線形回帰発表 台本(仮)グループ11

【スライド１】

山本：

グループ11の中間発表を始めます。山本と甲野藤が発表させていただきます。よろしくお願いします。

【スライド２】

山本：

今回私たちは、『バネの長さと重さの関係』を線形回帰で分析しました。  
バネは重りを吊るすと伸びますが、どのくらい伸びるのかを数式で予測できたら便利ですよね。  
そこで、実際のデータを使って、重さとバネの長さの関係性を統計的に分析しました。  
それを表したグラフが次のスライドになります。

【スライド３】

山本：

こちらが散布図と回帰直線です。

青い点が実測値であり、赤い線が予測のための回帰直線になります。

このグラフから見てわかる通り、実測値は全体的に直線状に分布しており、

点はおおむね直線的に並んでいて、大きく外れているものはありません。  
このことから、重さが増えるとバネは直線的に伸びる傾向があることがわかります。

~~この回帰直線はy= 0.914x + 7.115を使用しました。  
では、次にこの直線を求める数式について説明します。~~

バネの長さと重さの関係を最もよく表す『最適な直線』を求める際に用いたのが

\*\*『最小二乗法』\*\*です。

【スライド４】

最小二乗法は、実測値と回帰直線との『ズレ²（誤差の二乗）』の合計が最も小さくなるように、傾きと切片を計算する手法です。

これにより、データの全体的な傾向を最もよく表す直線を選び出します。

なぜ二乗するのか？ですが、上に外れた点は「＋」、下に外れた点は「−」になるので、そのまま足すと相殺されてしまいます。

そこで、それぞれのズレを二乗して、全部「正の数」に変えてから合計することで、

「全体としてどれだけズレているか」を正しく評価できるようにしています。

【スライド５】

山本：

続きまして「線形回帰で得られた数式とその意味」ですが、

スライド３にもありました回帰式はy=0.914x+7.115という式で計算した結果です。

甲野藤：

この式はy=ax+bという形をしており、この式の意味を簡単に説明すると、重さが1g増えるとバネが約0.914mm伸びることを示しています。  
また、重さが0gのとき、バネの長さは7.115mmということになります。

この式を作るために使ったのが、線形回帰の計算式です。

具体的には、最小二乗法という手法を用いています。

傾きのaの計算方法ですが

a = np.sum((x - x\_mean) \* (y - y\_mean)) / np.sum((x - x\_mean) \*\* 2)

を当てはめました。まず x は重さ、y はバネの長さを表しています。そして

x\_mean や y\_mean はそれぞれの平均値です。この (x - x\_mean) という部分は、

「それぞれの重さが平均からどれくらい離れているか」を表します。

同じく「y」も、長さの平均からのズレです。つまりこの式は、

重さと長さが一緒にどうズレるか？をデータ1つ1つで掛け合わせて、

それをすべて足し合わせているということになります。

なぜこれが「傾き」になるのか?という内容ですが、傾きは、x が 1 増えたとき、

y がどれくらい増えるかを示す指標でした。そのためには、

「x と y がどれだけ一緒に動いているか」を、

「x がどれだけばらついているか」で割ればいいのです。

同様に切片bの計算方法ですが、b = y\_mean - a \* x\_meanを使用します。

これは「全体の平均値から見たときのyのスタート地点」のようなイメージです。

この回帰式を元にxの重さを任意の数値を出してもyの長さが予測できます。

例をあげるとxが20gの重さの場合、y=0.914×20+7.115≒ 25.4mm

という計算結果でyの長さが予測できます。

【スライド６】

山本：

次に「決定係数で予測の精度で評価する」という内容ですが、

今回の式がどれくらい信頼できるかを評価する指標がこちら、決定係数です。

R²は、予測が実際のデータにどれだけ当てはまっているかを示すもので、  
0から1の間の数値で表され、1に近いほど予測精度が高いという意味になります。

今回のR²は0.913でした。

これは、91.3%の精度でバネの長さを説明できているということになります。  
逆に言えば、残りの8.7%くらいが外的な誤差やノイズというわけです。

この値が高いということは、今回の回帰式が非常に正確であることを意味します。

甲野藤：

実際、グラフでも点が直線にかなり近く、目で見ても当てはまりがよく分かります。

計算には以下の式を使いました。

ss\_total = np.sum ((y - y\_mean) \*\* 2)

ss\_residual = np.sum ((y - y\_pred) \*\* 2)

R² = 1 - (ss\_residual / ss\_total)

ss\_total は、yの全体のばらつき

ss\_residual は、予測値とのズレ（二乗誤差）

それらの比を使って、どれだけズレを減らせているかを計算します

つまり、R²はただの計算式ではなく、「このモデルは信用できるか？」を判断するための指標です。  
今回のように0.9を超えるような高い値が出ていると、安心して今後の予測や応用に使っていけると考えられます。」

ここで、予測精度の違いが一目でわかるように、決定係数が異なる2つのサンプルを用意しました。  
R²＝0.98と0.73の回帰直線を比較すると、前者は実測値が直線の近くに分布しており、精度が非常に高いことがわかります。御覧の通り98%は回帰直線のかなり近い所に実測値があります。

一方で、73%の方はばらつきが大きく、直線から離れている点も多いため、予測の信頼性が下がります。

このことから先ほど私たちの回帰直線のデータである91%(スライド２をみせ、スライド５にもどる)はかなり信頼度の高いデータであるものといえます。

【スライド７】

最後に、今回の分析をまとめます。

今回は、バネの長さと重さの関係を、線形回帰を使って数式で表すことに実証しました。  
得られた回帰式 y=0.914x+7.115は、重さが増えるにつれてバネの長さが直線的に伸びるという関係を示しており、  
傾きや切片の意味を理解することで、「どんな傾向か」「予測に使えるのか」を把握できました。

特に、傾き aは「xが変化したときにyがどれだけ動くか」、切片 b は「基準点（スタート位置）」として、平均からのズレをもとに計算されています。

さらに、決定係数 R² = 0.913 という高い値から、この回帰直線が実測データに非常によく当てはまっていることがわかります。

以上のことから、今回の線形回帰分析を通して、データの傾向を正確に捉え、将来の予測に活用できる有効な手法であることが確認できました。

感想:山本 最初は「線形回帰とは？」という状態だったんですが、調べていくうちにだんだん仕組みがわかってきて、最後にはちゃんと使えるようになったのが嬉しかったです。  
チームでの発表も楽しく、良い経験になりました。

「最小二乗法」の明示について：

傾きaや切片bを導き出す計算のロジックは詳細に説明されていますが、

この最適な直線を導き出す手法の名称である「最小二乗法」という言葉が明示されていません。

資料のテーマが「最小二乗法 最適な直線の求め方（傾き、切片）」 となっていることを考えると、「最小二乗法という手法を用いています。」のような感じで触れておきたいです。

雑感：

専門的な言葉や概念をわかりやすい言葉で説明していて、理解しやすかったです。

ただ、5分という短い時間で、これだけの詳細な内容を説明するのは難しい気がするので、

台本通りやるのであればストップウォッチなどを使った事前の練習が不可欠かと思われます。

スライドに書かれた数式は、読み上げるのではなく、「ご覧のスライドの通り」といった形で指し示す程度に留めて、口頭ではその意味や、なぜそうなるのか？という概念的な説明に集中すると良いかも。

強調すべきポイント:

「バネの長さと重さはおおよそ比例関係にある」 というデータの大まかな傾向。

「y=ax+b」という回帰式が「重さからバネの長さを予測できるモデル」として機能すること。

傾き a が「重さ1gあたりのバネの伸び」であること。

切片 b が「重さ0gのときのバネの長さ」（初期値）であること。

決定係数 R² = 0.913 が高い予測精度であることを示し、この式が信頼して予測に使えること。

このあたりの重要な点は、メリハリをつけて説明したいですね！