

# 確率・統計 後期 第12回

回帰直線

稻積 泰宏 (いなづみ やすひろ)

# 今日の内容

- 回帰分析とは
- 回帰直線の求め方
- 最小二乗法
- 回帰直線の利用（予測）
- 注意点

# 前回の復習

相関係数で2変数の関係の強さを測った

- 散布図で正の相関・負の相関を確認
- 相関係数 $r$ で関係の強さを数値化

今回やること

関係がわかったら、その関係を式で表せないか？

→ 回帰直線

# 回帰分析とは

回帰分析：2変数の関係を直線（式）で表す方法

目的：一方の変数から他方の変数を予測

## 例

- 勉強時間から成績を予測
- 気温からアイスの売上を予測
- 広告費から売上を予測

# 回帰直線

回帰直線：散布図のデータの傾向を表す直線

$$\hat{y} = ax + b$$

- $x$  : 説明変数（独立変数）
- $\hat{y}$  : 目的変数の予測値（従属変数）
- $a$  : 傾き（回帰係数）
- $b$  : 切片

# 回帰直線の求め方（最小二乗法）

**最小二乗法**：実際の値と予測値の差を最小にする方法

- 残差： $e_i = y_i - \hat{y}_i$ （実測値と予測値の差）
- 残差の二乗和を最小化

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ を最小に}$$

すべてのデータ点からの距離（残差）が最も小さくなる直線を求める

# 回帰係数の公式

$$a = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$b = \bar{y} - a\bar{x}$$

## 重要な性質

回帰直線は必ず点 $(\bar{x}, \bar{y})$ を通る

## 例題2)

# 注意点：回帰直線の限界

## 回帰直線を使う前に確認すべきこと

1. 線形関係のみを表現：曲線的な関係は表せない
2. 外れ値の影響：極端な値があると直線が歪む
3. 外挿の危険性：データの範囲外での予測は慎重に

散布図で関係を視覚的に確認してから使用すること

問2)

問3)

# まとめ

## 回帰直線

- 2変数の関係を式で表現 :  $\hat{y} = ax + b$
- 最小二乗法 : 残差の二乗和を最小化
- 回帰直線は点 $(\bar{x}, \bar{y})$ を通る

## 回帰直線の利用

- 一方の変数から他方の変数を予測できる

## 重要な注意点

- 散布図で関係を確認してから使用

次回までに問題集Basic92-94を解いておいてください

感想を会議のチャット欄へ