

相関と回帰の入口

1 導入

この講義で最重要なのは、2 変量の統計では、平均や分散を別々に見るだけでなく、「2 つが一緒にどう動くか」を見る必要があるということです。

2 用語と定義

散布図は、2 つの量の組を平面に打った図です。

相関係数は、2 変量の直線的な結びつきの強さを表す量です。

回帰直線は、散布図の傾向を 1 本の直線で近似するものです。

3 方針

まず散布図で全体の形を見て、そのあと相関係数で強さを、回帰直線で傾向を表します。

4 直感的な説明

身長と体重のように、一方が大きいと他方も大きくなりやすい組があります。この「一緒に増えるか」「逆に動くか」「関係が薄いか」を見るのが相関です。

5 厳密な説明

相関係数 r は -1 から 1 の間の値を取ります。 r が 1 に近ければ強い正の相関、 -1 に近ければ強い負の

相関、 0 に近ければ直線的な関係は弱いと見ます。

回帰直線は、各点から直線までのずれをできるだけ小さくするように引かれます。これによって、一方の

値から他方の傾向を読み取りやすくなります。

6 具体例

勉強時間と得点の散布図が右上がりなら、「長く勉強した人ほど高得点になりやすい」という傾向が見えます。

ただし、これだけで因果関係が証明されるわけではありません。

7 見分け方

- 2 種類の量が同時に与えられているなら、まず散布図を考える
- 一緒に増えるかどうかを見たいなら相関

- 値^{あたい}から傾向^{けいこう}を予測^{よそく}したいなら回帰^{かいき}

8 どこまで成り立つか^{な た}

相関^{そうかん}があることと因果関係^{いんがんけい}があることは別^{べつ}です。また、相関係数^{そうかんけいすう}が捉えるのは主^{おも}に直線的^{ちよくせんてき}な関係^{かんけい}です。

9 関連リンク^{かんれん}

→ [講義](#) [統計の基本](#) [lecture](#) [math](#) [statistics](#)
<https://study.bem130.com/lecture/math/statistics/統計の基本-講義/>

→ [講義](#) [確率と期待値](#) [lecture](#) [math](#) [probability](#)
<https://study.bem130.com/lecture/math/probability/確率と期待値-講義/>