



総務省統計局

社会人のためのデータサイエンス演習

第4週：課題の補講 Excel分析演習(回帰分析)

講師名：菅 由紀子

講座内容

1

- 課題の補講 Excel分析演習 回帰分析による予測モデル作成手順

2

- 課題の補講 Excel分析演習 各検定の実行方法

※Microsoft® , Microsoft® Office Excel® は、米国 Microsoft Corporation の、米国およびその他の国における登録商標または商標です。

※本資料は、Microsoft Corporation と提携しているものではなく、また、Microsoft Corporationが許諾、後援、その他の承認をするものではありません。

※本資料の本文では、©、®、™などの表記は割愛いたします。

※本資料ではデータ分析ツールとして Microsoft® Office Excel®2019を利用しています。

Microsoft® Office の他のバージョンや他の分析ツールを利用している場合は、ヘルプやインターネットなどで各自で調査し、該当機能に置き換えて参照してください。

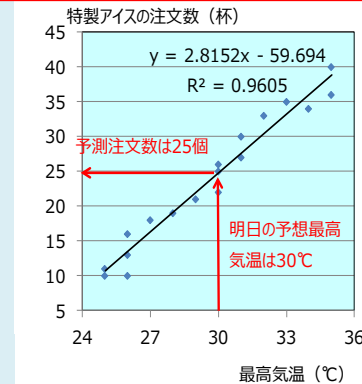
1. 第4週のまとめ

Analysis応用編

予測 / 分析結果の報告 / 機械学習

・単回帰での将来予測

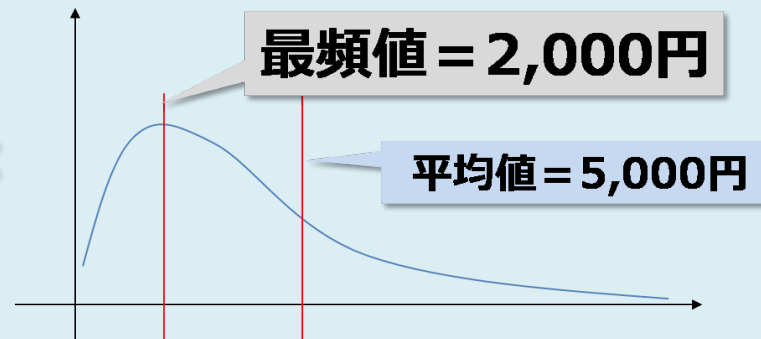
✓ 概念とエクセルでの実施



・分析報告の落とし穴

✓ 前提や指標を適切に提示

機械学習の応用先は幅広い



2. 回帰分析による予測モデル作成

両親の平均身長から子の身長を予測するモデルを作成します。

	A	B
1		(cm)
2	両親の平均身長	子の身長
3	185.0	183.0
4	161.1	167.4
5	166.2	175.0
6	161.0	164.2
7	166.6	173.9
8	164.5	170.0
9	180.5	179.3
10	169.1	172.0
11	170.9	171.0
12	160.0	166.3
13	160.0	184.0
14		

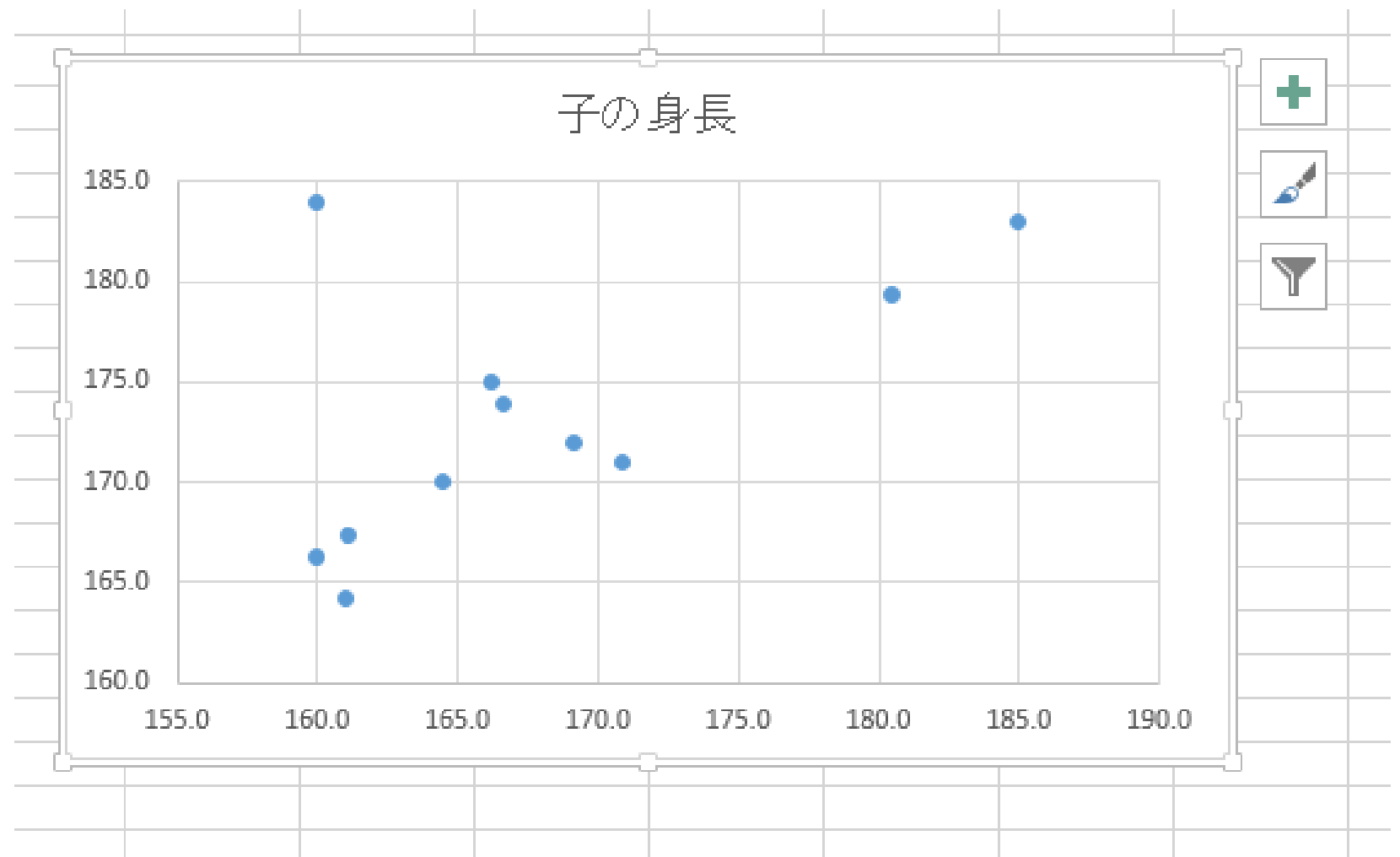
■ 演習ファイル名
hokou_data_W4.xlsx

2. 回帰分析による予測モデル作成

1) 下表から散布図を作成

(散布図の描画方法は、「第3週：課題の補講 Excel分析演習」参照)

	A	B
1	両親の平均身長	子の身長
2	185.0	183.0
3	161.1	167.4
4	166.2	175.0
5	161.0	164.2
6	166.6	173.9
7	164.5	170.0
8	180.5	179.3
9	169.1	172.0
10	170.9	171.0
11	160.0	166.3
12	160.0	184.0

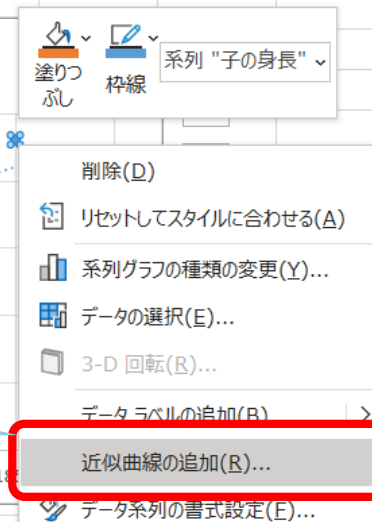


2. 回帰分析による予測モデル作成

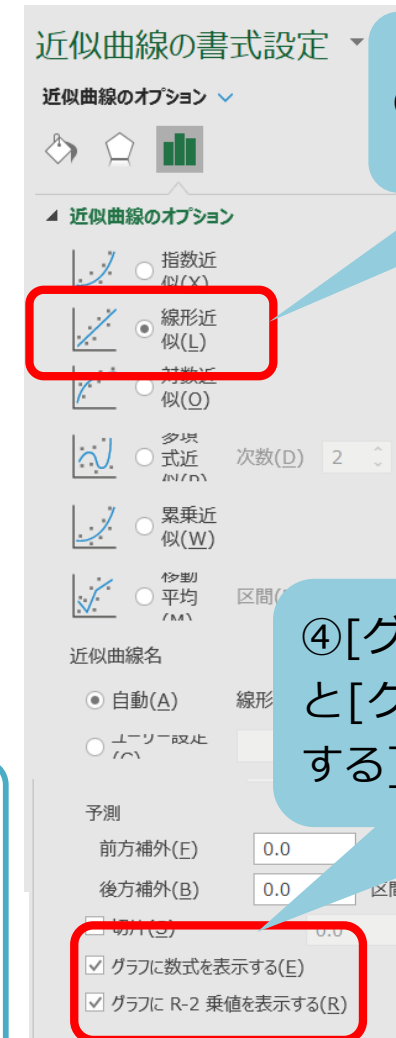
2) 散布図中のひとつのプロット上で右クリック
メニューを表示し、[近似曲線の追加]を選択

①プロット上で
右クリック

②[近似曲線の追加]
をクリック



3) 近似曲線のオプションで、近似または回帰の種類
を選択。（直線回帰の場合は線形近似を選択）



③ [線形近似] を選択

④[グラフに数式を表示する]
と[グラフにR-2乗値を表示
する]にチェック

補足

予測モデルは関数や「分析ツール」の「回帰分析」を使って算出することも可能。

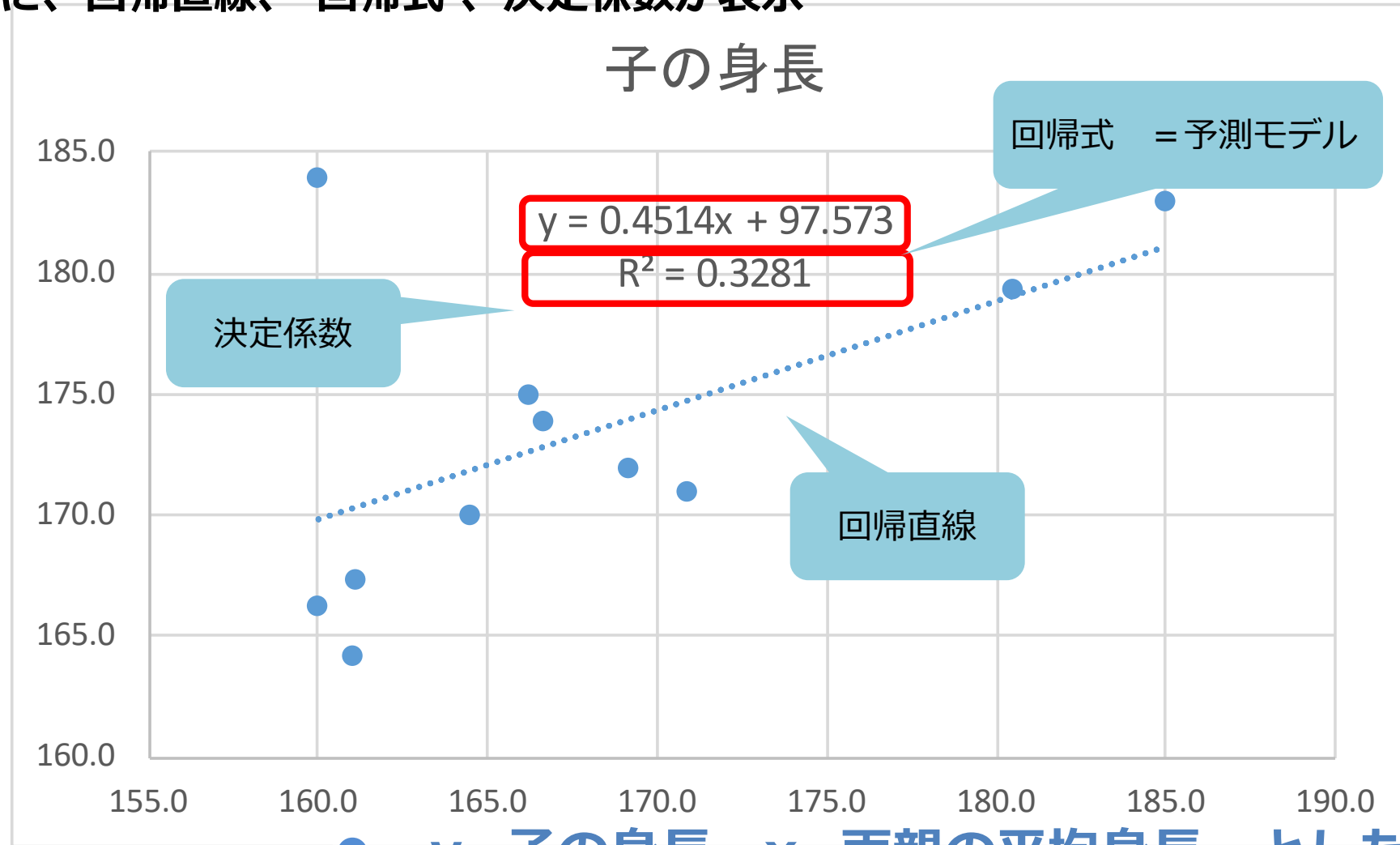
$Y = aX + b$ の場合

a (傾き) = SLOPE(変量Y、変量X)

b (切片) = INTERCEPT (変量Y、変量X)

2. 回帰分析による予測モデル作成

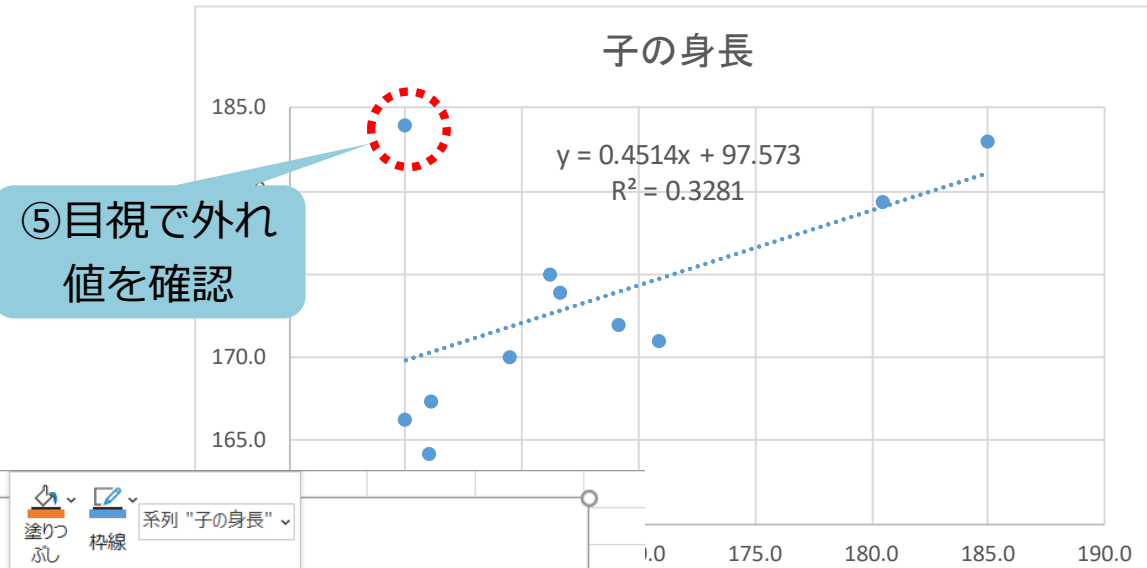
4) 散布図上に、回帰直線、回帰式、決定係数が表示



- y =子の身長、 x =両親の平均身長 としたとき
 $y = 0.4514x + 97.573$ の関係
- 決定係数 (R^2) \Rightarrow モデルのあてはまり指標は0.3程度と低い

2. 回帰分析による予測モデル作成

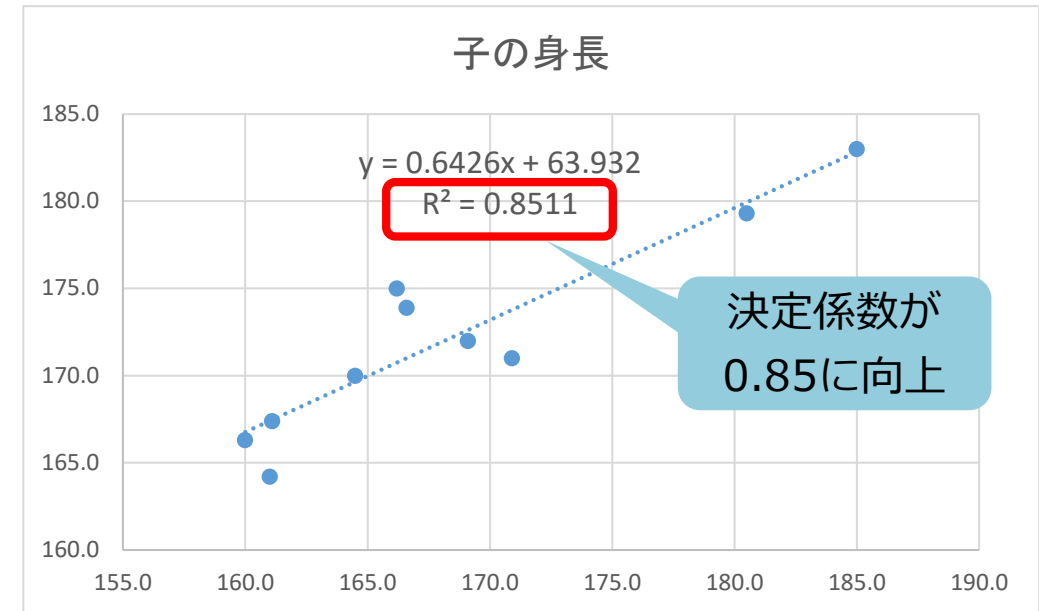
5) 外れ値を特定する



⑥外れ値のプロット（点）を選択し、右クリックバーから[データの吹き出しを追加]を選択

⑦X、Y座標値が表示されるので、当該値をデータから除外する

6) 外れ値を除外したモデルが表示



- 2変数の関係では、外れ値は散布図を描画し目視確認するのが簡便。
- 外れ値の除外判断は、分析目的やデータ特性を考慮する必要あり。