

「統計解析入門①：2変数の関係」



医学研究支援センター 医療統計室 室長・准教授 伊藤 ゆり

第4～6回は研究デザインについて紹介してきました。今回からは、統計解析についてご紹介していきます。ここまでで紹介してきた各研究デザインでは、曝露や介入(説明変数)とアウトカム(目的変数)の関係を見るのが主な目的になっていました。今回は二つの変数間の比較に関する統計解析について紹介します。

説明変数、目的変数それぞれの変数のデータの種類によって、分析の方法が変わります。説明変数、目的変数ともに連続値(量的変数)の場合、X軸に説明変数、Y軸に目的変数とした散布図を描きます。これらの関係性の強さを見るために相関係数を算出します。相関係数は-1～1の値を取り、1に近い場合、正の相関が高い、-1に近い場合、負の相関が高いといえます。どの程度の値だと、相関が高いといえるのかという質問をよく受けます。相関係数に関する検定は帰無仮説「相関係数が0である」と対立仮説「相関係数が0でない」に対して検定を行っています。そのため、検定結果が $p < 0.05$ であっても、推定された相関係数が0.1などのように、ほとんど0に近い場合もあります(計算に用いたデータ数が多いとそのようになります)。そのため、相関分析をした際には、相関係数の値自体をきちんと評価するようにしてください。 $p < 0.05$ だけで議論することの内容にしてください。相関分析は連続値の二変数の関係をサンプルに表現する方法です。原因→結果というような方向性のある関係性をより分かりやすく表現するのが回帰分析です。説明変数(x)と目的変数(y)の関係を式で表現します。

$$y = a + bx$$

目的変数 ← y x → 説明変数

このときのaは切片項で、bは直線の傾きとなります。このbは説明変数が1単位増えた場合に目的変数が増加する程度を示します。この回帰係数bから相関係数からも導くこともできます。

なお、説明変数が0または1のようなカテゴリカル(質的)の場合でもこの回帰分析を行うことができます。たとえば、降圧剤の投与あり=1、なし=0を説明変数(x)とした場合の血圧の変化を目的変数(y)とした場合の回帰式となります。その場合、aは降圧剤なしの群の血圧変化の平均値、a+bは降圧剤ありの群の血圧の平均値となります。回帰係数bは降圧剤ありの場合の血圧変化(つまり効果)の大きさとなります。

説明変数が質的変数、目的変数が量的変数の場合、上記の他に目的変数が正規分布を仮定する場合には、平均値の差の検定(ステューデントのt検定)があります。正規分布(やその他統計分布)を仮定できない場合、順位を用いた分析となり、代表地を中央値として報告し、二群の順位を比較するWilcoxonの順位和検定(Mann-WhitneyのU検定ともいいます)を行います。

説明変数も、目的変数も質的変数の場合にはクロス集計表を使って、各群の目的変数の割合の比較を行います。その際に用いる検定方法はカイ二乗検定があります。各セルの期待値が5を下回るようなセルが存在する場合、Fisherの正確検定を実行することが勧められています。

上記の検定は比較する群がそれぞれ独立している場合、つまり対応がない場合の方法になります。同じ集団の介入前後を比較するような場合には、各群独立とは言えないので、対応のあるデータとして扱います。それぞれ検定手法がありますので、表を参照の上、選択するようにしてください。

略歴

大阪大学大学院医学系研究科博士前期・後期課程卒業後、大阪府立成人病センター(現大阪国際がんセンター)リサーチ・レジデント、研究員、主任研究員を経て、大阪医科大学 医学研究支援センター 医療統計室 准教授(現職)。現在、がん疫学、健康格差、医療統計の研究に主に従事。

表. データのタイプに応じた検定手法

データ間	変数分類	正規性	検定方法
対応なし	量的	あり	スチューデントのt検定
		なし	Wilcoxonの順位和検定 (Mann-WhitneyのU検定)
	質的	—	χ^2 (カイ二乗) 検定
		—	Fisherの正確検定
対応あり	量的	あり	対応のあるt検定 (Paired t-test)
		なし	Wilcoxonの符号付順位検定
	質的	—	マクネマー検定