却して対立仮説を採用する→有意な差がある

0.05 以上→帰無仮説をできない→有意な差はない

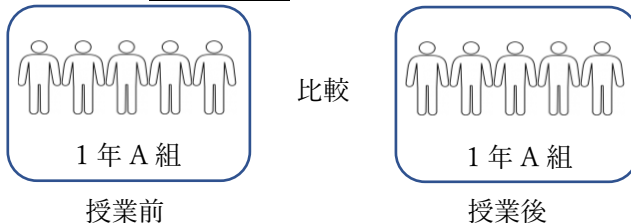
エクセルを使ったワーク④

ワークシート「インフルエンザと予防接種」「性別と運動」

(3) t 検定

t 検定は 2 つのデータの「平均の相違」を検定する際に用いる。大きく分けて 3 種類。

① t 検定：一対の標本による平均の検定



## エクセルを使ったワーク⑤

ワークシート「体重の変化」

データをクリック→データ分析→t検定：一対の標本による平均の検定

→変数1の入力範囲と変数2の入力範囲を指定→（必要に応じて）出力先（O）を指定→

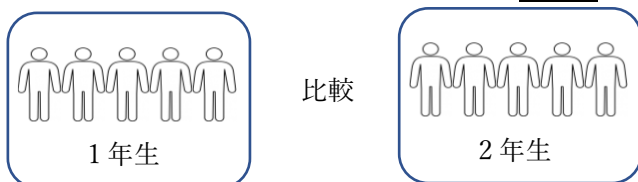
OKをクリック

※結果を見るポイント（とりあえず1つ、2つの変数の平均値に差があるかを見る）

P ( $T \leq t$ ) 両側 0.05 未満→有意な差がある、0.05 以上→有意な差はない

② t 検定：等分散を仮定した2標本による検定

③ t 検定：分散が等しくないと仮定した2標本による検定



（F 検定で 0.05 よりも大きかったら「2つのデータの分散は同程度」）

=F.TEST（範囲1，範囲2）

## エクセルを使ったワーク⑥

ワークシート「1年生と2年生のテスト（1）」「1年生と2年生のテスト（2）」

データをクリック→データ分析

→（F 検定） $>0.05$  → t 検定：等分散を仮定した2標本による検定

→（F 検定） $<0.05$  → t 検定：分散が等しくないと仮定した2標本による検定

→変数1の入力範囲と変数2の入力範囲を指定→（必要に応じて）出力先（O）を指定→

OKをクリック

#### (4) 相関

=PEARSON(範囲 1,範囲 2)

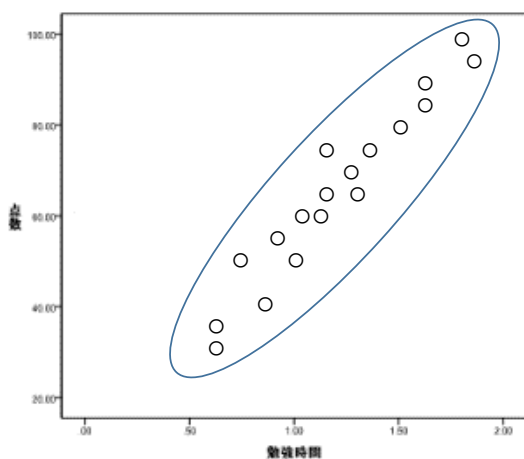
相関とは 2 変量の直線的な関係の強さを見る分析で、一般的にピアソンの積率相関で相関係数 ( r ) を求める。使える尺度は間隔尺度と比率尺度。

(ケンドール、スピアマンの順位相関は順序尺度で相関係数 ( r ) を求める方法もある)

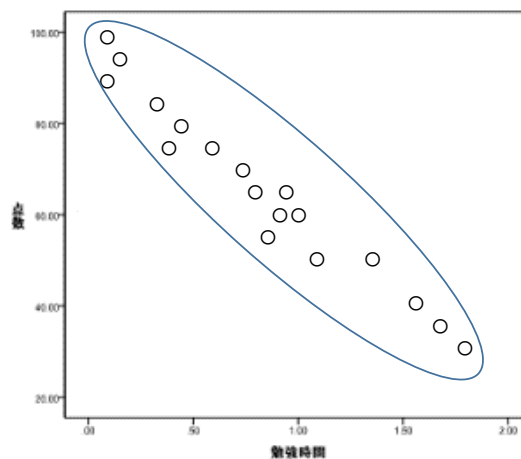
相関係数 ( r ) : 値の範囲 (-1~+1)

0.00~±0.20	ほとんど相関がない (0.00 は無相関)
±0.20~±0.40	低い (弱い) 相関がある
±0.40~±0.70	(中程度の) 相関がある
±0.70~±1.00	高い (強い) 相関がある (+1.00 は完全な正の相関、-1.00 は完全な負の相関)

#### 散布図



正の相関



負の相関

#### エクセルを使ったワーク⑥

ワークシート「勉強時間とテストの点数」

関数の挿入 → PEARSON → OK → 配列 1 と 配列 2 を指定 → OK

または

データ をクリック → データ分析 → 相関 → 入力範囲 (I) を指定 → (必要に応じて) 出力先 (O) を指定 → OK をクリック

## 散布図の作り方

データを選択し、**挿入**→**散布図**→**グラフ**をクリック→右上の+をクリック→**近似曲線**に  
を入れる

## 近似曲線

近似曲線は、グラフにおいてデータの推移を大まかに表した線

## 相関関係と因果関係

相関係数の値が大きくても、因果関係があるとは必ずしもいえず、変数 A と変数 B に相関がある時の可能性は、以下のようなこと考えられる。

- ①変数 A が変数 B の直接的な原因 ( $A \rightarrow B$ )
- ②変数 B が変数 A の直接的な原因 ( $A \leftarrow B$ )
- ③変数 A と変数 B が相互に影響し合う ( $A \rightarrow B, A \leftarrow B$ )
- ④間に変数 X が入ることによる間接的な原因  
(  $A \rightarrow X \rightarrow B, B \rightarrow X \rightarrow A$  )
- ⑤偶然 等々