

## 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

## A-1：第12回の立ち位置（運用）

---

【目的】「今日は何を理解する回か」を最初に固定する（最小二乗法＝線の決め方）。

【口頭補足】第11回は「使う（予測する）」回。第12回は「作り方（決め方）」の回。

【実習仕様】clean データで、 $\hat{y}$  → 残差 → 残差の二乗 → SSE（合計）を Excel で作る（提出なし）。

【運用】冒頭で確認：回帰は“判断”ではなく“予測”。（検定と混同させない）

## 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。

## A-2：例を継続する理由（口頭補足）

---

【目的】例（ハンバーガーショップ）を固定し、内容理解に集中させる。

【口頭補足】例が変わると「状況理解」からやり直しになる。店長ストーリーで一貫させる。

【実習仕様】 $x$ =バーガー数、 $y$ =ポテト数（clean / outlier の 2 データで扱う）。

【運用】ここで「今日は予測の話」と言い切る（検定の“判断”と混同しやすい）。

### 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。

## A-3：混同防止（短く強く）

---

【目的】検定と回帰の混同を最初に止める（留学生ほど混ざりやすい）。

【口頭補足】検定＝結論を“判断”。回帰＝値を“予測”。今日は  $p$  値は使わない。

【実習仕様】回帰では「ズレ（残差）」を作り、その二乗の合計（SSE）で線の良さを比べる。

【運用】確認質問：検定は何？ → 判断。回帰は何？ → 予測。今日のキーワードは？ → 残差・二乗・最小。



## 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

## B-4：問いの狙い（運用）

---

【目的】「回帰直線は 1 本に決まる」という前提を崩し、評価基準が必要だと気づかせる。

【口頭補足】学生に質問：「線 A・B・C どれが良い？」→理由も言わせる（“近い” “ずれが小さい” を引き出す）。

【実習仕様】後半で、Excel で「ズレ」を数値化して比べる（SSE まで作る）。

【運用】ここではまだ数式を出さない。まず“線の良さを決めたい”という欲求を作る。

## 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

## B-5：評価基準を先に固定（口頭補足）

---

【目的】回帰の本体は「線を決めるルール」だと理解させる。

【口頭補足】良い線＝“当たる保証”ではなく、“ズレが小さい線”。（未来の保証ではない）

【実習仕様】残差→残差の二乗→合計（SSE）を Excel で作り、線の良さを比較する。


【運用】次スライドにつなぐ一言：「ズレ＝残差。まず“残差って何？”を定義する」。

## 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。

## C-6：残差を「距離」として固定（運用）

---

【目的】残差を「式の引き算」ではなく「点と線の縦のズレ」として理解させる。

【口頭補足】横方向ではなく縦方向のズレを見る（ $x$  は説明変数として固定して考える）。

【実習仕様】Excel で各行に  $\hat{y}$ （予測）列を作り、残差列  $e = y - \hat{y}$  を作る。

【運用】一言確認：「残差って何？」→「実測－予測（縦のズレ）」。

## 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

## C-7：なぜ“そのまま足せない”のか（運用）

---

【目的】「相殺」という問題を先に理解させ、次の“二乗”が必要な理由を作る。

【口頭補足】ズレは“方向”ではなく“大きさ”を評価したい（だから符号は邪魔になる）。

【実習仕様】Excel で残差列を作った後、残差の二乗を作る（相殺を防ぐ準備）。

【運用】次につなぐ問い：「相殺を防ぐにはどうする？」→「二乗する」。



## 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。

## D-8：二乗への導入（運用）

---

【目的】「合計では評価できない」という問題意識を確定し、二乗の必然性を作る。

【口頭補足】ズレは“方向”ではなく“大きさ”を評価したい、という言い方に統一する。

【実習仕様】この直後に「残差の二乗」を作らせ、相殺が消えることを体感させる。


【運用】ここで問いかけ：「相殺しない方法は？」→ 次スライドで二乗を提示。

## 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。

## D-9：二乗の納得（口頭補足）

---

【目的】二乗を“暗記”ではなく“必要性”として納得させる。

【口頭補足】③は深掘りしない。「二乗だと数学的に最小を作りやすい」程度で止める。

【実習仕様】残差列と残差の二乗列を並べ、相殺が消えることを確認させる。

【運用】学生に一言で言わせる：「なぜ二乗？」→「相殺しない／大ズレを重くする」。

## 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。

## D-10：SSE を「1つの数」にする意味（運用）

---

【目的】多数の点のズレを「比較できる 1 つの数」にまとめる発想を固定する。

【口頭補足】SSE は「ズレの大きさの合計」。単位は  $y$  の二乗になるが、ここでは深掘りしない。

【実習仕様】Excel で残差の二乗を作り、SUM で SSE を作成（線ごとに SSE 比較も可能）。


【運用】次スライドにつなぐ問い：「良い線は？」→「SSE が最小の線」。

## 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

## D-11：結論の回収（運用）

---

【目的】 A-1 で先に出した結論を、SSE の定義を経て“納得”として回収する。

【口頭補足】 ここで微分や公式導出は不要。最小二乗法＝「SSE が最小」と言えれば十分。

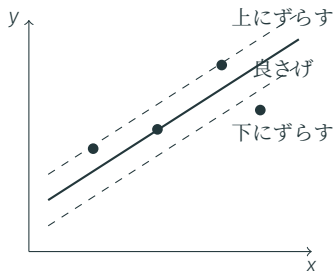
【実習仕様】 Excel で SSE を実際に計算し、線を変えると SSE が変わることを体感させる。

【運用】 ここで一言確認：「回帰直線はどう決まる？」→「SSE が最小の線」。

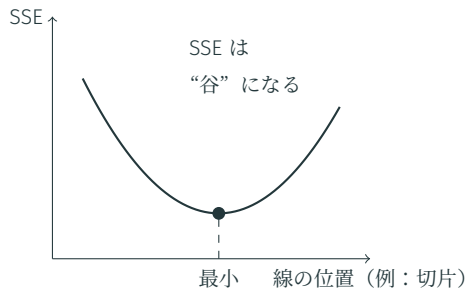


## 12. 直感：線を少し動かすと SSE が増える／減る（“最小” のイメージ）

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

## E-12：最小の直感を作る（運用）

---

【目的】「最小二乗法＝SSE 最小」を“谷のイメージ”で直感的に理解させる。

【口頭補足】右図は「SSE が最小になる場所がある」というイメージ図（厳密な導出ではない）。

【実習仕様】後で Excel で線（トレンドライン）を出す＝SSE 最小の線が自動で選ばれる、と接続する。

【運用】ここで確認：「回帰直線はどう決まる？」→「SSE が最小の線」。

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

3つの点があるとしします（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	

## E-13：SSE 比較を“手で一回だけ”やる理由（運用）

---

【目的】SSE が「点のズレの二乗の合計」で、比較に使えることを具体例で固定する。

【口頭補足】ここでは導出はしない。「SSE で比べると、良い線が決まる」を体感させる。

【実習仕様】Excel で同じことを大量の点で自動的にやっている（手作業はこの例だけ）。

【運用】計算は最小限：残差→二乗→合計、の流れだけを覚えさせる。

## 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

## E-14：Excel のブラックボックスを壊す（運用）

---

【目的】「トレンドラインの式＝勝手に出てくる数字」ではなく「SSE 最小の結果」と結びつける。

【口頭補足】Excel は“魔法”ではない。残差の二乗を全部足して最小になる線を選んでいるだけ。

【実習仕様】 散布図→トレンドライン→式/R の二乗表示→ $\hat{y}$  列→残差列→残差の二乗列→SSE (SUM)。

【運用】次の実習へ自然につなぐ：「じゃあ Excel で同じことをやってみよう」。

## 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

## F-15：実習①の位置づけ（運用）

---

【目的】「SSE 最小」が Excel の裏側のルールだと体験で納得させる。

【口頭補足】ここでは“導出”しない。SSE を作って比較するだけで十分。

【実習仕様】データは clean 版を使用（外れ値なしで成功体験を作る）。

【運用】このスライドを見せたら、すぐ Excel 操作へ移る（説明疲れ対策）。



## 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

## F-16：操作の落とし穴（運用）

---

【目的】  $\hat{y}$  列（予測値）を全員が確実に作れるようにする。

【口頭補足】  $\hat{y}$  は「点の中心を通る線が出す予測」。実測  $y$  とは一致しない。

【実習仕様】 (1) 散布図作成 → (2) 線形近似 → (3) 数式表示 → (4)  $a, b$  をセルへ → (5)  $\hat{y}$  列作成。

【運用】  $a, b$  を式に直書きさせるとミスが増えるので、セル参照（\$固定）を推奨。

## 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)

## F-17：ここで固定したい言葉（運用）

---

【目的】残差→二乗→合計（SSE）の意味を“列”として定着させる。

【口頭補足】残差の合計は相殺するのでダメ。二乗してから足す。

【実習仕様】全員が SSE を 1 つのセルで出せたら次へ。数値が合わない学生は残差列を確認させる。

【運用】<sup>（上付き）</sup>はエラーになりやすいので、必要なら POWER を推奨。

## 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

## F-18：実習①の締め（運用）

---

【目的】「最小」の意味を“数値の増減”で納得させる（説明ではなく体験）。

【口頭補足】少し動かすだけで SSE が増える＝“谷の底”にいた、ということ。

【実習仕様】(1) 元の SSE をメモ→(2) $a$  を少し変える→ SSE 再確認→(3) $b$  も少し変える。

【運用】変え幅が大きいと混乱するので「少しだけ」を強調。早い学生には「どちら方向に動かすと増え方が大きい？」を追加。

## 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?

## G-19：実習②の運用（疲れさせない）

---

【目的】「二乗＝外れ値に弱い」を体験で理解させ、現実データの扱いへつなぐ。

【口頭補足】外れ値 1 点の  $e^2$  が大きすぎて、他の点を押しつけて SSE を支配することがある。

【実習仕様】(1)outlier で実習①と同じ列作成→(2) 残差<sup>2</sup> が最大の行を特定→(3) メモ列で理由づけ。

【運用】時間が厳しければ「SSE 最大の行を探す」だけでも OK（全列を作らせなくてもよい）。



## 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）

## H-20：締めと今後接続（運用）

---

【目的】今日のキーワードを“3点セット”で固定し、次回の「読み取り・応用」に自然につなぐ。

【口頭補足】回帰は「線を引いて終わり」ではなく、残差と外れ値から“何が起きたか”を考える道具。

【実習仕様】提出なし。最後に全員で「外れ値はなぜ危ない？」を一言で言わせて終了。

【運用】次回の冒頭で、clean/outlierの違いを1枚で再確認すると接続が良い。