

12 単回帰（回帰の意味と散布図）

1. 回帰で「将来予測」を作る（判断ではなく予測）

今日やること

ハンバーガーショップのデータを使って、

「ポテトの注文数を増やすには、何に関係していそうか」を見つけ、
「このくらい売れそう」という予測を作ります。

- ・ 今日のキーワード：散布図・回帰直線・予測式 ($\hat{y} = ax + b$)
- ・ 回帰は、正しい／間違いを決める話ではなく、未来を予測する話
- ・ ゴールは計算ではなく、式の意味を言葉で説明できること

今日の到達目標

- ・ 散布図から「増える／減る／関係なし」を説明できる
- ・ 回帰直線と式を表示し、傾き a の意味を言える
- ・ 回帰式で「もし x がこうなら、 y はこのくらい」を予測できる

2. ストーリー：店長の次の目標は「ポテト売上 UP」


前回まで店長は、ポテトの重さのばらつきをデータを見て改善しました。
（「本当に減ったと言ってよいか？」を判断できるようになった）

次の目標（現場の目標）

- ・ ポテトは利益率が高い。
- ・ だから店長は ポテトの売上を伸ばすことを目標にしました。

そこで店長は考えます：

- ・ 「お客さんがハンバーガーと一緒にポテトを注文する数を増やしたい」
- ・ 「そのために、何が増えるとポテトも増えるのだろう？」

 今日はこの問いに対して、データから予測の式を作る。

3. 前回の検定は「判断」／今回の回帰は「予測」

前回までやってきたのは、次のような判断です。

「平均との差がある」だけではなく、ばらつきの中で珍しいかで決めました。

検定（前回）＝判断

珍しさ（p 値）を使って、

「偶然と言ってよいか／言いにくいかな」を決める。

一方、今日は次の目的です。

回帰（今日）＝予測

2 つの数値の関係をを使って、

「もし x がこうなら、 y はこのくらい」を予測する。

- ・ 検定は「棄却／棄却できない」で結論を出す
- ・ 回帰は「予測の式」を作り、使う（結論の形が違う）

4. 問いの定式化：ハンバーガーが増えるとポテトは増える？ ($x \rightarrow y$)

店長の目標は「ポテト売上 UP」でした。

そこで、ポテトの注文数に関係しそうなものを考えます。

今日の問い（回帰の問い）

ハンバーガー注文数が増える日ほど、ポテト注文数も増えるのか？

回帰では、2つの数値の役割を先に決めます。

- ・ x （説明変数）：ハンバーガー注文数（原因かもしれない側）
- ・ y （目的変数）：ポテト注文数（増やしたい結果の側）

📌 今日やることは、 x を見たら y を予測できるかを確認すること。

5. 散布図とは：2つの数値の関係を「点」で見る図（定義）

回帰を始める前に、まず**散布図**で関係を目で見ます。

散布図（定義）

1つのデータ（1日）を1つの点として、
横軸に x 、縦軸に y をとって並べた図。

今回の例では、点の意味はこうです。

- ・ ある1日：（ハンバーガー注文数 x , ポテト注文数 y ）
- ・ 30日分なら、点が30個できる

📌 散布図を見ると、増える／減る／関係なしが直感的に分かる。

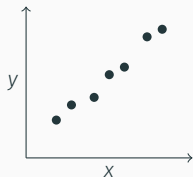
6. 散布図の読み方①：右上がり／右下がり／関係なし（3パターン）

散布図は、点の並び方で「関係の方向」を読みます。

（細かな計算の前に、まず目で判断できるようにする）

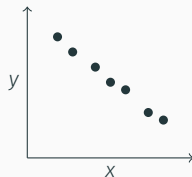
右上がり

（ x が増えると y も増えやすい）



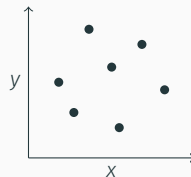
右下がり

（ x が増えると y は減りやすい）



関係なし

（ x が変わっても y が読めない）



🔗 今日のデータ（ハンバーガー → ポテト）は、**右上がり**になりそうかをまず確認する。

7. 散布図の読み方②：ばらつき・外れ値（現実データの特徴）

現実のデータは、きれいに一直線には並びません。
同じ x でも y が少し違うのが普通です。

重要：データには「ばらつき」がある

ばらつきがあるから、予測は **100%当たる** わけではない。
でも、**中心の傾向**をつかめば、だいたいの予測はできる。

さらに、**外れ値**が混ざることもあります。

- ・ 例：品切れでポテトが売れなかった日（普段より**極端に少ない**）
- ・ 例：キャンペーンでポテトが急に売れた日（普段より**極端に多い**）

 後半の実習では、外れ値があると**回帰直線**がどう変わるかを確認める。

8. 回帰の目的：点の雲を「一本の線」で要約し、予測に使う

散布図を作ると、点は雲（ばらついた集まり）になります。

このままだと、「次はどれくらい？」が言いにくい。

回帰の目的（理由）

点の雲を**一本の線**で要約して、
予測に使える形（式）にする。

回帰でやりたいことは、次の2つです。

- ・ 関係を要約する： x が増えると y はどう変わりやすいか
- ・ 予測する：もし x がこの値なら、 y はこのくらい (\hat{y})

✂️ ここでの「線」は、全部当てる線ではなく、**中心の傾向**を表す線。

9. 回帰直線（イメージ）：各 x に対して“中心”を通る予測の線

回帰では、点の雲の中心を通るような線を考えます。
この線を回帰直線と呼びます。

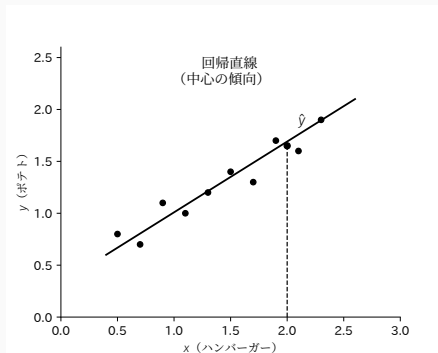
回帰直線（イメージ）

同じ x のときの y の“中心”を結んだような線。

この線上の値が、 y の予測値 (\hat{y})。

- ・ 点は線の周りに散らばる（ばらつき）
- ・ それでも線があると、だいたいの予測ができる

📌 点の中心を表す線だと思えばよい（まずはこれで OK）。



点が散らばっていても、線があれば
「この x なら \hat{y} はこのくらい」が言える

10. 回帰式： $\hat{y} = ax + b$ (\hat{y} = 予測値) を文章で理解する

回帰直線は、**式**で表せると便利です。

(Excel が計算して表示してくれる式を、**言葉で説明できる**のが今日の目標)

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

この式が言っていること (文章)

x (ハンバーガー注文数) が分かれば、
ポテト注文数の予測値 \hat{y} を計算できる。

- ・ \hat{y} (ハット付き) は**予測値** (実測の y とは違う)
- ・ 点はばらつくので、**実測 y は \hat{y} と一致しないことが多い**

 今日の実習は、Excel で式を出し、その意味を言葉で説明すること。

11. パラメータの意味：傾き a = 増え方 / 切片 $b = x = 0$ の予測

回帰式 $\hat{y} = ax + b$ の意味は、 a と b を見れば説明できます。

傾き a (いちばん大事)

x が 1 増えると、 \hat{y} が平均でどれくらい増えるか。

例： $a = 0.6$ なら、バーガーが 10 増えるとポテトは約 6 増える。

切片 b (注意して使う)

$x = 0$ のときの予測値。

ただし、現実には $x = 0$ (バーガー 0) が起きない範囲なら、

b の現実的な意味は薄いことがある。

🔑 実習では、まず a (増え方) を言葉で説明できることを優先する。

12. R^2 とは：線で説明できる割合の目安（正解率ではない）

回帰直線を引いたら、次に気になるのはこうです。

「この線は、どれくらいデータに合っているの？」

R^2 （決定係数）

回帰直線で、 y のばらつきのうち
どれくらいが x で説明できたかを表す目安。

- ・ R^2 は 0～1 の間の数（例：0.72）
- ・ 1 に近いほど、点が線に近く、説明できる割合が高い
- ・ 0 に近いほど、点が線から散らばり、説明できる割合が低い

ここが重要（誤解防止）

R^2 は 正解率ではない。

「未来が必ず当たる」ことは保証しない。

13. 注意：相関≠因果／回帰は「当てる保証」ではない

回帰ができると、つい次のように言いたくなります。

「バーガーが増えたから、ポテトが増えた！」

でも注意

散布図や回帰で分かるのは、基本的に**関係（相関）**であって、**原因（因果）**を確定することではない。

例：バーガーとポテトが同時に増える本当の理由は、別にあるかもしれません。

- ・ その日は**来客数**が多かった（両方増える）
- ・ **キャンペーン**をやっていた（セット注文が増える）
- ・ **品切れ**があった（片方だけ減る）

今日の結論の形

回帰で言えるのは、

「この範囲では、増える傾向がある／予測に使えるかもしれない」まで。

14. 実習で使うデータ：clean / outlier の2つ（目的が違う）

今日は同じテーマ（ポテト売上 UP）でも、目的の違う2種類のデータで学びます。
順番が大事です（まず成功→次に現実）。

データ A：clean（外れ値なし）

回帰の基本を「成功」させるためのデータ。

散布図が読みやすく、回帰直線がきれいに出来る。

- ・ 目的：散布図→直線→式→ R^2 →予測、をスムーズに体験
- ・ 「傾き a の意味」を言葉で説明できるようにする

データ B：outlier（外れ値あり）

現実のデータはズレることを体験するためのデータ。

品切れ／キャンペーンなどで点が極端に外れる日が混ざる。

- ・ 目的：外れ値が回帰直線・ R^2 ・予測に与える影響を知る
- ・ 「なぜズレたか」をメモ列から説明する

15. 実習のゴール：散布図→直線→式→ R^2 →予測→（外れ値でズレを考える）

実習では、回帰を「作って使う」までを一気にやります。
最後に、外れ値で**現実のズレ**も確認します。

実習の流れ（この順番で固定）

1. 散布図を作る（ x と y の関係を点で見る）
2. 回帰直線を引く（点の雲を一本の線で要約）
3. 回帰式 $\hat{y} = ax + b$ を表示する（ \hat{y} = 予測値）
4. R^2 を表示する（線で説明できる割合の目安）
5. 具体的な x を入れて **予測**する（ \hat{y} を計算）
6. （後半）outlier で **ズレの原因**を考える（メモ列と結びつける）

📌 「傾き a は何を意味する？」 / 「ズレた日はなぜ？」

16. 実習①-1 (clean) : 散布図を作る (x=バーガー数、y=ポテト数)

目的：回帰はまず**散布図**から始める。点の並びで「関係の方向」を読む。

やること (操作)

1. clean シートを開く (外れ値なし)
2. x (横軸) に **バーガー注文数**、y (縦軸) に **ポテト注文数** を選択
3. **散布図 (点)** を作る

必ずやる (見た目のルール)

- ・ グラフタイトル : 「**バーガー数とポテト数 (clean)**」
- ・ 横軸ラベル : **バーガー注文数 (件)**
- ・ 縦軸ラベル : **ポテト注文数 (件)**

 まずは**右上がり**／**右下がり**／**関係なし**のどれに近いかを言えるようにする。

17. 実習①-2 (clean) : 回帰直線・式・ R^2 を表示 (線形近似)

目的：点の雲を一本の線で要約し、式と R^2 を表示する。

やること (操作)

1. 散布図上の点をクリック
2. 近似曲線 (トレンドライン) を追加 (線形を選ぶ)
3. 「グラフに数式を表示する」を ON
4. 「グラフに R^2 値を表示する」を ON

表示場所の指示 (読みやすくする)

- ・ 式と R^2 は、点に重ならない場所へドラッグして移動
- ・ 点が見えにくい場合は、グラフを少し拡大して OK

 ここで得るものは2つ：回帰式 (予測ルール) と R^2 (当てはまりの目安)。

18. 実習①-3 (clean) : 読み取りと予測 (傾きの解釈+指定 x で \hat{y} 計算)

目的：出てきた式を「読める」「使える」にする。
(計算よりも意味を言葉で説明できることがゴール)

Step1 : 傾き a を言葉で説明する

回帰式 $\hat{y} = ax + b$ の a は、
バーガーが1増えたとき、ポテトが平均でどれくらい増えるか。

- ・ 例： $a = 0.6$ なら「バーガーが10増えると、ポテトは約6増える」

Step2 : 指定 x で予測値 \hat{y} を計算する

教師が指定する x (例： $x = 150$) を式に代入し、 \hat{y} を計算する。
線の上の値=予測値 (ハット付き)。

 口頭チェック： y は実測、 \hat{y} は予測 (同じではない)。

19. 実習② (outlier) : 外れ値入りで再実行 (線・式・ R^2 の変化を見る)

目的：現実のデータは「きれい」ではない。

外れ値が入ると、回帰直線・式・ R^2 ・予測がどう変わるかを体験する。

● やること (操作は実習①と同じ)

1. outlier シートを開く
2. 散布図を作る (軸ラベルも同じ)
3. 線形の回帰直線を追加し、式と R^2 を表示する

● 比較するポイント (ここが学び)

- ・ 線の向き・位置は変わったか？
- ・ 傾き a はどれくらい変わったか？
- ・ R^2 は上がった？ 下がった？
- ・ 同じ x (例： $x = 150$) で \hat{y} はどれくらい変わる？

● 理由づけ：メモ列で説明する

- ・ 外れている点を見つけたら、メモを見て理由を言葉にする。

例：品切れ (本当は売れるのに売れない) / キャンペーン (急に増える)

20. まとめ

今日の到達点（ここだけ）

- ・ 散布図で「関係」を見て、回帰直線で**一本の線**に要約した
- ・ 回帰式 $\hat{y} = ax + b$ を使って、**予測**を計算できるようになった
- ・ R^2 は「当てはまりの目安」で、**正解率ではない**と分かった

現実のポイント：ズレ（残差）が必ずある

実測 y と予測 \hat{y} は一致しないことが多い。
この**ズレ**を **残差**（誤差）と呼ぶ。

- ・ clean：ズレは小さめ → 回帰の基本が見える
- ・ outlier：ズレが大きくなる → 予測がぶれる／ R^2 も変わる

次回への：では、線はどうやって決めている？

Excel は勝手に線を引いているように見えるが、
実は「ズレ（残差）を小さくする」ルールで線を決めている。
次回はその決め方＝**最小二乗法**を学ぶ。