

## 13 単回帰（最小二乗法）＋回帰のまとめ

---

# 1. 回帰直線は「どう決める？」（最小二乗法）

## 今日の結論（先に言う）

回帰直線は、ズレ（残差）の二乗の合計が  
いちばん小さくなるように決められています。

- ・ 前回：散布図から回帰直線を引き、式と  $R^2$  を見て予測した
- ・ 今日：その「直線の決め方」を理解する（アルゴリズム）
- ・ ゴール：Excel の結果をブラックボックスにせず、なぜその線かを説明できる

## 到達目標

- ・ 残差 ( $y - \hat{y}$ ) を「点と線のズレ」として説明できる
- ・ 残差平方和（SSE）を作り、小さいほど良いと説明できる
- ・ 「最小二乗法＝SSE 最小の線」と言葉で言える

## 2. ストーリー：次の目標は「ポテト売上 UP」


店長は、ポテトのばらつきを改善したあと、次の目標を立てました。

店長の目標：

- ・ ポテトは利益率が高い
- ・ だから「**ポテトの売上を伸ばす**」ことを目標にする

そのために知りたいこと：

- ・ ハンバーガーの注文が増えると、ポテトの注文も増えるのか？
- ・ 関係があるなら、**来週の注文数を予測**できるか？

 今日の視点：関係を式（直線）にして予測に使う。  
その直線をどうやって決めるかが今日のテーマ。

### 3. 前回回収：検定は「判断」／回帰は「予測」（混同しない）

前回（検定）と今回（回帰）は目的が違います。

前回：仮説検定（判断）

- ・ 「平均が基準と違うと言ってよいか？」を珍しさ（ $p$  値）で判断する
- ・ 結論は「棄却／棄却できない」で書く

今回：回帰（予測）

- ・ 「 $x$ が増えると  $y$ はどう変わるか？」を式（回帰直線）で表す
- ・ 式を使って**未来の  $y$** を予測する

🔗 今日は  $p$  値は使わない。代わりに「ズレが小さい線」を考える。

## 4. 問い：同じ散布図に「線」は何本でも引ける

散布図を見ると「だいたい右上がり」に見えてきます。  
でも、直線は1本に決まっています。

### 問い

この散布図に引く直線として、  
どの線が「良い線」でしょうか？

- ・ 傾きが少し違う線／上下にずれた線……どれも「それっぽく」見える
- ・ だから、見た目ではなく評価基準（ルール）が必要

## 5. 評価の方針：良い線＝点に近い（ズレが小さい）を数で決める

良い線を決めるために、次の方針を使います。

📌 良い線とは、点と線のズレが全体として小さい線。

ズレを「数」で決めるために、次を行います。

- ・ 各点について「ズレ」を測る
- ・ 全部の点のズレをまとめて1つの数にする
- ・ その数がいちばん小さい線を選ぶ

次に出てくる言葉：


- ・ 点と線のズレを **残差（ざんさ）** と呼ぶ
- ・ 次スライドで、残差を**図と言葉と式**で定義する

## 6. 残差の定義：残差＝実測 $y$ - 予測 $\hat{y}$ （縦のズレ）

回帰直線（予測の線）を引くと、各点には「ズレ」が生まれます。  
このズレを**残差（ざんさ）**と呼びます。

$$\text{残差 } e = y - \hat{y}$$

- ・  $y$ ：実測値（実際に観測された値）
- ・  $\hat{y}$ ：予測値（直線が予測する値）
- ・ 残差  $e$  は、**点と線の「縦方向の距離」**（上下のズレ）

 回帰直線は「全部を完全に当てる線」ではない。  
だから、残差（ズレ）は必ず出る。

## 7. 残差の符号：線より上は+／下は-（そのまま足すと相殺する）

残差には符号（プラス／マイナス）があります。

- ・ 点が直線より上にある： $y > \hat{y}$  なので、残差は+
- ・ 点が直線より下にある： $y < \hat{y}$  なので、残差は-

ここで大事な問題が起きます。

✂ 残差をそのまま足すと、  
+と-が打ち消し合って（相殺して）小さく見えてしまう。

例（イメージ）

- ・ +5 と -5 を足すと 0（ズレがないように見える）
- ・ でも実際には、上下に大きくズレている




## 8. 問題：残差の合計は「+と-」で打ち消し合い、基準にならない

前スライドで見た通り、残差  $e = y - \hat{y}$  には+と-があります。  
そのため、残差をそのまま足すと「ズレの大きさ」を表せません。

### なぜダメか（相殺）

- ・ 例：+5 と -5 を足すと 0
- ・ 合計が 0 でも、実際には上下に大きくズレている


 だから、回帰では「符号付きの合計」ではなく、  
**ズレの大きさ**を足し合わせる方法が必要。

## 9. 二乗する理由：なぜ $(y - \hat{y})^2$ なのか

相殺を防ぐために、残差を二乗してから足し合わせます。

### 二乗する 3 つの理由（ここだけ覚える）

1. マイナスが消える（相殺しない）
2. 大きいズレを強く罰する（大外れを重く扱う）
3. 計算として扱いやすい（後で「最小」を作りやすい）

 回帰は「全部当てる」ではなく、ズレをできるだけ小さくすることが目的。


## 10. 残差平方和 (SSE) : ズレの大きさの合計を1つの数にする

残差を二乗して、全部足した数を作ります。

これが「線の良さ」を表す指標になります。

$$\text{SSE} = \sum (y - \hat{y})^2$$

- ・ 各点のズレ :  $(y - \hat{y})$
- ・ そのズレの大きさ :  $(y - \hat{y})^2$
- ・ それを全部足す :  $\sum (y - \hat{y})^2$


 SSE が小さいほど、点に近い（ズレが小さい）線。

## 11. 最小二乗法：SSE が最小になる直線が「回帰直線」

同じ散布図に引ける直線は何本もあります。  
その中から 1 本を選ぶルールが、最小二乗法です。

### 最小二乗法（言葉での定義）

- ・ 直線ごとに SSE（残差平方和）を計算する
- ・ SSE がいちばん小さい直線を選ぶ

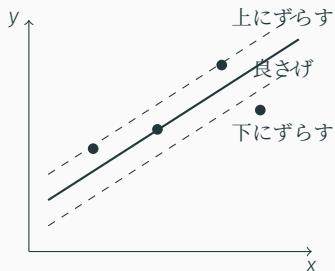
 つまり、回帰直線は  
「ズレの二乗の合計が最小」になるように決まる。

### 注意（ここで止める）

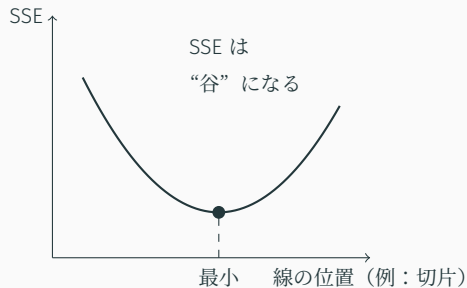
- ・ 今日は「最小にする考え方」を理解する回
- ・ 次の塊で「実際にどうやって最小を作るか（Excel で体験）」へ進む

## 12. 直感：線を少し動かすと SSE が増える／減る（“最小” のイメージ）

回帰直線は「当てずっぽう」で決めているのではなく、  
SSE（ズレの二乗の合計）を最小にするというルールで決まります。



直線を上下に動かすだけでも、  
点とのズレが変わる



SSE がいちばん小さい場所が 1 つ決まる

### 13. ミニ例（3点）：線 A と線 B で SSE を比べる（計算は最小限）

3つの点があるとして（散布図の一部だと思って OK）：

点	(x, y)	線 A の $\hat{y}$	線 B の $\hat{y}$
1	(1, 2)	2	1
2	(2, 2)	3	2
3	(3, 4)	4	3

ここで 2 本の候補：

- ・ 線 A： $\hat{y} = x + 1$
- ・ 線 B： $\hat{y} = x$

残差  $e = y - \hat{y}$  と SSE（残差の二乗の合計）を比べます。

点	$e_A$	$e_A^2$	$e_B$	$e_B^2$
1	$2 - 2 = 0$	0	$2 - 1 = 1$	1
2	$2 - 3 = -1$	1	$2 - 2 = 0$	0
3	$4 - 4 = 0$	0	$4 - 3 = 1$	1
合計 (SSE)	1		2	

## 14. Excel とつながる：トレンドラインの式＝SSE 最小の結果

Excel で散布図に「近似曲線（トレンドライン）」を追加すると、直線の式  $\hat{y} = ax + b$  が表示できます。

### ここがポイント

Excel のトレンドラインは、裏側で **SSE が最小になるように**  $a$ （傾き）と  $b$ （切片）を自動で決めている。

つまり：

- ・ 私たちがやってきた「SSE で良い線を決める」ルールを
- ・ Excel が**自動で実行**して、式を出している

実習でやる操作（次の塊につながる）

- ・ 散布図を作る（ $x$ =バーガー数、 $y$ =ポテト数）
- ・ 近似曲線（線形）を追加し、「**数式を表示**」「 **$R^2$  を表示**」
- ・ その式で  $\hat{y}$  を作り、残差と SSE を作る

## 15. 実習① (clean) : SSE を自分で作って「良い線」を理解する (全体手順)

Excel のトレンドラインは「SSE が最小の直線」を自動で出しています。  
実習①では、その中身を**自分の手で**再現します。

### 実習①のゴール (ここだけ)

予測値  $\hat{y}$  → 残差 → 残差<sup>2</sup> → 合計 (SSE) を作り、  
SSE が小さいほど良い線だと確かめる

1. 散布図から回帰式を表示し、傾き  $a$  と切片  $b$  を取り出す
2. 各行で  $\hat{y} = ax + b$  を計算して「予測値」列を作る
3. 残差  $e = y - \hat{y}$  列を作る (縦のズレ)
4. 残差<sup>2</sup> 列を作る (マイナスを消す+大ズレを強く罰する)
5. 残差<sup>2</sup> を合計して SSE を出す (SUM)

 最後に  $a$  や  $b$  を少し変えて、SSE が増えることを確認する



## 16. 実習①-1 (clean) : $\hat{y}$ (予測値) 列を作る (回帰式を使う)

まず、散布図にトレンドライン (線形) を追加し、  
数式  $\hat{y} = ax + b$  を表示して  $a$  と  $b$  をメモします。

- ・  $x$  : バーガー注文数 (説明変数)
- ・  $y$  : ポテト注文数 (目的変数)
- ・  $\hat{y}$  : 回帰式が出す「予測値」

### Excel で作る列 (例)

(例)  $a$  と  $b$  をセルに置く :

$a$  を H1、 $b$  を H2 に入力

予測値列 (例 : D2)

$=\$H\$1*A2 + \$H\$2$

 **ポイント** : \$ を付けて  $a, b$  のセルを固定すると、下までコピーできる

## 17. 実習①-2 (clean) : 残差・残差<sup>2</sup>を作り、SSEを計算する

残差は「実測」と「予測」のズレです。

回帰直線がどれだけ点に近いかを、残差で数にします。

### 作る列 (定義)

残差:  $e = y - \hat{y}$  (縦のズレ)

残差<sup>2</sup>:  $e^2$  (マイナスを消して足せるようにする)

SSE:  $\sum e^2$  (残差<sup>2</sup>の合計)

Excel 例 (列が違ってても OK)

- ・ 残差列 (例: E2):  $=B2 - D2$   $(y - \hat{y})$
- ・ 残差<sup>2</sup>列 (例: F2):  $=E2^2$  または  $=POWER(E2, 2)$
- ・ SSE (例: F1):  $=SUM(F2:F31)$

 SSE が小さいほど、点に近い (= 良い線)

## 18. 実習①-3 (clean) : $a$ や $b$ を少し変えて、SSE が増えることを確認する

ここが「体験で納得」パートです。

今の回帰式は、Excel が **SSE が最小**になるように選んだ直線でした。

### やること (超シンプル)

$a$  または  $b$  を **少しだけ**動かして、SSE がどう変わるかを見る  
(例)  $a$  を  $+0.1$  /  $b$  を  $+1$  など

### 確認のしかた

- ・  $a, b$  を変える  $\Rightarrow \hat{y}$  列が変わる
- ・  $\hat{y}$  が変わる  $\Rightarrow$  残差  $\cdot$  残差<sup>2</sup> が変わる
- ・ その結果、**SSE が増える** (多くの場合)

 **観察の結論** : 元の  $a, b$  が「SSE を最小にする」直線だった

## 19. 実習② (outlier) : 二乗は外れ値に弱い (線が引っ張られる)

実習① (clean) と同じ手順を、outlier データでもう一度やります。

ここで見たいのは、二乗 ( $e^2$ ) が外れ値に強く反応するという性質です。

### 今日の観察ポイント (ここだけ)

外れ値があると、1 点の大きなズレが  $e^2$  で何倍にも大きくなり、  
SSE を強く支配する → 回帰直線が外れ値に引っ張られる

### やること (実習①と同型)

1. outlier 版で散布図 → トレンドライン → 式 ( $a, b$ ) を表示
2.  $\hat{y}$  列 → 残差  $e$  列 → 残差<sup>2</sup> 列 → SSE を作る
3. clean 版と比べて、どこが大きく変わったかを見る

### 見るポイント (比較)

- ・  $a$  (傾き) や  $b$  (切片) はどれくらい変わった?
- ・  $R^2$  は上がった? 下がった?
- ・ 残差<sup>2</sup> 列で、異常に大きい行はどれ?

## 20. まとめ：回帰直線は「SSE 最小」で決まる／今後は“読み取りと応用”へ

今日のゴールは、回帰直線の裏側にあるルールを、式ではなく“列（手順）”で理解することでした。

- ・ 残差： $e = y - \hat{y}$ （実測と予測のズレ）
- ・ SSE： $\sum e^2$ （ズレの二乗の合計）
- ・ 回帰直線：SSE が最小になる直線（最小二乗法）
- ・ 二乗の性質：大きいズレを強く重視 → 外れ値に引っ張られやすい

### 今後への橋

今後は、回帰の結果を読む・使う練習を増やします。

（例）予測値の解釈／残差からの気づき／外れ値の扱い（データの意味づけ）

 同じ姿勢：データ → 要約（線）→ ズレ（残差）→ 解釈（理由）