

12. R²とは：線で説明できる割合の目安（正解率ではない）

回帰直線を描いた。実際に点になるのさ。この線は、どれくらいデータに合っているの？

R²（決定係数）
回帰直線が、yのばらつきのうちどれくらいが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、点が線から数らはり、説明できる割合が低い
ここが重要（理解ポイント）
R ² は正解率ではない。 「未来が必ず当たる」とは保証しない。

10/2020 データサイエンス 1/2

Slide12：R²の言い方を固定する（留学生対応）

【目的】R²を「当てはまり」や「正解率」と説明させない。意味を1次で固定する。
【理解ポイント】R²は「線で説明できる割合の目安」。未来が当たる保証ではない。
【実習ポイント】Excelの回帰診断オプションで「R²値を表示」をONにし、cleanとoutlierで線がどう変わるかを比較する。
【確認】学生に一言で説明する。「R²って何？」⇒「線で説明できる割合の目安」。

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。ついでに点に近づく。最後に、外れ値が増えたから、ポットが増えた！

でも注意
散布図や回帰で分かるのは、基本的な関係（相関）であって、原因（因果）を確定することではない。
例：バーガーとポットの関係が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（ポット注文が増える）
・ 品切れがあった（片方だけ減る）
今日の結論の形
回帰で表えるのは、 「この関係では、増える傾向がある／予測に使えるかもしれない」まで。

10/2020 データサイエンス 1/2

Slide13：相関と因果の線引き（やりすぎない）

【目的】相関＝因果の説明、と説明させない。特に留学生は因果関係に陥りやすい。
【理解ポイント】相関は「予測の道具」。因果は因果関係（原因・結果プロセス）が必要になることが多い。
【実習ポイント】outlierを削除し、メモを「原因（キャンペーン）」を記入し「因果関係がある」ことを確認する。
【確認】ここでは線引きしない。「相関＝因果」を言葉として1回だけ伝える。

24

25

26

27

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

今日は回帰データ（ポット売上）の2つでも、目的の違う2種類のデータで学びます。
線がどう変わる？（まず視覚的に見てみる）

データの差（視覚的に見てみる）
回帰の基準を「説明」させるためのデータ 散布図が読みやすい。回帰直線がきれいに出る。
・ 目的：散布図→直線→R ² →予測。をスムーズに体験
・ 「傾きαの意味」を言葉で説明できるようにする
データ差：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線から外れる目が混ざる。
・ 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」をメモから説明する

10/2020 データサイエンス 1/2

Slide14：2 データ構成の違い（疲れたない授業設計）

【目的】回帰の基本を「視覚的」で掴むから、現実（外れ値）に慣れさせる。留学生に。
【理解ポイント】同じデータでも、学ぶ目的が違う。cleanは「傾きを見る」、outlierは「現実の歪み」。
【実習ポイント】Excelファイル内でシートを切り替えて（例：clean / outlier）、回帰診断の結果を見て比較しやすくする。
【確認】学生に一言で「なぜ2つのデータが必要か」を説明する。例：「なぜズレる？」をみる。

15. 実習のゴール：散布図→直線→式→R²→予測→（外れ値でズレを考える）

実習は、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（この順番で確認）
1. 散布図を作る（xとyの関係の点で見る）
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（yを計算）
6. （後半）outlierでズレの原因を考える（メモ用と結びつける）
★「傾きαは何を意味する？」／「ズレた日はなぜ？」

10/2020 データサイエンス 1/2

Slide15：成果物の形（評価ポイントの明確化）

【目的】実習のゴールを「報告」ではなく「説明できる成果物」に置く。
【理解ポイント】計算はExcelがやる。大事なのは「結果を伝える」こと（α、y、R²、外れ値）。
【実習ポイント】散布図（散布図）の式とR²表示の予測値（αの値とy）と外れ値の説明（メモを準備）。
【確認】傾きが正しい場合：cleanでまだ正確。outlierは点の位置がずれただけでOK。

28

29

30

31

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

目的：目的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シートにバーガー数（x）とポット数（y）の散布図を描く
2. x（横軸）にバーガー数、y（縦軸）にポット数を指定
3. 散布図（点）を作る
必ずやる（見直しポイント）
・ グラフタイトルに「バーガー数とポット数（clean）」
・ 縦軸ラベルに「バーガー数（件）」
・ 横軸ラベルに「ポット数（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葉できるようにする。

10/2020 データサイエンス 1/2

Slide16：実習仕様（授業内での確認ポイント）

【目的】回帰の基本を「視覚的」で掴むから、現実（外れ値）に慣れさせる。留学生に。
【理解ポイント】点の並びがわかる。この時点で「関係」は読み取れる。
【実習ポイント】散布図を描く。回帰診断の結果を見て比較しやすくする。
【確認】学生に一言で「この散布図は3パターン（どれ？）」で説明できるようにする。

17. 実習①-2（clean）：回帰直線・式・R²を表示（線形近似）

目的：点の雲を一本の線で要約し、式とR²を表示する。

やること（操作）
1. 散布図上の点をクリック
2. 近似直線（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
表示場所の確認（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）。

10/2020 データサイエンス 1/2

Slide17：Excel操作で迷わせない（最短手順）

【目的】操作で迷わせる。目的が明確化された。手順を短くする。
【理解ポイント】Excelの「挿入」タブで「散布図」を選択し「散布図」を選択する。
【実習ポイント】散布図を描く。回帰診断の結果を見て比較しやすくする。
【確認】傾きが正しい場合：cleanでまだ正確。outlierは点の位置がずれただけでOK。

32

33

34

35

18. 実習①-3（clean）：読み取りと予測（傾きの解釈・指定αでy計算）

目的：指定αでyを「求める」「見える」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αを言葉で説明する
回帰式 $\hat{y} = ax + b$ のαは、 バーガーが1増えたとき、ポットが平均でどれくらい増えるか。
・ 例：α＝0.6なら「バーガーが10増えると、ポットは約6増える」
Step2：指定αで予測値yを計算する
教師が指定するα（例：α＝100）を式に代入し、yを計算する。 傾きの値＝予測値（ヒット付き）
★口頭チャット：yは実測、yは予測（同じではない）。

10/2020 データサイエンス 1/2

Slide18：授業内での確認（提出なし運用）

【目的】回帰の基本（傾きの解釈・予測）を「言葉」で伝えるようにする。
【理解ポイント】傾きαは「傾き」の単位。この時点で「関係」は読み取れる。
【実習ポイント】散布図を描く。回帰診断の結果を見て比較しやすくする。
【確認】学生に一言で「この散布図は3パターン（どれ？）」で説明できるようにする。

19. 実習②（outlier）：外れ値入りで再実行（傾・式・R²の変化を見る）

目的：現実のデータは「きれいな」ではない。
外れ値が入ると、回帰直線・式・R²・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（傾・式・R ² も同じ）
3. 回帰の回帰診断を追加し、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・傾きは変わったか？
・ 傾きαはどれくらい変わったか？
・ R ² は上がった？下がった？
・ 同じα（例：α＝100）でyはどれくらい変わる？
★理由づけ：メモで説明する
・ 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。
10/2020例：品切れ（点の位置がずれただけでOK）／キャンペーン（点が増える）

10/2020 データサイエンス 1/2

Slide19：外れ値を「発見→説明」させる運用

【目的】外れ値の影響（傾・式・R²・予測のズレ）を、言葉ではなく比較で理解させる。
【理解ポイント】傾きαは「傾き」の単位。この時点で「関係」は読み取れる。
【実習ポイント】散布図を描く。回帰診断の結果を見て比較しやすくする。
【確認】傾きが正しい場合：cleanでまだ正確。outlierは点の位置がずれただけでOK。

36

37

38

39

20. まとめ

今日の到達点（ここが！）
・ 散布図で「関係」を見て、回帰直線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（外れ）が必要
実測yと予測yは一致しないことが多い。 このズレを「傾き（傾斜）」と呼ぶ。
・ clean：ズレはあっても、回帰の基本が見える
・ outlier：ズレが大きくなる⇒予測がよくなる／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は勝手に線を描いているように見えるが、実は「ズレ（傾斜）」を小さくするメソッドを使っている。 今回はそのメソッドの「最小二乗法」を学ぶ。

10/2020 データサイエンス 1/2

Slide20：次回（最小二乗法）への接続を自然にする

【目的】回帰の「傾き」から「傾き」へ関心を移す。最小二乗法は「ズレを小さくするルール」として導入。
【理解ポイント】傾きαは「傾き」の単位。この時点で「関係」は読み取れる。
【実習ポイント】散布図を描く。回帰診断の結果を見て比較しやすくする。
【確認】学生に一言で「この散布図は3パターン（どれ？）」で説明できるようにする。

40

41