

13 単回帰（最小二乗法）＋回帰のまとめ

2.ストーリー：次の目標は「ボテト売上UP」

店長は、ボ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：

- ボテトは利益率が高い
- だから「ボ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、ボテトの注文が増えるのか？
- 関係があるなら、売上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。

10/20/2019, アルゴリズム 2.1/3 2 / 20

4

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「だいたいわ上がり」に見えます。
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、見た目でなく**評価基準（ルール）**が必要

10/20/2019, アルゴリズム 2.1/3 4 / 20

8

6.残差の定義：残差＝実測 y－予測 y（線のズレ）

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。

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「全部を完全に当ててる線」ではない。
だから、残差（ズレ）は必ず出る。

10/20/2019, アルゴリズム 2.1/3 6 / 20

12

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」を減しません。

なぜダメか（総論）

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

🔴だから、回帰では「四捨五入の合計」ではなく、**ズレの大きさを足し合わせる方法が必要**。

10/20/2019, アルゴリズム 2.1/3 8 / 20

16

10.残差平方和（SSE）：ズレの大きさの合計を1つの数にする

残差を単純して、全部足した数を作ります。
これが「**残差の合計**」を表す数値になります。

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

- 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。

10/20/2019, アルゴリズム 2.1/3 10 / 20

20

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）を継続し、内容理解を助ける。
【前提補足】例が使われる「双方向性」からやり直しになる。店長ストーリーで一旦きめる。
【前提補足】ハンバーガー数、yボテト数（mean / outlier の2パートで扱う）。
【総論】ここで「今日は予測の線」と言い切る（既定の「判断」と差別化する）。

B-4：問いの狙い（運用）

【目的】「回帰直線は1本に決まる」という前提を繰り返し、評価基準が必要だと気づかせる。
【前提補足】学習に慣習「線は1本しか引けない」→理解も深まる（「ズレ」が小さい）を促す。
【前提補足】後半で、Excel で「ズレ」を観測して比べる（SSE まで作る）。
【総論】ここではまだ式は出さない。まず「線の良さを見たい」という観点を育てる。

C-6：残差を「距離」として固定（運用）

【目的】残差を「式が引き算」ではなく「点と線の縦のズレ」として理解させる。
【前提補足】垂直方向ではなく縦方向のズレを見る（y は縦軸方向として同じで使える）。
【前提補足】Excel で実際に y（予測）列を作り、残差列 $e = y - \hat{y}$ を作る。
【総論】ここで例で例「残差って何？」→「実測－予測（縦のズレ）」。

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醸成し、二乗の必要性を作る。
【前提補足】ズレは「方向」ではなく「大きさ」を評価したい。という言い回しに統一する。
【前提補足】この段階に「残差の二乗」を呼ぶから、傾きが異なることを体感させる。
【総論】ここで例で例「傾きしないの直線？」→次スライドで二乗を提示。

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「計算できる1つの数」にまとめる意思を促す。
【前提補足】SSE は「ズレの大きさの合計」。単に y の二乗ではない。ここでは計算しにくい。
【前提補足】Excel で残差の二乗を作り、SUM で SSE を作る（縦ごとに SSE 計算も可能）。
【総論】次スライドに引く問い「良い線は？」→「SSE が最小の線」。

10/20/2019, アルゴリズム 2.1/3 20 / 20

21

1.回帰直線は「どう決める？」（最小二乗法）

今日の目標（先読み）
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まられています。**

- 前回：散布図から回帰直線を引く、式と R² を見て予測した
- 今日は：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 が最小の線」と言葉で言える

10/20/2019, アルゴリズム 2.1/3 1 / 20

2

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いいか？」を移しき（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 がわかると y はどう変わるか？を式（回帰直線）で表す
- 式を使って未知の 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。

10/20/2019, アルゴリズム 2.1/3 3 / 20

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「ズレ」で決めるために、次を行います。

- 各点について「ズレ」を観測
- 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**数と言葉と式**で定義する

10/20/2019, アルゴリズム 2.1/3 5 / 20

10

7.残差の符号：横より上は＋／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

10/20/2019, アルゴリズム 2.1/3 7 / 20

14

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（ここだけ覚えて）

- マイナスがなくなる（相殺しない）
- 大きいズレを大きく罰する（大抵を重く扱う）
- 計算として扱いやすい（後で「最小」を作りやすい）

🔴理由は「全部で見る」ではなく、**ズレをできるだけ小さくすることが目的**。

10/20/2019, アルゴリズム 2.1/3 9 / 20

18

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無限にあります。
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
- 次の晩で「実際にどうやって最小を作るか（Excel で体感）」へ進む

10/20/2019, アルゴリズム 2.1/3 11 / 20

A-1：第 12 回の立ち位置（運用）

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

【前提補足】第 11 回は「使う（予測する）」例、第 12 回は「作る方（決め方）」例。

【前提補足】2019 アドベツで、y＝残差＝残差の二乗＝SSE（合計）を Excel で作る（慣習なし）。

【総論】事前で解説「回帰は「判断」ではなく「予測」。（既定と判断を区別）」

A-3：混同防止（短く強く）

【目的】既定と回帰の目的を最初に定める（習学生は気づきやすい）。

【前提補足】既定＝結論を「判断」、回帰＝結果を「予測」、今日は結果は使わない。

【前提補足】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。

【総論】混同防止：既定は引く→判断、回帰は引く→予測。今日のキーワードは「二乗＝二乗＝最小」。

10/20/2019, アルゴリズム 2.1/3 3 / 20

7

B-5：評価基準を先に固定（口頭補足）

【目的】回帰の目的は「線を「決めたいルール」で決める」。

【前提補足】良い線＝「点と線のズレ」ではなく「ズレが小さい線」。（本来の解説ではない）

【前提補足】残差＝残差の二乗＝SSE（合計）を Excel で作り、線の良さを比較する。

【総論】次スライドにつなぐ「ズレ＝残差。まず「残差って何？」を定義する」。

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」が必要理由を作る。

【前提補足】ズレは「方向」ではなく「大きさ」を評価したい（y は縦軸方向として同じで使える）。

【前提補足】Excel で残差の二乗を作り、SUM で SSE を作る（縦ごとに SSE 計算も可能）。

【総論】ここから例「傾きしないの直線？」→「SSE が最小の線」。

D-9：二乗の納得（口頭補足）

【目的】二乗を「相殺」ではなく「定量化」して理解させる。

【前提補足】①は相殺しない（二乗だと数値的に罰金を作りやすい）例で定める。

【前提補足】相殺防止：残差の二乗＝SSE（合計）を Excel で作る（慣習なし）。

【総論】学生に「打ち消し合って（相殺）しないズレを小さくする」。

D-11：結論の回収（運用）

【目的】A4 で先に出した結論を、SSE の定義を踏まえて「判断」として回収する。

【前提補足】ここで検定や仮説検定は不要。最小二乗法（SSE が最小）と言えれば十分。

【前提補足】Excel で SSE を計算に慣習し、線を引ると SSE が変わることを体感させる。

【総論】ここで二乗＝二乗（回帰は最小の線）→「SSE が最小の線」。

10/20/2019, アルゴリズム 2.1/3 11 / 20

23

